Physicians Scheduling in Polyclinics

Mohammad Tohidi

A Thesis

In The Department

of

Mechanical, Industrial and Aerospace Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Industrial Engineering) at

Concordia University

Montréal, Québec, Canada

October 2018

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Mohammad Tohidi

Entitled: Physicians Scheduling in Polyclinics

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Industrial Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________________ Chair
Dr. Amin Hammad

______________________________ External Examiner
Dr. Nadia Lahrichi

______________________________ External to Program
Dr. Brigitte Jaumard

______________________________ Examiner
Dr. Ali Akgunduz

______________________________ Examiner
Dr. Hossien Hashemi Doulabi

______________________________ Co-supervisor
Dr. Ivan Contreras

______________________________ Co-supervisor
Dr. Masoumeh Kazemi Zanjani

Approved by ______________________________
Dr. Ali Dolatabadi, Graduate Program Director

December 10, 2018 ______________________________
Dr. Amir Asif, Dean,
Gina Cody School of Engineering and Computer Science

ABSTRACT

Physicians Scheduling in Polyclinics

Mohammad Tohidi, Ph.D.

Concordia University, 2018

Physician scheduling is an important part of hospital operation management. Fatigue, nervousness, high levels of stress and depression are common negative effects of inappropriate work schedules on physicians. A robust and automated personnel scheduling system, which satisfies physicians' preferences, not only improves the quality of life for physicians but also helps to provide a better care for patients and potentially makes significant savings in time and cost for hospitals. Polyclinics reduce the burden on hospitals and help bridge the gap between primary and secondary care. They provide various hospital services such as X-rays, minor surgeries and out-patient treatment and gather several practices under one roof to cooperate, interact and share available resources. In addition, this structure provides an opportunity for physicians of different disciplines to work together and enables patients with chronic and complex conditions to visit multiple clinics at the same place during the same visit. Our problem of interest is mainly motivated by an extension of physician scheduling problems arising in ambulatory polyclinics, where the interaction of clinics and its consequences in terms of sharing their scarce resources introduce new constraints and add complexity to the problem.

In the first part of this thesis, we present an integrated physician and clinic scheduling problem arising in ambulatory cancer treatment polyclinics, where patients may be assessed by multiple physicians from different clinics in a single visit. The problem focuses on assigning clinic sessions and their associated physicians to shifts in a finite planning horizon. The complexity of this problem stems from the fact that several interdisciplinary clinics need to be clustered together, sharing limited resources. The problem is formulated as a multi-objective optimization problem. Given the inherent complexity for optimally solving this problem with a standard optimization software, we develop

a hybrid algorithm based on iterated local search and variable neighborhood descent methods to obtain high quality solutions.

In the second part we propose a comprehensive bi-level physicians planning framework for polyclinics under uncertainty. The first level focuses on clinic scheduling and capacity planning decisions, whereas the second level deals with physicians scheduling and operational adjustments decisions. In order to protect the generated schedules against demand uncertainty, the first level is modeled as an adjustable robust scheduling problem which is solved using an ad-hoc cutting plane algorithm. To cope with variability in patients' treatment times, we formulate the second level as a two-stage stochastic problem and use a sample average approximation scheme to obtain solutions with small optimality gaps. Moreover, we use a Monte-Carlo simulation algorithm to demonstrate the potential benefits of using our planning framework.

In the last part of this thesis we investigate on the impact of physicians work schedules on patient wait-time under uncertain arrival pattern and treatment time of patients. We provide a methodology that combines discrete-event simulation with an optimization search routine to minimize patient wait-time and physician overtime subject to several scheduling/resource restrictions. We indicate the significant impact of adopting the proposed simulation optimization framework for physician scheduling on reducing the aforementioned key performance measures.

Acknowledgments

First and foremost, I wish to thank my advisors Dr. Masoumeh Kazemi Zanjani and Dr. Ivan Contreras for giving me the opportunity to do a PhD under their supervision. They have been supportive since the day one. They have supported me not only with their immense knowledge and invaluable advice, but also emotionally through the rough road to finish this thesis. It would not have been possible to complete my PhD without their invaluable advice over the past four years.

A special thanks to my fellow lab-mates and great friends at Concordia University particularly, Omid Sanei, Mohammad Jeihoonian, Ehsan Rezabeigi, Alireza Zandi Karimi, Dua Weraikat and Alireza Ebrahim Nejad.

In addition, I am grateful to my friend, Dr. Isabelle Lachance who was the only candle of hope in my darkest days of frustration.

Last but not least, I would like to thank my family: my parents, my brothers and sister for supporting me throughout my life. Words cannot express how grateful I am to you.

To my beloved parents and siblings

Fayegh, Parvin, Vahid, Baharak, Siamak, Siavash

Preface

This thesis has been prepared in "Manuscript-based" format under the co-supervision of Dr. Masoumeh Kazemi Zanjani and Dr. Ivan Contreras from the department of Mechanical, Industrial and Aerospace Engineering, Concordia University. This research was supported by Le Fonds de Recherche du Québec-Nature et technologies (FRQNT) and the Natural Sciences and Engineering Research Council of Canada (NSERC). All the articles presented in this thesis were co-authored and reviewed prior to submission for publication by Dr. Masoumeh Kazemi Zanjani and Dr. Ivan Contreras. The author of this thesis acted as the principal researcher and performed the mathematical models development, programming of the solution algorithms, analysis and validation of the results, along with writing the first drafts of the articles.

The first article entitled "Integrated physician and clinic scheduling in ambulatory polyclinics", co-authored by Dr. Masoumeh Kazemi Zanjani and Dr. Ivan Contreras was published in *Journal of the Operational Research Society* in February 2018.

The second article entitled "A physicians planning framework for polyclinics under uncertainty", co-authored by Dr. Masoumeh Kazemi Zanjani and Dr. Ivan Contreras was submitted to *IIE Transactions on Healthcare Systems Engineering* in October 2018.

The third article entitled "Simulation optimization for physicians scheduling in polyclinics", co-authored by Dr. Masoumeh Kazemi Zanjani and Dr. Ivan Contreras was submitted to *Computers & Operations Research* in September 2018.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Hospitals have been under increasing pressure to reduce their operating costs and to change their reimbursement policies. According to Erhard et al. [1], in the United States, on average more than 50% of hospital costs are workforce related. In order to minimize the operational expenses and to maintain quality of care, researchers and practitioners have used advanced operations research (OR) techniques to plan more efficient work schedules for nurses and physicians. In general, Pinedo [2] defines workforce allocation and personnel scheduling as creating work schedules and allocating staff to shifts in order to cover the demand. Gendreau et al. [3] define *physician scheduling problems* (PSPs) as creating work schedules for physicians in a pre-determined planning horizon such that in every given shift there are enough physicians to satisfy demand, while abiding to several rules and regulations. Such rules and regulations have made PSPs complex optimization problems. The complex nature of physician scheduling hinders in the creation of a generic model as in the case with nurse scheduling. Regional labor contracts, governing authorities, seniority levels, experience levels, and training prevent generalization [4]. Typically nurses' work contracts are ruled by collective agreements, whereas physicians' contracts are usually set up individually. In physician scheduling, one of the main goals of the schedule is the satisfaction of physicians; studies on career satisfaction among physicians have had an increasing trend recently, since it has been shown that career dissatisfaction is increasing among physicians [5]. Numerous evidences confirm that there is a relation between physician satisfaction and care quality. When physicians are satisfied with their career, quality of care improves [6]. Therefore, physician scheduling has a great importance in the sense that a proper schedule not only improves the career satisfaction among physicians, but also helps patients to enjoy a higher quality care service. Moreover, physicians represent one of the most valuable and expensive resources in hospitals, and any improvement in physicians costs usually results in significant cost-savings for the management of hospitals [7].

In recent years, outpatient services have become an important component in health care systems due to the focus on preventive care and patients' shorter lengths of stay [8]. Outpatient polyclinics are an attempt for moving some care out of hospitals into the community, where it is more convenient for patients. They reduce the workload of hospitals and fill the void between primary and secondary care. Polyclinics provide some hospital services such as X-rays, minor surgery and outpatient treatment. The simplest model involves several practices under one roof, sharing many available resources. Polyclinics provide a better structure and work environment for physicians of different disciplines to work together and enable patients with chronic and complex conditions to be assessed by physicians from different disciplines during a single visit.

Motivated by the additional requirements and constraints in physicians scheduling problems of polyclinics, the focus of this thesis is on designing physicians work schedules applicable in the context of outpatients polyclinics. Planning physicians work schedules in polyclinics is a complex problem particularly attributed to the various types of constraints introduced by clinics and shared resources. On the other hand, PSPs deal with a high degree of variability and uncertainty in terms of the number of arriving patients, patients arrival pattern, and patients processing time, which further complicate the planning of a robust schedule. Even though in recent years a rigorous stream of academic research has been dedicated to the study of PSPs in deterministic settings, to the best of our knowledge, including uncertainty in processing times or patients' arrivals has never been investigated. This is the critical point which distinguishes this manuscript from previous studies. More precisely, this thesis contributes to the existing literature through addressing the problem of scheduling physicians in polyclinics by integrating the clinic scheduling and physicians scheduling as a single optimization problem in a deterministic setting. In addition, we introduce new models in which the number of arriving patients to the center and patients processing times are stochastic. Finally, we investigate the impact of physicians work schedules on patients wait-time.

In what follows, we first provide some background and brief description of the studied problems. We then present our research scope and objectives. Finally, we describe the outline of this thesis.

1.2 Problem description

The polyclinic of interest is an ambulatory cancer treatment of the McGill University Health Centre (MUHC) in Montreal, Canada. It consolidates 13 cancer clinics (i.e., breast, urology, hematology, gynecology, hepatobiliary, lung, musculoskeletal, melanoma, upper gastrointestinal, pain, cancer rehab, colorectal and brain metastases). These clinics are divided into three categories, according to their operations and scheduling requirements: high throughput, interdisciplinary and multidisciplinary clinics. High throughput clinics function with a high tempo akin to a manufacturing plant. An arriving patient enters the clinic in the same manner as every other patient, and they usually receive standardized services.

