2018-07-19

Understanding and Improving Patient Flow in Outpatient Clinics and Emergency Departments

Zaerpour, Farzad

UNIVERSITY OF CALGARY

Understanding and Improving Patient Flow in Outpatient Clinics and Emergency Departments

by

Farzad Zaerpour

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MANAGEMENT

CALGARY, ALBERTA

JULY, 2018

Abstract

Improving patient flow is a critical aspect of quality management in emergency departments and other healthcare settings. By improving the flow of patients in healthcare facilities, we can decrease wait times and boost patient and staff satisfaction. Many patients face physical pain and suffering while waiting for treatment in healthcare facilities. Long wait times may also result in treatable illnesses and injuries becoming chronic conditions. This dissertation includes three main chapters, corresponding to three essays on understanding and improving patient flow in outpatient clinics and emergency departments. In some outpatient clinics, lab tests must be completed before the clinic appointment, as doctors need to have the test results when seeing a patient. Achieving this tight coordination of a patient's testing and his or her subsequent doctor's appointment may be difficult in a facility where many physicians share the same testing resources. The second chapter presents a mixed-integer programming (MIP)-based approach to reduce the likelihood of a patient not completing testing in time for the clinic appointment. In the third chapter, we focus on improving patient flow in emergency departments by looking at the physician scheduling problem. We show that the scheduling of physicians has a direct impact on the waiting time of patients. Chapter 4 presents a new crowding measure in emergency departments that is based on patient volume and mix of patients. We assess the relevance and significance of the proposed measure.

Acknowledgements

This dissertation would not have been possible without the guidance and the support of my supervisors, Professor Diane Bischak and Professor Mozart Menezes. Thank you for your extensive personal and professional guidance that has taught me what a good scientist (and person) should be. You have been a great source of inspiration to me. Thanks for Being Awesome.

I am also grateful to all of those with whom I have had the pleasure to work during this and other related projects. I would especially like to thank Dr. Marco Bijvank. I would also like to thank my committee members, Professor Michael Carter, Dr. Alireza Sabouri, and Dr. Ari Pandes. Thank you for agreeing to be a part of my defense committee and your valuable feedbacks.

I would like to show my gratitude to staff at Haskayne School of Business, especially Lesley Diamrzo who has always been helpful during my Ph.D. program.

Nobody has been more important to me in the pursuit of this work than the members of my family. I would like to thank my parents, Ahmad and Safoura. Your love and guidance are with me now and forever. Thank you for everything. I am exceedingly grateful to my brothers, Nima and Masoud. Thank you for always being there for me. Most importantly, I wish to thank my loving and supportive wife, Bahareh. Thank you for your support, patience, encouragement, love, and thanks for being beside me during this journey.

To my wonderful wife Bahareh

Table of Contents

List of Figures and Illustrations

List of Tables

List of Symbols, Abbreviations and Nomenclature

Symbols and abbreviations are introduced in each Chapter separately.

Chapter 1

Introduction

Healthcare systems and healthcare operations have received substantial attention lately, but many times not for good reasons. For example, long wait times have become a serious global phenomenon in many healthcare facilities. Healthcare operations problems are not constrained to developing countries. Over the past decade, for example, Canada has always ranked below other members of the Organisation for Economic Co-operation and Development (OECD) on several key indicators of patient flow and the related performance measures. This thesis focuses on patient flow, and its main goal is the improvement of flow in healthcare facilities, so that they do not experience unnecessary waits.

In order to further motivate the need for such study, we call attention to a 2017 report released by the Canadian Institute for Health Information (CIHI (2017)) comparing the healthcare performance of 11 countries which suggests that Canada has the highest patient wait times for visiting family doctors, specialists, and emergency department physicians. Moreover, the CIHI (2017) reports that: less than half (43%) of Canadians could get a same- or next-day appointment with their family doctor or at their regular place of care the last time they needed medical attention, compared with top-performing countries like New Zealand (76%) and the Netherlands (77%); access to after-hours care is also more difficult in Canada than in most other countries, with only 1 out of 3 patients being able to receive medical care in the evenings, on weekends or on holidays without going to the emergency department; and, Canadians visit emergency departments more often than people in other countries, and have the longest reported waiting time for service.

Still in CIHI (2017), it seems that the healthcare service design is corrupted by its own performance problems. For example, more than 40% of Canadians said that the last time they visited an emergency department, it was for a condition that could have been treated by their regular providers if they had been available. That is not surprising given that Canadians also report the longest wait times for specialists, with more than half (56%) of treatment seekers waiting longer than 4 weeks to see a specialist, compared with the international average of 36%.

A way for improving patient flows is to increase the availability of resources (medical doctors, nurses, healthcare equipment, etc.). Unfortunately, there are important costs associated with healthcare resources which creates pressure for a careful and wise use of them. That is where Operations Management and Operations Research can help to make the right trade-off between cost and quality of service, in order to ensure that resource availability, usually constrained by a limited budget, is of the right kind at the right place and closely matching patterns of patient demand. There are different levels of decisions making for improving performance and patient flow. These levels are classified as strategic, tactical, and operational level. At the strategic level, the interest is centred on determining the best number of healthcare resources, while at the tactical level the focus is on about making better use of the existing resources. That latter is the focus on chapters 2 and 3 of this thesis, although in two different healthcare settings. Finally, strategic, tactical, and operational levels need deeper understanding of current and future demand. In order to improve the flow of patients, we first need to know what goes into the healthcare facility on a daily basis and second examine the relationships between patient flow and performance metrics. That is the focus on chapter 4 in this thesis.

1.1 Patient flow and the related literature

In this section, we discuss patient flow and the related literature. First, we review studies that have addressed flow of patients and the consequences of delay. In the second part, we review studies that have addressed the problem at the strategic level to identify and remove bottlenecks. In the third part, we concentrate on the tactical level and review studies that have focused on a better allocation of available capacity to meet patient demands. Finally, the fourth part of this section concentrates on the operational level and patient flow management.

1.1.1 Patient flow and the consequences of delay

Reducing delays experienced by patients can produce dramatic improvements in access to service, treatment outcomes, and patient satisfaction. This section presents what is known about the patient flow and consequences of delays for patients.

Decreased access to health care

The first consequence of delays is that patients' ability to access care will be compromised. Access is influenced by many factors, but from the patient's perspective, the most important measure is how long they must wait to receive care. For example, Sills et al. (2011) did a cross-sectional study from November 1, 2007, to October 31, 2008, at a single, academic children's hospital emergency department. They sought to determine if ED crowding is associated with decreased access to care for children. Good access to care was defined as receipt of an indicated process within 1 hour of arrival. Nine crowding measures were used. The results suggested that two measures of ED crowding (total patient-care hours and number arriving in 6 hours) are consistently associated with access to care in the ED for pediatric patients.

The first consequence of the long wait times in emergency departments is that a significant proportion of patients may leave the hospital without being seen by a physician. Ding et al. (2016) suggested that EDs should make every effort to reduce the LWBS rate because these are the patients who are the least likely to receive care elsewhere. Rowe et al. (2006) studied the acuity level, reasons, and outcomes of patients who leave without being seen (LWBS). The study took place at the University of Alberta Hospital and Stollery Children's Hospital EDs in Edmonton, Alberta, Canada. A total of 711 (4.5%) of 15,660 registered emergency patients left without being seen. The majority of LWBS patients (73%) had triage level 4 (Less urgent) and 5 (Non-urgent). The major reason was identified as long waiting time (44.8%). Overall, 60% of LWBS cases sought medical attention within one week, 14 patients were hospitalized, and one required surgery. A study done at a public hospital in Los Angeles county showed that 46% of the patients who left the ED without visiting a physician were assessed to need immediate evaluation, and 11% were hospitalized during the subsequent week (Stock et al. (1994)).

Compromised quality of care

Patients value timely diagnostics, as early diagnosis will result in better treatment outcomes and reduce mortality rates. Neal et al. (2015) did a systematic literature review to examine whether the increased time to diagnosis and treatment in symptomatic cancer is associated with poorer outcomes. They found that the diagnosis of symptomatic cancer is likely to have benefits for patients in terms of improved survival, earlier-stage diagnosis and improved quality of life, although these benefits vary between cancers. The cancers with more reports of an association between shorter times to diagnosis and more favourable outcomes were breast, colorectal, head and neck, testicular and melanoma.

Several studies have shown that the odds of a patient experiencing a medical error increase when EDs are more crowded. For example, Epstein et al. (2012) examined the association between ED crowding and preventable medical errors for three diagnostic categories: acute myocardial infarction, asthma exacerbation, and dislocation requiring procedural sedation. They found that those patients seen during higher levels of ED crowding, the occurrence of medical errors was more than twofold higher. The relationship was shown to be non-linear, with most errors occurring at the highest crowding level. Shayne et al. (2009) studied the effect of ED crowding on undergraduate and graduate emergency medicine (EM) education. They reviewed possible positive and negative effects on each of the six recognized Accreditation Council for Graduate Medical Education (ACGME) core competencies. They found that less faculty supervision of learners leading to increased errors and decreased patient safety.

Lower patient satisfaction

Several studies have reported a strong and inverse relationship between patient satisfaction and wait times in different healthcare facilities.

In outpatient clinics, for example, Kreitz et al. (2016) examined how patient wait time relates to level of satisfaction and likelihood to recommend an orthopedic clinic to others. Data was collected at a single orthopedic clinic from June 2011 through October 2014. They reported that minimizing wait times in the orthopedic clinic may improve patient satisfaction but may not affect their likelihood of recommending the practice to others. Zopf et al. (2012) studied factors related to patient/family satisfaction in an academic pediatric otolaryngology clinic. Patients and families were surveyed following clinic encounters between May, 2010 and April, 2011. They found that examination room wait times and reception area wait times are associated with overall satisfaction and should be minimized.

In the ED setting, according to Gallup's 2002 patient loyalty database, patients who said they were "very satisfied" with their emergency department experiences reported waiting an average of 16 minutes. "Satisfied" patients waited an average of 35 minutes. Patients who were "somewhat dissatisfied" with their experiences waited 72 minutes on average, and "very dissatisfied" patients waited 92 minutes. Gallup research concluded that a wait time of about one hour or less is necessary to achieve some level of patient satisfaction. Thompson et al. (1996) assessed the impacts of actual waiting time, perception of waiting time, information delivery, and expressive quality on satisfaction scores. During a 12-month study period, a questionnaire was administered by telephone to a random sample of patients. Respondents were asked several questions concerning waiting times (ie, time from triage until examination by the emergency physician and time from triage until discharge from the ED), information delivery (eg, explanations of procedures and delays), expressive quality (eg, courteousness, friendliness), and overall patient satisfaction. They found that providing information, projecting expressive quality, and managing waiting time perceptions and expectations are more effective strategies to achieve improved patient satisfaction in the ED than decreasing actual waiting time.

1.1.2 Improving patient flow at the strategic level

An often suggested approach to improve patient flow in healthcare facilities is to increase associated resources for providing the care, such as number of hospital beds and number of physicians and specialist physicians. At the strategic level, healthcare decision-makers need to find the best number of resources to use in their healthcare facility.

Several studies have addressed improving patient flow at the strategic level. For example, Dreesch et al. (2005) suggested that health planners and decision-makers have to ensure that the right number of people, with the right skills, is at the right place at the right time to deliver health services for the population needs, at an affordable cost. They recommended that the methods chosen to estimate human resources requirements must reflect the political and economic choices and social values that underlie a particular health care system. They reviewed several approaches used to estimate requirements for human resources for health, such as needs-based, target-based, and utilization-based. In the needs-based approach, the number and type of health services to be delivered are estimated based on the health needs of the population. In the target-setting approach, the number and types of services are set by health authorities as specific targets, at various levels of care, considering the current level of technology, the demand of the population for certain services, and the various services already performed by health workers. The utilization-based approach usually takes the current level of health services utilization as being appropriate to meet the health needs and projects the future requirements of the health workforce based on future changes in the structure of the population.

Lamarche et al. (2011) examined whether a patient's experience with primary care services and use of services vary with the availability of health resources. The study captured the experience of 3,319 primary care users in five Quebec administrative regions in Canada. The patients' experience of care was recorded through a questionnaire that rated the accessibility, continuity and responsiveness of their primary care services and gathered their self-reported utilization of health services. In no case were positive ratings of services and greater use of them associated with greater resource availability. They concluded that simply adding resources runs the risk of diminishing, rather than improving, users' healthcare experience. Rabinowitz et al. (1999) studied the direct and long-term impacts of increasing the number of family physicians in rural and underserved areas. They reported that policy-makers and medical schools can have a substantial impact on the shortage of physicians in rural areas.

1.1.3 Improving patient flow at the tactical level

The tactical level suggests making better use of the resources already available and targets changes to the organization of health systems and the delivery of services. There are sometimes more effective and less expensive approaches to improve patient flow at the tactical level. Adding resources will have a marginal effect on improving patient flow if changes have not first been made to the organization of the systems (Romanow (2002)).

At the tactical level, there are several analysis tools than can be used to make better use of the resources, such as simulation models, queueing analyses, and optimization models. Below we review several studies that have used these tools at the tactical level.

Simulation models

Simulation models have been widely used to evaluate proposed changes in the delivery of health care. For example, Hung et al. (2016) constructed a discrete event simulation model to test different pediatric emergency department staffing scenarios. Simulation of an addition of a hospital volunteer and a second triage nurse demonstrated reductions in pretriage waiting time and the proportion of patients waiting longer than 30 or 60 minutes for pretriage. Simulation of an extra physician shift to the staff schedule demonstrated reductions in length of stay for patients of all triage categories. Rohleder et al. (2008) applied discrete event simulation modelling to test different improvement alternatives for reducing wait time and congestion at an orthopedic outpatient clinic. They examined optimized staffing levels, better patient scheduling, and an emphasis on staff arriving promptly. Coelli et al. (2007) used discrete event simulation in the analysis of a mammography clinic for defining optimal operating conditions, indicating the most adequate capacity configurations and equipment maintenance schedules. The results showed that a large impact over patient waiting time would appear in the smaller capacity configurations. Cótẽ (1999) developed a discrete event simulation model to investigate the relationship between examining room capacity and patient flow across four clinic-based performance measures. They showed that increased resource utilization does not necessarily imply longer waiting lines nor longer patient flow times.

Queueing analyses

Queueing analysis is also a key tool in healthcare resource plannings. For example, Cochran and Roche (2009) used queueing models to find the staffing levels of different emergency department areas. Bruin et al. (2007) used queuing theory to investigate the bottlenecks in the emergency care chain of cardiac in-patient flow. The goal was to determine the optimal bed allocation over the care chain given a maximum number of refused admissions. They found that refused admissions are primarily caused by unavailability of beds downstream the care chain.

Queuing analyses and simulation models each have their advantages. Queuing models are easier to use and require less data. However, discrete event simulation allows us to model the details of patient flows in more complex healthcare settings. Simulation is used in chapter 3 of this thesis to model patient flow in an ED facility. However, we will not explore queueing analyses further in this thesis.

Mathematical models: The technique used in chapters 2-3

A key element in improving patient flow through healthcare systems is that the schedule of care providers matches with the time-varying patient demand. Mathematical modelling is a commonly used technique to find the best schedule of care providers, such as nurses and physicians. For example, El Adoly et al. (2018) proposed a mathematical model for the nurse scheduling problem. The objective was minimizing the overall hospital cost while taking into consideration the governmental rules and hospital standards. The results showed the superiority of the obtained schedule to those generated manually by the supervisor head nurse in terms of improving nurses' satisfaction as well as reducing the overall overtime cost. Azaiez and Al Sharif (2005) suggested a zero-one linear goal programming (LGP) approach for scheduling a number of nurses in a 4-week period. The scheduling problem contained a total of 11 scheduling sets of constraints. These include balanced schedules, fairness considerations, and nurses' preferences, in addition to ergonomic considerations, and staffing requirements. The objective function was to minimize the sum of the weighted deviations from the constraints. Beaulieu et al. (2000) proposed a mathematical programming approach to construct a six-month schedule of physicians in the emergency room. They partitioned the constraints of the model into four following categories: compulsory constraints, ergonomic constraints, distribution constraints, and goal constraints. The objective of the proposed model was to minimize the overall deviations from the constraints.

We use a mathematical modelling approach at the tactical level in chapter 2 and chapter 3 of this thesis, although in two different healthcare facilities. In chapter 2, we develop a mixed-integer programming (MIP)-based approach for assigning time slots to the physician clinics in outpatient clinics. In chapter 3, we propose a two-stage stochastic mixed-integer programming (MIP) for scheduling physicians to improve patient flow through emergency departments.

1.1.4 Improving patient flow at the operational level

Improving patient flow at the operational level has extensively studied in the literature. For example, Barak-Corren et al. (2017) suggested an automatic hospitalisation prediction model to reduce delays in transferring an admitted patient from the ED to an inpatient department. They found early clinical decisions such as testing for calcium levels to be highly predictive of hospitalisations. They estimated that the use of the prediction system in the studied ED can save more than 250 patient hours per day. Nguyen et al. (2011) studied the problem of missed patient appointments in a resident continuity clinic to determine patient characteristics and healthcare outcomes. They found that the disproportionate frequency of missed appointments in the resident continuity clinic is explained by patient factors and practice discontinuity, and that patients with frequent missed appointments demonstrated worse health care outcomes. Min and Yih (2010) addressed the problem of scheduling patients with different priorities in a surgical facility with finite capacity. They formulated this problem by using a stochastic dynamic programming model. A structural analysis of the proposed model was conducted to understand the properties of an optimal schedule policy. They found that the consideration of patient priority results in significant differences in surgery schedules from the schedule that ignores the patient priority. Wilper et al. (2008) studied the trends and predictors of wait times in U.S. emergency departments between 1997 and 2004. They found ED closures and an increase in total ED visits as the main reasons of longer ED waits. Between 1994 and 2004, the number of U.S. ED visits increased from 93.4 million to 110.2 million annually, while the number of EDs fell by as much as 12.4 percent. They suggested that prolonged ED waits have serious impacts on the quality of care, such as prolonged pain and suffering, patients leaving without seeing a physician, and dissatisfaction with care.

When making resource planning for several years, it is necessary to consider how patient flow may change over that period of time. That is a necessary input for all decision levels and will be the focus on chapter 4 of the thesis.

1.2 Outline of the dissertation

This dissertation includes three main chapters, corresponding to three essays. Table 1.1 positions the three chapters of this thesis in relation to the three decision levels for improving patient flow.

Chapters of the thesis	Objective	Decision level(s)	Healthcare setting	Analysis tool
Chapter 2 & Chapter 3	Making better use of the existing resources	Tactical level	Outpatient clinic & Emergency department	Combinatorial optimization & Stochastic programming and Simulation
Chapter 4	Deeper understanding of current and future demand	Strategic, tactical, and operational levels	Emergency department	Regression analysis

Table 1.1: Improving patient flow from three decision levels

At the tactical level, we address two problems: 1) A doctor-clinic assignment problem in medical outpatient clinics (Chapter 2), and 2) A tactical scheduling problem of physicians in emergency departments (Chapter 3). In both problems, we consider the strategic level decision as a constraint, as it is the result of several internal and external factors like political and economic choices and social values. However, we still provide some insights to identify and remove bottlenecks at the strategic level. Finally, chapter 4 concentrates on emergency department patient flow and propose a new crowding measure that is a necessary input for all decision levels. We summarize each chapter as follows.

Chapter 2: Coordinated lab-clinics: A tactical assignment problem in healthcare

In some medical outpatient settings, it is desirable to perform patient diagnostic testing just before the appointment with the physician, effectively linking the testing to the clinic appointment. If testing resources are shared by several physicians, it becomes difficult to assure that testing is completed in time (with some probability) due to the variation in testing requirements across patients and types of clinics held concurrently. To address this tactical-level doctor-clinic assignment problem, we develop a mixed-integer programming (MIP)-based approach for assigning time slots to the physician clinics. The approach maximizes the minimum service level across blocks of time to reduce the likelihood of a patient not completing testing in time for their clinic appointment. A branch-and-price heuristic procedure is proposed to solve practical problem instances, and numerical examples are presented to show the efficiency of this model. Two mini-cases based on clinics' actual operations are provided. The results of the mini-cases suggest that the proposed scheduling method will bring important improvements to these systems. This chapter is based on Zaerpour et al. (2017).

Chapter 3: Physician scheduling to improve emergency department patient flow

Emergency department (ED) crowding has become a serious concern worldwide that may impact treatment outcomes, patient satisfaction, and access to high-quality medical care. In this chapter, we develop a two-stage stochastic mixed-integer programming (MIP) for scheduling physicians to improve patient flow through emergency departments. Scheduling ED physicians has traditionally been built around physicians' preferences, regulatory and work constraints. A more effective method is a procedure to achieve an overall hourly balance between patient arrivals and ED physicians' productivity while taking into account the traditional constraints. We define an ED physician's productivity as the number of new patients seen per hour by that physician. The study of ED physicians demonstrates significant differences among physicians in terms of their productivity. Furthermore, the analysis of hourly productivity rates suggests that ED physicians evaluate significantly more patients early in a shift, and few at the end of a shift. We found no significant relationship, however, between ED crowding and physicians' productivity. Finally, we provide a case study and develop a simulation model to evaluate the impact of the near-optimal stochastic MIP solution on reducing patient wait times. The results of the case study suggest that our proposed schedule will reduce the average wait time of patients. This chapter is based on Zaerpour et al. (2018b).

Chapter 4: A new definition of crowding in emergency departments: patient volume and patient mix

In this chapter, we present a new crowding measure for emergency departments (EDs) that is based on patient arrivals and the mix of patients. The most commonly used crowding measure is patient volume, which ignores the mix of patients visiting EDs. However, identifying patient mix is of great importance for strategic and operational planning in EDs. We define ED crowding as the number of patients arriving to EDs from each group. To capture the patient mix arriving to EDs, we propose a patient classification system based on patient demographics and diagnostic information.

A crowding measure is usable and meaningful if, first, we can estimate the arrival patterns of its suggested variables and, second, it has significant impacts on ED performance metrics. To examine these two conditions for the suggested crowding definition, we use data from five hospitals in February, July and November for the years of 2007, 2012, and 2017 in the city of Calgary, Alberta, Canada. Our primary goal is to examine arrival patterns of all patients as well patients at each group. We find meaningful arrival time patterns of patients as well as types of patients in EDs. To show the relevance of our suggested ED crowding metric, we then examine the relationships between the two crowding variables and performance metrics. The performance metric dimension is based on commonly-used metrics of system performance, including both time-based and quality measures. The regression results suggest that number of ED arrivals is the main predictor of time-based ED performance measures. Patient mix is, however, the key predictor of quality of care in EDs. The regression results show the relevance and significance of our crowding definition that is based on both patient arrivals and mix of patients. This chapter is based on Zaerpour et al. (2018a).

Chapter 5: Summary and conclusion

Finally, chapter 5 summarizes the results of previous chapters and provides potential future research questions.

Chapter 2

Coordinated lab-clinics: A tactical assignment problem in healthcare

2.1 Introduction

In some medical outpatient settings, it is desirable to perform diagnostic tests on a patient just before the patient's appointment with a physician, effectively linking the testing to the appointment. For example, a patient who has an appointment with a respirologist will undergo pulmonary function testing by a respiratory therapist in a pulmonary function laboratory. Ideally, the testing is performed only just prior to the patient's appointment with the respirologist, as the physician needs to have current test results in hand when seeing a patient, but patients should not have to arrive a long time before their appointment in order to have testing done.

Achieving this tight coordination of a patient's testing and his or her subsequent doctor's appointment may be difficult in a facility where many physicians share the same testing resources. During a week in such a facility, each physician holds one or more clinics, which are scheduled blocks of time during which a physician sees patients by appointment. The physical space available to see patients limits the number of clinics that can run at the same time at the facility, but since the weekly clinic schedule is based to some degree on physician preferences, there may be more clinics held on some half-days than on others. Some physicians may hold general clinics where they see any type of patient within their specialty, while other physicians hold clinics for patients with a particular condition that is in their subspecialty area of expertise. Physicians also vary on the number of new and follow-up patients they want to see during a clinic. Different physicians thus see different numbers and types of patients in a given half-day clinic. However, the variety in patient types can mean that each patient requires a different type of pre-appointment testing that takes a different amount of time to complete. The amount of, and variation in, the testing requirements generated by patient appointments across concurrently scheduled clinics may not be considered when the week's schedule of clinics is set up. A poor clinic schedule can yield an uneven workload on the testing facility, making it more difficult to complete testing in time for the patient's appointment. This can lead to lower quality of care as well as result in friction between physicians and testing staff. An approach to clinic scheduling that takes into account the varying testing requirements generated by concurrently scheduled clinics, as well as the limited availability of the resources required to complete that testing, is therefore needed.