The musculoskeletal clinic is an example of a high throughput clinic. Interdisciplinary clinics integrate separate discipline assessments into a single consultation session by having a group of physicians from different disciplines examine a patient. Lung and cancer rehab clinics are examples of interdisciplinary clinics. Integration of multiple assessments into a single session creates cluster of clinics which need to be scheduled in the same shift for a minimum number of times during the planning horizon. Multidisciplinary clinics are comprised of cross-functional physicians, who work in the same environment. For instance, the urology clinic is a multidisciplinary clinic whose doctors work independently with little coordination and individual appointments. Given that the available resources (e.g. examination rooms, waiting rooms, etc.) at the polyclinic shared among all clinics are rather limited, it is not possible to host all clinics in any given shift of the planning horizon. This makes the design of physicians' schedules even more involved, as it needs to be integrated with the assignment of clinic sessions to shifts in the polyclinic. The main problem that the hospital manager must deal with is thus to schedule clinics' sessions and to assign physicians to shifts, complying with various rules and constraints. Decomposing those problems and solving them independently may result in sub-optimal solutions. The *physician and clinic scheduling problem* (PCSP) considers the design of a master schedule for a polyclinic by assigning clinics to shifts and specifying the on-duty physicians for every allocated shift. The scheduling decisions of clinics and associated physicians incorporate the capacity limitations of treatment and waiting rooms, various clinics' requirements, and physicians' preferences into a single optimization problem.

In many cases, uncertain events disturb planned operations in health services. Variations in the number of arriving patients to the polyclinic and patients' processing times are two major sources of uncertainty in the context of outpatients clinics. In the presence of uncertainty, the PSCP becomes a three-level planning problem. The first level corresponds to the *clinic scheduling and capacity planning problem* (CSCPP), a long-term (strategic) problem to determine the number of patients to admit, the total working hours of physicians, and the number of required examination rooms. In this level, clinic schedules are determined according to a weekly demand forecast. Nevertheless, over a long term planning horizon (e.g., one year) the weekly forecast can fluctuate within a given uncertain interval. Hence, hospital managers seek a robust strategic planning tool to maximize the number of patients who can be served on a weekly basis, even in the presence of worst-case (extremely high) demand scenarios, while determining the optimal personnel and resource requirements at each shift of the planning horizon. The second level of the planning process, i.e., PSP, involves a tactical assignment of physicians to the established shifts while taking into account the requirements of the shifts and physicians preferences. In this level, schedules are affected by the uncertainty in patients' treatment times. On the one hand, if the actual treatment times are longer than the estimated ones, the number of scheduled physicians might not be enough to serve all patients during the regular shifts. In this case, either patients would need to be rescheduled on another session or extra resources, such as on-call doctors or overtime shifts, would need to be deployed. On the other hand, if treatment times are shorter that the estimated ones, physicians would be idle which is undesirable for the administration. These scenarios lead to the third level of the physicians planning procedure, which is to perform last-minute operational adjustments to the planned schedules (i.e., prior to each shift). In this operational planning level, proper corrective (recourse) actions must be foreseen in order to minimize the average expected cost of schedules in the presence of uncertainty in treatment times.

Although in recent years hospitals have been forced to reduce the expenses while improving service quality, the inclusion of patients satisfaction in PSPs have been neglected.

The stochastic scheduling of physicians, described earlier, is extended, where the decision-makers aim for the simultaneous minimization of resource costs and patient wait-time, under uncertain arrival pattern and treatment times in the polyclinic. In this settings, studying the problem provides the opportunity to understand the impact of physicians work schedules on patients' satisfaction.

1.3 Scope and objectives

To fill the void in the existing literature, the main contribution of this thesis is to design a comprehensive physician scheduling framework that will aid in developing of efficient work schedules in the context of polyclinics. Given the problem description, the specific objectives are summarized as follows.

1. To extend physician scheduling in the context of polyclinics and include clinic requirements in the problem.

2. To formulate the deterministic PCSP as a mathematical programming model to determine clinics and physicians work schedules while satisfying physicians' preferences.

3. To develop a math-heuristic algorithm to solve the PCSP.

4. To explicitly incorporate the uncertainty on the number of arriving patients in the clinic scheduling problem and to formulate the problem as a robust optimization problem.

5. To develop an efficient algorithm for solving the robust counterpart of the CSCPP.

6. To explicitly incorporate the uncertainty of patients processing time in the PSP and to formulate it as a two-stage stochastic model.

7. To develop a sample average approximation scheme to obtain high quality solutions for the stochastic PSP.

8. To investigate the impact of physicians work schedules on patients wait-time, and to include patient-wait time and physicians' overtime in the PSP.

9. To apply a simulation-optimization scheme for optimizing the non-analytic performance measures in the previous objective.

10. To investigate the tractability of the proposed models and the performances of the solution methods based on a real-life case study.

1.4 Organization of the thesis

This manuscript has five chapters organized as follows. Chapter 2 addresses the fundamental problem of this study that is integrating physicians scheduling with clinic scheduling in the context of polyclinics. The decisions to be made are the assignment of clinics and physicians to the shifts such that clinic and physicians' requirements are satisfied, and physicians' preferences are optimized. To this end, the problem is formulated as a MIP model in which the objective is to maximize the preferences of physicians. In order to solve the proposed mixed integer program (MIP) in a reasonable amount of time, an iterated variable neighborhood descent math-heuristic is developed. Furthermore, we shed light on the impact of the integrating clinics' requirements on physicians' work schedules. In Chapter 3, through modeling uncertainty in the number of arriving patients to the polyclinic and randomness in the patients' processing times, the deterministic problem is extended into the strategic CSCPP and the tactical/operational stochastic PSP. The former is formulated as robust optimization problem, and the latter is formulated as a two-stage mixed-integer stochastic program to minimize the expected cost. Given the large number of scenarios, an implementor/adversary (I/A) algorithm is adapted to solve the robust model, and a sample average approximation (SAA) scheme is applied to the stochastic program. Moreover, some insight on the level of conservativeness in the robust model is provided. Chapter 4 presents a simulation-optimization model to address physicians' overtime and patients wait-time in PSPs. The objective is to minimize the two key performance indicators (KPIs) while satisfying clinics' and physicians' requirements. The stochastic factors are the patients' arrival pattern and the patients' processing time. The large-scale optimization problem is then solved by an enhanced simulated annealing (SA) algorithm in which the initial solution is provided by solving a reformulated approximation model of the problem. Finally, Section 5 summarizes concluding remarks in addition to providing several avenues for future research.

Chapter 2

Integrated physician and clinic scheduling in ambulatory polyclinics

Abstract

This paper presents an integrated physician and clinic scheduling problem arising in ambulatory cancer treatment polyclinics, where patients may be assessed by multiple physicians from different clinics in a single visit. The problem focuses on assigning clinic sessions and their associated physicians to shifts in a finite planning horizon. The complexity of this problem stems from the fact that several interdisciplinary clinics need to be clustered together, sharing limited resources. The problem is formulated as a multi-objective optimization problem. Given the inherent complexity for optimally solving this problem with a standard optimization software, we develop a hybrid algorithm based on iterated local search and variable neighborhood descent methods to obtain high quality solutions. Computational results using a set of instances inspired from a case study in a hospital in Canada along with some managerial insights are reported and analyzed.

2.1 Introduction

Physician scheduling is an important class of planning problems in hospital operations management. Numerous studies concerning the effects of work schedules on physical and mental wellbeing show that fatigue, nervousness, high level of stress, and depression are common problems among physicians [9]. The use of a robust and automated personnel scheduling system, capable of generating schedules satisfying physicians' preferences, helps to improve their quality of life which in turn, aids providing a better care for patients. It also has a significant impact on time and cost savings.

PSPs consist of creating work schedules for physicians in a pre-determined planning horizon such that in every given shift there are enough physicians to satisfy the demand [3]. Creating physicians' work schedules must be done while abiding to several rules and regulations. Gendreau et al. [3] provide four categories of rules commonly considered in PSPs:

supply and demand (e.g. assigning physicians within their availability or preference), workload (e.g. setting the number of working shifts according to physicians' contracts), fairness (e.g. distributing weekend or night shifts evenly), and ergonomic (e.g. restricting consecutive working hours) constraints. These rules are frequently in conflict with one another, causing difficulties to create a schedule that jointly satisfies all rules. Depending on the problem, rules are classified as either soft or hard. Soft rules can be violated, whereas hard ones must be satisfied. PSPs are difficult combinatorial optimization problems and finding a feasible solution that satisfies all the physicians' requests is a challenging task [10]. The objective in PSPs is often to create physicians' work schedules according to their preferences. This is an important difference with respect to other personnel scheduling problems arising in health care, such as nurse scheduling, in which the labor cost reduction is as significant as nurses' preferences [11, 12]. Given that physicians' work contracts are usually set up individually, they introduce more conflicting constraints to the problem, whereas in nurse scheduling work contracts follow the alignments of a single collective agreement and they are more general [4].

In this paper, we study an extension of PSPs arising in ambulatory polyclinics. This problem is inspired by a real case study in an ambulatory cancer treatment polyclinic of the McGill University Health Centre (MUHC) in Montreal, Canada. Polyclinics are facilities which consolidate multiple multidisciplinary, interdisciplinary and high throughput clinics that differ in terms of patient flow and treatment time. As a result, polyclinics allow patients to visit more than one clinic during the same session when needed. Clinics in polyclinics usually co-operate and interact in the assessment and treatment of patients and thus, predefined clusters of clinics need to be assigned to the same sessions a minimum number of times. An important consequence of the consolidation of clinics in a single facility is that the resources, such as the waiting and treatment rooms must be shared among all clinics. Given that such resources are rather limited, it is not possible to host all clinics in any given shift of the planning horizon. This makes the design of physicians' schedules even more involved, as it needs to be integrated with the assignment of clinic sessions to shifts in the polyclinic. The main problem that the hospital manager must deal with is thus to schedule clinics' sessions and to assign physicians to shifts, complying with various rules and constraints. Decomposing those problems and solving them independently may result in sub-optimal solutions. Most of the literature in physician scheduling revolves around physicians of a single department without considering resource capacity constraints.

We present the integrated PCSP in the context of polyclinics, in which clinics' requirements along with physicians' preferences are explicitly considered. The PCSP consists of designing work schedules of clinics with respect to clinics' requirements (hard constraints), and of the assignment of physicians to corresponding clinics' work sessions taking into account physicians' preferences (soft constraints) over a finite planning horizon. The objective is to minimize the violation of physicians' preferences while ensuring that clinics are clustered into common shifts whenever required and that capacity constraints associated with the polyclinic resources are satisfied.