This research is motivated by the example of outpatient respirology facilities in hospitals that are part of Alberta Health Services in Alberta, Canada. Currently, administrators divide a period (usually a week) into blocks of time of a few hours each day (half-days) and within each block schedule clinics of particular types to be run by assigned physicians. For the sake of stability, these periods have similar assignments for each block for a relatively long time. Hence, decisions are made with a long horizon in mind. Although at the operational level there are adjustments with respect to the order in which physicians see patients and the type of patient (new or returning) a physician sees, administrators would like to have schedules set up in such way that there would be a high probability that the resulting mix of patients that have appointments in each block of time would be able to finish the required lab tests before their appointment. This is the problem we address.

It is important to stress that the problem under consideration is that of assigning doctors to clinics and clinics to blocks of time in a period (usually a week). The assignments are repeated period after period until there is a major reason for reviewing the assignments (possibly a new hire or the introduction of a new clinic). This is a tactical level problem and should not be confounded with the operational level problem of scheduling patients in a given block (ordering the patients to be treated).

More specifically, our research problem is the following: given the available testing time in each block of the period, what is the assignment of physicians' clinics to each block that will maximize the minimum service level over all blocks in a period while satisfying physicians' preferences for types and blocks for their clinics? Although there are different perceptions of service level, which will be reviewed in the next section, we define the service level for a block as the probability that the sum of all patients' testing times assigned to a block is less than or equal to the available testing time. This performance metric assures that the probability of not completing any given patient's testing prior to their appointment will be as similar, and as small, as possible.

Although this study is motivated by the problems that we have observed at respirology lab-clinics, and for ease of exposition and consistency we will continue with that example throughout, more generally our study provides an approach for assigning servers in any type of demand centers that share a predetermined amount of a service resource that provides a time-sensitive activity to those demanders, who send independent requests for varying amounts of service. Outside the healthcare setting there are several variations of the problem, for example, cold temperatures and deicing problems can cause delays in airports. A limited amount of deicing equipment is available for airfield pavement and aircraft anti-icing operations. To avoid or reduce delays due to the need for anti-icing, it would be valuable to assign flights to runways, taxiways, and time slots such that all planes can be anti-iced prior to their scheduled departure time. The human resources (HR) department of a particular business school assigns blocks for dealing with specific tasks (e.g., safety and ergonomic related concerns are only dealt with from Tuesday to Thursday during morning times, while paycheck related issues are dealt with every afternoon) and the HR department wants to choose what specific tasks should be assigned to each block in order to minimize the chance that requests are not met within the block allotted time - the backroom of the HR department is analogous to a testing-lab in the respirology clinic case. In some tax revenue agencies in France, different types of "tax-payer requests" are assigned to different blocks where the tax-payer deals first with a general administrative worker who proceeds with assessment, organization of documents, checking basic validity of claims etc., and then, when the general pre-work is done, sends the client to talk to the particular field expert related to the request. Back to the healthcare setting, for small surgical interventions and some MRI exams where recovery is relatively fast compared to the length of a block, the problem of assigning doctors-clinics to particular blocks of time will create needs for recovery rooms.

In this study, we address a tactical-level problem for assigning time slots to the physician clinics. First, we propose a MILP with the objective of maximizing the minimum service level across blocks of time. As a second step, a branch-and-price heuristic procedure is proposed to solve practical problem instances.

This study has three key contributions. Firstly, it shows the potential of operations research (OR)-based studies as helpful decision making tools for real-world healthcare problems. Secondly, it demonstrates and evaluates the implementation of the branch-and-price algorithm for the doctors' assignment problem or any assignment problems in any type of demand centers that share a predetermined amount of a service resource. The third contribution is a method for facility managers to decide on the required lab capacity or the common resource at the strategic level.

The rest of this chapter is organized as follows. In section 2.2, we define more precisely the coordinated lab-clinics problem and review the relevant literature. In section 2.3, we describe the properties of the problem. In section 2.4, the mathematical formulation for the model is first developed, and then the complexity of the problem is discussed. A branch-and-price heuristic is proposed to solve practical examples in section 2.5. In section 2.6, two mini-case studies from two respirology clinics within the Alberta Health Services (AHS) are provided. Finally, we conclude in section 2.7 with a discussion and further research suggestions.

2.2 The coordinated lab-clinics problem and related literature

In healthcare, many scheduling studies are focused on outpatient appointment scheduling, with the goal of finding an appointment system for which a particular measure of performance is optimized (e.g., Baron et al. (2016); Samorani and LaGanga (2015); Zacharias and Pinedo

(2013); White et al. (2011); Kucukyazici et al. (2011)). The lateness and interruption level of doctors are two doctor-related parameters studied (see, e.g., Klassen and Yoogalingam (2008); Cayirli and Veral (2003)). However, the outpatient appointment scheduling problem is focused on the operational level of scheduling, whereas the coordinated lab-clinics problem addresses the tactical level. A problem that is more closely related to the lab-clinics scheduling problem is medium-term surgical block scheduling, where a cyclic schedule is determined to assign doctors to blocks over a time period (see e.g., Guerriero and Guido (2011); Fei et al. (2009); Gupta (2008)). In both problems, the strategic level decision acts as a constraint, as the number of physicians and the available time in each block has already been decided. The operational level of both scheduling problems, where the sequence of patients will be determined, occurs after the tactical level. Figure 2.1 summarizes the main problems on each decision level in the surgical block and the lab-clinics scheduling processes.

	The coordinated lab-clinics problem	**The surgical block problem**
Strategic Level	Determining the number of physicians and testing capacity	Determining the total OR time dedicated to each surgical specialty
Tactical Level	Assigning doctors to clinics to maximize service level	Allocating surgical specialties to ORs to maximize utilization
Operational Level	Sequencing patient testing and clinic appointments	Sequencing patients in each OR

Figure 2.1: Stages in the surgical and the lab-clinics scheduling processes

Van Houdenhoven et al. (2007) applied bin-packing and portfolio techniques to assign surgeries to blocks with the objective of minimizing the probability of overtime. The bin-packing algorithm suggested in Van Houdenhoven et al. (2007) can also be used to solve heuristically our assignment problem for a given service level. Although real-sized problems are solvable when applying the exact solution proposed herein, the heuristic approach can be, with minor modification, used after applying our method for obtaining the time distribution. Latorre-Núñez et al. (2016) proposed an integer linear programming model to address the surgery scheduling problem considering not only the assignment of surgeries to operating rooms, but all the resources required for each surgery (human and material), and the recovery beds. They then developed a metaheuristic based on a genetic algorithm to deal with largersized instances. Silva et al. (2015) addressed surgery scheduling and resources assignment in an operating room. They applied an integer model and integer programming based heuristics to maximize the use of the operating rooms. Min and Yih (2010) proposed a stochastic optimization method to find an optimal surgery schedule of elective surgery patients with uncertain surgery operations.

The surgery scheduling problem has been studied with varying objective function formulations. Two typical objective functions, however, are minimizing patient waiting time before undergoing major or minor operations and maximizing the utilization of the operating theatre (see, e.g., Saadouli et al. (2015); Choi and Wilhelm (2014); Hans et al. (2008); Lamiri et al. (2008)). Hans and Vanberkel (2012) used frequency of overtime to compute the operating room utilization level. The frequency of overtime definition is very similar to the service level defined in this study. They defined the frequency (or accepted risk) of overtime as the probability that the total surgery time duration is greater than the amount of allotted time.

Nurse scheduling problems and physician scheduling problems are also closely related to this research, because they deal with personnel scheduling. For example, Azaiez and Al Sharif (2005) suggested a zero-one linear goal programming (LGP) approach for scheduling a number of nurses in a 4-week period. The scheduling problem contained a total of 11 scheduling sets of constraints. These include balanced schedules, fairness considerations, and nurses' preferences, in addition to ergonomic considerations, and staffing requirements. The objective function was to minimize the sum of the weighted deviations from the constraints. Carter and Lapierre (2001) interviewed physicians from six hospitals located in greater Montreal, Canada, in order to understand the characteristics of the schedules. They then developed a mathematical programming technique to produce better schedules and reduce the time needed to build them. Beaulieu et al. (2000) proposed a mathematical programming approach to construct a six-month schedule of physicians in the emergency room. They partitioned the constraints of the model into four following categories: compulsory constraints, ergonomic constraints, distribution constraints, and goal constraints. The objective of the proposed model was to minimize the overall deviations from the constraints.

In the nurse scheduling problems, maximizing personnel satisfaction and minimizing salary cost are often considered as two objectives to achieve simultaneously (See e.g., Dowsland (1998)), while emergency physician schedules are driven by personal preferences (See e.g., Ferrand et al. (2011)). However, the coordinated lab-clinics problem studied in this chapter is focused on providing a proper service level.

Our coordinated lab-clinics problem can be cast as an assignment problem in which doctors are assigned to run clinics in specific blocks of time. Each doctor assigned to a clinic will see a certain number of patients with different testing requirements, and the type of clinic run by each doctor in a given block determines the total testing time required. Assignment problems are a well-studied topic in combinatorial optimization. The original version of the assignment problem is to find a one-to-one assignment of a number of tasks, jobs, or activities to an equal number of agents, machines, or resources in order to minimize the total cost of the assignments. However, many variations of the original problem have been developed since it was introduced by Votaw and Orden (1952) (see. e.g., Pentico (2007) for a review on different versions of the assignment problem), and it has numerous application areas. For example, Domenech and Lusa (2016) developed a mixed integer linear programming model to assign teachers to courses, taking their preferences into consideration. Each teacher can choose his/her working time preference (the hours of every day of the week that s/he can teach). The model is used to solve the teachers-courses assignment problem in order to balance teachers' load and maximizing teachers' working time preferences. Gelareh et al. (2015) addressed a truck dock assignment problem with operational time constraint. In the truck dock assignment problem, our goal is to find an assignment of the trucks to the docks such that the number of deliveries being processed is maximized while meeting the arrival and the departure times of every truck (as well as capacity of the cross dock). They developed a branch-and-cut algorithm to solve real-life size instances in a reasonable time. Carello and Lanzarone (2014) considered the problem of assigning patients to nurses for home care services. They developed a cardinality-constrained robust assignment model that takes into account uncertainty in patients' demands as well as continuity of care, which requires that a patient is always cared for by the same nurse. Tànfani and Testi (2010) addressed the problem of determining the assignment among wards and operating rooms during a given planning horizon, together with the subset of patients to be operated on during each day.

They presented a mixed integer linear programming model to minimize a cost function based upon a priority score, that takes into account both the waiting time and the urgency status of each patient. The version of the assignment problem that is most closely related to the lab-clinics problem is a three-dimensional assignment problem, where we assign n tasks to m agents over a set of t time periods; during each time period, each task has to be assigned to one agent, and each agent can process at most one task (see e.g., Ćustić et al. (2015); Hahn et al. (2008)). In the lab-clinics problem, a set of clinic types has to be assigned to a set of doctors over a planning period. Each doctor has a certain demand for each clinic type. If a doctor $d \in \{1, \ldots, D\}$ is assigned to clinic type $c \in \{1, \ldots, C\}$ in block b, he/she must visit a certain number of patients. This group of patients induces a service time in the lab, coming from a known distribution. However, the performance measure studied in this study is service level, a non-standard measure in the assignment literature.

Another problem closely related to the one presented herein is the Multiple Subset Sum Problem (MSSP) (e.g., Caprara et al. (2000)) - in section 2.4.2, the MSSP is shown to be a special case of the coordinated lab-clinics problem and therefore, any polynomial time algorithm that could solve the latter problem could be used for solving the former. The goal of the MSSP is to select a subset of items with different weights from a set of given items and pack them into knapsacks such that the overall weight of the items packed in the knapsacks is maximized. The MSSP is a special case of the multiple identical knapsack problem in which the value of each item is equal to its weight and all knapsacks have the same capacity. Martello and Toth (1984) proposed a mixed dynamic programming and branch-and-bound approach for the subset sum problem with only one knapsack. Caprara et al. (2003) suggested a polynomial-time 3/4-approximation algorithm to solve the MSSP. They show that the running time of the proposed method is linear in the number of items and quadratic in the number of knapsacks.

2.3 Problem description

The objective of the coordinated lab-clinics scheduling problem (CLCSP) is to maximize the minimum service level over all blocks, i.e., the *max-min* objective. Alternatively, we could change the objective to maximize the overall service level over all blocks, i.e., a *max-sum* objective. The main difference between the two solutions would be on the blocks with worst service level, as the max-sum solution to this problem may leave blocks where too many patients are quite dissatisfied with the service level provided. That is, although the overall sum of the service level is maximized and thus better than that obtained by a max-min objective, the number of unhappy patients may be a too high price to pay to achieve it. Note that, in our problem, patients care strongly if they receive the full service in the same block, giving less weight to the total amount of time to finish the service if that does not imply returning another day. Table 2.1 summarizes the notation used in the description of the problem. A block b is one of the time-divisions of the work week, which are non-overlapping and cover the whole working period. Typically, there are different types of clinics held in a given block. Doctors assigned to clinics in a particular block visit a certain number of patients. This group of patients induces a total testing time in the lab with an expected service time and variance. The mean and variance of testing time in a block depends on doctors, but also on clinics in that particular block due to the fact that the number of patients a doctor visits in a clinic type might be different from another doctor-clinic pair. Each block may have a different available testing time because of different amounts of personnel and testing lab resources available in that block.

Table 2.1: Notation for the lab-clinics problem

d	$\in \{1, \ldots, D\}$	index of doctors, where D represents number of doctors
b	$\in \{1, \ldots, B\}$	index of blocks, where B represents number of blocks
c	$\in \{1, \ldots, C\}$	index of clinics, where C represents number of clinic types
T_b		the available testing time in block b
$\bar{t}_{dc}$		the mean of testing time induced by doctor d in clinic type c
v_{dc}		the variance of testing time induced by doctor d in clinic type c
μ_b		the mean of total testing time induced in block b
W_{dc}		the demand of doctor d for clinic type c
R_c		the requirement of clinic type c over all blocks
V		the testing time variance set
S		the testing time standard deviation set
y_{bi}		1 if element i is chosen from sets V and S for block b, 0 otherwise
Z_{db}		1 if doctor d has a preference to work during block b, 0 if indifferent
x_{dbc}		1 if doctor d is assigned to clinic c in block b, 0 otherwise

There is a pre-agreement on the number of new and follow-up patients seen by doctor d in clinic type c before the doctors' schedule is made.These numbers are quite stable over a long period of time. The number and type of patients seen by doctor d in clinic type c determine the testing time requirement for that particular doctor-clinic combination, which is the convolution of some random variables. As a numerical example, suppose we have doctor d_1 visiting 1 new and 2 follow-up patients in clinic type c_1. Furthermore, the mean and variance of the required testing times are (80, 900) and (60, 800) for the new and follow-up patients, respectively. We can now compute the mean, $\bar{t}_{d_1c_1}$, and the standard deviation, $\sqrt{v_{d_1c_1}}$, of testing time induced to the test room by doctor d_1 in clinic type c_1, which is $\bar{t}_{d_1c_1} = 80 + 2 * 60 = 200$ and $\sqrt{v_{d_1c_1}} = \sqrt{900 + 2 * 800} = 50$.

Having the expected testing times of all doctor-clinic combinations, one can now compute the total expected time needed in a block. For example, if in a given block doctor d_1 performs clinic c_1 and doctor d_2 performs clinic c_2, then the total expected time needed for testing would be $\bar{t}_{d_1c_1} + \bar{t}_{d_2c_2}$. Note that a doctor-clinic combination provides a certain number of patients to the lab regardless of the block in which the clinic takes place.

Any block has a limited number of possible doctor-clinic combinations. Each doctor operates at most C clinics, and a block has at most D doctors. Thus, $\sum_{i=0}^{D} \binom{D}{i} C^i$ is the (theoretical) maximum number of possible doctor-clinic combinations in a block. Each possible doctor-clinic combination induces a total testing time that is the summation of random variables.

Hence, we can pre-compute an ordered set $V = \{V_1, \ldots, V_{u \leq \sum_{i=0}^{D} \binom{D}{i} C^i}\}$ that collects the different possible testing time variance, where u represents the cardinality of the set. From V one may obtain set $S = \{S_1, \ldots, S_u\}$ with a corresponding order to V such that $V_i = S_i^2$; that is, set S collects the possible standard deviations. These two constructs will be very helpful for the formulation below. As a simple numerical example, suppose we have two doctors running two types of clinic. Table 2.2 gives $\bar{t}_{dc}$ and v_{dc}, the mean and the variance of testing time for each doctor-clinic pair.

d \ c	1	2
1	(3,4)	(5,5)
2	(2,7)	(6,9)

Table 2.2: The mean and the variance of testing time for each doctor-clinic combination, ($\bar{t}_{dc}$,v_{dc})

The total number of doctor-clinic combinations in a block is equal to $\sum_{i=0}^{2} \binom{2}{i} 2^i = 9$. We can now compute and order the variance set V for all these 9 doctor-clinic combinations, which is $V = \{0, 4, 5, 7, 9, 11, 12, 13, 14\}$. From V one may obtain set S, by taking the square roots; that is $S = \{0, 2, 2.2, 2.6, 3, 3.3, 3.4, 3.6, 3.7\}$.

A major issue with these two sets is that it could be possible, just by coincidence, that there are cases of times resulting in equal values for the mean but different variances. We are assuming that times are coming from real numbers and such coincidences would be unlikely, and in the event it happened then it is always possible to differentiate times after some numbers of decimal places so that for all practical purposes the results would not suffer.

2.4 Mathematical formulation

In this section, we formulate the problem of maximizing the service level, given the problem constraints. The service level is the probability that the sum of all patients' testing times assigned to block b is less than or equal to the available testing time, that is $\Pr(\mu_b + \omega S_b \leq T_b)$. This value of ω tells us the number of standard deviations that T_b is away from the mean. Calculating the service level requires the mean and the standard deviation of the total testing time induced to the testing lab, which are the results of our model. However, for any given service level we can compute the corresponding value of ω and therefore, maximizing the minimum service level within all blocks is the equivalent of maximizing the deviation from the mean. In order to formally state the problem mathematically, we define x_{dbc} to be 1 if doctor d is assigned to clinic c in block b, and zero otherwise. The binary decision variable y_{bi} is equal to 1 if element i is chosen from sets V and S for block b, and zero otherwise.

The model presented below is a typical *max-min* type of formulation, and the variable ω represents the least number of standard deviations away from the mean that we can guarantee. We want to determine the schedule of clinics for B blocks of time that maximizes the minimum service level while meeting all constraints as explained below.

$$
\begin{aligned}
\text{CLCSP: Maximize} \quad & \omega && && (2.1)\\
s.t. \quad & && && \\
\sum_{d=1}^{D}\sum_{c=1}^{C} x_{dbc}\bar{t}_{dc} &= \mu_b\,; && b = 1,\ldots,\ B && (2.2)\\
\sum_{d=1}^{D}\sum_{c=1}^{C} x_{dbc}v_{dc} &= \sum_{i} y_{bi}\, V_i\,; && b = 1,\ldots,\ B && (2.3)\\
\mu_b + \sum_{i=1}^{u} \omega\, y_{bi}\, S_i &\leq T_b; && b = 1,\ldots,\ B && (2.4)\\
\sum_{c=1}^{C} x_{dbc} &\leq 1\,; && d = 1,\ldots,D;\ b = 1,\ldots,\ B && (2.5)\\
\sum_{b=1}^{B} x_{dbc} &\geq W_{dc}\,; && d = 1,\ldots,\ D; c = 1,\ldots,C && (2.6)\\
\sum_{c=1}^{C} x_{dbc} &\geq Z_{db}\,; && d = 1,\ldots,\ D; b = 1,\ldots,B && (2.7)\\
\sum_{d=1}^{D}\sum_{b=1}^{B} x_{dbc} &= R_c\,; && c = 1,\ldots,C && (2.8)\\
\sum_{i=1}^{u} y_{bi} &= 1 && b = 1,\ldots,\ B && (2.9)\\
x_{dbc},\ y_{bi} &\in \{0,1\} && && (2.10)\\
\omega &\leq \omega^{(max)} && && (2.11)
\end{aligned}
$$

Constraint (2.2) and constraint (2.3) calculate the mean and the variance of total testing time induced to the test room for each block b. Constraint (2.4) links the objective function and constraints. This constraint assures the deviation from the expected testing time required for each block b is less than or equal to the available testing time for that given block. Constraint (2.5) states that each doctor can run a maximum of one clinic type in each block. Constraint (2.6) guarantees that the number of times a doctor performs a particular clinic over all blocks is at least equal to the demand of the doctor for that particular clinic.

Constraint (2.7) ensures that doctor d works in block b if he/she has a preference to work in that block ($Z_{db} = 1$). If a doctor has a preference not to work in a block, we can add a constraint to our model to ensure that no clinic is assigned to the doctor in that particular block. The requirement for the total number of clinics of type c is guaranteed by constraint (2.8). Finally, constraint (2.9) says that each block must have a single mean and variance.

We set an upper bound to the value of ω to $\omega^{(max)}$ beyond which the problem is, for all practical matters, unconstrained. For example, a setting of $\omega^{(max)} = 4$ gives an upper bound of at most four times the standard deviation from the mean. Note that the assumption of normality of testing times makes the translation of ω into service level very easy, which would not be true if the lab time distribution was coming from a different distribution. Nonetheless, maximizing ω is equivalent to maximizing service level independent of the distribution it is coming from.

2.4.1 Linearization of the quadratic constraint

In the formulation above, constraint (2.4) is quadratic, where each term in the summation is a product of ω and y_{bi}, two decision variables. We therefore propose the following formulation that includes the linearization of the quadratic constraints, a standard approach for products of a continuous and a binary variable. We reformulate the problem exactly as a MIP by replacing constraint (2.4) with the following linearization constraints:

$$\begin{aligned}
\mu_b + \sum_{i=1}^{u} y'_{bi}\, S_i &\leq T_b; \qquad b = 1, \ldots, \ B && (2.12)\\
y'_{bi} &\leq \omega^{(max)} y_{bi} && (2.13)\\
y'_{bi} &\leq \omega && (2.14)\\
y'_{bi} &\geq \omega - \omega^{(max)}(1 - y_{bi}) && (2.15)\\
x_{dbc}, y_{bi} &\in \{0, 1\} && (2.16)\\
y'_{bi} &\geq 0 && (2.17)\\
\omega &\leq \omega^{(max)} && (2.18)
\end{aligned}$$

In constraint (2.12), $\omega\, y_{bi}$ is replaced by a non-negative variable y'_{bi}. If y_{bi} is zero, then by constraints (2.13) and (2.17) $y'_{bi} = 0$. If $y_{bi} = 1$, then by constraints (2.14) and (2.15) $y'_{bi} = \omega$. Hence this is an exact reformulation of the problem with linear constraints.

2.4.2 Problem complexity

CLCSP can be regarded as a generalization of the Multiple Subset Sum Problem (MSSP) which is defined as: given a set of items I, each item $i \in I$ with weight w_i, and a set of identical knapsacks K with capacity $\mathcal{C}$, determine items to be packed in each knapsack such that the sum of the item weights in every knapsack does not exceed the knapsack capacity c and the overall sum of the weights of the items packed is as large as possible.

Starting from the parameters given for an instance of problem MSSP, we assign values I to be the doctor-clinic combination (dc), setting $\bar{t}_{dc}$ to zero for all combinations of dc, and the length of each testing block $\mathcal{C}$. Remove all constraints related to doctors' preferences and apply an algorithm that solves the CLCSP. Use the returned solution as the solution to Problem MSSP, thus demonstrating that the CLCSP is as complex as the MSSP. Garey and Johnson (1979) have shown that the decision version of the MSSP is NP-complete in the strong sense. Clearly, allowing for stochastic weights for each item and heterogeneous sizes of knapsacks (blocks) only further complicates the problem.

Direct solution methods, such as the mixed-integer programming method, are able to solve small instances of the lab-clinics scheduling problem. For real-sized problems we propose and apply in our numerical tests a branch-and-price procedure which integrates column generation with branch-and-bound. We next focus on our proposed column generation approach, after which the branch-and-bound procedure on columns to obtain an integer solution will be explained.

2.5 A column-generation based approach

To deal with the complexity of the problem and solve practical examples, a Column-Generation (CG) based algorithm will be proposed. The column generation approach was first introduced by Dantzig and Wolfe (1960). CG-based algorithms have often been developed for set-partitioning constraints. Therefore, the mixed-integer programming model constructed in section 2.4 will be first reformulated as a set-partitioning problem.

2.5.1 Suitable model for Column Generation (set-partitioning model)

Let us define a column j as an assignment of doctors to clinic slots. Furthermore, a feasible column must respect the constraint that each doctor can run a maximum of one clinic type. We also define a binary decision variable x_j to be 1 if column j is accepted, and zero otherwise. The rest of the notation for the set-partitioning model is summarized in Table. 2.3.

Table 2.3: Notation for the set-partitioning model

$E(T_j)$	the mean of total testing time induced by column j
σ_j	the standard deviation of total testing time induced by column j
Ξ	Set of all feasible columns
p_{bj}	1 if plan j is scheduled in block b, 0 otherwise
e_{dcj}	1 if doctor d is assigned to clinic c by column j, 0 otherwise

The set-partitioning model of the coordinated lab-clinics scheduling problem is presented below:

I) The master problem

$$\text{Maximize} \quad \omega \tag{2.19}$$

$$s.t.$$

$$\sum_{j \in \Xi} e_{dcj}\, x_j \geq W_{dc}\,; \qquad d = 1, \ldots, D; c = 1, \ldots, C \tag{2.20}$$

$$\sum_{j \in \Xi} x_j\, p_{bj}\, E(T_j) = \mu_b\,; \qquad b = 1, \ldots, B \tag{2.21}$$

$$\mu_b + \sum_{j \in \Xi} \omega\, x_j\, p_{bj}\, \sigma_j \leq T_b; \qquad b = 1, \ldots, B \tag{2.22}$$

$$\sum_{j \in \Xi} x_j\, p_{bj} = 1\,; \qquad b = 1, \ldots, B \tag{2.23}$$

$$\sum_{j \in \Xi} \sum_{d=1}^{D} e_{dcj}\, x_j = R_c\,; \qquad c = 1, \ldots, C \tag{2.24}$$

$$x_j \in \{0, 1\} \tag{2.25}$$

The problem is again maximizing the service level given the problem constraints, but it is now expressed in terms of the new x_j decision variable. Constraint (2.20) ensures that the number of times a doctor performs a particular clinic over all blocks is at least equal to the demand of the doctor for that particular clinic. Constraint (2.21) calculates the mean of the total testing time induced in block b. Constraint (2.22) assures that the deviation from the expected testing time required for each block b is less than or equal to the available testing time for that given block. Constraint (2.23) ensures that only one plan is assigned to each block b. Finally, the requirement for the total number of clinics of type c is guaranteed by constraint (2.24).

Note that constraint (2.22) is again quadratic, because each term in the summation is a product of ω and x_j. We now apply the same approach proposed in the previous section to linearize the quadratic constraints.

$$\text{Maximize} \quad \omega \tag{2.26}$$

$$s.t.$$

$$\sum_{j \in \Xi} e_{dcj}\, x_j \geq W_{dc}\,; \qquad d = 1, \ldots, D; c = 1, \ldots, C \tag{2.27}$$

$$\sum_{j \in \Xi} x_j\, p_{bj}\, E(T_j) = \mu_b\,; \qquad b = 1, \ldots, B \tag{2.28}$$

$$\mu_b + \sum_{j \in \Xi} x'_j\, p_{bj}\, \sigma_j \leq T_b; \qquad b = 1, \ldots, B \tag{2.29}$$

$$\sum_{j \in \Xi} p_{bj}\, (x'_j - \omega^{(max)} x_j) \leq 0; \qquad b = 1, \ldots, B \tag{2.30}$$

$$\sum_{j \in \Xi} p_{bj}\, x'_j - \omega \leq 0; \qquad b = 1, \ldots, B \tag{2.31}$$

$$\omega + \sum_{j \in \Xi} p_{bj}\, (\omega^{(max)} x_j - x'_j) \leq \omega^{(max)}; \qquad b = 1, \ldots, B \tag{2.32}$$

$$\sum_{j \in \Xi} x_j\, p_{bj} = 1\,; \qquad b = 1, \ldots, B \tag{2.33}$$

$$\sum_{j \in \Xi} \sum_{d=1}^{D} e_{dcj}\, x_j = R_c\,; \qquad c = 1, \ldots, C \tag{2.34}$$

$$\omega \leq \omega^{(max)} \tag{2.35}$$

$$x'_j \geq 0 \tag{2.36}$$

$$x_j \in \{0,1\} \tag{2.37}$$

In our column generation approach, we make use of the linearization of the master problem presented above.

II) The subproblem

In order to ensure the feasibility of each column in the master problem, the following constraints must be respected:

$$\sum_{c=1}^{C} e_{dcj} \leq 1; \qquad d = 1,\ldots, D; j \in \Xi \tag{2.38}$$

$$\sum_{c=1}^{C} p_{bj}\, e_{dcj} \geq Z_{db}; \qquad d = 1,\ldots, D; b = 1,\ldots, B; j \in \Xi \tag{2.39}$$

$$\sum_{b=1}^{B} p_{bj} = 1; \qquad j \in \Xi \tag{2.40}$$

Constraint (2.38) states that each doctor can run a maximum of one clinic type in each column. Constraint (2.39) ensures that each column j assigned to block b respects the working time preferences for that given block. Finally, constraint (2.40) ensures that each column generated in each iteration is assigned to only one block.

2.5.2 Initial restricted master problem

In the above, we have presented the set partitioning formulation. The column generation method starts with an initial feasible solution to the LP relaxation of the restricted master problem that is only a subset of the columns. Note that the LP relaxation of the restricted master problem can be obtained by replacing the following constraint with constraint (2.37):

$$0 \leq x_j \leq 1 \tag{2.41}$$

Selection of an appropriate initial solution has an important impact on the efficiency of the column generation procedure. We propose a binary integer programming model to find a set of initial columns for the LP relaxation of the restricted master problem.

$$\sum_{c=1}^{C} x_{dbc} \leq 1; \qquad d = 1, \ldots, D;\ b = 1, \ldots,\ B \tag{2.42}$$

$$\sum_{b=1}^{B} x_{dbc} \geq W_{dc}; \qquad d = 1, \ldots,\ D; c = 1, \ldots, C \tag{2.43}$$

$$\sum_{c=1}^{C} x_{dbc} \geq Z_{db}; \qquad d = 1, \ldots,\ D; b = 1, \ldots, B \tag{2.44}$$

$$x_{dbc} \in \{0, 1\} \tag{2.45}$$

The solution can be regarded as a set of initial feasible plans that respects the demand of each doctor for every clinic type as well as their working time preferences.

2.5.3 Pricing problem

We now solve the restricted master problem with the initial columns to obtain the objective value ω and dual multipliers π. Given an optimal solution to the restricted master problem, dual variables $(\pi^1_{dc}, \pi^2_b, \pi^3_b, \pi^4_b, \pi^5_b, \pi^6_b, \pi^7_b, \pi^8_c)$ are associated with constraints (2.27) to (2.34) in the respective order. Then, the reduced cost of a new column j is given by:

$$\bar{C_j} = C_j - Z_j =$$

$$0 + \sum_{d=1}^{D}\sum_{c=1}^{C} \pi^1_{dc}\, e_{dcj} - \sum_{b=1}^{B} \pi^2_b\, p_{bj}\, E(T_j) - \sum_{b=1}^{B} \pi^3_b\, p_{bj}\, \sigma_j - \sum_{b=1}^{B} \pi^t_b\, p_{bj} - \sum_{c=1}^{C} \pi^8_c \sum_{d=1}^{D} e_{dcj} \tag{2.46}$$

where $\pi^t_b = \sum_{i=4}^{7} \pi^i_b; \quad b = 1, \ldots,\ \ B$, which is the sum of dual prices of restrictions (2.30) to (2.33).

Since only a subset of the columns is available and the enumeration of all feasible columns and calculation of $C_j - Z_j$ is a time consuming task, this cannot be checked explicitly. Thus, we may instead solve the following pricing problem to generate a new column with positive reduced cost.

$$\text{Maximize} \quad \bar{C}_j \tag{2.47}$$

$$s.t.$$

$$\sum_{c=1}^{C} e_{dcj} \leq 1; \qquad d = 1, \ldots, D; j \in \Xi \tag{2.48}$$

$$\sum_{c=1}^{C} p_{bj}\, e_{dcj} \geq Z_{db}; \qquad d = 1, \ldots, D; b = 1, \ldots, B; j \in \Xi \tag{2.49}$$

$$\sum_{b=1}^{B} p_{bj} = 1; \qquad j \in \Xi \tag{2.50}$$

Note that the pricing problem aims to determine values of e_{dcj}, p_{bj}, $E(T_j)$, and σ_j for one column. However, constraint (2.49) is quadratic, because each term in the summation is a product of variables e_{dcj} and p_{bj}. To linearize this constraint, we decompose the above pricing problem into B subproblems, each one corresponds to one block. For example, pricing problem 1 is related to block 1 implying that the new generated column is assigned to block 1, that is $p_{1j} = 1$. As an example, the pricing problem 1 is formulated below.

$$\text{Maximize} \quad \bar{C}_j \tag{2.51}$$

$$s.t.$$

$$\sum_{c=1}^{C} e_{dcj} \leq 1; \qquad d = 1, \ldots, D; j \in \Xi \tag{2.52}$$

$$\sum_{c=1}^{C} p_{1j}\, e_{dcj} \geq Z_{d1}; \qquad d = 1, \ldots, D; j \in \Xi \tag{2.53}$$

The set of pricing problems can now be solved separately for each block and the column with highest reduced cost enters the master problem. In other words, each pricing problem generates the best feasible column for each block b. Now the goal is to select the column with the highest reduced cost over all blocks and add the new column to the restricted master problem. The column generation algorithm terminates when no more columns with a positive reduced cost can be found.

2.5.4 Branching

If the values of all decision variables are integral, the optimal solution is obtained for the master problem. If the optimal solution satisfies all the constraints except the binary one, it can be regarded as an upper bound for the master problem. In this case, branching on the decision variable x_j has to be established. Branching on column variables can be achieved in our problem because of solving the pricing problem for each block individually and including constraint (2.33) in the master problem. To speed-up the process, we can stop branching when the gap between the upper and lower bounds becomes smaller than a certain threshold. The lower bound is obtained by taking the maximum of the optimal objective values of all of the current leaf nodes. This speed-up technique can greatly reduce the computations required.

One can now define the branch-and-price algorithm as follows:

The branch-and-price algorithm

Step 0: Generate a feasible set of columns by solving the proposed binary integer programming model (Constraints (2.42) to (2.45)).

If no feasible solution is found, the branch-and-price procedure is ended because no feasible weekly schedule can be obtained.

Step 1: Solve the LP relaxation of the restricted master problem with the initial columns to obtain the objective value ω and dual multipliers π.

Step 2: Solve the pricing problem suggested in section 2.5.3 to find a column with positive reduce cost. If such a column is found, then add the column to the restricted master problem and return to **Step 1**. Otherwise, go to **Step 3**.

Step 3: Check for the integrality constraint. If one is violated, then perform the branching operation suggested in section 2.5.4. Create two new branches on the largest fractional column. Add the branching constraints to the restricted master problem and return to **Step 1**.

2.5.5 Computational experiments

In this section, we test the computation time of the column generation based method through randomly generated experiments, comparing the branch-and-price method with the direct approach using a MILP solver. All experiments were performed on a Lenovo Y50 Laptop (Intel Core i7 / 8GB RAM/ Windows 10). The branch-and-price algorithm was written in Eclipse Java Mars 4.5. The optimization software used in branch-and-price and MILP is CPLEX optimization studio 12.6.3.

For the randomly generated experiments, we solved 4 different combinations of doctors, blocks and clinics, and for each combination we generated 25 random instances, for a total of 100 instances solved. Table 2.4 shows the combinations of doctors, blocks, and clinics in the first column. The second column presents in two sub-columns the service level and CPU time for the branch-and-price approach. The third column presents similar sub-columns for the MILP optimization. The last column presents the average optimality gap. We note that the average optimality gap is less than 0.003, and the results do not suggest any particular trend of the gap as a function of the size of the problem.

Size of problem	**Branch-and-Price**		**MILP**		**Average gap***
	Service level %	**Time**	**Service level %**	**Time**	
$2 \times 5 \times 2$	99.97	less than 1s	99.98	2s	0.0001
$3 \times 5 \times 2$	99.13	less than 1s	99.4	14s	0.0027
$4 \times 6 \times 3$	99.64	less than 1s	99.74	53s	0.001
$5 \times 7 \times 4$	99.93	less than 2s	99.96	1220s	0.0003

*Average gap=$\sum$ Optimality gap/n, where optimality gap is the difference between the MILP and the branch-and-price results.

Table 2.4: Comparison of the two approaches

If only a minimum service level suggested by the clinic administrator needs to be satisfied, we can add a stopping criteria to stop the proposed algorithm when there is a feasible schedule with higher service level than the suggested one. This probably could reduce the computations required. Note that there is, however, a possibility of finding no feasible schedule if the suggested service level is greater than the optimal service level.

2.6 Case study: two respirology lab-clinics facilities

2.6.1 Mini-Case I: Hospital A

Our first study is based on data gathered from the operations of coordinated respirology lab-clinics at a hospital that is part of the AHS system. Respirology clinics are scheduled for a half-day (3 to 4 hours) and separately for each physician. There are currently nine physicians running three types of clinics. Table 2.5 gives the demand of each doctor for each clinic type (W_{dc}), as well as the mean and standard deviation of testing time induced by doctor d in clinic type c, that is ($\bar{t}_{dc}$,$\sqrt{v_{dc}}$). Clinics 1, 2, and 3 represent the general respirology, pulmonary hypertension, and neuromuscular clinics, respectively. Each doctor runs at most two types of clinics in all blocks. The maximum theoretical available pulmonary function testing time remains the same Monday through Friday for all blocks and is nine hours.

d \ c	1	2	3
1	4, (95, 32.26)	-	-
2	2, (114, 30.83)	-	-
3	1, (135, 32.06)	-	-
4	2, (142, 35.29)	1, (151, 38.27)	-
5	1, (112, 29.17)	2, (124, 31.08)	-
6	2, (126, 30.82)	-	1, (146, 33.12)
7	1, (117, 35.39)	-	1, (146, 33.12)
8	1, (140, 32.35)	2, (134, 32.64)	-
9	3, (119, 30.46)	-	-

Table 2.5: The demand of each doctor for each clinic type W_{dc}, as well as the mean and standard deviation of testing time ($\bar{t}_{dc}$,$\sqrt{v_{dc}}$)

Currently the clinic schedule at Hospital A is as shown in Table 2.6. Given that there is a constant nine hours of lab time available on each half-day, the current clinic schedule might be expected to be somewhat more evenly distributed across the week. However, physicians have other professional commitments that limit their availability to run clinics at the hospital. The current schedule may thus reflect physicians' preferences for which half-days to run their clinics, although they may be open to schedule changes. The current service level and expected utilization of the test room in each block are also shown in Table 2.6.

b c	Mon Morn	Mon Aft	Tue Morn	Tue Aft	Wed Morn	Wed Aft	Thur Morn	Thur Aft	Fri Morn	Fri Aft
Clinic 1	d1,d2	d2,**d3**,d4	d6			d1,d5	d6,d7,d9	d9	d1,d4,d8	d1,d9
Clinic 2		**d5**			d4,d5,d8	**d8**				
Clinic 3				d6,**d7**						
Objective value(ω)	4	0.38	4	2.02	2.20	1.41	4	4	4	4
Service level %	100	64.8	100	97.83	98.61	92.07	100	100	100	100
Expected utilization %	38.70	95.37	23.33	54.07	75.74	63.14	67.03	22.03	70.18	39.62

Table 2.6: Current schedule at hospital A

We are interested in finding a schedule of the clinics that maximizes the minimum service level over all blocks given the available laboratory testing time for each block. We will compute the solution without considering doctors' preferences. This gives us the improvement possible if doctors' schedules were completely flexible. Since the problem could not be solved using a MILP solver directly, we have used the column-generation based approach discussed previously. The current schedule implemented at the hospital sets a lower bound on the optimal solution. Service level is measured by its proxy: the minimum number of standard deviations above the mean testing time. Table 2.7 summarizes the optimal solution with the clinic scheduling for each doctor. The differences between the current and the optimal schedules are shown in bold in Tables 2.6 and 2.7.

b c	Mon Morn	Mon Aft	Tue Morn	Tue Aft	Wed Morn	Wed Aft	Thur Morn	Thur Aft	Fri Morn	Fri Aft
Clinic 1	d1,d2	d2,d4	d6	**d3**		d1,d5	d6,d7,d9	d9	d1,d4,d8	d1,d9
Clinic 2			**d5**		d4,d5,d8					**d8**
Clinic 3				d6				**d7**		
Objective value(ω)	4	4	4	4	2.20	4	4	4	4	4
Service level %	100	100	100	100	98.61	100	100	100	100	100
Expected utilization %	38.70	47.40	46.29	52.03	75.74	38.33	67.03	49.07	70.18	64.44

Table 2.7: Optimal solution with the clinic scheduling for each doctor

Solving the problem with the current schedule at Hospital A shown in Table 2.6 gives a service level of 0.38, or 64.8%. However, the optimal schedule can dramatically improve the service level. The optimal solution shown in Table 2.7 leads to a service level of 2.20, or 98.61%. Table 2.7 also shows expected utilization of the test room in each block. A comparison of the current schedule with the optimal schedule reveals that only four changes of clinic times, involving doctors 3, 5, 7, and 8, would be required. In reality, it may be difficult for hospital administrators to persuade all four physicians to move a clinic. However, even if fewer than four physicians agree to do so, there will be marked improvement in the achieved service level. For example, moving doctor 5's clinic from Monday afternoon to Tuesday morning will by itself increase the service level to 92.07%. Figure 2.2 shows the improvement possible due to moving one, two, three, or four of these clinics.

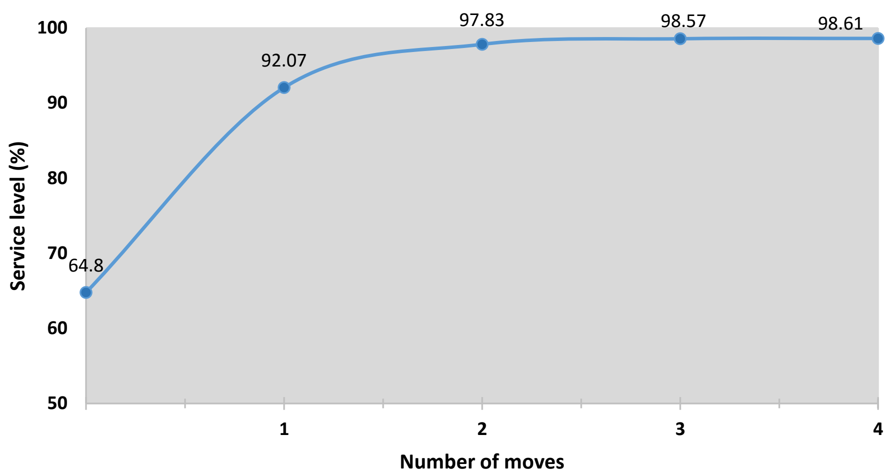

Figure 2.2: Possible improvements due to shifting one, two, three, or four clinics

Figure 2.3 shows the test room expected utilization of the current and optimal schedules in each block. The average utilization over all blocks is 54.92% for both schedules. The figure shows that the variability of the utilization rates is lower for the optimal schedule.

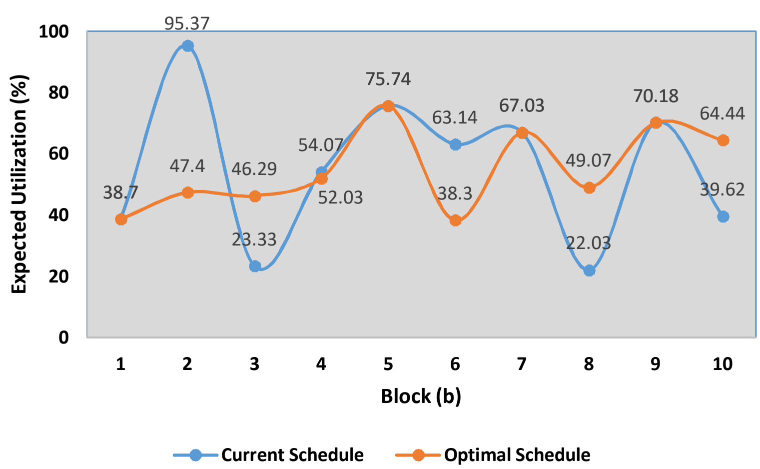

Figure 2.3: Expected utilization rates of the test room

The exercise of optimizing service level is that of reducing utilizations' departure from the mean utilization. A side effect of improving service level is the homogenization of doctor/clinics' workload.

2.6.2 Mini-Case II: Hospital B

Here we use data gathered from the operations of the coordinated respirology lab-clinics at another AHS hospital. We perform sensitivity analysis on the fraction of the theoretical time a lab is available (α), as will be explained below.

Again, the respirology clinics need to be scheduled for a half-day and separately for each physician, with up to nine clinics held on the same half-day (one block of time). Three types of clinics are held. Table 2.8 gives the demand of each doctor for each clinic type (W_{dc}), as well as the mean and standard deviation of testing time induced by doctor d in clinic type c, that is ($\bar{t}_{dc}$,$\sqrt{v_{dc}}$). Each doctor runs at most two types of clinics in all blocks, where clinics 1, 2, and 3 represent the oncology, general respirology, and interstitial lung diseases clinics, respectively.

d \ c	1	2	3
1	2, (122, 30.51)	-	-
2	3, (138, 32.38)	2, (130, 38.50)	-
3	1, (122, 30.51)	2, (86, 29.10)	1, (162, 31.11)
4	-	5, (118, 30.28)	-
5	-	4, (132, 34.49)	1, (102, 31.16)
6	-	1, (122, 30.51)	-
7	-	1, (118, 30.28)	-
8	2, (100, 28.30)	-	5, (158, 38.19)
9	1, (88, 30.09)	2, (127, 36.94)	-

Table 2.8: The demand of each doctor for each clinic type W_{dc}, as well as the mean and standard deviation of testing time ($\bar{t}_{dc}$,$\sqrt{v_{dc}}$)

Table 2.9 shows the maximum theoretical available testing time of each test room in every block b, T_b. There are currently 4 test rooms working at the same time at Hospital B. We define α as the fraction of the theoretical available testing time that is usable testing time. That is, α times the theoretical time available gives the potential actual time availability.

b	Mon Morn	Mon Aft	Tue Morn	Tue Aft	Wed Morn	Wed Aft	Thur Morn	Thur Aft	Fri Morn	Fri Aft
	3.5	4.5	3.5	4.5	3	4	3.5	4.5	3.5	4.5

Table 2.9: The maximum theoretical available testing time (hours) in block b (T_b)

Currently the clinic schedule at Hospital B is as shown in Table 2.10.

b c	Mon Morn	Mon Aft	Tue Morn	Tue Aft	Wed Morn	Wed Aft	Thur Morn	Thur Aft	Fri Morn	Fri Aft
Clinic 1		d1	d2	d2	d1	d3	d2	d9	d8	d8
Clinic 2	d2,d4	d2,d4	d3,d4	d6,d7	d5,d9	d5,d9	d3,d4	d4	d5	d5
Clinic 3		d3	d8	d8	d8	d8	d8	d5		
Expected utilization %	29.52	49.35	59.52	49.62	74.86	56.14	59.52	28.51	27.62	21.48

Table 2.10: Current schedule at hospital B

The proposed branch-and-price method is applied to solve the problem with the maximum theoretical available testing time ($\alpha = 1$). Table 2.11 summarizes the optimal solution when all theoretical available testing time can be used.

b c	Mon Morn	Mon Aft	Tue Morn	Tue Aft	Wed Morn	Wed Aft	Thur Morn	Thur Aft	Fri Morn	Fri Aft
Clinic 1		d1	d2		d1	d3	d2	d9	d8	d2,d8
Clinic 2	d2, d4,d5	d2, d4	d3, d4	d6, d7	d9	d5	d9,d3	d4	d4, d5	d5
Clinic 3		d3	d8	d8	d8	d8	d8	d5		
Expected utilization %	45.23	49.35	59.52	36.85	56.52	42.39	60.59	28.51	41.66	34.25

Table 2.11: Optimal solution ($\alpha = 1.0$)

The solution leads to a doctor scheduling scheme that guarantees a service level of 3.65 over all blocks. Solving the problem with the current schedule shown in Table 2.10 gives the service level of 2.56. For all practical purposes both solutions provide a service level near 100%. However, the fraction of the theoretical available testing time may be substantially lower than 1.0 due to factors such as setup times between patients, patients who have poor mobility or need to repeat a test procedure, language comprehension difficulties, etc.