The main contributions of this work are the following. We introduce a new PSP which integrates the scheduling of clinics and physicians in ambulatory polyclinics. We show how this problem can be stated as a multi-objective optimization model, where the violation of a set of conflicting soft constraints is minimized. Using the weighted sum method, a single-objective MIP formulation is presented to solve the PCSP. Given that standard optimization softwares fail to optimally solve the problem in reasonable CPU times, we develop a hybrid solution algorithm based on iterated local search and variable neighborhood descent methods to quickly obtain feasible solutions of high quality.

The remainder of this paper is structured as follows. Section "Literature review" reviews the relevant literature on the PCSP. A detailed definition of the problem and a multi-objective mathematical programming formulation is given in Section "The physician scheduling problem". An IVND algorithm is presented in Section "A heuristic algorithm for the PCSP". The results of computational experiments performed on a set of instances as well as some managerial insights are given in Section "Computational results". Conclusions follow in Section "Concluding remarks".

2.2 Literature review

Workforce allocation and personnel scheduling problems commonly arise in the service industry (e.g., telephone operators, flight crews, bus drivers, doctors and nurses, etc.). Comprehensive surveys in the area of personnel scheduling are found in Ernst et al. [13] and Van den Bergh et al. [14]. A wide variety of analytical methods such as mathematical programming, constraint programming, heuristic methods, and discrete-event simulation have been widely utilized to tackle these problems. In health care systems and hospital operations management, nurse scheduling problems have been extensively studied [15]. However, PSPs have received less attention in the literature. Extensive literature reviews on physicians and nurse scheduling problems can be found in Erhard et al. [1] and Burke et al. [16], respectively.

As Erhard et al. [1] indicated, no prior work in the literature investigates PSP in the context of polyclinics. In other words, the majority of existing studies in this area focused on physicians of a single department of hospitals, without taking into account any kind of resource capacity limitations that must be shared among different departments. Beaulieu et al. [17] propose the first MIP formulation to solve a PSP arising in the emergency room (ER) in a major hospital in Montreal. They divide hospital rules into two main categories: compulsory and flexible rules. They use a partial branch and bound algorithm to solve the model. Topaloglu [18] addresses a resident scheduling problem in a pulmonary unit of a hospital and formulates the problem as a multi-objective MIP for a six-month planning horizon. The authors consider residents with four levels of seniority that need to cover the demand over weekday and weekend shifts. The sequential method and the weighted sum method are applied to the multi-objective model with a single set of weights for the objectives. Topaloglu [19] studies a resident scheduling problem in the ER of a hospital in Turkey. The problem deals with the assignment of residents grouped into three seniority levels to three different shift types. The problem is formulated as a goal-programming model in which the objective function is a weighted function of the deviation variables. The weights are generated through pairwise comparison. Song et al. [20] also use goal-programming in combination with discrete-event simulation to determine physicians inquiry start time in physical examination services in Taiwan. The authors consider multiple objectives such as maximizing physician utilization and minimizing patient waiting time in their study. Raouf and Ben-Daya [21] determine staffing level of physicians in a single department outpatient clinic in Saudi Arabia by the aid of discrete-event simulation.

Brunner et al. [4] study the scheduling of physicians in an anesthesia department of a German hospital for a two-week planning horizon. The authors apply a flexible shift scheduling approach, in which the shifts have variable starting times and durations. They consider the objective function as to minimize the cost of personnel which is a function of paid time, overtime and outside physician hours. The MIP model is first decomposed into one-week problems and then sub-problems are solved by a commercial solver. Stolletz and Brunner [22] reformulate the problem investigated in [4] as a set covering problem. The authors include some additional ergonomic and distribution constraints in their reformulation. Brunner [15] propose a MIP for scheduling physicians with multiple experience levels for the same case study in [4] for a one-year planning horizon. The authors decompose the problem into weekly sub-problems and implement a column generation-based heuristic for solving each sub-problem. Brunner et al. [23] tackle the same problem as [4] by applying a branch and price algorithm. The authors are able to get an exact solution for planning horizons up to six weeks while incorporating seniority rules as well as fair distribution of holiday shifts in their model.

Rousseau et al. [10] argue that a combination of constraint programming with local search can be a promising generic method to a wide variety of PSPs. They apply their method to two case studies which consider physicians in a single department. Topaloglu and Ozkarahan [24] focus on a resident scheduling problem in a university hospital in Turkey. All constraints are considered as hard constraints except for the demand coverage constraints. The model is solved by a constraint programming-based column generation technique. Carter and Lapierre [11] study physicians of ERs at six major hospitals in Canada and proposed a tabu search algorithm for generating work schedules in two case studies for ER departments in Montreal. Puente et al. [25] apply a genetic algorithm in order to solve a PSP in the ER of a hospital for temporary and fulltime physicians. The considered planning horizon is one month, with three regular shifts and one observation shift each day. Priority of soft constraints is determined by assigning a score, which is calculated by using the Delphi method. Bruni and Detti [12] investigate the scheduling of physicians of two departments in a hospital in Rome. The work schedule is created for 32 physicians of four different groups. The authors do not consider any sort of resource capacity constraint and the two investigated departments are completely disjoint. The model is solved for a one-year planning horizon using a commercial solver.

To the best of our knowledge, only few papers have studied integrated PSPs with other types of decision problems. In the case of surgery scheduling, Gunawan and Lau [26] and Van Huele and Vanhoucke [27] design work schedules of physicians associated with multiple departments. The authors consider resource capacity constraints (e.g. available operation rooms, available recovery beds). Gunawan and Lau [26] propose a multi-objective optimization model in order to plan the full day to day range of physicians' duties for a one-week planning horizon in a surgery department. They consider the number of unassigned duties and the number of non-preferred assigned duties as the objectives of the model. The authors include resource capacity constraints alongside physician scheduling constraints. However, they simplify the problem by assuming that the resources are not shared among the duty types. The model is solved with a commercial solver for small size instances, and a local search is proposed for solving larger size instances. Van Huele and Vanhoucke [27] integrate the PSP and the surgery scheduling problem. The authors propose a mathematical programming formulation which includes most commonly used constraints of the surgery scheduling problem along with the PSP. They consider the minimization of overtime in operation rooms. Experimental analyses demonstrate that resource capacity constraints (e.g. number of available beds) have a significant impact in terms of solution quality and computational time. Furthermore, the results show that certain specific physicians' preferences have the most impact on the operational surgery schedule. Roland and Riane [28] also integrate the PSP into the surgery scheduling problem. Instead of using the conventional objective function in surgery scheduling problems (i.e. minimizing the cost of operation rooms), the authors formulate the problem as a multi-objective model that minimizes the cost and maximizes surgeons preferences simultaneously. Roland et al. [29] tackle the problem of surgery scheduling combined with medical staff availability constraints over a one-week planning horizon. The focus of their study is on the surgery scheduling problem since the objective function of the problem is to minimize the cost of operation rooms. Moreover, the authors propose a genetic algorithm for solving large scale instances.

We would like to highlight that none of these PSP incorporating resource capacity constraints consider similar clustering constraints of subsets of clinics to the ones present in the PCSP. This is actually one of the distinguishing features of our problem, and as it will be shown in Section ”Computational results”, this makes the problem actually very challenging to solve.

In this section, we describe polyclinics and their distinguishing features and define our scheduling problem in details. After that, we formulate the problem as a multi-objective optimization problem and explain components of the model. Finally, we determine the importance of each objective through a priori articulation approach.

2.3 Problem definition

Polyclinics are an attempt for moving some care out of hospitals into the community, where it is more convenient for patients. They reduce the burden in hospitals and helps bridge the gap between primary and secondary care. Polyclinics provide some hospital services such as X-rays, minor surgery and outpatient treatment. The simplest model involves several practices under one roof, sharing many services. Polyclinics provide a better structure for physicians of different disciplines to work together and enable patients with chronic and complex conditions to visit multiple clinics at the same place during the same visit.

A concrete application for the PCSP arises in an ambulatory cancer treatment polyclinic of the MUHC in Montreal. The polyclinic consolidates 13 cancer clinics (i.e., breast, urology, hematology, gynecology, hepatobiliary, lung, musculoskeletal, melanoma, upper gastrointestinal, pain, cancer rehab, colorectal and brain metastases). These clinics are divided into three categories, according to their operations and scheduling requirements: high throughput, interdisciplinary and multidisciplinary clinics. High throughput clinics function with a high tempo akin to a manufacturing plant. An arriving patient enters the clinic in the same manner as every other patient and they usually receive standardized services. The musculoskeletal clinic is an example of a high throughput clinic. Interdisciplinary clinics integrate separate discipline assessments into a single consultation session by having a group of physicians from different disciplines examine a patient. Lung and cancer rehab clinics are examples of interdisciplinary clinics. Integration of multiple assessments into a single session creates cluster of clinics which need to be scheduled in the same shift a minimum number of times during the planning horizon. Multidisciplinary clinics are comprised of cross-functional physicians, who work in the same environment. For instance, the urology clinic is a multidisciplinary clinic whose doctors work independently with little coordination and individual appointments. Given that the available resources (e.g. examination rooms, waiting rooms, etc.) at the polyclinic are shared among all clinics, the allocation decisions of clinics to shifts cannot be made independently.

The PCSP considers the design of a master schedule for a polyclinic by assigning clinics to shifts and specifying the on-duty physicians for every allocated shift. We assume disjoint shifts which is a realistic assumption in the polyclinic under investigation and is also common in the literature of healthcare personnel scheduling [e.g., 17, 26, 28, 30, 31]. Further, in the same context, it is very common for physicians to be affiliated to more than one hospital. They frequently have more than one contractual agreement with hospitals, especially for physicians working in ambulatory polyclinics. As a result, these physicians dedicate a portion of their time to diagnose and treat patients of these clinics. On the contrary, they have other duties at larger hospitals and even teaching responsibilities at universities. As a result, it is rather common for physicians to work in a given week more or less shifts than the ones stipulated in their contracts when demand fluctuates from one week to another. The scheduling decisions of clinics and associated physicians incorporate: *(i)* the capacity limitations of treatment and waiting rooms, *(ii)* various clinics' requirements, and *(iii)* physicians' preferences. The first two points correspond to hard constraints whereas the third point are the soft constraints. We assume that the planning horizon is one week (Monday-Friday), and every day consists of two four-hour shifts (morning and afternoon). The objective of the PCSP is to minimize the violation of the soft constraints. We list the assumptions considered in the PCSP below.