Figure 2.4 shows the service levels of the current and the optimal schedules for values of $0.5 \leq \alpha \leq 1.0$ (in steps of 0.05). The service level for each α is obtained by using the corresponding ω from a standardized normal distribution. In Figure 2.4 it can be seen that the performance of the optimal schedule is quite robust, but relatively small departures from the full theoretical amount of time available for testing leads to poor service levels with the current schedule that can be dramatic for values of $\alpha \leq 90\%$.

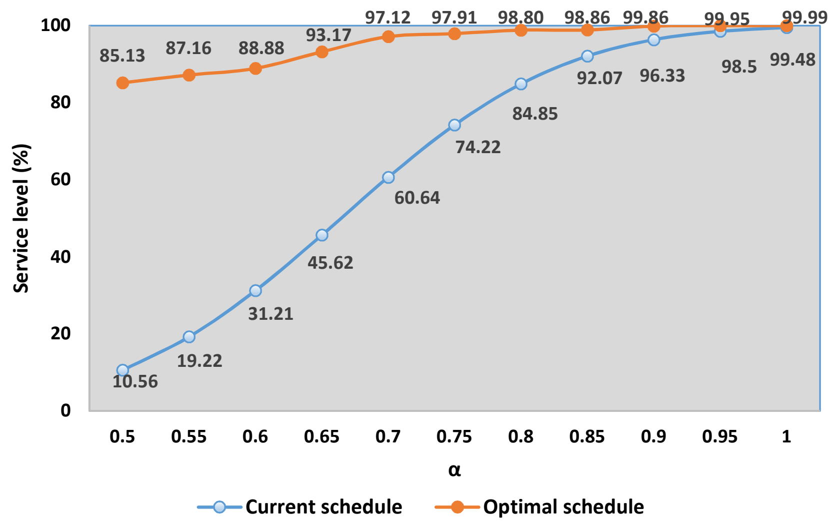

Figure 2.4: Sensitivity analysis on the fraction of theoretical available testing time.

A potential insight from this case is that lab capacity may be unnecessarily high to cope with the effects of the lack of a systematic approach to optimize the scheduling of doctors and clinics. Figure 2.5 shows the service level that can be achieved with a smaller amount of testing time. Even if testing time is reduced by 70%, a service level of at least 97% can still be achieved.

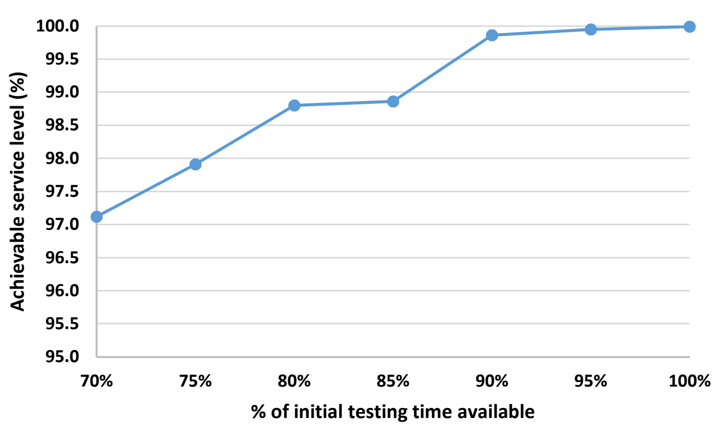

Figure 2.5: Required testing time in the lab.

Although this study addresses a tactical level problem, clinic administrators may use the results obtained in this phase to decide on the lab capacity at the strategic level.

2.7 Conclusion and future research

In this chapter, we have formulated Problem CLCSP and provided a MILP formulation. The optimization version of the problem was shown to be NP-hard, and a branch-and-price approach was proposed for solving larger instances. We then applied our approach to two real situations.

Our two case studies have implications for the implementation of our assignment approach by facility managers in practice. Since the scheduling method allows clinic management to demonstrate to physicians the improvements in testing performance that would result if changes in the clinic schedule were made, it is a management tool that has the potential to persuade physicians to improve the coordination of their clinics. The first case study suggests that in some situations the performance of the current schedule can be greatly improved with only minimal changes. This is extremely important, since doctors are often constrained by a schedule of teaching, research, or service requirements in addition to their clinical activities (and their personal lives). Hence in many settings it is difficult for most doctors to be flexible concerning the timing of their clinics. If moving only one or two clinics, out of many, can be shown to achieve a significantly higher standard of service for patients seen within a shared facility, there is both a strong argument and a strong incentive for doctors to collaborate to find a way to make those changes. In the other case study, we have shown that an optimal schedule can robustly preserve a very high service level when lab capacity is decreased. Managers who are looking for ways to reduce spending without damaging either the patient experience, the quality of care delivered, or the quality of lab staff work-life can achieve greater efficiencies while reducing pressure on lab staff and providing greater job satisfaction for doctors who rely upon timely lab results for better diagnosis and treatment of their patients. Since the model proposed in this research only covers the tactical phase of the coordination problem, further studies can address the problem of patient sequencing in the test center at the operational level.

Chapter 3

Physician scheduling to improve emergency department patient flow

3.1 Introduction

Emergency Department (ED) crowding has become a serious and growing international crisis that may impact the quality and access of health care (Hoot and Aronsky (2008)). Crowding in EDs is associated with long wait times, which result in decreased quality of care, decreased patient satisfaction, and higher number of patients who leave without being seen (LWBS). Patients who go to EDs need medical examination by a physician, and a faster diagnosis can result in better treatment outcomes. Evidence shows that patients who wait for an average of 30 minutes or less tend to be satisfied (McKinnon et al. (1998)). Shaikh et al. (2012) found that 50% of patients are willing to wait up to 2 hours for a care provider before leaving the ED without being seen. A study at a public hospital in Los Angeles county showed that 46% of the patients who left the ED without visiting a physician needed immediate evaluation, and 11% was hospitalized during the subsequent week (Stock et al. (1994)).

ED wait times can get reduced when sufficient numbers of resources are available to meet the demand for emergency care. In this chapter, we focus on the physicians who assess and treat the patients at EDs since they predominantly determine the productivity of an ED. We define productivity as the speed at which new patients are seen by physicians. In particular, we study the assignment of individual physicians to the shifts of a given schedule. Traditionally, physician schedules in EDs are built around physician preferences, regulatory constraints and work constraints. One important aspect that is ignored in the literature is that the productivity of individual physicians is different. Consequently, the assignment of physicians to shifts has a direct impact on the waiting time of patients to be assessed by a physician. We take this aspect into consideration when solving the physician scheduling problem.

To illustrate the relevance of our work, let us consider a case study for one of the four EDs in the city of Calgary (AB), Canada. Canada does not perform very well when it comes to ED wait times. In fact, it ranked last among 11 OECD countries in 2013, with 26% of patients who waited over four hours (Health at a Glance 2013: OECD Indicators). For our case study, we have data available over a two-year time period from August 2013 to July 2015. The hourly trend of patient wait times for each day of the week is presented in Figure 3.1. Each blue dot represents the actual wait time of an ED patient for the initial assessment by a physician, while an orange dot represents the average patient wait time during a given hour of the day. The hourly trend shows a 24-hour cycle with a peak after midnight that decreases toward the morning and rises again in the afternoon. It is interesting to note that there is no significant difference in the pattern of wait times during weekdays and the weekend.

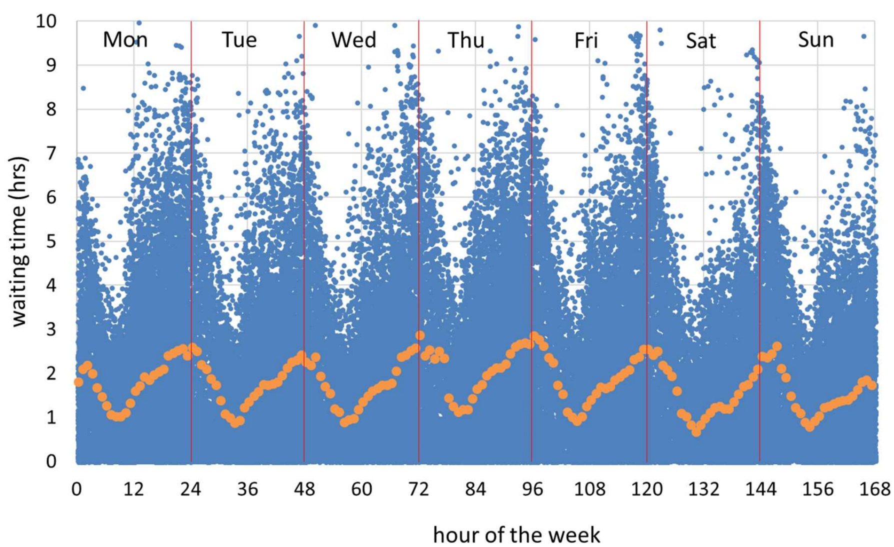

Figure 3.1: ED patient wait times (hourly trend)

The amount of time a patient waits to be seen by a physician varies significantly over different time periods of the day. Therefore, we have a closer look at the hourly demand for emergency care as well as the number of physicians scheduled per hour of the day (there are 13 shifts scheduled every day of the week). Similar to the waiting time, we do not see any significant differences between the number of arrivals over the days in a week (see Figure 3.2).

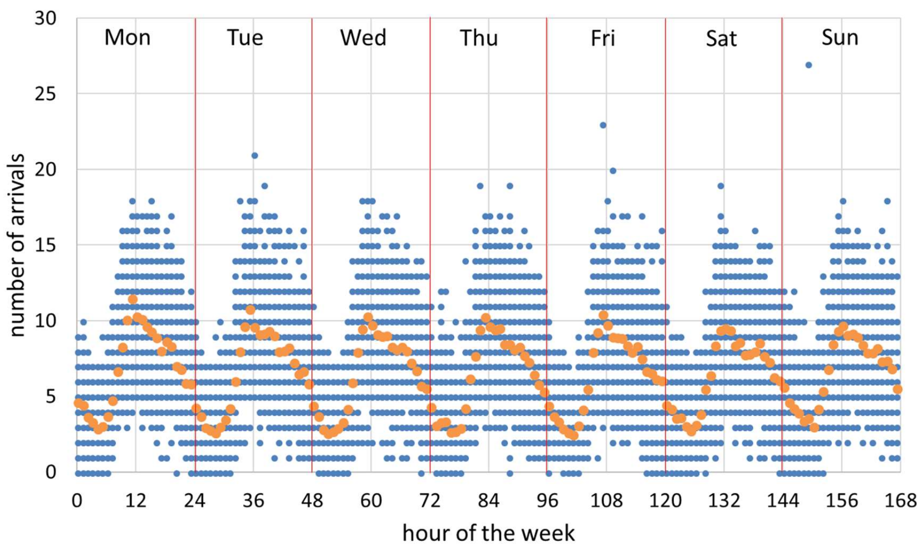

Figure 3.2: ED arrivals (hourly trend)

When looking at the number of physicians scheduled at each hour, there are always 13 shifts scheduled in a day with the same starting and ending hour. Consequently, we only report the average arrival rate per hour over all days in Table 3.1 as well as the number of physicians working at the ED for each hour of a day. Furthermore, this table includes the ratio of these two numbers since this indicates the average demand to assess new patients for each emergency physician.

	nr physicians working	nr arrivals/hr	New Patient:Physician Ratio
12AM-1AM	4	4.62	1.16
1AM-2AM	3	3.95	1.32
2AM-3AM	3	3.46	1.15
3AM-4AM	2	3.23	1.61
4AM-5AM	2	2.88	1.44
5AM-6AM	2	2.95	1.48
6AM-7AM	2	3.27	1.64
7AM-8AM	2	4.25	2.12
8AM-9AM	3	5.92	1.97
9AM-10AM	3	7.61	2.54
10AM-11AM	5	9.28	1.86
11AM-12PM	5	10.31	2.06
12PM-1PM	6	9.78	1.63
1PM-2PM	5	9.35	1.87
2PM-3PM	5	9.15	1.83
3PM-4PM	5	8.98	1.80
4PM-5PM	5	8.51	1.70
5PM-6PM	5	8.02	1.60
6PM-7PM	5	8.25	1.65
7PM-8PM	5	8.12	1.62
8PM-9PM	5	7.25	1.45
9PM-10PM	4	6.85	1.71
10PM-11PM	4	6.22	1.56
11PM-12AM	4	5.78	1.45

Table 3.1: Average arrivals per hour vs. number of physicians working in each hour

Even though the number of physicians scheduled to work in a particular hour of the day is always the same, the productivity of the ED is not. This is because the productivity of the ED equals the sum of the productivities of the physicians who are scheduled to work during a particular hour. Figure 3.3 shows the productivity for all physicians who have worked at least 90 shifts over the two-year time period.

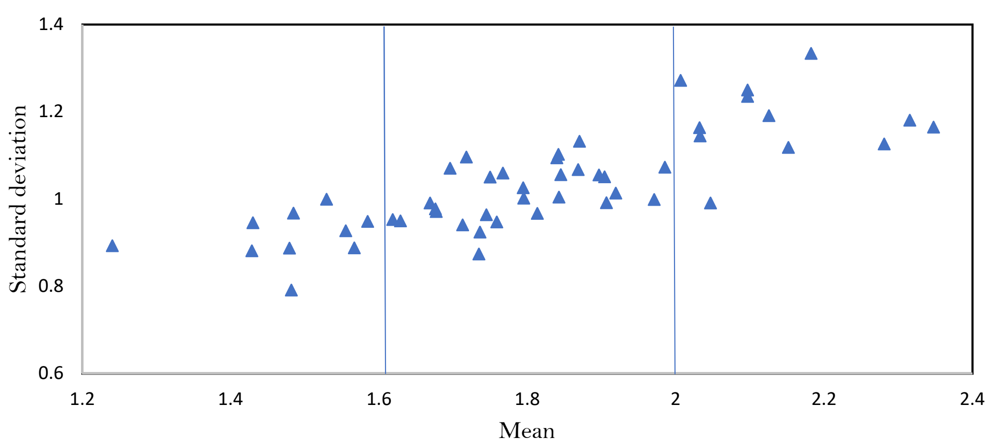

Figure 3.3: ED physicians' productivity (means and standard deviations)

When the heterogeneity among physicians is not taken in consideration in the physician scheduling problem, this can have serious consequences for the observed wait times by patients who need emergency care. Besides including the heterogeneity among physicians, we also include the stochastic nature of the ED patient arrivals and the physicians' productivity. In section 3.2, we review the related literature. Section 3.3 discusses the physician productivity in our case study. In section 3.4, we introduce our notations and develop a stochastic mixed-integer programming (MIP) formulation of the problem. In section 3.5, we first propose an L-Shaped heuristic to solve the physician scheduling problem, followed by a bounding technique to assess the solution quality. In section 3.6, we use the data from an emergency department within the Alberta Health Services (AHS) for our numerical study. Finally, we conclude in section 3.7 with a discussion and further research suggestions.

3.2 Review of related literature

There exists a substantial body of literature on the scheduling problem of ED physicians. To the best of our knowledge, no one has studied the heterogeneity among physicians to meet the time-varying patient demand. Most models in the literature focus on the problem formulation with the inclusion of certain constraints (related to physician preferences, regulatory constraints, work constraints, etc.) and the authors propose different solution techniques for their models. For example, Carter and Lapierre (2001) presented the characteristics of the ED physician scheduling problem based on six hospitals in Montreal, Canada. Then, they developed a mathematical programming model to construct the schedule of ED physicians. The objective function was to find a feasible solution or to minimize the cost of violating constraints if no feasible solution found. Beaulieu et al. (2000) proposed a mathematical programming approach to construct a six-month schedule of physicians in EDs. They partitioned all constraints into the four following categories: compulsory constraints, ergonomic constraints, distribution constraints, and goal constraints. The objective function was to minimize the overall deviations from the four types of constraints. Gendreau et al. (2007) first proposed a series of generic constraints to describe physician scheduling problem in EDs and then discussed four solution techniques for this problem, namely tabu search, column generation, mathematical programming, and constraint programming. Ferrand et al. (2011) devised a cyclic schedule for emergency department physicians using an integer programming approach. The main contribution of this study was formulating specific constraints associated with holidays, work assignments, and vacation requests. They also surveyed physicians after implementation of the new schedule to assess the quality of their suggested schedule. They reported that the new schedule provided predictability and well-balanced work patterns. Cabrera et al. (2001) applied an agent-based simulation to find the optimal ED staff configuration, i.e. doctors, triage nurses, and admissions personnel. Patient wait time and patient throughput were proposed and tested as two performance metrics. Brunner et al. (2009) proposed a mathematical model to address the shift scheduling problem in EDs where shifts can start at any pre-defined period in the planning horizon. The objective function of the proposed model was to cover the demand with regular hours first, then with overtime, and finally with outside hours when the first two options are exhausted.

Another stream of literature that is related to our study addresses the workforce scheduling problem with heterogeneous workers. Workforce scheduling is a common problem which has to be performed in any service or product based company. The problem is to schedule a number of employees to a set of tasks or work shifts while taking into account employees' skills and preferences, demand needs, and other applicable requirements. In the case of heterogeneous workers, the processing time of a task is dependent to the worker assigned to the task or each employee can perform only some of the tasks. Recently, there have been more studies in the field of operations research that acknowledge a heterogeneous workforce to narrow the gap between research and practice. For example, Benavides et al. (2014) propose a scatter search-based heuristic to solve the flow shop scheduling problem with heterogeneous workers, where the processing time of an operation depends on the assigned worker, e.g., an established agent processes a task (on average) faster than a new agent. Stolletz and Zamorano (2014) develop a rolling planning horizon-based heuristic to solve a tour scheduling problem for agents with multiple skills who work at check-in counters in airports. Agents with lower skill levels cannot be assigned to cover higher skill requirements, e.g., only experienced agents are allowed to operate different computer systems at check-in counters. Maenhout and Vanhoucke (2013) studied an integrated nurse staffing and shift scheduling problem, where each nurse has different skill categories. Hojati and Patil (2011) studied the scheduling of heterogeneous, part-time employees of a service organization. Each employee is only available during certain hours of certain days, and have skills to perform only specific tasks.

To the best of our knowledge, the workforce scheduling problem that is most relevant to our problem is the construction of multi-skilled staff schedules in call centers. The heterogeneous workforce scheduling problem in a call center aims to find the schedule of agents to match a time-varying customer demand while keeping cost under control. There is a vast body of literature on personnel scheduling in call centers. However, most studies assume that agents are homogeneous with a common service rate. For example, Excoffier et al. (2016) applied a joint chance-constrained program to the workforce scheduling problem in a single-class, single-skill call center with uncertain call arrivals. Robbins and Harrison (2010) developed a mixed-integer stochastic program for scheduling call center agents to meet a global service level constraint while the arrival rate is uncertain. Mehrotra et al. (2010) proposed an intra-day schedule updating methodology to manage the trade-off between labor costs and service levels. The underlying business context was a call center, where all agents have the skills to handle only a single type of call. However, several studies address the heterogeneity among agents in call centers. For example, Ibrahim et al. (2016) conducted an empirical study of service times in a large call center with heterogeneous agents and multiple call types. They report that, for a given call type, the service-time distribution depends on the individual agent. Furthermore, they found that the average service time for a given agent and a given call type varies significantly over time. Mehrotra et al. (2012) examined several routing strategies in a call center for determining which calls should be handled by which agents, where each agent has different performance metrics in terms of average call handling time and call resolution. Gans et al. (2012) performed an empirical analysis to quantify different aspects of agent heterogeneity such as learning effects, agent by-agent differences, shift fatigue, and system congestion. Their proxy for heterogeneity in this study was the agents' service times.

It is a common presumption that higher productivity can only be achieved by a sacrifice in patient satisfaction. However, a study conducted by Boffeli et al. (2012) suggests that physicians can in fact be productive while at the same time creating a satisfying patient experience. They measured the productivity of 22 physicians as well as their patient experience scores. They found that a relatively equal number of physicians fell into each of 4 quadrants – strong productivity/strong satisfaction, strong productivity/weak satisfaction, weak productivity/strong satisfaction and weak productivity/weak satisfaction. Each individual emergency physician has a specific patient per hour (PPH) rate that is dependent on different factors, such as resident subspecialty and years of training (Brennan et al. (2007); Dowd et al. (2005); Deveau et al. (2003)). Batt and Terwiesch (2014) studied the effects of crowding on service time in emergency department and Fast Track (FT) area. The FT area is an express lane for low acuity patients. The service time of a patient was defined as the time from when a patient is placed in a treatment bed to when treatment in the ED is complete. They focused solely on the waiting room census as the measure of ED load. They found that the slowdown effects tend to dominate in the emergency department, while in the FT area, the effects of slowdown and speed up balance out. In the ED, they showed that service time first increases then decreases with load as the relative strength of speed up and slowdown mechanisms shifts. In the next section, we test the relationship between physicians' productivity and crowdedness using our dataset from the case study. Furthermore, we know that physicians' productivity decreases significantly on an hourly basis over the course of a shift (Joseph et al. (2018); KC and Terwiesch (2009); Caldwell (2001)).

3.3 Physician productivity in our case study

The most relevant discriminant between physicians to decide which physician to assign to a particular shift is their individual productivity. Therefore, we first study this characteristic in more detail based on our case study before we mathematically formulate our physician scheduling problem. In particular, we examine if physicians' productivity decreases over the course of a shift and whether physicians become more productive when an ED is more crowded.

3.3.1 Decay of productivity over time

As mentioned at the end of section 3.2, physicians will assess less new patients as they progress through their shift (i.e., physicians are more productive at the beginning of a shift compared to the end). This is also observed in our case study. Figure 3.4 displays Box-and-Whisker plots of the hourly shift productivity from 52 physicians in 7-hour shifts. It is apparent that physicians show a significant decline in productivity between the first and second hour of their shift, and in the final hour of their shift. They have the highest average productivity of 3.53 patients in the first hour of the shift. This number decreases to an average of 0.91 patients during the last hour of the shift.

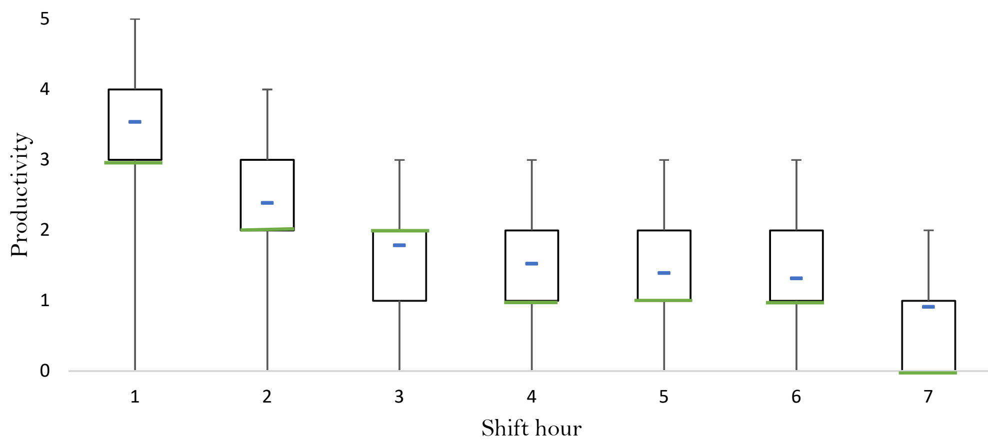

Figure 3.4: Hourly shift productivity. Median in green and Mean in blue

3.3.2 Relation between productivity and crowdedness

Next, we examine the relationship between ED crowdedness and physician productivity. In the ED, there are several crowding measures, such as waiting room census, ED in-service census, and waiting for admission census. The waiting room census represents the number of patients in a common room waiting to visit an ED physician. The ED in-service census is the number of patients under treatment. Finally, the waiting for admission census represents the number of patients admitted to hospital, but must wait in the ED for an available inpatient bed. As suggested by Batt and Terwiesch (2014), we focus only on the waiting room census as the measure of ED crowding in our analysis. Batt and Terwiesch (2014) suggested that ED physicians focus only on this crowding measure as the waiting room census is visible to ED physicians on electronic dashboards. Another reason for choosing waiting room census is that it has no upper bound compared to in-service and waiting for admission census measures. The later two crowding measures are limited by the number of available beds in the ED and as a result show lower variabilities.

Figure 3.5 plots the productivity in the first hour for one of the physicians who was scheduled for the highest number of shift during the studied period (based on 237 shifts) versus the ED crowdedness which is measured as the number of patients in the waiting area. Each blue dot represents the starting productivity of the physician, while an orange dot represents the average productivity when ED crowdedness is within a certain range.