Clinics requirements

- Interdisciplinary clinics that interact must be scheduled together a minimum number of shifts over the planning horizon.

- In order to level the utilization of the waiting room over the planning horizon, the fluctuation in the utilization of the waiting room from one shift to another must be limited to a certain magnitude. For simplicity, we consider that the number of visiting patients in every shift represents the utilization of the waiting room.

- The total number of patients in a shift cannot exceed the capacity of the waiting room. We assume that there must be a reserved spot at the waiting room queue for every patient scheduled to be visited during a shift. This is similar to '*environmental factors*' that are considered in appointment scheduling problems and determines the '*clinic size*' [32].

- The total number of on-duty physicians in a shift cannot exceed the number of available examination rooms. Also, it is assumed that each physician is only affiliated with one clinic of the polyclinic under investigation. In addition, we assume that every physician needs one examination room to be able to assess the patients. As we pointed out earlier, the treatment rooms are shared among physicians of different clinics.

- For each clinic, the total number of on-duty physicians in all shifts during the week must be enough to serve the weekly demand of patients.

Physicians requirements

- Physicians should be assigned to at most one shift per day.

- Each clinics' workload should be distributed *fairly* among their physicians.

- A physician must not be assigned to those shifts requested as shifts-off.

- Physicians prefer to work in shifts which are scheduled on consecutive days. In other words, they do not want any off-duty days between two on-duty days.

- Each physician has a set of *preferred* shifts in a week.
- Each physician has a number of *tokens* that can be spent to be assigned to their preferred shifts.

2.3.1 A mathematical programming formulation

We next state the PCSP as a multi-objective optimization problem, where each objective corresponds to the computation of the violation of each soft constraint. We use the following mathematical notation to represent the problem.

Sets

$\boldsymbol{J}$ set of days

$\boldsymbol{K}$ set of shifts per day

$\boldsymbol{C}$ set of clinics

$\boldsymbol{I}$ set of physicians

$\boldsymbol{I}_c$ set of physicians of clinic c

$\boldsymbol{NJ}_i$ set of days on which physician i is not available to work

$\boldsymbol{T}$ set of sub-sets of clinics (clusters) that must be scheduled together

Parameters

B number of available rooms

NWV_c number of weekly patients of clinic c

VPS_i number of patients that physician i can assess during a shift

D capacity of the waiting room

F^t minimum number of times that cluster t must be scheduled together.

H_i number of shifts that physician i should work per week according to his/her contract

TK_i number of tokens that physician i can spend to be assigned to preferred shifts

P_{ijk} 1 if physician i prefers to work on day j, shift k, 0 if they are indifferent

WU target level of fluctuation in waiting room's utilization between any pair of shifts

To model the problem we use the following set of decision variables. For each $i \in \boldsymbol{I}$, $j \in \boldsymbol{J}$, and $k \in \boldsymbol{K}$, we define the binary decision variable x_{ijk} equal to 1 if and only if physician i is assigned to shift k on day j. For each $c \in \boldsymbol{C}$, $j \in \boldsymbol{J}$, and $k \in \boldsymbol{K}$, we define the binary decision variable y_{cjk} equal to 1 if and only if clinic c is assigned to shift k on day j. For each $i \in \boldsymbol{I}$, we define the binary decision variable m_i equal to 1 if and only if physician i is assigned to no more than one shift during the planning horizon. Finally, for each $j \in \boldsymbol{J}$, $k \in \boldsymbol{K}$, and $t \in \boldsymbol{T}$, we define the binary decision variable f_{tjk} equal to 1 if and only if cluster t is assigned to shift k on day j. Let $(x)^+ = \max\{0, x\}$. The PCSP considers the following six objectives:

Objective 1: minimize the maximum number of unspent tokens among all physicians

$$g_1(x) = \max_{i \in \boldsymbol{I}} \left(TK_i - \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} P_{ijk} x_{ijk} \right)^+ .$$

By considering the number of available tokens, this objective maximizes the number of shifts assigned to physicians according to their preferences.

Objective 2: minimize the total maximum difference between physicians of the same clinic in terms of number of working shifts

$$g_2(x) = \sum_{c \in \boldsymbol{C}} \max_{(i,i') \in \boldsymbol{I}_c \times \boldsymbol{I}_c} \left(\sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} x_{ijk} - \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} x_{i'jk} \right)^+ .$$

This objective aims at satisfying a fair distribution of shifts among physicians.

Objective 3: maximize physicians' preference

$$g_3(x) = - \sum_{i \in \boldsymbol{I}} \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} P_{ijk} x_{ijk}.$$

Objective 4: minimize the total number of assigned shifts in non-consecutive days patterns

$$\begin{aligned} g_4(x) = & \sum_{c \in \boldsymbol{C}} \sum_{i \in \boldsymbol{I}_c} \sum_{k' \in \boldsymbol{K}} \left(- \sum_{k \in \boldsymbol{K}} x_{i(|J|-1)k} + x_{i|J|k'} - m_i \right)^+ \\ & + \sum_{c \in \boldsymbol{C}} \sum_{i \in \boldsymbol{I}_c} \sum_{k' \in \boldsymbol{K}} \left(- \sum_{k \in \boldsymbol{K}} x_{i2k} + x_{i1k'} - m_i \right)^+ \\ & + \sum_{c \in \boldsymbol{C}} \sum_{i \in \boldsymbol{I}_c} \sum_{j \in 2 \ldots |\boldsymbol{J}|-1} \sum_{k' \in \boldsymbol{K}} \left(- \sum_{k \in \boldsymbol{K}} x_{i(j+1)k} - \sum_{k \in \boldsymbol{K}} x_{i(j-1)k} + x_{ijk'} - m_i \right)^+ . \end{aligned}$$

This objective is used to assign the physicians as much as possible to shifts on consecutive days. In practice, if some physicians are better assigned as compared to others in two different planning horizons, this objective could be less penalized for them during the upcoming planning horizon.

Objective 5: minimize the total number of times a physician works two shifts on the same day

$$g_5(x) = \sum_{i \in \boldsymbol{I}} \sum_{j \in \boldsymbol{J}} \left(\sum_{k \in \boldsymbol{K}} x_{ijk} - 1 \right)^+ .$$

This objective aims at satisfying the condition that a physician is assigned to only one shift per day. As mentioned, in the context under investigation, physicians work on a part-time basis in ambulatory polyclinics and usually have a contract with another hospital and/or university.

Objective 6: minimize the total number of shifts that physicians work above their target weekly load

$$g_6(x) = \sum_{i \in \boldsymbol{I}} \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} (x_{ijk} - H_i)^+ .$$

Objective 7: minimize the total number of shifts that physicians work under their target weekly load

$$g_7(x) = \sum_{i \in \boldsymbol{I}} \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} (H_i - x_{ijk})^+ .$$

It should be noted that despite setting the number of working hours a priori in physicians' contracts, over-time and under-time might may also occur as a result of pursuing the goal of a fair distribution of shifts among physicians and the limited number of shared resources in polyclinics . Nevertheless, objectives 6 and 7 aim at minimizing over-time and under-time, respectively. That is, the model is more likely to select a schedule that such costly and inconvenient circumstances (i.e. over-time and under-time) do not occur.

Using the above objectives, the PCSP can be stated as the following multi-objective optimization problem:

$$\text{minimize } G(x) = [g_1(x), \ldots, g_7(x)] \tag{1}$$

$$\text{subject to } \sum_{i\in \boldsymbol{I}_c}\sum_{j\in \boldsymbol{J}}\sum_{k\in \boldsymbol{K}} VPS_i x_{ijk} \geq NWV_c \qquad \forall c \in \boldsymbol{C} \tag{2}$$

$$\sum_{i\in \boldsymbol{I}} x_{ijk} \leq B \qquad \forall j \in \boldsymbol{J},\ \forall k \in \boldsymbol{K} \tag{3}$$

$$\sum_{i\in \boldsymbol{I}} VPS_i x_{ijk} \leq D \qquad \forall j \in \boldsymbol{J}, \forall k \in \boldsymbol{K} \tag{4}$$

$$\sum_{c\in C|c\in t} y_{cjk} \geq |t| f_{tjk} \qquad \forall t \in \boldsymbol{T},\ \forall j \in \boldsymbol{J},\ \forall k \in \boldsymbol{K} \tag{5}$$

$$\sum_{j\in \boldsymbol{J}}\sum_{k\in \boldsymbol{K}} f_{tjk} \geq F^t \qquad \forall t \in \boldsymbol{T} \tag{6}$$

$$\sum_{i\in \boldsymbol{I}} \left(\frac{VPS_i x_{ijk}}{D} - \frac{VPS_i x_{ij'k'}}{D}\right) \leq WU \forall j, j' \in \boldsymbol{J}, \forall k, k' \in \boldsymbol{K} \tag{7}$$

$$\sum_{i\in \boldsymbol{I}_c} x_{ijk} \geq y_{cjk} \qquad \forall c \in \boldsymbol{C}, \forall j \in \boldsymbol{J},\ \forall k \in \boldsymbol{K} \tag{8}$$

$$\sum_{j\in \boldsymbol{J}}\sum_{k\in \boldsymbol{K}} x_{ijk} - 1 \leq |J||K|(1 - m_i) \qquad \forall i \in \boldsymbol{I} \tag{9}$$

$$x_{ijk} = 0 \qquad \forall i \in \boldsymbol{I},\ \forall j \in \boldsymbol{NJ}_i,\ \forall k \in \boldsymbol{K} \tag{10}$$

$$x_{ijk}, y_{cjk} \in \{0,1\}\ \forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I},\ \forall j \in \boldsymbol{J},\ \forall k \in \boldsymbol{K} \tag{11}$$

$$f_{tjk} \in \{0,1\} \qquad \forall t \in \boldsymbol{T},\ \forall j \in \boldsymbol{J},\ \forall k \in \boldsymbol{K} \tag{12}$$

$$m_i \in \{0,1\} \qquad \forall i \in \boldsymbol{I}. \tag{13}$$

Constraints (2) ensure that the overall number of assigned physicians to different shifts is enough to assess all patients visiting each clinic during the week. Constraints (3) limit the number of assigned physicians at any given shift to the number of available examination rooms in the polyclinic, whereas constraints (4) represent the capacity limitations on the number of available spots in the waiting room. Constraints (5) and (6) work in a joint fashion to guarantee that the interdisciplinary clinics are scheduled simultaneously. In particular, constraints (5) force f_{tjk} to take value 1 if and only if all clinics in cluster t are assigned to shift k on day j, and constraints (6) ensure that all clinics in cluster t are scheduled simultaneously at least F^t shifts in a week. Constraints (7) control the fluctuation of waiting room's utilization in different shifts by ensuring that the maximum difference in waiting room's utilization between every pair of shifts over the planning horizon is below the threshold WU. Constraints (8) state that an assignment of a clinic to a shift exists only when at least one physician of that clinic is assigned to such a shift. It is noteworthy that constraints (2) and (8) force each clinic to occur a specific number of times in order to satisfy the demand. Constraints (9) make the block assignment constraints redundant, if the physician works less than two shifts during the planning horizon. Constraints (10) forbid the assignment of physicians to the shifts that they are not available to work at. Finally, constraints (11)-(13) are the standard integrality conditions on the decision variables.