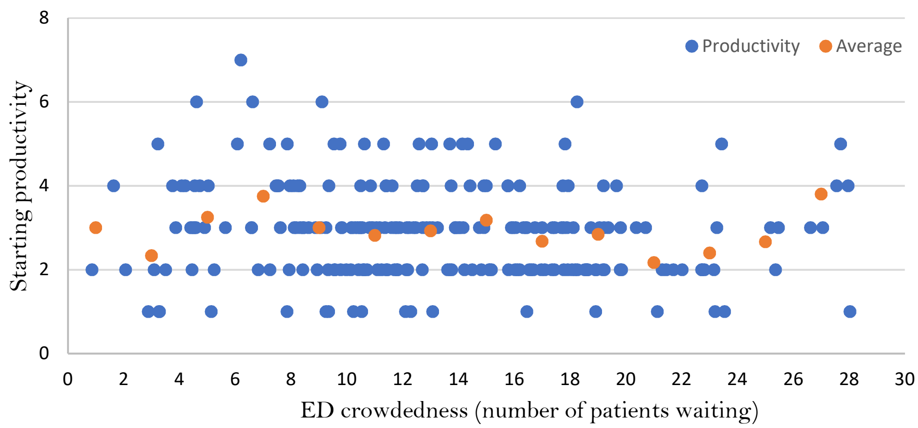

Figure 3.5: ED crowdedness vs. physician's productivity

Based on this figure, the physician's productivity does not seem to be dependent on the ED crowdedness. It is important to stress that Batt and Terwiesch (2014) studied the effects of crowding on service time, however we are examining the impacts of crowding on physician's productivity.

To capture the waiting room census in any given hour, we divide the study period into 1-hour intervals. During the same hours, we then record the actual productivity of physicians who were scheduled in those hours. For our analysis, we select 10 physicians who were scheduled for almost an equal number of shifts (90-100 shifts). First, we perform a multiple linear regression analysis with two set of independent variables, namely waiting room census variable and physicians' indicator variables. We find that the model predicts 42.0% of the variance in productivity (p-value=0.0001). Table 3.2 displays coefficient values of the the measure of ED crowdedness and physicians' indicator variables with their significance levels. The results suggest that the best predictor of productivity is the physician indicator variable. The ED crowding measure is not significantly associated with the productivity. Based on these results, we can conclude that physician productivity is at least not linearly associated with ED crowdedness.

Variable	Coefficient value β (SE)
Intercept	4.72*** (.12)
Waiting room census	-.007(.006)
MD1	-1.03***(.15)
MD2	-1.33***(.16)
MD3	-.61***(.15)
MD4	-1.01***(.14)
MD5	-.84***(.15)
MD6	-1.41***(.15)
MD7	-1.11***(.16)
MD8	-2.10***(.15)
MD9	-.43**(.15)

***Significant at 1% level. **Significant at 5% level. *Significant at 10% level

Dependent Variable: Productivity

Table 3.2: The relationship between productivity and crowdedness

However, Batt and Terwiesch (2014) suggested that the ED first speeds up and then slows down as load increases from low to high. They used an accelerated-failure-time (AFT) model with a log-normal distribution that relates the log of service time for a patient to the load variable and control variables such as the patient-visit specific covariates (e.g., age, gender, race, triage level, and chief complaint). Next, we use the proposed model by Batt and Terwiesch (2014) to examine the relationship between ED crowding measure and productivity. The model predicts 40.2% of the variance in productivity (p-value=0.0001). Table 3.3 displays coefficient values of the the measure of ED crowdedness and physicians' indicator variables with their significance levels. Again, the results suggest that the best predictor of productivity is the physician indicator variable. The ED crowding measure is not significantly associated with the productivity. Based on these results, we can conclude that physician productivity is not associated with ED crowdedness. Note that Batt and Terwiesch (2014) studied service time that is different from our productivity variable. Service time of a patient is the total time to complete all tasks that includes all laboratory, radiology, and medication orders. In terms of productivity, they also suggested that physicians are rarely seen running through the halls of the ED or performing specific procedures faster when ED is crowded. We should also note that physicians know that ED has limited capacity in terms of testing equipment and inpatient bed availability. Therefore, physicians are aware that they will only shift the bottlenecks form the wait room to other ED areas by adjusting their clinical behavior and become more productive when ED is crowded.

Variable	Coefficient value β (SE)
Intercept	1.50*** (.04)
Waiting room census	-.002(.002)
MD1	-0.25***(.04)
MD2	-0.33***(.05)
MD3	-0.13***(.04)
MD4	-0.25***(.04)
MD5	-0.20***(.04)
MD6	-0.38***(.04)
MD7	-0.27***(.05)
MD8	-0.59***(.04)
MD9	-.12*** (.04)

***Significant at 1% level. **Significant at 5% level. *Significant at 10% level

Dependent Variable: ln(Productivity)

Table 3.3: The relationship between productivity and crowdedness

3.3.3 Modeling of physician productivity

From the data we observe the productivity (or PPH rate) of each physician for each hour of the shift. Based on this, we were not able to fit a theoretical distribution that describes the likelihood to assess a certain number of new patients in each hour. Therefore, we decided to investigate the time between initial assessments during each hour of the shift. It turns out that a lognormal distribution provides a statistically significant fit to describe the length of this time period. Consequently, we can derive a probability distribution for the productivity in a certain hour as follows: Let $X_{it}^{(k)}$ denote the time (in minutes) between the initial assessment of patient i and the previous patient assessed by physician k during the t-th hour of her shift. As mentioned before, this random variable follows the lognormal distribution. Furthermore, let $Y_t^{(k)}$ represent the productivity of physician k during the t-th hour of her shift. Consequently,

$$P(Y > y) = P(\sum_{i=1}^{y+1} X_i \leq 60) \qquad \text{and} \qquad P(Y = y) = P(Y > y-1) - P(Y > y) \tag{3.1}$$

The sum of lognormally distributed random variables with the same parameters is reasonably approximated by a lognormal distribution again. We set the first and second moment of this lognormal distribution such that the first and second moment of Y (according to Eq. (3.1)) results in the same numbers as observed in the data. This means that we have an empirical distribution for Y, such that our problem formulation in the next section should be distribution free as well.

3.4 Problem description

In this section, we formulate the physician scheduling problem. The objective function of the physician scheduling problem is to minimize the total hourly mismatch between patient demand and physicians' productivity.

3.4.1 Introduction of notation

The model presented in this study incorporates uncertainty associated with ED arrivals and physician productivity for finding the optimal assignment of physicians to a given schedule that minimizes the overall hourly mismatch between time-varying patient arrivals and physician-dependent productivity. We do this through a stochastic programming formulation. Before entering the mathematical formulation, we introduce the notation as summarized in Table 3.4.

Table 3.4: Input parameters and decision variables

i	$\in \{1, \ldots, I\}$	index of physicians
j	$\in \{1, \ldots, J\}$	index of days over the planning period
k	$\in \{1, \ldots, K\}$	index of shifts
t	$\in T = \{0, \ldots, 23\}$	index of time dividing the day (in hourly intervals)
s	$\in \{1, \ldots, S\}$	index of scenarios
p_s		probability of scenario s
W_i		maximum number of shifts physician i can work during the planning period
A_{sjt}		number of patients arriving in interval t of day j in scenario s
λ_{jt}		average number of arrivals in interval t of day j
P_{sin}		productivity of physician i during the nth hour of the shift in scenario s
f_{kt}		n if nth hour of shift k is at time t
M_{sjt}		mismatch between arrivals and productivity in interval t of day j in scenario
Z_{ij}		1 if physician i has a preference to work on day j, 0 if indifferent
x_{ijk}		1 if physician i is assigned to shift k of day j, 0 otherwise

Furthermore, each shift belongs to one of the following sets: K_D, the day shifts, K_E, the evening shifts, K_N, the night shifts.

3.4.2 Mathematical Formulation

Next, we formulate the problem of minimizing the total hourly mismatch between patient arrivals and physician productivity over the planning period, respecting the problem constraints.

$$\text{Minimize} \sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{t \in T} p_s \; M_{sjt} \tag{3.2}$$

$$s.t.$$

$$A_{s10} - \sum_{k=1}^{K} \sum_{i=1}^{I} x_{i1k} \, P_{sif_{k0}} \quad \leq \quad M_{s10} \, ; \quad s \in S \tag{3.3}$$

$$A_{sjt} + M_{sj(t-1)} - \sum_{k=1}^{K} \sum_{i=1}^{I} x_{ijk} \, P_{sif_{kt}} \quad \leq \quad M_{sjt} \, ; \, j = 1, \ldots, J; t \in T - \{0\}; s \in S \tag{3.4}$$

$$A_{sj0} + M_{s(j-1)23} - \sum_{k=1}^{K}\sum_{i=1}^{I} x_{ijk}\, P_{sif_{k0}} \leq M_{sj0}\,; \quad j = 2, \ldots, J; s \in S \tag{3.5}$$

$$\sum_{i=1}^{I} x_{ijk} = 1\,; \quad j = 1, \ldots, J; k = 1, \ldots, K \tag{3.6}$$

$$\sum_{k=1}^{K} x_{ijk} \leq 1\,; \quad i = 1, \ldots, I; j = 1, \ldots, J \tag{3.7}$$

$$\sum_{j=1}^{J}\sum_{K=1}^{K} x_{ijk} \leq W_i\,; \quad i = 1, \ldots, I \tag{3.8}$$

$$\sum_{k=1}^{K} x_{ijk} \geq Z_{ij}\,; \quad i = 1, \ldots, \quad I; j = 1, \ldots, J \tag{3.9}$$

$$\sum_{k \in K_E} x_{i(j-1)k} + \sum_{k \in K_N} x_{i(j-1)k} + \sum_{k \in K_D} x_{ijk} \leq 1\,; \quad i = 1, \ldots, I; j = 2, \ldots, J \tag{3.10}$$

$$\sum_{k \in K_N} x_{i(j-1)k} + \sum_{k \in K_N} x_{ijk} \leq 1\,; \quad i = 1, \ldots, I; j = 2, \ldots, J \tag{3.11}$$

$$M_{sjt} \geq 0\,; \quad j = 1, \ldots, J;\; t \in T;\; s = 1, \ldots, S \tag{3.12}$$

$$x_{ijk} \in \{0, 1\} \tag{3.13}$$

Constraints (3.3) to (3.5) calculate the hourly mismatch between patient demand and physicians' productivity. The left-hand side of the constraints (3.3) to (3.5) represent the patient demand minus the number of patients treated in time period t, where the patient demand is the sum of new patients arriving in the time period t and patients still waiting from the previous time period $t-1$. Constraint (3.6) ensures that exactly one physician is assigned to each shift of the period. Constraint (3.7) makes sure that a physician is assigned to at most one shift per day. Constraint (3.8) states that each physician i works at most W_i shifts over the planning period. Constraint (3.9) ensures that physician i works on day j if she has a preference to work on that day ($Z_{ij} = 1$). If a physician has a preference not to work on a day, we can add a constraint to our model to ensure that no shift is assigned to the physician on that particular day. Constraint (3.10) guarantees that a physician who is assigned to an evening or a night shift will not be assigned to a day shift on the next day. Finally, constraint (3.11) states that a physician assigned to a night shift must not be assigned to a night shift on the next day.

Our model with a finite number of scenarios can form a full deterministic mixed integer program. However, with many realizations, the problem becomes quite large. Therefore, algorithms such as branch and bound become inefficient for solving such a large-scale stochastic program. For real-sized problems, we propose and apply a L-Shaped procedure that benefits from the decomposition algorithm of Dantzig and Wolfe (1960). The L-shaped method is a technique to solve a stochastic program where random variables take finitely many values.

3.5 The L-Shaped Algorithm

The L-Shaped method was first introduced by Van Slyke and Wets (1969) for two-stage stochastic integer programs. Our scheduling problem can be formulated as a two-stage stochastic programming problem with binary variables in the first stage and continuous variables in the second stage This allows us to separate the variables into two sets of deterministic (first stage) and stochastic (second stage). Therefore, the MILP constructed in the previous section will be first decomposed into a master problem (MP) where a feasible physician scheduling is made, and a series of sub-problems (SP) where the mismatch is calculated for each scenario.

I) Master problem (first-stage)

The master problem is the LP relaxation of the original problem with only deterministic variables. Let v denote the major iterations of the algorithm. In the formulation below, variable θ, along with the optimality cut (constraint (3.15)), are used as appropriate approximations of the second stage.

$$\text{(MP): Minimize} \quad \theta^v \tag{3.14}$$

$$s.t.$$

$$E^v X^v + \theta^v \geq e^v \tag{3.15}$$

$$\sum_{i=1}^{I} x_{ijk}^v = 1\,; \quad j = 1, \ldots, J; k = 1, \ldots, K \tag{3.16}$$

$$\sum_{k=1}^{K} x_{ijk}^v \leq 1\,; \quad i = 1, \ldots, I; j = 1, \ldots, J \tag{3.17}$$

$$\sum_{j=1}^{J} \sum_{K=1}^{K} x_{ijk}^v \leq W_i\,; \quad i = 1, \ldots, I \tag{3.18}$$

$$\sum_{k=1}^{K} x_{ijk}^v \geq Z_{ij}\,; \quad i = 1, \ldots, \quad I; j = 1, \ldots, J \tag{3.19}$$

$$\sum_{k \in K_E} x_{i(j-1)k}^v + \sum_{k \in K_N} x_{i(j-1)k}^v + \sum_{k \in K_D} x_{ijk}^v \leq 1\,; \quad i = 1, \ldots, I; j = 2, \ldots, J \tag{3.20}$$

$$\sum_{k \in K_N} x_{i(j-1)k}^v + \sum_{k \in K_N} x_{ijk}^v \leq 1\,; \quad i = 1, \ldots, I; j = 2, \ldots, J \tag{3.21}$$

$$0 \leq x_{ijk}^v \leq 1 \tag{3.22}$$

II) The subproblem (second-stage)

$$\text{(SP): Minimize} \quad \sum_{t \in T} \sum_{j=1}^{J} M_{sjt} \tag{3.23}$$

$$s.t.$$

$$A_{s10} - \sum_{k=1}^{K} \sum_{i=1}^{I} x_{i1k}\, P_{sif_{k0}} \leq M_{s10}\,; \quad s \in S \tag{3.24}$$

$$A_{sjt} + M_{sj(t-1)} - \sum_{k=1}^{K} \sum_{i=1}^{I} x_{ijk}\, P_{sif_{kt}} \leq M_{sjt}\,; j = 1, \ldots, J; t \in T - \{0\}; s \in S \tag{3.25}$$

$$A_{sj0} + M_{s(j-1)23} - \sum_{k=1}^{K} \sum_{i=1}^{I} x_{ijk}\, P_{sif_{k0}} \leq M_{sj0}\,; \quad j = 2, \ldots, J; s \in S \tag{3.26}$$

$$M_{sjt} \geq 0\,; \quad j = 1, \ldots, J; \; t \in T \tag{3.27}$$

We now define the binary L-shaped algorithm as follows:

The binary L-shaped algorithm

Step 0: set $v = 0$.

Step 1: Solve the MP and let (x^v, θ^v) be an optimal solution.

If $v = 0$ (no optimality cuts), remove θ from (MP) and set $\theta^0 = -\infty$.

Step 2: For $s = 1, \ldots, S$ solve the SP.

Let π_s^{Iv}, π_{tjs}^{IIv}, and π_{js}^{IIIv} be the dual multipliers of the SP optimal solution associated with constraints (3.24), (3.25), and (3.26), respectively. Define

$$E^v X^v =$$

$$\textstyle\sum_{s=1}^{S} p_s(\pi_s^{Iv} \sum_{k=1}^{K} \sum_{i=1}^{I} x_{i1k} P_{sif_{k0}} + \sum_{j=1}^{J} \sum_{t=1}^{23} \pi_{sjt}^{IIv} \sum_{k=1}^{K} \sum_{i=1}^{I} x_{ijk} P_{sif_{kt}} + \sum_{j=2}^{J} \pi_{sj}^{IIIv} \sum_{k=1}^{K} \sum_{i=1}^{I} x_{ijk} P_{sif_{kt}}) \quad (3.28)$$

$$e^v = \sum_{s=1}^{S} p_s(\pi_s^{Iv} A_{s10} + \sum_{j=1}^{J} \sum_{t=1}^{23} \pi_{sjt}^{IIv} A_{sjt} + \sum_{j=2}^{J} \pi_{sj}^{IIIv} A_{sj0}) \quad (3.29)$$

Let $\omega^v = e^v - E^v X^v$. If $\theta^v \geq \omega^v$, then stop with the optimal solution given by x^v. Otherwise, set $v = v + 1$, add optimality cut to equation (3.15), and return to **Step 1**.

Step 3: Begin branch-and-cut algorithm. Create two new branches; Check if there are any violated cuts and return to **Step 1**.

Assessing the quality of a candidate solution plays a fundamental role in stochastic optimization. We apply a Monte Carlo bounding technique described in Mak et al. (1999) to evaluate the quality of our candidate solution. Let's consider the following stochastic program

$$SP : z^* = \min_{x \in X} \mathbb{E} f(x, \xi), \quad (3.30)$$

where f is a real-valued function that determines the cost of operating with decision x under a realization of the random vector ξ. Denote an optimal solution and the optimal value of (SP) as x^* and z^*, respectively. Let $\xi^1, \xi^2, ..., \xi^n$ be an independent and identically distributed (iid) sample from the distribution of ξ. A sampling approximation of (SP) is given by

$$SP_n : z_n^* = min_{x \in X}\, 1/n \sum_{i=1}^{n} f(x, \xi^i) \tag{3.31}$$

Mak et al.(1999) showed that a lower bound on z^* can be obtained by interchanging minimization and expectation

$$\mathbb{E} z_n^* \leq z^* \tag{3.32}$$

Hence, the lower bound on the true solution, z^*, can be obtained by solving the original problem, SP_n, for multiple times, each with independently generated scenarios.

$$\bar{L}_(n_l) = 1/n_l \sum_{j=1}^{n_l} z_n^{*j} \tag{3.33}$$

where $z_n^{*j} = min_{x \in X}\, 1/n \sum_{i=1}^{n} f(x, \xi^{ij})$.

Given a candidate solution $\hat{x}$ and a sample size n for (SP_n), an upper bound on the optimal value of (SP) can be obtained by:

$$z^* \leq \mathbb{E} f(\bar{x}, \xi) \tag{3.34}$$

The upper bound in Eq. (3.35) comes from suboptimality of $\bar{x}$. A straightforward estimate of $\mathbb{E} f(\hat{x}, \xi)$ is the standard sample mean estimator shown below:

$$\bar{U}_(n_u) = 1/n_u \sum_{i=1}^{n_u} f(\hat{x}, \xi^i) \tag{3.35}$$

We can now define an approximate $(1 - 2\alpha)$-level confidence interval for the optimality gap at $\hat{x}$ as

$$[0, [\bar{U}_(n_u) - \bar{L}_(n_l)]^+ + \hat{\epsilon}_u + \hat{\epsilon}_l] \tag{3.36}$$

where $\hat{\epsilon}_u$ and $\hat{\epsilon}_l$ represent the standard errors.

3.6 Computational Results

Our case study is based on data gathered from an emergency department that is part of the Alberta Health Services (AHS) system in Alberta, Canada. we have data available over a two-year time period from August 2013 to July 2015. We are interested in finding the assignment of physicians to shifts that minimizes the total mismatch over all days in a planning period of four weeks considering all constraints discussed in the mathematical formulation section. To capture physicians' productivity, we include 52 physicians who were scheduled for more than 90 shift during the studied period. There are currently 13 shifts scheduled every day of the week in the studied ED. Therefore there are in total 28 x 13 = 364 shifts in the planning period of four weeks. To build a fair schedule, we allow all 52 physicians to work for the same number of shifts during four weeks which means 364/52 = 7 shifts per each physician. We will compute the solution without considering physician preferences. This gives us the improvement possible if physician schedules were completely flexible.

3.6.1 Solution quality and optimal schedule

In this section, we first evaluate the quality of stochastic solutions in order to find a reasonable number of scenarios. Afterwards, we compare the solution of the proposed stochastic programming model to that of the deterministic model for our case study.

We can obtain a better solution if the number of scenarios increases, but the computational cost of finding that solution increases as well. We evaluate the quality of three candidate solutions for 10 batches of sampled scenarios with sizes of 20, 40, and 60. For each candidate solution, we calculate the 95% confidence interval on its optimality gap based on the bounding technique discussed in the previous section. The upper bounds for the three candidate solutions $\bar{x}^{20}$, $\bar{x}^{40}$, $\bar{x}^{60}$ are computed by solving the deterministic versions of the MIP model presented in section 3.4.2. Table 3.5 displays the quality of the three candidate solutions. It can be seen from this table that we can improve the solution quality and obtain tighter confidence intervals for the optimality gap if we increase the number of scenarios. Therefore, we will solve the stochastic programming model using 60 sample realizations of the physician productivities and patient arrivals.

Candidate solution	$\bar{x}^{20}$	$\bar{x}^{40}$	$\bar{x}^{60}$
Number of scenarios	20	40	60
Number of batches	10	10	10
Confidence interval (95%)	[0,3701]	[0,1905]	[0,811]

Table 3.5: Optimality gap for the three candidate solutions

Since the problem could not be solved using a MILP solver directly, we have used the L-shaped approach discussed in Section 3.5. The L-Shaped algorithm model was coded in Eclipse Java Mars 4.5 and solved with CPLEX Optimization Studio 12.6.3 on a Lenovo Y50 Laptop (Intel Core i7/ 8GB RAM/ Windows 10). The solution leads to a physician scheduling scheme that gives the total mismatch of 21,420 over the planning period of four weeks. To emphasize the negative consequences of not considering the physicians' productivity in the physician scheduling problem, we construct a random feasible physician schedule. Solving the problem with the random schedule gives the total mismatch of 34,532 over the same period, an increase of 61% compared to the optimal schedule.

3.6.2 Simulation

Next, we use a simulation model to determine if considering physicians' heterogeneity in the scheduling problem has a meaningful impact on the wait times. We develop a simulation model in Java to investigate the impact of the near-optimal stochastic MIP solution on reducing patient wait times. We run the simulation over 280 days using the same distributions for patient arrivals and physicians' productivity. We consider the initial two days as the warm-up period of the simulation model. In particular, we are interested to compare the average wait times of three schedules: the optimal schedule, the random schedule, and deterministic schedule. The deterministic schedule is constructed by solving the deterministic version of the problem where all stochastic variables are fixed at their mean values.

The average hourly waits for the three schedules are shown in Figure 3.6. Comparing the optimal schedule to the random schedule we see that ignoring the heterogeneity among physicians in constructing ED physicians' schedule leads to increased unnecessary wait times experienced by patients. The simulation results suggest that the optimal schedule can reduce the average wait time over all hours by over half an hour compared to the random schedule. When compared to the deterministic schedule, the optimal schedule can decrease the overall average wait time over 10 minutes which illustrates the importance of the stochastic model.

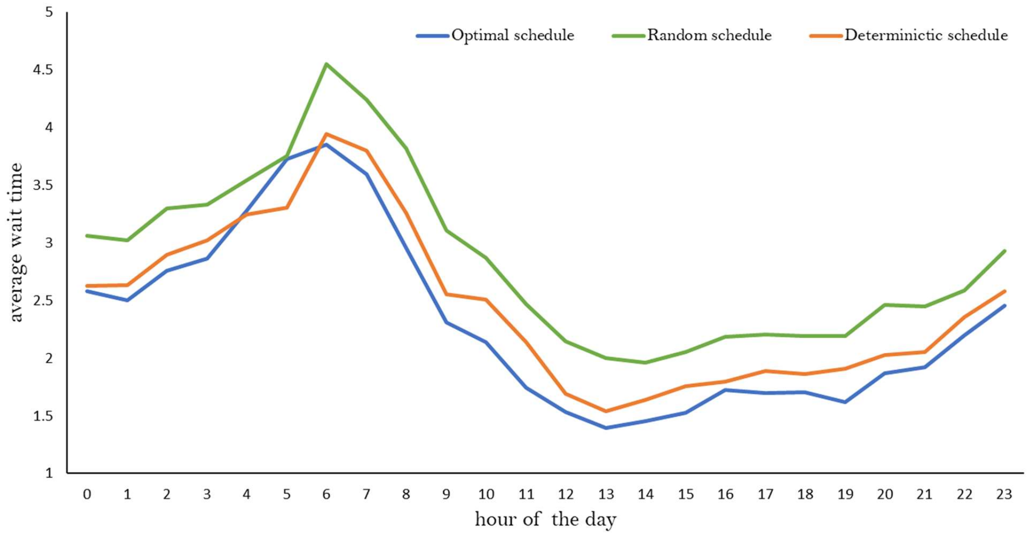

Figure 3.6: Average hourly wait times for the three schedule

3.6.3 More insights: physician clustering

Finally, we examine ED physicians clustering into some groups based on their average productivity rates. In reality, it may be difficult for hospital administrators to implement the optimal schedule if we assume that each individual physician is different than others. ED physicians often have teaching, research, and family responsibilities that reduce their availability. Hence it might be difficult for most ED physicians to be flexible concerning the timing of their working shifts. Clustering ED physicians can improve the flexibility in selecting among ED physicians when constructing the schedule. In this section, we examine the effects of having two, three, five, and fifty-two clusters on reducing the average wait time. For each clustering scenario, we first find the optimal schedule of physicians using the proposed two-stage stochastic program and second determine the average hourly wait times using the simulation model.

In the case of two clusters, we divide our physicians into two equal groups of slow and fast physicians as shown in Figure 3.7. The productivity of physicians in each cluster is a random variable whose mean and standard deviation are calculated using the productivity of the twenty-six physicians belonged to that cluster. In the case of fifty-two clusters, we assume each physician has a specific productivity performance that is a random variable with a known distribution discussed before. Furthermore, we examine the impacts of considering three clusters (slow, medium, and fast) and five clusters (very slow, slow, medium, fast, and very fast) on reducing wait times. The three clusters are also shown in Figure 3.7.

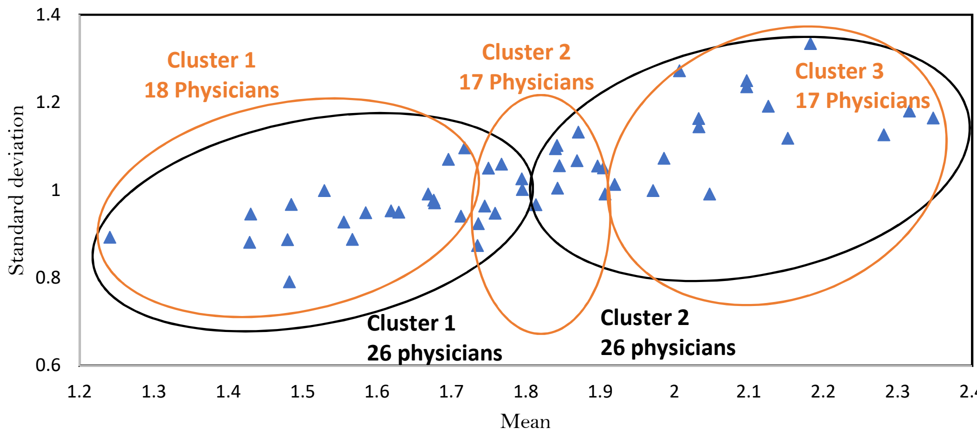

Figure 3.7: Two or three clusters of ED physicians in our case study

Figure 3.8 shows the average hourly wait times for the different number of clusters. We find the highest average wait times over all hours in the case of two clusters. The wait times are decreased when the three clusters of physicians is considered. However, the average wait time is still higher (over 13 minutes) compared to the fifty-clusters. The results of the five clusters is, however, roughly the same as the wait times of the fifty-two clusters. These suggest that hospital administrators do not need to consider the productivity of each individual physician in constructing the ED schedule.

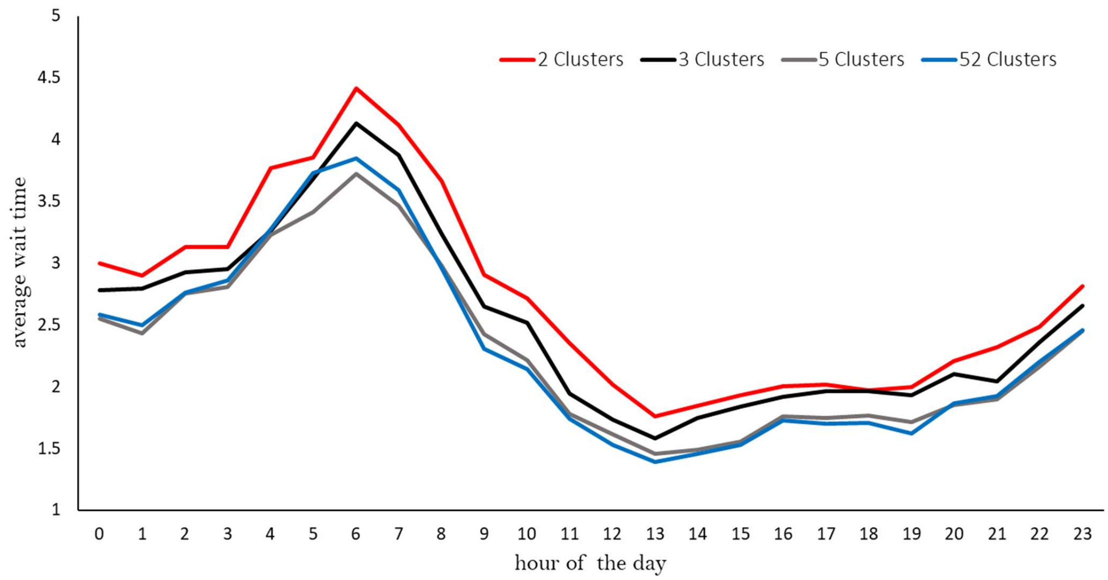

Figure 3.8: Average hourly wait times

3.7 Conclusion and future research

In this study, we have formulated the scheduling problem of physicians to minimize the total hourly mismatch between patient arrivals and physicians' productivity. We defined productivity as the speed at which new patients are seen by physicians. The scheduling of physicians has a direct impact on the waiting time of patients because of the heterogeneity among physicians in terms of their productivity. The analyses of hourly productivity rates suggested that ED physicians evaluate significantly more patients early in a shift, and few at the end of a shift. We found no significant relationship, however, between ED crowdedness and physicians productivity. Besides including the heterogeneity among physicians, we also include the stochastic nature of the ED patient arrivals and the physicians' productivity in our formulation.

Next, we developed a simulation model to investigate the impact of the stochastic MIP schedule on reducing wait times. The results showed that the optimal solution can reduce average wait time compared to the schedule constructed based on the existing physician scheduling literature. We concluded that the wait times can be reduced when the productivity of the ED lines up with the numbers and timing of when patients present to the emergency department. Finally, we examine the effects of physicians clustering in reducing wait times. We found that clustering ED physicians based on their productivity rates can improve flexibility in building the schedule while optimizing the flow of patients.

In this research, we only discussed the physician scheduling problem. Further studies can address the staffing problem and the shift-scheduling problem at the strategic level.

Chapter 4

A new definition of crowding in emergency departments: patient volume and patient mix

4.1 Introduction

Although there is a large body of literature on the causes and effects of crowding in emergency departments (EDs), there is no standard definition of crowding. The most commonly crowding measure used in the literature is the number of patient arrivals (Hwang et al. (2011)). However, the number of ED arrivals may remain the same over years, while the patient mix changes through time. A new crowding measure that includes both patient arrivals and patient mix can provide important information to predict ED resource requirements more accurately.

In this chapter, we present a more comprehensive crowding measure that includes both number of patients and patient mix. We define ED crowding as the number of ED patients from each group. To classify emergency department visits for identifying the patient mix, we propose a patient classification system (PCS). The proposed PCS separates the patient population into 880 homogeneous groups using four types of variables: age, gender, triage level, and disease. An ED crowding measure becomes applicable if, first, the arrival patterns of its suggested variables can be predicted and, second, it is relevant. Thus, forecasting ED crowding at different hours of the day is our primary goal in this study. To demonstrate the relevance and significance of the proposed crowding metrics, we then examine the effects of the two crowding variables on ED performance metrics. Patient wait time for the initial assessment by a physician, total length of stay, and number of re-visits within 72 hours are our variables for the ED performance metrics. The first two variables are commonly used time-based ED performance metrics, while the last one is a quality-based ED performance measure. The suggested ED crowding variables and the proposed performance measures are summarized in Table 4.1.

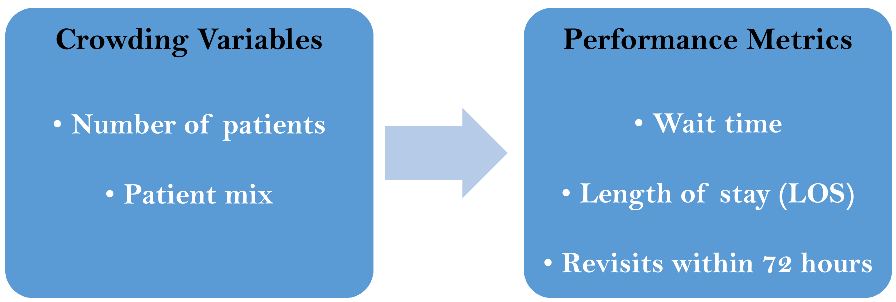

Table 4.1: Crowding variables and performance measures

We therefore address the following research questions:

Research Question 1: *Can we forecast ED crowding?*

We can further divide the first research question into the following two sub-questions.

Research Question 1a: *Is there an arrival time pattern of patients in EDs? Has it changed through time?*

Research Question 1b: *Is there an arrival time pattern of types of patients in EDs? Has it changed through time?*

Research Question 2: *Are there any relationships between crowding variables and performance metrics? Should the crowding measure be redefined by adding the patient mix?*

To address the above research questions, we use data on all admissions to emergency departments in February, July and November for the years of 2007, 2012, and 2017 in the city of Calgary, Alberta, Canada. The dataset includes five hospitals as case study sites, where four are adult EDs and one provides care for sick children. One of the adult-focused EDs was opened in 2013. Data has been extracted from two data sets: (i) National Ambulatory Care Reporting System (NACRS); (ii) Discharge Abstract Database (DAD). NACRS collects demographic, administrative, clinical and service-specific data for ED visits, and DAD contains demographic, administrative and clinical data for hospital inpatient discharges. Our final database includes the following information on 237,366 patient visits:

i. Patient information: age, gender, date and time of triage

ii. Diagnostic information: Canadian triage and acuity scale (CTAS) and international classification of diseases (ICD10)

iii. Performance metrics: wait time (the time between triage and the physician's initial assessment), total length of stay, and whether this is a re-visit within 72 hours of last ED discharge

The results of this study will be of interest to health providers as well as the operations research community. Although strategic and operational planning in EDs are normally based on averages, the changes in the profile of patients are also of great importance. This study is the first large scale analysis of its type, and findings can be applied for developing optimisation models, in special queueing-based models, on emergency departments. This research has three key contributions. Firstly, we propose a patient classification system for classifying emergency department visits. Secondly, we examine arrival patterns to find whether the arrivals of patients as well as each group of patients can be predicted in a specific time period. The third contribution is an exploration of the relationships between ED crowding variables and ED performance metrics. The remainder of this chapter is organized as follows. We review the related literature in section 4.2. Section 4.3 presents the patient classification system (PCS), which is followed in section 4.4 by application of the PCS to the case study and the findings. Finally, we conclude in section 4.5 with a discussion and further research suggestions.

4.2 Literature review

In this section, we review two strands of literature that are relevant to the two dimensions shown in Table 4.1: studies on the ED crowding that address patient arrival patterns and/or changes in the mix of arrivals over time, and literature on the ED crowding and its impact on the performance measures.

Number of patients and patient mix are our two proposed variables for the ED crowding. Identifying variables that contribute to ED crowding help policy makers to better understand and manage crowding. Hwang et al. (2011) conducted a systematic review of crowding measures in EDs. They found that the most commonly crowding measures used are numerical counts or percentage of patients (as arrivals, in the waiting room, at triage or registration, by acuity, etc.). To the best of our knowledge, no one has suggested a crowding measure that is based on both patient arrivals and mix of patients. The most related crowding measure in the literature to our proposed metric is the number of patients at each acuity level (See e.g., Bullard et al. (2009)). However, the ED acuity scores alone do not provide important information about patients other than their levels of urgency. Our proposed ED crowding includes patient volume and patient mix.

Classifying emergency department visits in several homogeneous groups is the initital step in determining the mix of ED patients. Williams and Crouch (2006) conducted a systematic literature review of emergency department patient classification systems (PCS) for the years 1985-2004. The purpose of these PCSs was to group patients based on their nursing care needs. 12 ED PCSs were identified in this systematic review. However, only three classification systems appeared to the nurses to be valid tools and provided consistent results across different EDs. These three PCSs were the ED patient classification matrix (Butler (1986)), the Conners tool (Conners (1994)), and the Jones dependency tool (JDT) (Crouch and Williams (2006)). All these classification tools are, however, developed based on nurse staffing requirements to predict nursing workload. In this study, we develop a PCS based on patient demographics (age and gender) and diagnostic information (triage level and disease code) that has more general applicability. The characteristics of patients and their reasons for ED visits are important to address a wide range of strategic decision problems in emergency departments, such as resource and financial planning. The proposed PCS allows us to identify the mix of patients visiting EDs in a specific time period.

As already discussed in the previous section, the primary goal of this research is to examine the arrival patterns of the proposed ED crowding metrics. There are several studies that have examined patterns of patient arrivals in emergency departments. For example, Carvalho-Silva et al. (2017) used two years of ED arrivals data to first examine arrival patterns and then build forecasting models based on time series, such as moving average, exponential smoothing, Holt-Winters, and autoregressive integrated moving average (ARIMA). They analyzed the arrivals to the ED regarding the month of the year, the day of the week and the time of the day. The results suggested that the greatest demand is in February (533.7 arrivals per day) and on Monday (567.6 arrivals per day), and during the day there are two peaks for increased demand: at the start of the morning and early afternoon. The best forecasting model for the test period was the ARIMA based on the mean absolute percentage error. McCarthy et al. (2008) analyzed one year of ED arrival data from an academic ED. The results showed that hour of the day is the most important predictor in ED arrival rates. Seasonal factors, however, did not statistically influence patient arrivals to the ED. Some studies addressed the changes in patterns of patient arrivals over a longer time period. For example, McNaughton et al. (2015) used a sample of U.S. emergency department data from 2006 to 2012 to identify hypertension-related ED visits and evaluate the changes over time. They reported that 23.6% of all adult ED visits that occurred during the 7-year study period were hypertension-related visits, with an annual percentage growth rate of 5.2%. Lewis et al. (2015) focused on U.S. ED toothache visits by 20-29 year-olds during 2001-2010. They used data from the National Hospital Ambulatory Medical Care Survey (NHAMCS). The average annual increase in ED visit rates was 6.1% for toothache. The results also suggested that 42% of all 20-29 year-olds ED visits were toothache visits. They concluded that younger adults with a chief complaint of toothache increasingly rely on EDs. Morganti et al. (2013) studied the evolving role of EDs in the U.S. health care system using a mix of quantitative and qualitative methods. They analyzed ten years of data to evaluate the changing patterns of health care utilization and delivery in 60 communities nationwide. The main finding was that office-based physicians are increasingly directing complex patients to EDs rather than managing these patients themselves. Herring et al. (2013) examined the increasing use of the ED by patients with life-threatening illness in California from 2002 through 2009. They reported that total annual ED visits increased by 25%, from 9.0 million to 11.3 million, but high-intensity ED visits increased by 87% from 778,000 to 1.5 million per year. These results suggested that more complex patients are going to EDs as an alternative to care outside EDs. All of these studies have examined the arrival patterns of the total patient population or only a specific group of patients. However, it is also important to study the patient mix to identify what types of patients arrive to EDs. In this study, we investigate whether there exists a pattern of arrivals for a group of patients as well.

Finally, we examine the effects of the two crowding variables on ED performance metrics to show the significance of the proposed crowding definition. Over the last decade, there have been several studies on crowding and its impact on performance metrics in EDs. For example, Ajmi et al. (2015) performed a continuous 10-day mapping exercise to record patient paths through a pediatric ED. The objective was to identify crowded situation indicators and bottlenecks. They identified the wait time for a hospital bed as the greatest source of delay in patient flow. Martin et al. (2011) performed a patient flow mapping through an Australian ED for a continuous 84-hour (3.5 days) period. The goal was to identify bottlenecks that contribute to overcrowding. The long waiting time for a bed request for a patient to exit from the ED was the main reason for overcrowding. Welch et al. (2007) designed an integrated tracking system to track patients' progress throughout their ED visit. They analyzed the data from 39,704 ED encounters for a 1-year period to determine relevant patterns that might affect staffing and operational efficiencies. They found that patients seen during less busy times, in the middle of the night, had a higher acuity. Radiology and laboratory utilization were highly correlated with ED arrivals, and the higher the acuity, the greater the utilization. These existing efforts to understand and improve ED performance metrics usually occur within a relatively short period of time and in one ED. It is unclear whether their findings would be valid in a longer time period and across conditions within EDs. In this study, we use the data from several EDs over several years, which may provide results that are more generalizable. We should also note that the definition of crowding in this study is different than the above references.

4.3 Characterising the patient

The initial step in identifying patient mix is to classify patients into homogeneous sets such that patients that belong to the same group have similar characteristics and diagnoses. How many attributes should be used to classify emergency department visits? Having a huge number of attributes may weaken the practicality of the classification and reduce its usability, and some classes may contain very few cases. If, on the other hand, very few attributes are included, each group has significant variability. Thus, there is always a trade-off to find the best number of attributes. We separate patients into a reasonable number of groups based on patient demographics (age and gender) and diagnostic information (CTAS and disease code). Our patient database includes various items of patient information. We selected age and gender as the most salient attributes, as summarized in Table 4.2 below.

Patient information	
Age	**Gender**
1- 0-1 year 2- >1-18 years 3- >18-65 years 4- >65 years	1- male 2-female

Table 4.2: ED patient demographics

There are many different reasons for an ED visit, but for all patients coming to the emergency room, a rapid assessment will be done to determine the urgency level of that patient. A triage nurse initially asks an ED patient about his/her complaint and takes the patient's vital signs. The triage nurse will then assign the patient a triage level, which is a proxy of waiting time for that patient (Horwitz et al. (2010)). We use the Canadian triage and acuity scale (CTAS) and the international classification of diseases (ICD-10 codes) shown in Table 4.3 to identify the ED patient's diagnostic information. Appendix A lists the 22 ICD-10 codes with their definitions. We should note that a patient can have multiple diagnoses in a single visit. However, we only consider the primary diagnosis in our analysis. There will be somewhat similar resource consumption for patients with the same combination of triage level and diagnosis, although the diagnoses within a CTAS or ICD10 category can be quite wide-ranging.

Diagnostic information	
CTAS	**Disease code**
1-Resuscitation 2-Emergent 3-Urgent 4-Less urgent 5-Non-urgent	Refer to ICD-10 codes list (22 broad categories)

Table 4.3: ED patient diagnostic information

The proposed PCS assigns a numeric code number to each patient which provides distinct information about that patient. The code begins with a digit representing the age group of the patient, followed by a digit that is the gender code, followed by a digit to identify the CTAS of the patient, and finally two digits (01-22) representing the primary diagnosis code of the patient. For example, assume a patient has arrived in the ED with the following attributes: 15 years old (age group 2), female (gender group 2), triage code 3, and is primary diagnosed with respiratory infection (diagnosis group 10). Here the coding number would be 22310.

4.4 Application of the PCS to the case study

A total of 237,366 patients arrived to seek care in the Calgary EDs during the nine months examined (February, July and November for the years 2007, 2012, and 2017). Demographic characteristics of the patients and their diagnostic information are shown in Table 4.4.

Variable	Total % within group	Variable	Total % within group
Age		**CTAS**	
0-1 year	7.3	CTAS 1	1.1
>1-18 years	20.3	CTAS 2	27.8
>18-65 years	56.4	CTAS 3	47.3
>65 years	15.9	CTAS 4	20.9
Gender		CTAS 5	2.9
male	50.6	**Disease codes**	
female	49.4	1) A00-B99	4.6
		2) C00-D49	0.5
		3) D50-D89	0.4
		4) E00-E89	1.3
		5) F01-F99	5.7
		6) G00-G99	1.8
		7) H00-H59	1.2
		8) H60-H95	1.3
		9) I00-I99	4.4
		10) J00-J99	8.3
		11) K00-K95	6.9
		12) L00-L99	2.7
		13) M00-M99	4.5
		14) N00-N99	4.7
		15) O00-O9A	2.0
		16) P00-P96	0.3
		17) Q00-Q99	0.1
		18) R00-R99	21.9
		19) S00-T88	24.5
		20) U00–U99	0.0
		21) V00-Y99	0.0
		22) Z00-Z99	2.8

Table 4.4: ED patient demographic characteristics and their diagnostic information (n=237,366)

There are 4 age groups, 2 dimensions for gender, 5 CTAS levels, and 22 disease codes. Therefore there are in total 4 x 2 x 5 x 22 = 880 groups. Patients in each group share common properties, while members in different groups tend to be dissimilar to each other in one dimension or more. During the three months examined in 2007, 576 of 880 patient types visited EDs in the city of Calgary. The number of groups with a nonzero arrival rate increased to 637 during the same period in 2017, an increase of about 10%. The histograms of the number of groups with various number of ED arrivals in the three months of the years of 2007, 2012, and 2017 are shown in Figure 4.1.

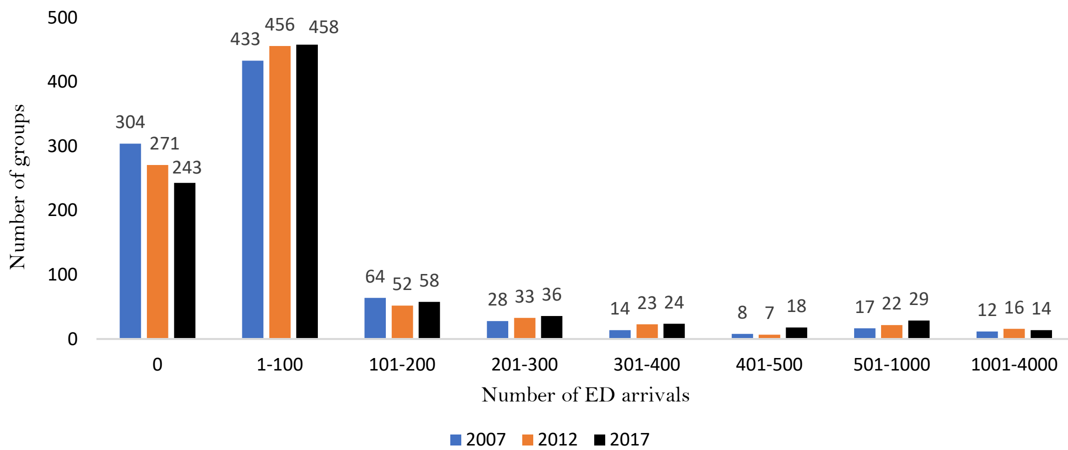

Figure 4.1: Histograms of groups by number of arrivals during the research period

It is important to know what types of patients arrive to emergency departments. Identifying the most frequent ED users, their characteristics and diagnostic information help hospital administrators in the processes of resource planning to meet the needs of different segments of patients. Figure 4.2 is a multi-level hierarchical pie chart that separates the patient population based on gender, age, CTAS, and disease attributes. Each level is represented by one circle, with each pie wedge size indicating the relative size of a dimension. The most frequent ED visits belongs to patients with the coding numbers of 31318 (>18-65 years, male, urgent triage level, and disease code 18: "symptoms, signs and abnormal clinical and laboratory findings") and 32319 (>18-65 years, female, urgent triage level, and disease code 19: "injury, poisoning and certain other consequences of external causes") with 8,376 and 7,348 patients in the population, respectively. The most requested specialties for patients with disease codes 18 and 19 were cardiologists and orthopaedic surgeons, respectively. Electrophysiologic test, chest X-ray, and abdominal ultrasound are the popular procedures required for patients with disease codes 18. Electrophysiologic test, repair injured skin, and X-ray are the common procedures needed for patients with disease codes 19.

Figure 4.2: Separation of patients in the research period

Next, we examine the relationships of patient groups with sizes greater than 100 patients to some commonly used ED evaluation variables shown in Table 4.5. Examinations of the evaluation variables indicate that patients with classification code 32415 (>18-65 years, female, resuscitation triage level, and disease code 15: "pregnancy, childbirth and the puerperium") have shown the highest percentage of revisits. The highest averages of length of stay and lab tests turnaround time belong to those patients with 32205 (>18-65 years, female, emergent triage level, and disease code 05: "mental, behavioral and neurodevelopmental disorders") and 21310 (>1-18 years, male, urgent triage level, and disease code 10: "diseases of the respiratory system") codes, respectively. The descriptions of diagnosis codes and total population results are also summarized in Table 4.5. Those with classification code 42201 (>65 years, female, emergent triage level, and disease code 01: "certain infectious and parasitic diseases") have the highest percentage of inpatients.

Evaluation variable \ Patients group	Group with highest	Value of variable	Diagnosis code description	Total population Value of variable
Revisits (%)	32415	51.8	Pregnancy, childbirth and the puerperium	7.8
Average length of stay(mins)	32205	458	Mental and behavioral disorders	361
Average lab tests turnaround time(mins)	21310	295	Respiratory diseases	110
Inpatients (%)	42201	69.9	Certain infectious and parasitic diseases	11.5

Table 4.5: ED evaluation variables and patient groups

From the case study analysis, we can address the research questions stated above.