2.3.2 A weighted sum approach

The weighted sum method is a classical approach when dealing with multiple objectives [33, 34]. It can be used for obtaining multiple solution points by varying the weights as there are usually infinite number of Pareto optimal solutions for a multi-objective problem. When all objective functions and constraints are convex, it has been shown that every optimal solution of a set of positive weights is a Pareto optimal point [34, 35]. Studies with a posteriori articulation of preferences focus on providing the Pareto optimal set [33].

The weighted sum method can also be used to provide a single solution when a single set of weights reflects preferences. In this work, we incorporate a priori articulation of preferences, in which preferences for each objective are computed before optimizing the problem. We thus define a new objective function as a linear combination of the original objectives as follows:

$$F(x) = \sum_{i=1}^{7} w_i g_i(x),$$

where w_i represents the weight given to objective i. Given that $G(x)$ and $F(x)$ are composed of several non-linear objective functions, we first need to linearize them. To do so, we use additional sets of decision variables and constraints as follows.

Objective 1: Let tu be a continuous decision variable denoting the maximum number of unspent tokens among all physicians. This constraint aims for a fair distribution of shifts among different physicians according to their preferences. We then have that $g_1(x) = tu$, and

$$tu \geq \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} TK_i - P_{ijk} x_{ijk} \qquad \forall i \in \boldsymbol{I} \tag{14}$$

$$tu \geq 0. \tag{15}$$

Objective 2: For each $c \in \boldsymbol{C}$, we define the continuous decision variable pu_c equal to the maximum difference of assigned shifts among physicians of clinic c. We have that $g_2(x) = \sum_{c \in \boldsymbol{C}} pu_c$, and

$$pu_c \geq \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} x_{i'jk} - \sum_{j \in \boldsymbol{J}} \sum_{k \in \boldsymbol{K}} x_{ijk} \qquad \forall c \in \boldsymbol{C},\ \forall i, i' \in \boldsymbol{I}_c \tag{16}$$

$$pu_c \geq 0 \qquad \forall c \in \boldsymbol{C}. \tag{17}$$

Objective 4: For each $i \in \boldsymbol{I}$ and $j \in \boldsymbol{J}$, we define the continuous decision variable n_{ij} equal to the number of work days that have not been assigned according to a consecutive days pattern. We have that $g_4(x) = \sum_{i\in \boldsymbol{I}} \sum_{j\in \boldsymbol{J}} n_{ij}$, and

$$\sum_{k\in \boldsymbol{K}} x_{i(j+1)k} + \sum_{k\in \boldsymbol{K}} x_{i(j-1)k} - x_{ijk'} + n_{ij} + m_i \geq 0 \tag{18}$$

$$\forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I}_c,\ \forall j = 2 \ldots |\boldsymbol{J}| - 1,\ \forall k' \in \boldsymbol{K}$$

$$\sum_{k\in \boldsymbol{K}} x_{i2k} - x_{i1k'} + n_{i1} + m_i \geq 0$$

$$\forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I}_c,\ \forall k' \in \boldsymbol{K} \tag{19}$$

$$\sum_{k\in \boldsymbol{K}} x_{i(|\boldsymbol{J}|-1)k} - x_{i|\boldsymbol{J}|k'} + n_{i|\boldsymbol{J}|} + m_i \geq 0$$

$$\forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I}_c,\ \forall k' \in \boldsymbol{K} \tag{20}$$

$$n_{ij} \geq 0 \qquad \forall i \in \boldsymbol{I},\ \forall j \in \boldsymbol{J}. \tag{21}$$

Objective 5: For each $i \in \boldsymbol{I}$ and $j \in \boldsymbol{J}$, we define the binary variables l_{ij} equal to 1 if and only if physician i is assigned to more than 1 shift on day j. We have that $g_5(x) = \sum_{i\in \boldsymbol{I}} \sum_{j\in \boldsymbol{J}} l_{ij}$, and

$$\sum_{k\in \boldsymbol{K}} x_{ijk} - l_{ij} \leq 1 \quad \forall i \in \boldsymbol{I},\ \forall j \in \boldsymbol{J} \tag{22}$$

$$l_{ij} \in \{0, 1\} \qquad \forall i \in \boldsymbol{I},\ \forall j \in \boldsymbol{J}. \tag{23}$$

Objectives 6 and 7: For each $i \in \boldsymbol{I}$, we define the continuous decision variables h_i^1 and h_i^2 equal to the number of shifts that physician i works above or below, respectively, his/her target weekly load. Then, we have $g_6(x) = \sum_{i\in \boldsymbol{I}} h_i^1,\ g_7(x) = \sum_{i\in \boldsymbol{I}} h_i^2$, and

$$\sum_{j\in \boldsymbol{J}} \sum_{k\in \boldsymbol{K}} x_{ijk} - h_i^1 + h_i^2 = H_i \quad \forall i \in \boldsymbol{I} \tag{24}$$

$$h_i^1, h_i^2 \geq 0 \qquad \forall i \in \boldsymbol{I}. \tag{25}$$

We can now formulate the PCSP as the following MIP:

$$\text{minimize } w_1 tu + \sum_{c\in \boldsymbol{C}} w_2 pu_c - \sum_{i\in \boldsymbol{I}}\sum_{j\in \boldsymbol{J}}\sum_{k\in \boldsymbol{K}} w_3 P_{ijk} x_{ijk} + \sum_{i\in \boldsymbol{I}}\sum_{j\in \boldsymbol{J}} w_4 n_{ij}$$
$$+ \sum_{i\in \boldsymbol{I}}\sum_{j\in \boldsymbol{J}} w_5 l_{ij} + \sum_{i\in \boldsymbol{I}} w_6 h_i^1 + \sum_{i\in \boldsymbol{I}} w_7 h_i^2$$
$$\text{subject to } (2) - (25) \qquad .$$

We use the analytical hierarchy process (AHP) to determine the weights according to the objectives' priorities. AHP has been successfully applied to other personnel scheduling problems arising in health care [18, 19]. The first step in the AHP is to structure the problem as a hierarchy and to arrange the factors in layers descending from an overall goal. In our study, the overall goal is to plan the best schedule for clinics and physicians, and the medical staff and administrator are influencing factors in the lower layer of the overall goal. The objectives are placed in the successive layer as the bottom layer's alternatives.

The second step is the elicitation of pairwise comparison judgments in each layer of the structure. Saaty [36] suggests that pairwise comparisons should be done by rating in the range $[1/9, 9]$. A rating of 9 indicates that one objective is extremely more important than the other one. The rate of 7 is given, when an objective is strongly more important than the other one, and 5 shows that an objective is more important than the other one. A rate of 1 represents equal importance between two objectives. Rates of 2, 4, 6 and 8 indicate intermediate values in order to reflect fuzzy inputs. Reciprocal values are used for reflecting dominance of second alternative compared with the first one. The final step in the AHP is to establish the composite global priorities of each objective. In our study, we assume that the relative preference for each objective among medical staff and administrator is the same; therefore, the weight of each objective is determined by the head physician using the pairwise comparison matrix in Table 2.1. In order to calculate the relative importance weights, we first need to sum the values in each column of the pairwise comparison matrix and then divide each value in the matrix by its column total. After that, the average of each row gives us the relative importance of the corresponding objective. It is noteworthy that the objectives are normalized prior to assignment of the AHP weights. We normalize each objective by applying the method suggested by Grodzevich and Romanko [37], in which each objective is divided by the difference of the Nadir and Utopia points.

Table 2.1: Pairwise comparison of objectives

Objective	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)Fair distribution of shifts ($\sum_{c\in \boldsymbol{C}} pu_c$)	1	1/9	1	1/9	1/8	1/3	1
(2)Maximum unused tokens (tu)	9	1	9	1	2	8	9
(3)Preferred assignment ($\sum_{i\in \boldsymbol{I}} \sum_{j\in \boldsymbol{J}} \sum_{k\in \boldsymbol{K}} P_{ijk}x_{ijk}$)	1	1/9	1	1/9	1/8	1/3	1
(4)Block assignment ($\sum_{i\in \boldsymbol{I}} \sum_{j\in \boldsymbol{J}} n_{ij}$)	9	1	9	1	2	8	9
(5)Two shifts in one day ($\sum_{i\in \boldsymbol{I}} \sum_{j\in \boldsymbol{J}} l_{ij}$)	8	1/2	8	1/2	1	8	9
(6)Overtime ($\sum_{i\in \boldsymbol{I}} h_i^1$)	3	1/8	3	1/8	1/8	1	9
(7)Under time ($\sum_{i\in \boldsymbol{I}} h_i^2$)	1	1/9	1	1/9	1/9	1/9	1

2.4 A heuristic algorithm for the PCSP

In this section we present an approximate solution algorithm for the PCSP. Given the fact that the aforementioned model cannot be solved to optimality by a commercial solver in reasonable time, our main motivation behind devising a heuristic algorithm relies on developing a simple procedure that is capable of finding high quality solutions in reasonable CPU times. The proposed heuristic is a hybrid algorithm which combines two well-known metaheuristics: iterated local search (ILS) and variable neighborhood descent (VND). From now on, we refer to this hybrid algorithm as an iterated variable neighborhood descent (IVND) procedure. On the one hand, ILS is a procedure that builds a sequence of solutions generated by a heuristic, usually a simple local search, which can lead to better solutions than repeated random trials of that heuristic [38]. On the other hand, VND is a procedure that is based on a systematic exploration of a set of neighborhoods that modifies the structure of the solution space [39]. There are four components that need to be considered when designing an ILS algorithm: an initial solution, an embedded local search, a perturbation strategy, and an acceptance criterion. Algorithm 1 shows a generic ILS procedure. In what follows, we explain the components of our IVND procedure for the PCSP in details.