4.4.1 RQ1: *Can we forecast ED crowding?*

Having the arrival patterns is necessary from an operational viewpoint to adjust staffing levels such that schedules are appropriate for what might be significantly different patterns of patient arrivals over time. We use patient arrivals and the patient mix to examine the patient flow in EDs through time. The first research question can be therefore divided into two sub-questions.

RQ1a: ***Is there an arrival time pattern of patients in EDs? Has it changed through time?***

Figure 4.3 displays the total number of arrivals into Calgary emergency departments in February, July and November for the years 2007, 2012, and 2017. The results show an increase of 47% in the average number of ED visits from 2007 to 2017. Furthermore, there seems to be a seasonal pattern in the arrivals data, with the highest number of arrivals belonging to the month of July and the lowest in the month of February.

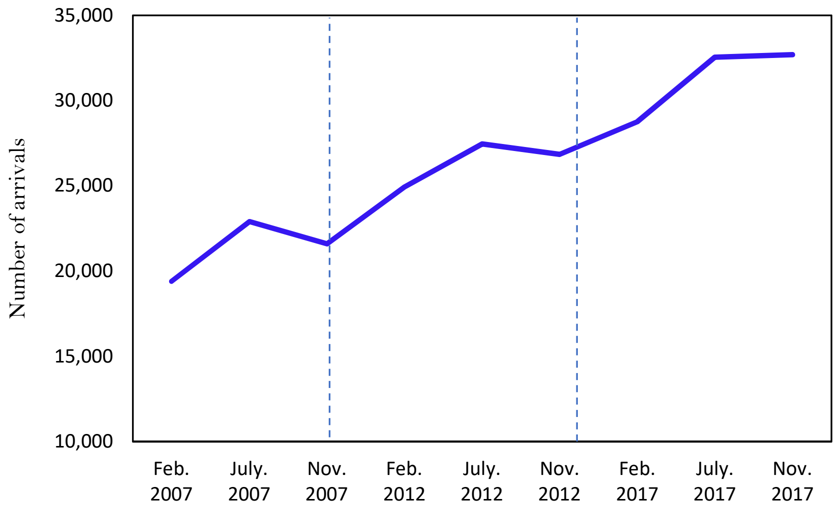

Figure 4.3: ED patient arrivals by month for 2007, 2012, and 2017

Next, we examine daily arrival pattern to find which hour of the day results in the highest arrivals. The hourly trend of arrivals for each day of the week is plotted in Figure 4.4.

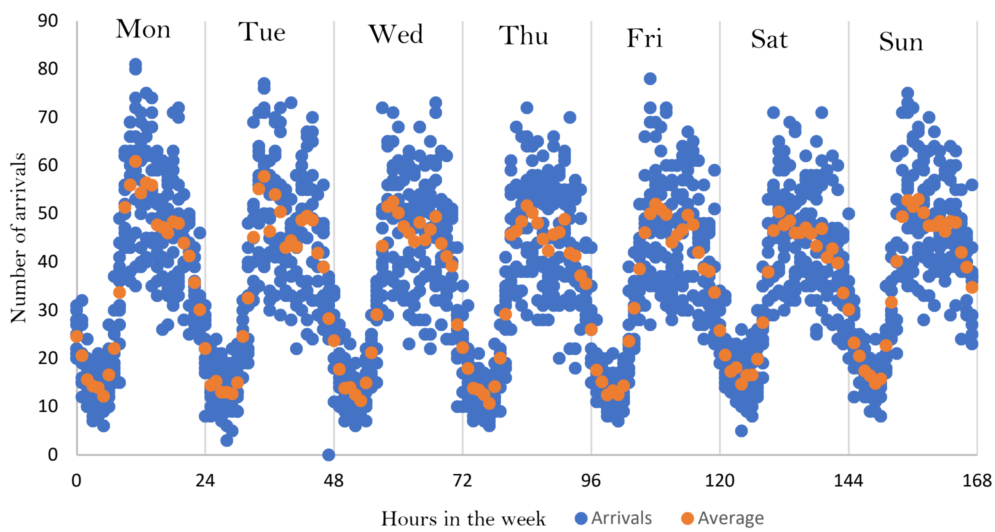

Figure 4.4: ED patient arrivals (hourly trend)

Figure 4.4 shows a 24-hour cycle with a peak around midday that decreases toward the evening and rises again throughout the morning. It is interesting to note that there is no major difference between the pattern of arrivals during the week and on the weekend, although the average number of ED visits is higher on Mondays, followed by Sundays.

Forecasting demand for ED resources is contingent upon the arrival distribution of incoming patients. Most queuing formulas are based on the assumption that the arrival times of patients follow a Poisson process because of its mathematical convenience. We examine this assumption by using a chi-square goodness-of-fit test. We calculate the hourly arrival rates over a time span of 24 hours, as the data suggests that hour of the day is the most important predictor in ED arrival rates. To remove the seasonal component in the data, we have chosen one month, February. We hypothesize that the number of patient arrivals in each hour follows a Poisson distribution whose mean is the average number of arrivals during that hour of all days in February. This gives a total of 28 arrival data points for each hour of the day in February in a given year.

H0a: The number of arrivals per hour follows a Poisson distribution

The chi-square goodness of fit is used to test the null hypothesis for each hour in each of the three years of data, for a total of 72 tests. The computed P-values are greater than the significance level α=0.05 in 93% of the tests (67 out of 72) indicating that one can not reject the null hypothesis. Time intervals h1 and h8 in 2007, h13 in 2012, and h22 and h23 in 2017 did not pass the chi-square test. The conclusion is that the Poisson process is a reasonable assumption for predicting hourly arrival distributions of incoming patients. Figure 4.5 compares the average hourly arrival rates of February in the years 2007, 2012, and 2017, showing an increasing similar trend on arrival rates per hour over years. The overall average hourly patient arrivals has increased from 28 patients in February of 2007 to 36 patients in February of 2012 and 43 patients in February of 2017.

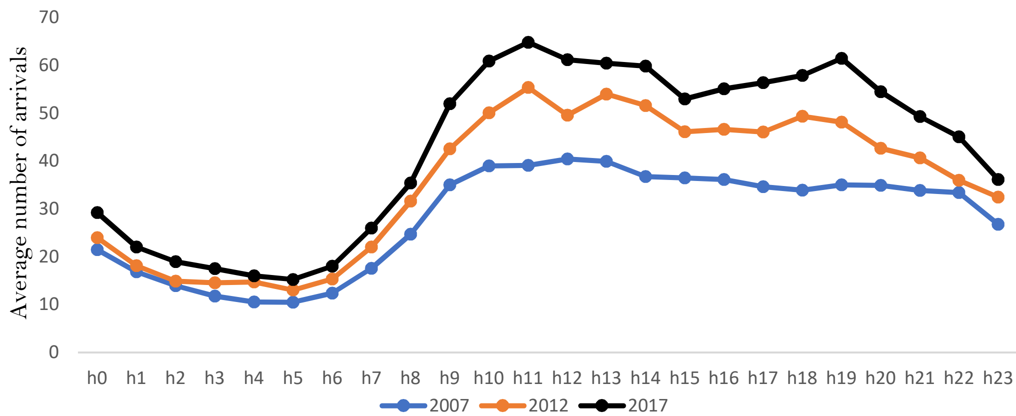

Figure 4.5: Average hourly arrivals in February of 2007, 2012, and 2017

ED average hourly visit rates with the sample standard errors (SE) calculated across the days of February in the years 2007, 2012, and 2017 are summarized in Table 4.6.

Time	Average ED visits (SE) February 2007	Average ED visits (SE) February 2012	Average ED visits (SE) February 2017
h0	21.1 (0.8)	24 (3.5)	26.5 (2.6)
h1	16.1 (3.4)	19.5 (2.4)	20.6 (3.1)
h2	14.3 (2.2)	14.8 (1.3)	17.1 (3.1)
h3	12 (2.0)	15.7 (2.0)	17.6 (1.6)
h4	11.2 (1.9)	15.4 (1.9)	17.2 (2.1)
h5	10.8 (1.6)	13.1 (2.6)	12.8 (2.5)
h6	12.6 (1.7)	14.9 (2.0)	16.6 (3.0)
h7	16.7 (2.2)	21.7 (1.8)	25.6 (4.3)
h8	23.2 (3.8)	30.4 (2.3)	34.8 (3.3)
h9	32.4 (3.4)	42.9 (4.5)	51.4 (4.1)
h10	37.3 (2.9)	48 (3.6)	59.5 (4.5)
h11	38.7 (2.9)	54.4 (3.3)	60.6 (6.1)
h12	39.2 (3.3)	49.3 (4.9)	60.4 (3.8)
h13	39.2 (3.2)	50.1 (4.3)	59.6 (3.8)
h14	36.8 (3.3)	50.8 (4.1)	55.9 (5.2)
h15	36.8 (4.0)	44.1 (4.3)	49.4 (4.9)
h16	35.1 (3.0)	44.2 (3.4)	54.5 (4.7)
h17	34.6 (3.6)	46.7 (3.2)	56.1 (3.2)
h18	31.8 (3.6)	47.9 (3.3)	56.7 (3.0)
h19	35.1 (2.7)	48.8 (2.1)	63.1 (4.2)
h20	33.8 (3.7)	43.8 (2.8)	52.2 (4.2)
h21	31.2 (4.0)	39.5 (2.5)	50 (4.3)
h22	33.8 (3.5)	35.8 (3.3)	41.5 (5.0)
h23	28.1 (2.3)	31.7 (3.6)	32.7 (3.5)

Table 4.6: Emergency departments hourly visit rates with standard errors

RQ1b: ***Is there an arrival time pattern of patient types in EDs? Has it changed through time?***

There are 880 patient types based on the PCS proposed in the previous section. Figure 4.6 displays the number of patient types that visited Calgary emergency departments in February, July and November for the years 2007, 2012, and 2017. Similar to the ED arrival pattern, there seems to be a seasonal trend in the number of patient mix that visits EDs. The peak occurs in July, while the lowest belongs to the month of November. This suggests that the arrival time pattern of patient mix can be different than the arrival pattern of overall patients.

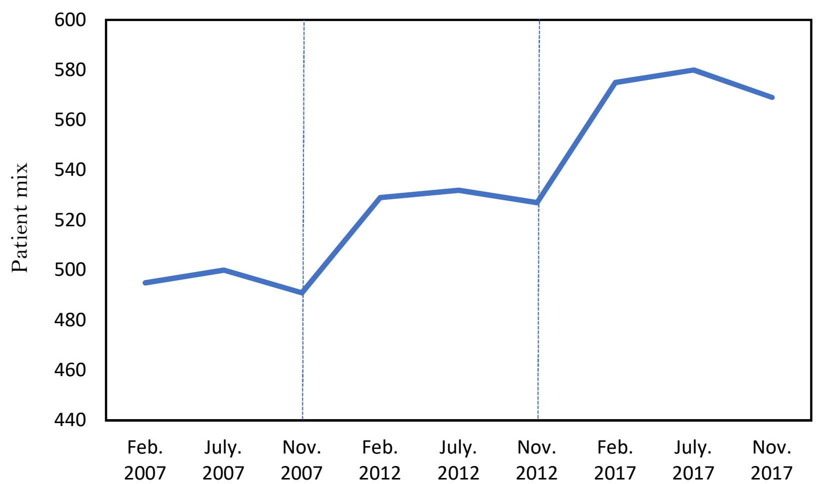

Figure 4.6: Patient mix by month for 2007, 2012, and 2017

Next, we examine daily arrival patterns for an example patient group to find which hour of the day results in the highest arrivals. We select the patient group with the coding numbers of 31318 (>18-65 years, male, urgent triage level, and disease code 18: "symptoms, signs and abnormal clinical and laboratory findings"), which has the most frequent ED visits during the studied period. The hourly trend of arrivals for each day of the week for patients from this group is plotted in Figure 4.7. Similar to ED arrival pattern, Figure 4.7 shows a 24-hour cycle with a peak around midday that decreases toward the evening and rises again throughout the morning. But dissimilar to the total arrivals, the average number of arrivals for patients in this group is higher on Tuesdays, followed by Mondays. This figure suggests that hour of the day is the most important predictor of arrivals even for patients from one group.

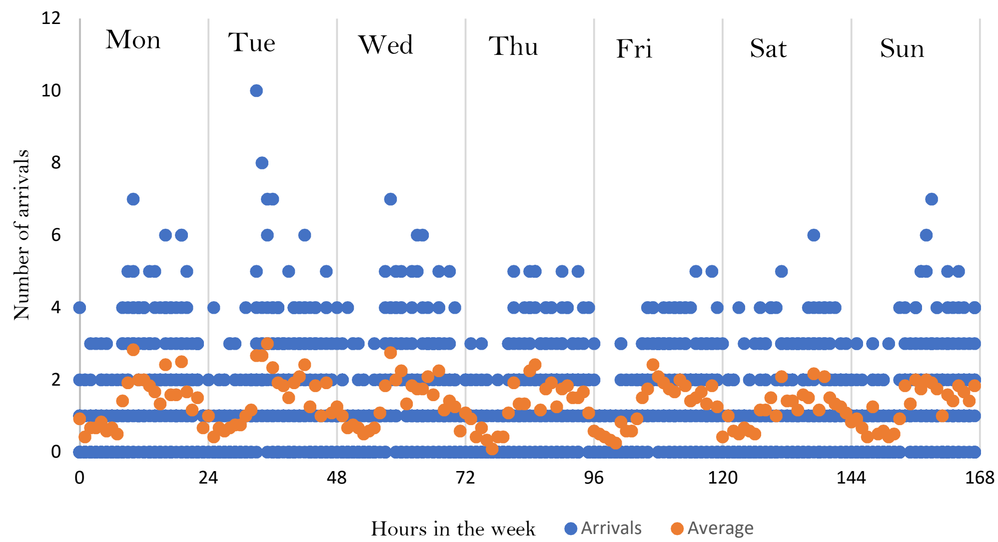

Figure 4.7: ED arrivals for group with the coding numbers of 31318 (hourly trend)

In the previous section, we showed that the Poisson process is a reasonable assumption for predicting hourly arrival distributions of incoming patients. In this section, we examine whether Poisson distributions can be used to predict hourly arrivals of patients from one group. This is important in understanding the variation of arrivals even in a given hour of the day.

H0b: The arrivals of a particular patient group in a given hour of the day follows a Poisson distribution

To test the null hypothesis, we examine hourly patient arrivals of the 5 largest groups in February 2007, 2012, and 2017 (5 groups x 24hrs x 3yrs = 360 chi-square tests). Table 4.7 lists times of the day that we have to reject the null hypothesis (27 out of 360). The last column of Table 4.7 then shows the number of times that the Poisson distribution fits the hourly arrivals for each group. The results suggest that using the Poisson distribution is a reasonable assumption even in predicting hourly arrivals for a particular group of patients.

Group \ Year	2007	2012	2017	Passed the test (out 0f 72 times)
31318	h4, h13, h23	h5, h9	h2, h3, h5	64
32319	h4, h6, h18, h21	h13, h14, h17	h2	64
32318	h1, h12	h6	-	69
31319	h1, h9	h16	h7, h14	67
31218	h4, h20	h9	-	69

Table 4.7: Times of the day that the null hypothesis is rejected

4.4.2 RQ2: *Are there any relationships between crowding variables and performance metrics? Should the crowding measure be redefined by adding the patient mix?*

In the previous section, we presented the two ED crowding variables and examined their patterns of arrivals. In this section, we first present and examine the changes in the values of performance metrics in five hospitals as case study sites. Four are adult EDs, whereas ED 5 provides care for sick children. Hospital 4 opened its adult-focused emergency department in 2013. In the second part of the section, we then assess the relationships between the two crowding variables and the suggested performance metrics to show the significance of the proposed crowding metrics.

ED performance metrics

The first performance measure suggested in this study is the average percentages of revisits within 72 hours of the last ED discharge, which is a proxy of quality of care in each ED. Over the three months examined in 2007, 2012, and 2017, we find a decrease of 1.9% in average percentages of revisits (from 9.5% in 2007 to 7.6% in 2012) followed by a 0.9% decrease in 2017 (6.7% in 2017).

The next two performance measures studied over the same time period are average length of stay and average wait time. The length of stay (LOS) is the difference between the first contact (triage time) and admission time for inpatients or departure time for all others. The wait time is, however, calculated by the difference between the first contact (triage time) and the time of initial assessment by a physician. Figure 4.8 compares the average LOS with average wait time in all EDs over the three months examined in 2007, 2012, and 2017. The results show opposite trends between average LOS and average wait time over the examined period. It seems that opening of the new hospital could change the upward trends in waiting times in 2017. However, it is interesting that opening the new hospital did not decrease average patients' LOS in 2017 compared to 2012, as might be expected.

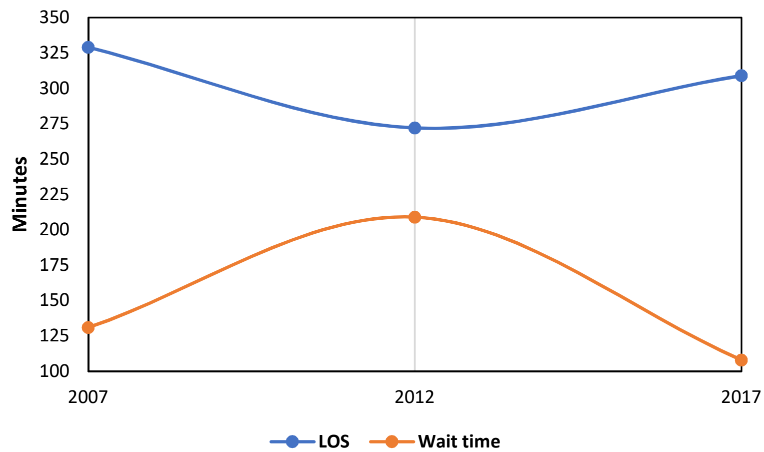

Figure 4.8: Average length of stay (LOS) vs. average wait time

The relationships between the crowding variables and suggested ED performance metrics

So far, we have presented the proposed crowding variables and performance metrics. In this section, we perform a series of multiple regression analyses to assess the relationships between crowding variables and ED performance metrics. The data from four adult EDs in February 2017 is used to examine the relationships across conditions/locations of different EDs (sample size = 4(EDs) x 28(days) = 112). We use the data from February since the number of arrivals and the month indicator variables were highly correlated. The predictor variables and multiple regression models are then validated using the data from November 2017. First, we examine how well number of patients, patient mix, and location of hospital's ED predict patient wait times. The wait time is the time of triage to the physician initial assessment time.

H0c: There is no relationship between wait time and the number of patients, patient mix, and ED location

A series of multiple linear regression analyses are carried out to test the null hypothesis. If a predictor in regression is statistically insignificant, we remove that variable and repeat the test. This procedure continues until all predictors are statistically significant at the 5% level. Table 4.8 displays the coefficient values of the model predictors and their significance levels for the first and final runs, coming from both the original and validation sets. The final regression results of both the original and validation samples suggest that the best predictor of wait time is the number of arrivals. Patient mix and ED conditions/locations are not significantly associated with the wait time. The original model explains 95.5% of the variation in wait time (F(1,111)=2980.8 and p-value of less than 0.001). The validation model explains 91.7% of the variation in wait time (F(1,111)=1399.2 and p-value of less than 0.001).

	Original model		Validation model	
Variable	Coefficient value β First run	Coefficient value β (SE) Final run	Coefficient value β First run	Coefficient value β (SE) Final run
Intercept	19.86	–	-56.24	–
Arrivals	0.42***	0.53*** (0.2)	0.57*	0.55*** (0.3)
Patient mix	0.02	–	0.46	–
Hospital 1	-5.73	–	5.48	–
Hospital 2	3.26	–	6.51	–
Hospital 3	-6.62	–	4.45	–

***Significant at 1% level. **Significant at 5% level. *Significant at 10% level

Table 4.8: Coefficient values and standard errors (SE) for wait time predictors

Second, we examine whether the patient LOS depends on the number of patients, patient mix, and ED location. The patient LOS is the total time a patient spends in the ED from the time of triage to the time that a disposition decision is made.

H0d: There is no relationship between LOS and the number of patients, patient mix, and ED location

The final regression results of the original model shown in Table 4.9 suggest that the key predictors of LOS are number of arrivals and ED location. ED location in a particular is the only predictor of LOS in the validation model. The original model explains only 19.8% of the variation in LOS (F(3,108)=10.1 and p-value of less than 0.001). The validation model also explains only 24.2% of the variation in LOS (F(3,108)=12.8 and p-value of less than

0.001).

	Original model		Validation model	
Variable	Coefficient value β First run	Coefficient value β (SE) Final run	Coefficient value β First run	Coefficient value β (SE) Final run
Intercept	221.71***	203.03*** (61.3)	208.61***	276.89*** (12.2)
Arrivals	0.48**	0.38** (0.3)	0.21	–
Patient mix	-0.36	–	0.26	–
Hospital 1	26.00**	30.64*** (14.9)	34.02***	38.33*** (17.0)
Hospital 2	11.80	15.50** (14.6)	47.09***	50.67*** (17.0)
Hospital 3	-4.00	–	17.92	24.40*** (17.0)

***Significant at 1% level. **Significant at 5% level. *Significant at 10% level

Table 4.9: Coefficient values and standard errors (SE) for LOS predictors

Finally, we assess whether the number of revisits within 72 hours, which is a proxy for quality of care, depends on the number of patients, patient mix, and ED condition/location.

H0e: There is no relationship between quality and the number of patients, patient mix, and ED location

Table 4.10 displays coefficient values of the model predictors and their significance levels for the first and final runs, coming from both the original and validation sets. The final regression results of both the original and validation samples suggest that the best predictor of revisits are patient mix and ED condition. The original model explains 88.8% of the variation in revisits($F(4,108)=243.8$ with p-value of less than 0.001). The validation model explains 88.8% of the variation in revisits($F(4,108)=247.3$ with p-value of less than 0.001).

	Original model		Validation model	
Variable	Coefficient value β First run	Coefficient value β (SE) Final run	Coefficient value β First run	Coefficient value β (SE) Final run
Intercept	1.19	–	-0.72	–
Arrivals	4.38E-03	–	0.01	–
Patient mix	0.03	0.05*** (0.01)	0.05	0.05*** (0.01)
Hospital 1	0.97	1.19** (1.14)	1.07	1.22*** (1.19)
Hospital 2	1.36*	1.53** (1.15)	1.62**	1.74*** (1.20)
Hospital 3	1.54*	1.70*** (1.17)	1.45	1.62*** (1.21)

***Significant at 1% level. **Significant at 5% level. *Significant at 10% level

Table 4.10: Coefficient values and standard errors (SE) for quality of care predictors

The regression results suggest that the patient mix can not be ignored in defining an ED crowding measure. Thus, we define ED crowding as the number of patients arriving into EDs from each group. As already discussed, our proposed ED crowding metrics can be predicted by the Poisson distributions.

4.5 Conclusion and future research

In this study, we introduced a new crowding measure for emergency departments (EDs) as functions of patient arrivals and mix of patients. The most commonly used crowding measure in the literature is patient volume that ignores the mix of patients visiting EDs. To show the applicability and significant of the suggested ED crowding variables, we first examined the arrival patterns of incoming patients and types of patients and second assessed the relationships between crowding variables and performance metrics. We used data from five hospitals in February, July and November for the years of 2007, 2012, and 2017 in the city of Calgary, Alberta, Canada.

In terms of ED patient arrivals, we found a seasonal pattern in the arrivals data with the highest number of arrivals in the month of July and the lowest in the month of February. We found that the hour of the day is the key predictor of ED arrivals. However, there was no major difference between the pattern of arrivals during the week and on the weekend. Moreover, the statistical tests showed that the Poisson process is a reasonable assumption for predicting hourly arrival distribution of incoming patients. In terms of patient mix, the data showed a seasonal trend in the patient mix arrivals with the peak occurring in July. However, the lowest number of patient mix belonged to the November that was dissimilar to the trend of arrivals. Next, we examined the arrival rate probability distribution as a function of patient type. The results suggested that the Poisson distribution is a reasonable assumption even in predicting hourly arrivals for a particular group of patients.

To examine the significance of the crowding metrics, we then assessed the relationship between the crowding metrics and performance metrics. We found the following: i) the number of ED arrivals is a predictor of wait time and total length of stay, and ii) patient mix is a predictor of quality of care in EDs.