Algorithm 1 Iterated local search

$\boldsymbol{A}^0$ = GenerateInitialSolution
$\boldsymbol{A}^*$ = LocalSearch($\boldsymbol{A}^0$)
while termination condition has not been met **do**
  $\boldsymbol{A}^0$ = Perturbation($\boldsymbol{A}^*$, memory)
  $\boldsymbol{A}'$ = LocalSearch($\boldsymbol{A}^0$)
  $\boldsymbol{A}^*$ = AcceptanceCriterion($\boldsymbol{A}^*$,$\boldsymbol{A}'$,memory)

We first construct an initial feasible solution by solving a strong relaxation of the PCSP, denoted as PCSP-R, in which only the second objective is disregarded, and solve it using a standard optimization software. The idea behind this comes from the fact that preliminary experiments showed that the second objective, associated with the balancing of workload between physicians of the same clinic, is one of the most difficult objectives to optimize. This is partially attributed to the fact that a large number of dense constraints (16)–(17) are needed to linearize the minmax objective $g_2(x)$. When these constraints (and their associated objective) are removed from the MIP formulation, a standard optimization software is capable of solving the resulting relaxation much faster as compared to the original problem. Moreover, we note that relaxing these constraints do not actually lead to an infeasible solution associated with the hard constraints, but only to unbalanced workloads between physicians. Now, given that an optimal solution to this relaxation may not necessarily be a good solution to the original problem, we use a time limit, denoted as TL, for solving this relaxation with the commercial solver; our main goal is to have an initial feasible solution for the PCSP, but not necessarily an optimal solution to the relaxation.

The embedded VND is used to improve the solution obtained by solving the relaxed problem. The VND is applied by systematically searching in a set of 10 neighborhoods N_1, N_2, $\ldots$, N_{10} of the current feasible schedule. Given that the order of exploring the neighborhoods can have a substantial impact on the final results, we use the approach suggested in Burke et al. [30] and Burke et al. [31] for the case of nurse rostering problems, which consider the exploration of neighborhoods in a sequential order starting from N_1. The VND algorithm is applied using a first improvement strategy; it explores neighborhoods N_1, ..., N_{10} sequentially, until an improved solution is found. Each time the search finds a better solution than the current solution, it updates the current solution and restarts from neighborhood N_1.

The neighborhoods in our VND algorithm consist of swapping consecutive shifts among physicians. Let $\boldsymbol{S} = \{1, \ldots, |\boldsymbol{J}||\boldsymbol{K}|\}$ be the set of shifts in the entire planning horizon (one week). In what follows, schedules are represented with an $|\boldsymbol{I}| \times |\boldsymbol{S}|$ matrix $\boldsymbol{A}$, where $a_{is} = 1$ if physician $i \in \boldsymbol{I}$ works on shift $s \in \boldsymbol{S}$, and $a_{is} = 0$ otherwise. In our implementation of the VND algorithm, we explore $|\boldsymbol{J}||\boldsymbol{K}|$ types of neighborhood structures. In particular, for each $k = 1, \ldots, |\boldsymbol{J}||\boldsymbol{K}|$, N_k consists of solutions which can be reached by swapping k consecutive shifts having the same starting shift s_r between any two physicians. That is,

$$
\begin{aligned}
&N_k(\boldsymbol{A}) = \{\, \boldsymbol{A}^{'} : \exists i_1, i_2 \in I, i_1 \neq i_2, \text{ and } s_r \in \{1, \ldots, |\boldsymbol{J}||\boldsymbol{K}| - (k-1)\}\,, \\
&a^{'}_{i_1} = (\ldots, a_{i_1 s_r - 1}, a_{i_2 s_r}, a_{i_2 s_r + 1}, \ldots, a_{i_2 s_r + k}, a_{i_1 s_r + k + 1}, \ldots)\,, \\
&\qquad\qquad\qquad a^{'}_{i_2} = (\ldots, a_{i_2 s_r - 1}, a_{i_1 s_r}, a_{i_1 s_r + 1}, \ldots, a_{i_1 s_r + k}, a_{i_2 s_r + k + 1}, \ldots)\}\,.
\end{aligned}
$$

We note that a relevant feature of these neighborhoods is that all hard constraints are always satisfied if the swaps are performed among physicians of the same clinic. For any swaps among the physicians of two different clinics, we do not consider those that violate the feasibility of hard constraints. Figure 2.1 illustrates a schedule for 3 physicians and possible swaps in N_3 and N_{10}. A new schedule in $N_3(\boldsymbol{A})$ is obtained by the swap movements depicted with solid lines between physicians 1 and 3 with $s_r = 1$ and $k = 3$. That is, in the initial solution $\boldsymbol{A}$, physician 1 works the first shift but not the second and third ones, whereas physician 3 does not work on the first shift but does work on the second and third shifts. In the new schedule $\boldsymbol{A}^{'}$, after the swap is performed, physician 1 works in the second and third shift and physician 3 works only on the first shift, and the rest of the solution remains the same. Also, it should be noted that the movements change physicians' total number of working shifts. Dashed lines depict the only possible swap between physician 1 and 3 in $N_{10}(\boldsymbol{A})$, as in this neighborhood the entire weekly work schedules are exchanged between any two physicians.

	Mon		Tue		Wed		Thu		Fri	
Phys 1	1	0	0	1	1	0	0	0	1	0
Phys 2	1	0	1	1	0	0	0	0	0	1
Phys 3	0	1	1	0	1	0	1	1	0	0

Figure 2.1: VND neighborhood structures

The perturbation procedure is a diversification mechanism to move away from a local optimal solution. Our procedure works by partially destroying the current solution by removing some of its elements, and obtaining a new solution by repairing the perturbed solution using an MIP. Let ***Pt*** denotes the perturbation set that corresponds to the set of elements $(i, s) \in \boldsymbol{I} \times \boldsymbol{S}$ of $\boldsymbol{A}$, which values will be removed from the current solution. In our perturbation strategy, we do not perturb the improved part of the solution by the VND in the current iteration, hence we define ***Fe*** as the set of elements whose values have changed during the VND procedure in the current iteration. During the perturbation phase, a percentage of elements ($prec\%$) of the current solution $\boldsymbol{A}$ are selected randomly and if they do not belong to set ***Fe***, they will be added to set ***Pt***. We now define set

$$\boldsymbol{R} = \{(i, s) \in \boldsymbol{I} \times \boldsymbol{S} : (i, s) \notin \boldsymbol{Pr}\},$$

as the set of elements whose values will be fixed to their current values in the repairing phase.

For instance, if in the current solution $a_{is} = 1$ and $(i, s) \notin \boldsymbol{Pr}$, then we set $x_{ij(s)k(s)} = 1$

in the repairing procedure, where $j(s)$ and $k(s)$ denote the day and shift of such day associated with shift s, respectively. For repairing the perturbed solution, we solve a reduced version of PCSP-R in which the following constraints are added in order to fix the elements of $\boldsymbol{R}$ to their current values:

$$x_{ij(s)k(s)} = a_{is} \quad \forall (i,s) \in \boldsymbol{R}. \tag{26}$$

Note that partially fixing a subset of variables not only preserves a favorable part of the current solution obtained by the VND, but also considerably reduces the solution space without changing the structure of the model. The repairing process is done within a certain time limit *repair_time_limit*, since the main goal is to obtain a new trial feasible solution.

The acceptance criterion determines not only if a solution $\boldsymbol{A}'$ is accepted or not as the new current solution but also plays the role of controlling the balance between diversification and intensification of the search. We conduct a random walk with a very limited usage of memory as our acceptance criterion. That is, we apply the perturbation procedure to the most recently visited local optimal solution. However, we conduct a backtracking procedure and restart the search from the incumbent solution if no improved solution has been found in a given number of iterations, denoted as NIL. Let nit_{last} be the last iteration where a better solution was found, nit be the iteration counter, $F(\boldsymbol{A})$ be the objective value of solution $\boldsymbol{A}$, and $\boldsymbol{A}^*$ be the incumbent solution. The acceptance criterion is defined as

$$Accept(\boldsymbol{A}^*, \boldsymbol{A}, memory) = \begin{cases} \boldsymbol{A}^*, & \text{if } F(\boldsymbol{A}) \geq F(\boldsymbol{A}^*) \text{ and} \\ & nit_{last} - nit > NIL, \\ \boldsymbol{A}, & \text{otherwise.} \end{cases}$$

Finally, we use a time limit $maxtime$ as the termination criterion for our IVND procedure. The overall IVND procedure is outlined in Algorithm 2.

Algorithm 2 IVND procedure for the PCSP

```
Solve PCSP-R to obtain initial solution Ā
A* ← Ā
nit ← 0
while currenttime < maxtime do
    r ← 1
    while r ≤ |S| do
        Explore N_r(Ā) to obtain a local solution A'
        if F(A') < F(Ā) then
            Ā ← A'
            r ← 1
            Update Fe with changes in A'
        else
            r ← r + 1
    if F(Ā) < F(A*) then
        A* ← Ā
    else
        nit++
    if nit > NIL then
        Ā ← A*
        nit ← 0
    Update Pt and R
    Solve PCSP-R with (26) to obtain a new trial solution Ā
return A*
```

2.5 Computational results

In this section, we present the results of computational experiments we have run in order to compare and analyze the performance of the formulation and the proposed solution algorithm. We first describe the benchmark instances we have used. We then evaluate the impact of clinic scheduling constraints on physicians work schedules to provide some managerial insights into the added value of integrating physician and clinic scheduling decisions. Finally, we give numerical results to analyze the computational performance and limitations of the proposed formulation when solved with a standard optimization software and of our proposed IVND algorithm.