ED crowding measures have been widely studied in the literature. Despite the negative impact of patient mix on the quality of care discussed in this study, we find no measure in the ED literature of crowding that includes the patient mix. We concluded that a new definition of crowding should include both patient arrivals and patient mix, and therefore we defined ED crowding as the number of patients arriving into EDs from each group. Further studies can develop a quantitative ED crowding measure that is based on both patient arrivals and the patient mix.

Chapter 5

Summary and Conclusion

5.1 Summary

Improving patient flow and providing high service quality are the major challenges in healthcare facilities. It is important to define and evaluate service quality, although patients might have different perceptions and expectation of service quality. The aim of this dissertation was to understand, evaluate, and improve the flow patients as well as the quality of services in emergency departments and outpatients clinics. We addressed two tactical-level problems in chapters 2 and 3 of to improve patient experiences in healthcare services. In chapter 2, we studied a doctor-clinic assignment problem in medical outpatient clinics, and in chapter 3, we focused on a scheduling problem of physicians in emergency departments. At the operational decision level, we introduced a new crowding measure and showed the importance of applying that in emergency departments. We summarize each chapter as follows.

In chapter 2, we addressed the doctor-clinic assignment problem in coordinated lab-clinics settings in order to improve the service level. we defined the service level as the probability that the sum of all patients' testing times is less than or equal to the available testing time. We provided two mini-cases based on clinics' actual operations. The results showed that our proposed scheduling technique can bring important improvements to these systems. We understand that doctors are often constrained by a schedule of teaching, research, or service requirements in addition to their clinical activities. However, the results suggested that moving only one or two clinics, out of many, can bring a significantly higher standard of service for patients. The optimal schedule was also more robust to changes in lab capacity.

In chapter 3, we studied the scheduling problem of physicians to improve patient flow through emergency departments. We developed a two-stage stochastic mixed-integer programming to formulate the problem. We then presented a case study and a simulation model to evaluate the impact of the near-optimal stochastic MIP solution on reducing patient wait times. We showed that our proposed schedule can reduce wait times by matching hourly patient arrivals and ED physicians' productivity during each hour.

In chapter 4, we presented a new crowding measure in emergency departments based on patient volume and mix of patients. To show the significance of this measure, we examined the relationships between the two crowding variables and some commonly used ED performance metrics. The results suggested that number of ED arrivals is the key predictor of time-based ED performance measures. Patient mix is, however, the key predictor of quality of care in EDs.

5.2 Future outlook

In this section, I highlight some potential research questions and briefly picture the future works.

While we have studied the tactical level of coordinated lab-clinics setting in chapter 2, several research questions remain open. Further studies can address the problem of patient sequencing in the test center at the operational level. At the strategic level, one can find the optimal number of physicians and the available time required in the test center. Implementation of the findings is our next step in this research. We are working closely with respiratory clinics at a hospital in the city of Calgary, Alberta, Canada to improve service quality. Patient pulmonary function testing (PFT) is linked to the clinic appointment in that different types of tests must be completed prior to the patient's appointment with the respirologist.

In chapter 3, it is also interesting to study the shift scheduling problem and staffing at the strategic level by considering the differences among ED physicians in terms of their productivity rates. We are working with an emergency department scheduler in the city of Calgary to implement the findings of this chapter.

In chapter 4, we have discussed the negative impact of patient mix on the quality of care. Further studies can develop a quantitative ED crowding measure in EDs that is based on both patient arrivals and the patient mix.

Appendix A: ICD-10-CM Codes

ICD-10 Codes

1. A00-B99 Certain infectious and parasitic diseases
2. C00-D49 Neoplasms
3. D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4. E00-E89 Endocrine, nutritional and metabolic diseases
5. F01-F99 Mental, Behavioral and Neurodevelopmental disorders
6. G00-G99 Diseases of the nervous system
7. H00-H59 Diseases of the eye and adnexa
8. H60-H95 Diseases of the ear and mastoid process
9. I00-I99 Diseases of the circulatory system
10. J00-J99 Diseases of the respiratory system
11. K00-K95 Diseases of the digestive system
12. L00-L99 Diseases of the skin and subcutaneous tissue
13. M00-M99 Diseases of the musculoskeletal system and connective tissue
14. N00-N99 Diseases of the genitourinary system
15. O00-O9A Pregnancy, childbirth and the puerperium
16. P00-P96 Certain conditions originating in the perinatal period
17. Q00-Q99 Congenital malformations, deformations and chromosomal abnormalities
18. R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19. S00-T88 Injury, poisoning and certain other consequences of external causes
20. U00-U99 Codes for special purposes
21. V00-Y99 External causes of morbidity
22. Z00-Z99 Factors influencing health status and contact with health services

Appendix B: Copyright Permissions

Copyright Clearance Center RightsLink®

Home | Create Account | Help

Title: Coordinated lab-clinics: A tactical assignment problem in healthcare

Author: Farzad Zaerpour,Diane P. Bischak,Mozart B.C. Menezes

Publication: European Journal of Operational Research

Publisher: Elsevier

Date: 16 November 2017

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials.
Already **a RightsLink user** or want to learn more?

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK | CLOSE WINDOW

Bibliography

Ajmi, I., Zgaya, H., Gammoudi, L., Hammadi, S., Martinot, A., Beuscart, R., and Renard, J. M. 2015. Mapping patient path in the Pediatric Emergency Department: A workflow model driven approach. *Accident and Emergency Nursing*, 54, 315–328.

Azaiez, M. N. and Al Sharif, S. S. 2005. A 0-1 goal programming model for nurse scheduling. *Computers and Operations Research*, 32 (3), 495–507.

Barak-Corren, Y., Israelit, S., and Reis, B. Y. 2017. Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. *Emergency Medicine Journal*, 34 (5), 308–314.

Baron, O., Berman, O., Krass, D., and Wang, J. 2016. Strategic idling and dynamic scheduling in an open-shop service network: case study and analysis. *Manufacturing Service Operations Management*, forthcoming.

Batt, A. J. and Terwiesch, M. 2014. Doctors Under Load: An Empirical Study of State Dependent Service Times in Emergency Care. Working paper, The Wharton School.

Beaulieu, H., Ferland, J. A., Gendron, B., and Michelon, P. 2000. A mathematical programming approach for scheduling physicians in the emergency room. *Health Care Management Science*, 3 (3), 193–200.

Benavides, A. J., Ritt, M., and Miralles, C. 2014. Flow shop scheduling with heterogeneous workers. *European Journal of Operational Research*, 237 (2), 713–720.

Boffeli, T. J., Thongvanh, K. L., Evans, S. J., and Ahrens, C. R. 2012. Patient experience and physician productivity: debunking the mythical divide at Health Partners clinics. *The Permanente Journal*, 16 (4), 19–25.

Brennan, D. F., Silvestri, S., Sun, J. Y., and Papa, L. 2007. Progression of emergency medicine resident productivity. *The Permanente Journal*, 14 (9), 790–794.

Bruin, A. M. de, Rossum, A. C. van, Visser, M. C., and Koole, G. M. 2007. Modeling the emergency cardiac in-patient flow: an application of queuing theory. *Health Care Management Science*, 10 (2), 125–137.

Brunner, J. O., Bard, J. F., and Kolisch, R. 2009. Flexible shift scheduling of physicians. *Health Care Management Science*, 12 (3), 285–305.

Bullard, M. J., Villa-Roel, C., Bond, K., Vester, M., Holroyd, B. R., and Rowe, B. H. 2009. Tracking emergency department overcrowding in a tertiary care academic institution. *Healthcare Quarterly*, 12, 99–106.

Butler, W. R. 1986. ED patient classification matrix: development and testing of a tool. *Journal of Emergency Nursing*, 12, 279–285.

Cabrera, E., Taboada, M., Iglesias, M. L., Epelde, F., and Luque, E. 2001. Optimization of Healthcare Emergency Departments by Agent-Based Simulation. *Procedia Computer Science*, 4, 1880–1889.

Caldwell, J. A. 2001. The impact of fatigue in air medical and other types of operations: A review of fatigue facts and potential countermeasures. *Air Medical Journal*, 20 (1), 25–32.

Caprara, A., Kellerer, H., and Pferschy, U. 2000. The multiple subset sum problem. *SIAM Journal on Optimization*, 11 (2), 308–319.

— 2003. A 3/4-approximation algorithm for multiple subset sum. *J. Heuristics*, 9 (2), 99–111.

Carello, G. and Lanzarone, E. 2014. A cardinality-constrained robust model for the assignment problem in Home Care services. *European Journal of Operational Research*, 236 (2), 748–762.

Carter, M. and Lapierre, S. D. 2001. Scheduling emergency room physicians. *Health Care Management Science*, 4 (4), 347–360.

Carvalho-Silva, M., Monteiro, M. T. T., De Sá-Scares, F., and Dória-Nóbrega, D. 2017. Assessment of forecasting models for patients arrival at emergency department. *Operations Research for Health Care*, forthcoming.

Cayirli, T. and Veral, E. 2003. Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, 12 (4), 519–549.

Choi, S. and Wilhelm, W. E. 2014. An approach to optimize block surgical schedules. *European Journal of Operational Research*, 235 (1), 138–148.

CIHI 2017. How Canada Compares: Results From The Commonwealth Fund's 2016 International Health Policy Survey of Adults in 11 Countries. *Accessible Report*, Ottawa, ON: CIHI; 2017.

Cochran, J. K. and Roche, K. T. 2009. A multi-class queueing network analysis methodology for improving hospital emergency department performance. *Computers and Operations Research*, 36, 1497–1512.

Coelli, F. C., Ferreira, R. B., Almeida, R. M., and Pereira, W. C. 2007. Computer simulation and discrete-event models in the analysis of a mammography clinic patient flow. *Pediatric Emergency Care*, 87 (3), 201–207.

Conners, A. M. 1994. Patient classification system in a rural emergency department. *Accident and Emergency Nursing*, 2, 7–20.

Cótẽ, M. J. 1999. Patient flow and resource utilization in an outpatient clinic. *Socio-Economic Planning Sciences*, 33 (3), 231–245.

Crouch, R. and Williams, S. 2006. Patient dependency in the emergency department (ED): reliability and validity of the Jones dependency tool (JDT). *Accident and Emergency Nursing*, 14, 219–229.

Ćustić, A., Klinz, B., and Woeginger, G. J. 2015. Geometric versions of the three-dimensional assignment problem under general norms. *Discrete Optimization*, 18 (C), 38–55.

Dantzig, G. B. and Wolfe, P. 1960. Decomposition principle for linear programs. *Operations Research*, 8 (1), 101–111.

Deveau, J. P., Lorenz, J. E., and Hughes, M. J. 2003. Emergency medicine resident work productivity and procedural accomplishment. *Journal of the American Osteopathic Association*, 103 (6), 291–296.

Ding, R., McCarthy, M. L., Li, G., Kirsch, T. D., Jung, J. J., and Kelen, G. D. 2016. Patients who leave without being seen: their characteristics and history of emergency department use. *Annals of Emergency Medicine*, 48 (6), 686–693.

Domenech, B. and Lusa, A. 2016. A MILP model for the teacher assignment problem considering teachers' preferences. *European Journal of Operational Research*, 249 (3), 1153–1160.

Dowd, M. D., Tarantino, C., Barnett, T. M., Fitzmaurice, L., and Knapp, J. F. 2005. Resident efficiency in a pediatric emergency department. *Academic Emergency Medicine*, 12, 1240–1244.

Dowsland, K. A. 1998. Nurse scheduling with tabu search and strategic oscillation. *European Journal of Operational Research*, 106 (2-3), 393–407.

Dreesch, N., Dolea, C., Dal Poz, M., Goubarev, A., Adams, A., Aregawi, M., Bergstrom, K., Fogstad, H., and Sheratt D. Linkins, J. 2005. An approach to estimating the human resource requirements to achieve the MDGs. *Health Policy and Plan*, (20), 267–276.

El Adoly, A. A., Gheith, M., and Fors, N. M. 2018. A new formulation and solution for the nurse scheduling problem: A case study in Egypt. *Alexandria Engineering Journal*, (20).

Epstein, S. K., Huckins, D. S., Liu, S. W., Pallin, D. J., Sullivan, A. F., Lipton, R. I., and Camargo, C. A. 2012. Emergency department crowding and risk of preventable medical errors. *Internal and Emergency Medicine*, 7 (2), 173–180.

Excoffier, M., Gicquel, C., and Jouini, O. 2016. A joint chance-constrained programming approach for call center workforce scheduling under uncertain call arrival forecasts. *Computers and Industrial Engineering*, 96, 16–30.

Fei, H., Chu, C., and Meskens, N. 2009. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research*, 166 (1), 91–108.

Ferrand, Y., Magazine, M., Rao, U. S., and Glass, Y. F. 2011. Building cyclic schedules for emergency department physicians. *Interfaces*, 41 (6), 521–53.

Gans, N., Liu, N., Mandelbaum, A., Shen, H., and Ye, H. 2012. Service times in call centers: Agent heterogeneity and learning with some operational consequences. *A Festschrift for Lawrence D. Brown, IMS Collections*, 6, 99–123.

Garey, M. R. and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman, New York.

Gelareh, S., Monemi, R. N., Semet, F., and Goncalves, G. 2015. A branch-and-cut algorithm for the truck dock assignment problem with operational time constraints. *European Journal of Operational Research*, 249 (3), 1144–1152.

Gendreau, M., Ferland, J., Gendron, B., Hail, N., Jaumard, B., Lapierre, S., Pesant, G., and Soriano, P. 2007. Physician Scheduling in Emergency Rooms. *Practice and Theory of Automated Timetabling VI. Springer, Springer-Verlag Berlin Heidelberg*,

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

Gupta, D. 2008. Surgical suites' operations management. *Production and Operations Management*, 16 (6), 689–700.

Hahn, P. M., Kim, B. J., Stützle, T., Kanthak, S., Hightower, W. L., Samra, H., Ding, Z., and Guignard, M. 2008. The quadratic three-dimensional assignment problem: Exact and approximate solution methods. *European Journal of Operational Research*, 184 (2), 416–428.

Hans, E., Wullink, G., Van Houdenhoven, M., and Kazemier, G. 2008. Robust surgery loading. *European Journal of Operational Research*, 185 (3), 1038–1050.

Hans, E. W. and Vanberkel, P. T. 2012. *Operating theatre planning and scheduling*. In (ed. R. Hall) Handbook of Healthcare System Scheduling, Springer.

Herring, A. A., Ginde, A. A., and Fahimi, J. 2013. Increasing Critical Care Admissions From US Emergency Departments, 2001-2009. *Critical Care Medicine*, 41 (5), 1197–1204.

Hojati, M. and Patil, A. S. 2011. An integer linear programming-based heuristic for scheduling heterogeneous, part-time service employees. *European Journal of Operational Research*, 209 (1), 37–50.

Hoot, N. R. and Aronsky, D. 2008. Systematic review of emergency department crowding: causes, effects, and solutions. *Annals of Emergency Medicine*, 52 (2), 126–136.

Horwitz, L. I., Green, J., and Bradley, E. H. 2010. United States emergency department performance on wait time and length of visit. *Annals of Emergency Medicine*, 55 (2), 133–141.

Hung, G. R., Whitehouse, S. R., O'Neill, C., Gray, A. P., and Kissoon, N. 2016. Computer modeling of patient flow in a pediatric emergency department using discrete event simulation. *Pediatric Emergency Care*, 23 (1), 5–10.

Hwang, U. et al. 2011. Measures of crowding in the emergency department: a systematic review. *Academic Emergency Medicine*, 18, 527–538.

Ibrahim, R., L'Ecuyer, P., Shen, H., and Thiongane, M. 2016. Inter-dependent, heterogeneous, and time-varying service-time distributions in call centers. *European Journal of Operational Research*, 250 (2), 480–492.

Joseph, J. W., Davis, S., Wilker, E. H., Wong, M. L., Litvak, O., Traub, S. J., Nathanson, L. A., and Sanchez, L. D. 2018. Modelling attending physician productivity in the emergency department: a multicentre study. *Emergency Medicine Journal*, 35 (5), 317–322.

KC, D. S. and Terwiesch, C. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management science*, 55 (9), 1486–1498.

Klassen, K. J. and Yoogalingam, R. 2008. An assessment of the interruption level of doctors in outpatient appointment scheduling. *Operations Research*, 1 (2), 95–102.

Kreitz, T. M., Winters, B. S., and Pedowitz, D. I. 2016. The Influence of Wait Time on Patient Satisfaction in the Orthopedic Clinic. *Journal of Patient Experience*, 3 (2), 39–42.

Kucukyazici, B., Verter, V., and Mayo, N. 2011. An analytical framework for designing community-based care delivery processes for chronic diseases. *Production and Operations Management*, 20 (3), 474–488.

Lamarche, P. A., Pineault, R., Gauthier, J., Hamel, M., and Haggerty, J. 2011. Availability of healthcare resources, positive ratings of the care experience and extent of service use: an unexpected relationship. *Internal and Emergency Medicine*, 6, 46–56.

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

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

Lewis, C. W., McKinney, C. M., Lee, H. H., Melbye, M. L., and Rue, T. C. 2015. Visits to US Emergency Departments by 20 to 29 Year-olds with Chief Complaint of Toothache during 2001-2010. *Journal of the American Dental Association*, 146 (5), 295–302.

Maenhout, B. and Vanhoucke, M. 2013. An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems. *Omega*, 41 (2), 485–499.

Martello, S. and Toth, P. 1984. A mixture of dynamic programming and branch-and-bound for the subset-sum problem. *Management Science*, 30 (6), 765–771.

Martin, M., Champion, R., Kinsman, L., and Masman, K. 2011. Mapping patient flow in a regional Australian emergency department: a model driven approach. *International Emergency Nursing*, 19 (2), 75–85.

McCarthy, M. L., Zeger, S. L., Ding, R., Aronsky, D., Hoot, N. R., and Kelen, G. D. 2008. The challenge of predicting demand for emergency department services. *Academic Emergency Medicine*, 15 (4), 337–346.

McKinnon, K., Champion, P. D., and Edwards, R. H. T. 1998. The outpatient experience: Results of a patient feedback survey. *International Journal of Health Care Quality Assurance*, 11 (5), 156–160.

McNaughton, C. D., Self, W. H., Zhu, Y., Janke, A. T., Storrow, A. B., and Levy, P. D. 2015. Incidence of hypertension-related emergency department visits in the United States, 2006 to 2012. *American Journal of Cardiology*, 116 (11), 1717–1723.

Mehrotra, V., Ozlük, O., and Saltzman, R. 2010. Intelligent procedures for intra-day updating of call center agent schedules. *Production and Operations Management*, 19 (3), 353–367.

Mehrotra, V., Ross, K., Ryder, G., and Zhou, Y. 2012. Routing to manage resolution and waiting time in call centers with heterogeneous servers. *Manufacturing Service Operations Management*, 14 (1), 66–81.

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

Morganti, K. G., Bauhoff, S., Blanchard, J. C., Abir, M., Iyer, N., Smith, A., Vesely, J. V., Okeke, E. N., and Kellermann, A. L. 2013. The Evolving Role of Emergency Departments in the United States. *RAND Health Quarterly*, 3 (2).

Neal, R. D. et al. 2015. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *British Journal of Cancer*, 112 (1), 92–107.

Nguyen, D. L., DeJesus, R. S., and L., Wieland. M. 2011. Missed appointments in resident continuity clinic: patient characteristics and health care outcomes. *Journal of Graduate Medical Education*, 3 (3), 350–355.

Pentico, D. W. 2007. Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, 176 (2), 774–793.

Rabinowitz, H. K., Diamond, J. J., Markham, F. W., and Hazelwood, C. E. 1999. A program to increase the number of family physicians in rural and underserved areas: impact after 22 years. *Internal and Emergency Medicine*, 281 (3), 255–260.

Robbins, T. R. and Harrison, T. P. 2010. A stochastic programming model for scheduling call centers with global Service Level Agreements. *European Journal of Operational Research*, 207 (3), 1608–1619.

Rohleder, T. R., Lewkonia, P., Bischak, D. P., Duffy, P., and Hendijani, R. 2008. Using simulation modeling to improve patient flow at an outpatient orthopedic clinic. *Health Care Management Science*, 14 (2), 135–145.

Romanow, R. 2002. Building on values: the future of health care in Canada. *The Future of Health Care in Canada*, Final Repor. Regina, SK.

Rowe, B. H., Channan, P., Bullard, M., Blitz, S., Saunders, L. D., Rosychuk, R. J., Lari, H., Craig, W. R., and Holroyd, B. R. 2006. Characteristics of patients who leave emergency departments without being seen. *Academic Emergency Medicine*, 13 (8), 848–852.

Saadouli, H., Jerbi, B., Dammak, A., Masmoudi, L., and Bouaziz, A. 2015. A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department. *Computers and Industrial Engineering*, 80 (0), 72–79.

Samorani, M. and LaGanga, L. R. 2015. Outpatient appointment scheduling given individual day-dependent no-show predictions. *European Journal of Operational Research*, 240 (1), 245–257.

Shaikh, S. B., Jerrard, D. A., Witting, M. D., Winters, M. E., and Brodeur, M. N. 2012. How long are patients willing to wait in the emergency department before leaving without being seen? *Western Journal of Emergency Medicine*, 13 (6), 463–467.

Shayne, P., Lin, M., Ufberg, J. W., Ankel, F., Barringer, K., Edwards, S. M., DeIorio, N., and Asplin, B. 2009. The effect of emergency department crowding on education: blessing or curse? *Academic Emergency Medicine*, 16 (1), 76–82.

Sills, M. R., Fairclough, D., Ranade, D., and Kahn, M. G. 2011. Emergency department crowding is associated with decreased quality of care for children with acute asthma. *Academic Emergency Medicine*, 57 (3), 191–200.

Silva, T., De Souza, M., Saldanha, R., and Burke, E. 2015. Surgical scheduling with simultaneous employment of specialised human resources. *European Journal of Operational Research*, 245 (3), 719–730.

Stock, L. M., Bradley, G. E., Lewis, R. J., Baker, D. W., Sipsey, J., and Stevens, C. D. 1994. Patients who leave emergency departments without being seen by a physician: magnitude of the problem in Los Angeles County. *Annals of Emergency Medicine*, 23 (2), 294–298.

Stolletz, R. and Zamorano, E. 2014. A rolling planning horizon heuristic for scheduling agents with different qualifications. *Transportation Research Part E: Logistics and Transportation Review*, 68, 39–52.

Tànfani, E. and Testi, A. 2010. A pre-assignment heuristic algorithm for the Master Surgical Schedule Problem (MSSP). *Annals of Operations Research*, 178 (1), 105–119.

Thompson, D. A., Yarnold, P. R., Williams, D. R., and Adams, L. S. 1996. Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. *Annals of Emergency Medicine*, (28), 657–665.

Van Houdenhoven, M., Van Oostrum, J. M., Hans, E. W., Wullink, G., and Kazemier, G. 2007. Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. *Anesthesia and Analgesia*, 105 (3), 707–714.

Van Slyke, R. M. and Wets, R. 1969. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17 (4), 638–663.

Votaw, D. F. and Orden, A. 1952. The personnel assignment problem, symposium on linear inequalities and programming. *SCOOP 10, U.S. Air Force*, 155–163.

Welch, S. J., Jones, S. S., and Allen, T. 2007. Mapping the 24-Hour Emergency Department Cycle to Improve Patient Flow. *The Joint Commission Journal on Quality and Patient Safety*, 33 (5), 247–255.

White, D. L. F., Craig, M., and Klassen, K. J. 2011. The effect of integrated scheduling and capacity policies on clinical efficiency. *Production and Operations Management*, 20 (3), 442–455.

Williams, S. and Crouch, R. 2006. Emergency department patient classification systems: A systematic review. *Accident and Emergency Nursing*, 14 (3), 160–70.

Wilper, A. P., Woolhandler, S., Lasser, K. E., McCormick, D., Cutrona, S. L., Bor, D. H., and Himmelstein, D. U. 2008. Waits to see an emergency department physician: U.S. trends and predictors, 1997-2004. *Health Affairs*, 27 (2), 84–95.

Zacharias, C. and Pinedo, M. 2013. Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, 23 (5), 788–801.

Zaerpour, F., Bischak, D. P., and Menezes, M. B. C. 2017. Coordinated lab-clinics: A tactical assignment problem in healthcare. *European Journal of Operational Research*, 263 (1), 283–294.

— 2018a. A descriptive study of emergency departments. Working paper.

Zaerpour, F., Bijvank, M., and Sun, Z. 2018b. Physician scheduling to improve emergency department patient flow. Working paper.

Zopf, D., Joseph, A. W., and Thorne, M. C. 2012. Patient and family satisfaction in a pediatric otolaryngology clinic. *International Journal of Pediatric Otorhinolaryngology*, 76 (9), 1339–1342.