All experiments were run on a Dell station with an Intel(R) Core(TM) CPU i7-4790 processor at 3.60 GHz and 16 GB of RAM under Windows 7 environment. The formulations and algorithms were coded in C++, and the associated MIPs were solved using the Concert Technology of CPLEX 12.6.3. Table 2.2 shows CPLEX parameters used with non-default values in the computational study.

Table 2.2: CPLEX non-default settings

Parameter Name	Value
CPX_PARAM_THREADS	8
CPX_PARAM_MIPEMPHASIS	4
CPX_PARAM_TILIM	7200
CPX_PARAM_EPGAP	0.01

Table 2.3: Characteristics of benchmark instances

Class	#Days	#Shifts	# Clinics	# Physicians	# Clusters
A	5	2	5	15-30	3
B	5	2	10	60-100	5
C	5	2	13	133	5

2.5.1 Instance generation

A set of benchmark instances was generated using data obtained from a case study at MUCH. This set is divided into three classes of instances of increasing size with respect to the number of clinics and physicians. Table 2.3 gives details on some of the inputs of the problem. The largest class C corresponds to the real case study at MUCH and the other two were obtained by considering a subset of clinics and physicians of C.

Each class contains 10 instances that vary with respect to the number of arriving patients, number of available rooms, capacity of the waiting room, and physicians' preferred shifts. This information was generated as follows.

The weekly number of patients was randomly generated through the following uniform distribution:

$$NWV_c \sim U\left[N\bar{W}V_c(1-\beta), N\bar{W}V_c(1+\beta)\right],$$

where $N\bar{W}V_c = \sum_{i \in \boldsymbol{I}_c} H_i VPS_i$, and VPS_i is the average of the number of patients that physician i assessed during one shift calculated from historical data. $0 < \beta < 1$, is a parameter that controls the percentage of number of patients over or under the expected number. The rest of the parameters are generated proportionate to the number of weekly arriving patients and the number of available physicians in order to make resource capacity constraints tight enough. The number of examination rooms as

$$B \sim \left\lceil \frac{\sum_{c \in \boldsymbol{C}} \frac{NWV_c}{VPS_c}}{|\boldsymbol{J}||\boldsymbol{K}|} \right\rceil$$

and the waiting room capacity is defined as

$$D \sim \left\lceil \frac{\sum_{c \in \boldsymbol{C}} NWV_c}{|\boldsymbol{J}||\boldsymbol{K}|} \right\rceil.$$

The number of times that a cluster $t \in \boldsymbol{T}$ has to be scheduled together is randomly generated as

$$F_t \sim U \left[0, \min \left\{ |\boldsymbol{J}||\boldsymbol{K}|, \min_{c \in \boldsymbol{C} | c \in t} \left\{ \left\lceil \frac{NWV_c}{\sum_{i \in \boldsymbol{I}_c} VPS_i} \right\rceil \right\} \right\} \right].$$

Finally, the target level of fluctuation in the waiting room's utilization was set to $WU = 0.2$, since the management of the hospital would not prefer high level of fluctuations in the waiting room utilization over the week.

2.5.2 The analysis of integrating clinic scheduling with physician scheduling problems

We next investigate the impact of combining clinic scheduling with physician scheduling (denoted as integrated approach) compared to schedules obtained from solving the aforementioned problems separately (denoted as sequential approach). Afterwards, we analyze the impact of clinics' and administration's constraints on physicians' work schedules.

In the sequential approach, the clinic scheduling problem is solved with respect to the demand, resource, administration's and clustering constraints (i.e. (2)-(8)). Any feasible solution of the clinic scheduling problem is also feasible to the PSP; therefore, the solution of the clinic scheduling problem is directly plugged into the PSP. After that, the PSP is solved to optimality. Figure 2.2 depicts the average percent improvement in the seven objectives obtained from solving 10 small instances (in which the optimal solution can be obtained) as an integrated problem versus solving them based on the sequential approach. As the figure shows, the highest improvements are in the objectives related to physicians' preferences and ergonomic aspects (i.e. objectives 1, and 3-5); the demand related objectives (i.e. objectives 2, 6, and 7), on the contrary, get least improved. The reason is the demand constraints are included in the clinic scheduling problem. In brief, the above experiments clearly indicated the positive impact of incorporating physician scheduling problem into the clinic scheduling one particularly in terms of physicians' work schedule.

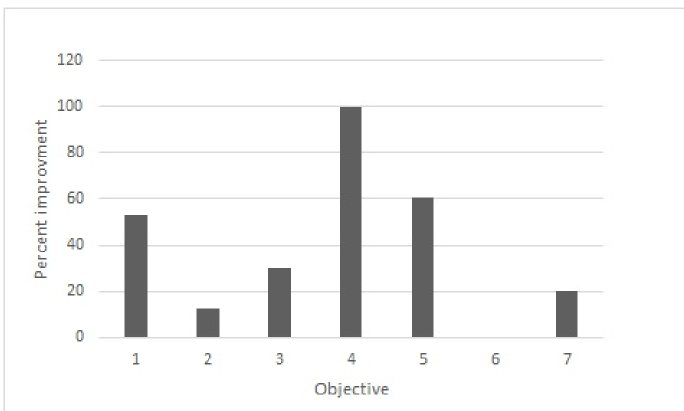

Figure 2.2: Comparison of the integrated versus sequential scheduling approaches

In order to analyze the impact of clinics' and administration' s constraints on physicians' work schedule, we first solve the PCSP formulation without considering clinics' and administration's constraints. We denote this formulation as PSP. Afterwards, we add each set of the aforementioned constraints, one at a time, and resolve the resulting problems. Table 2.4 shows the clinic's constraints considered in this experiment.

The analysis was conducted on those instances, where the solver can obtain the optimal solution. The results of this analysis are given in Figure 2.3. We use two different sets of

Table 2.4: Clinic's scheduling constraints

Constraint	Name	Description
(5), (6), (8)	ctT	simultaneous scheduling interdisciplinary clinics
(7)	ctWU	fluctuation level in waiting room utilization
(3)	ctB	examination rooms availability
(4)	ctD	waiting room availability

right-hand-side (RHS) values for exam rooms and waiting room capacity constraints: the nominal case and one in which the RHS or the available resource is increased by 20%. The horizontal axis represents each set of the constraints, while the vertical axis shows the percent increase in the objective function value of the PSP when each set of constraints is added.

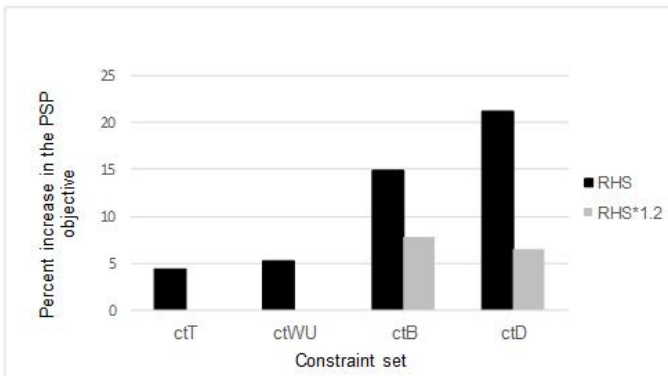

Figure 2.3: Analysis of adding clinic scheduling constraints to the PSP

The results given in Figure 2.3 demonstrate that incorporating the clinics' requirements in the PSP has a major impact on the objective function value. That is, physician preferences tend to be more violated when adding clinic constraints to the PSP model. On average, the waiting room capacity has the most impact in the deterioration of the objective function, followed by the number of available examination rooms. When each of those resources is increased by 20%, the objective function value improves significantly. The simultaneous scheduling requirements of interdisciplinary clinics and the fluctuation level in waiting room utilization seem to have a lower impact on the objective of the PSP.

The observed impact of clinic constraints on the PSP provides insights to the hospital administrator about the significance of each resource in improving physicians work schedules. In this study, counter intuitively, the waiting room is the crucial resource. In other words, increasing the capacity of the waiting room or limiting the number of admitted patients to a certain level are the choices with the highest potential improvement. Another observation is the amount of improvement gained by increasing one unit of each resource or relaxing a clinic scheduling constraint. For instance, we observe almost 7% improvement in physicians work schedules by 20% growth in the number of exam rooms. Furthermore, these experiments confirm that increasing the available resources for a fixed weekly demand provides more flexibility in the assignment of physicians that leads to higher quality schedules.

2.5.3 Analyzing the computational performance of the PCSP formulation and IVND algorithm

We next present the results of solving the considered instances with CPLEX and the proposed IVND algorithm. The goal is to analyze the limitations of the proposed PCSP formulation and to evaluate the performance and quality of the solution obtained with the IVND algorithm. Table 2.5 summarizes the values of the parameters of the IVND used in these experiments. It should be noted that the values were obtained through extensive preliminary experiments in order to fine-tune our solution methodology.

Table 2.5: The values of parameters of the IVND

$perc\%$	TL (sec.)	$repair\_time\_limit$ (sec.)	NIL
0.70	100	100	10

Table 2.6 represents the results for the set of class A instances. The first column of this table corresponds to the identification of each instance, where the first number presents the instance number within its class, and the second number after the dot implies the set of

weights that have been assigned to each objective in the objective function. The first set of weights have the distribution of $[1, 100, 1, 100, 90, 10, 1]$ for $w1 - w7$, and the objectives are not normalized prior to assignment of the weights. The second set of weights are the AHP-based weights, and the objectives are normalized. The second and third columns present the CPU time in seconds required to obtain the optimal solution by CPLEX and the IVND algorithm, respectively. The last column shows the percent deviation of the incumbent solution obtained in the first iteration of the IVND from the optimal solution, and it is calculated according to $100 * |Incumbent - Optimum|/|Optimum|$.

Table 2.6: Comparison of the IVND and solver on class A instances

Instance	CPLEX Time (sec.)	IVND Time (sec.)	VND %Dev.
1.1	2	12	0
2.1	<1	<1	0
3.1	<1	<1	0
4.1	<1	<1	0
5.1	<1	<1	0
6.1	6	106	6.4
7.1	<1	<1	0
8.1	9	213	11.9
9.1	6	100	10.2
10.1	7	156	7.3
1.2	3	17	0
2.2	<1	<1	0
3.2	<1	<1	0
4.2	7	<1	0
5.2	<1	<1	0
6.2	9	19	0
7.2	<1	<1	0
8.2	5	<1	0
9.2	4	15	0.8
10.2	<1	<1	0

The results of Table 2.6 demonstrate that the proposed formulation and the IVND are capable of finding the optimal solution for small instances of class A. As the VND results show, it is remarkable that the single iteration VND is not able to find the incumbent solution in all the test instances, and the iterated procedure significantly improves the VND. It is worth to note that class A instances are generated in order to demonstrate the advantage of implementing the iterative procedure, and also to test if the IVND heuristic is capable of obtaining the optimal solution.

Table 2.7 and 2.8 show the results for classes B and C instances. In both tables, the results are grouped in three categories: CPLEX, IVND, and the incumbent solution. More specifically, for CPLEX results, *%D1H* provides the percent deviation of the best solution (an upper bound (UB) on the optimal objective value) by the solver in 1 hour of CPU time, from the best solution (incumbent) known for the corresponding problem instance, which is obtained either by the solver or by the IVND in 2 hours of CPU time. This deviation is calculated according to $|UB - Incumbent|/|Incumbent| * 100$. *%D2H* is calculated the same as *%D1H* for an upper bound obtained after 2 hours of CPU time. It should be noted that *%D1H* and *%D2H* show the quality of solutions provided by CPLEX in 1 and 2 hours time limit compared to incumbent solutions. Note that an *N.A.* entry indicates that CPLEX was not able to find a feasible solution within the time limit. *%D1H'** and *%D2H'** are the percent deviation of the best solution from the incumbent solution, found by the IVND in 1 and 2 hours over 5 runs, respectively. They are calculated similarly to *%D1H* and *%D2H* according to the bounds obtained by the IVND algorithm. $\%\bar{D}1H'$ and $\%\bar{D}2H'$ are the percent deviation of the average solutions from the incumbent solution, obtained over 5 runs, in 1 and 2 hours. $STD.$ is the standard deviation of the percent deviation from the incumbent solution for the solutions, obtained in 2 hours over 5 runs. *Time* demonstrates the total time (in seconds) that has been spent for solving the MIP model during the iterative process. *Value* shows the objective function value of the incumbent solution. *Fair* is the measure of fairness in the incumbent solution which is calculated by $\left\lceil \frac{\sum_{c \in C} p\bar{u}_c}{|C|} \right\rceil$ where $p\bar{u}_c$ are the solution values of pu_c in the incumbent solution. *Ergo* is the measure of ergonomy, calculated as $\sum_{i \in I} \sum_{j \in J} \bar{l_{ij}}$ where $\bar{l_{ij}}$ are the solution values of l_{ij} in the incumbent solution. Finally, the last row presents the average values of corresponding columns.

Table 2.7: Comparison of the IVND and CPLEX on class B instances

Inst.	CPLEX		IVND						Incumbent		
	%DH1	%DH2	%D1H′*	%D2H′*	%D̄1H′	%D̄2H′	STD.	Time	Value	Fair	Ergo
1.1	5.55	0.07	0.14	0.00	0.19	0.07	0.05	6611	1405.00	2	1
2.1	31.58	10.50	0.10	0.00	0.24	0.10	0.12	6320	1010.00	1	2
3.1	0.10	0.00	0.10	0.10	0.17	0.12	0.05	6838	965.00	1	0
4.1	N.A.	N.A.	0.00	0.00	0.11	0.02	0.04	6543	1062.00	1	2
5.1	7.09	7.09	0.09	0.00	0.14	0.03	0.08	6747	1157.00	1	2
6.1	10.73	10.63	0.00	0.00	0.04	0.04	0.06	6673	960.00	1	0
7.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6915	955.00	1	0
8.1	0.11	0.11	0.11	0.00	0.13	0.11	0.07	6721	949.00	1	0
9.1	0.13	0.13	0.00	0.00	0.00	0.00	0.00	6730	747.00	1	0
10.1	9.63	1.78	0.00	0.00	0.20	0.02	0.04	6817	1121.00	1	2
1.2	1.56	1.56	1.56	0.00	1.56	1.25	0.70	6604	0.02	1	5
2.2	0.86	0.86	0.07	0.00	0.32	0.07	0.05	6747	0.27	1	7
3.2	0.91	0.84	0.00	0.00	0.45	0.07	0.09	6713	0.29	1	0
4.2	1.27	1.09	0.00	0.00	0.04	0.04	0.09	6305	0.17	1	2
5.2	0.12	0.12	0.10	0.00	0.10	0.08	0.04	6800	0.26	1	0
6.2	2.27	2.08	0.10	0.00	0.12	0.08	0.05	6640	0.20	1	0
7.2	3.37	0.00	0.27	0.04	0.50	0.18	0.13	6638	0.09	1	1
8.2	10.20	10.20	0.00	0.00	0.06	0.06	0.08	6254	0.16	1	0
9.2	3.39	3.39	0.00	0.00	0.44	0.25	0.14	6648	0.06	1	0
10.2	2.18	0.03	0.23	0.00	0.27	0.14	0.12	6741	0.09	1	1
Ave.	4.79	2.66	0.14	0.01	0.25	0.14	0.10	6650.24		1.05	1.25

From the results provided in *%D1H′* and *%D1H* columns of Tables 2.7 and 2.8, we can observe that in 1 hour CPU time, the proposed IVND algorithm is able to outperform CPLEX by finding high quality solutions with the average percent deviation of 0.14 and 0.82 compared to 4.79 and 20.71, respectively, for class B and C instances. When the time limit is increased to 2 hours, by comparing the average of column *%D2H′* with *%D2H*, it is apparent that the IVND algorithm finds notably better solutions than CPLEX, with the average percent deviation of 0.01 for class B instances and 0.00 for class C instances compared to 2.66 and 19.58, respectively. It is noteworthy that in 38 instances out of all 40 instances, the incumbent solution is found by the IVND. The average values of %D̄1H′ and %D̄2H′ indicate the consistency of the solution methodology in providing high quality solutions in different runs. On average over all the instances, 1 hour average solution is only 0.25 and 2.55 percent away from the incumbent solution for class B and class C,

Table 2.8: Comparison of the IVND and CPLEX on class C instances

Inst.	CPLEX		IVND						Incumbent		
	%DH1	%DH2	$\%D1H'^*$	$\%D2H'^*$	$\%\bar{D}1H'$	$\%\bar{D}2H'$	STD.	Time	Value	Fair	Ergo
1.1	54.58	54.58	0.89	0.00	8.97	5.51	5.58	5206	1354.00	2	2
2.1	21.46	21.46	0.07	0.00	6.16	1.41	2.69	5203	1421.00	2	2
3.1	30.78	30.78	0.08	0.00	0.25	0.17	0.13	5907	1293.00	2	2
4.1	40.10	39.94	0.00	0.00	5.84	5.67	3.18	5607	1232.00	2	0
5.1	44.28	44.17	0.00	0.00	0.49	0.42	0.24	5707	901.00	1	2
6.1	24.38	24.38	6.84	0.00	8.57	3.89	3.49	5607	1214.00	2	2
7.1	24.05	24.05	6.90	0.00	7.16	5.64	3.15	5681	1231.00	2	1
8.1	15.70	15.70	0.00	0.00	0.34	0.27	0.17	5808	1312.00	2	2
9.1	44.28	28.00	0.00	0.00	0.15	0.11	0.25	6008	1093.00	2	1
10.1	30.51	30.51	0.00	0.00	1.91	0.36	0.22	5931	1003.00	1	1
1.2	28.74	27.94	0.00	0.00	6.52	4.87	4.48	5210	0.04	2	2
2.2	6.68	6.57	0.85	0.00	0.91	0.66	0.38	5324	0.18	2	2
3.2	4.15	4.09	0.12	0.00	0.26	0.18	0.14	5823	0.16	2	5
4.2	9.00	8.72	0.00	0.00	0.14	0.08	0.06	5616	0.18	1	6
5.2	6.04	5.98	0.13	0.00	0.51	0.24	0.31	5933	0.15	1	4
6.2	7.18	6.97	0.25	0.00	1.50	0.39	0.39	5422	0.24	1	6
7.2	8.44	3.97	0.00	0.00	0.27	0.11	0.08	5722	0.16	1	5
8.2	2.44	2.44	0.34	0.00	0.68	0.38	0.39	6022	0.17	2	2
9.2	6.11	5.98	0.00	0.00	0.17	0.09	0.15	5933	0.16	1	6
10.2	5.27	5.27	0.00	0.00	0.12	0.10	0.08	6012	0.16	1	3
Ave.	20.71	19.58	0.82	0.00	2.55	1.53	1.28	5684.09		1.60	2.80

respectively. In 2 hours time limit, the average of $\%\bar{D}2H'$ shows the average solution deviated 0.14 and 1.53 from the incumbent solution for class B and C, respectively. The average of STD. demonstrates that percent deviation of solutions obtained in 2 hours time limit are close to the average solution percent deviation, as for class B and class C they are merely 0.10 and 1.28, respectively. We observe almost 1000 seconds increase in the average time spent on the local search phase during the iterative process from class B to class C instances. The average values of Fair and Ergo columns are proper evidence of the quality of the incumbent schedules. The average values of Fair columns show that there is at most 1.05 and 1.60 shifts difference in terms of number of assigned shifts among physicians of each clinic in class B and class C, respectively. The average of Ergo column implies that the average number of times that a physicians work two shifts per day is only 1.25 for class B and 2.80 for class C, respectively.

2.6 Concluding remarks

Inspired by the challenge of incorporating physicians' preferences and availabilities into the scheduling of clinic sessions in a real ambulatory care polyclinic, we proposed a novel multi-objective mixed-integer programming model for an integrated physician and clinic scheduling problem (PCSP). More specifically, the model aims at assigning physicians associated with different multidisciplinary, interdisciplinary, and high-throughput clinics to treatment sessions such that all previously booked patients are assessed. Along with the demand satisfaction constraints, the limited capacity of resources, such as waiting and treatment rooms that must be shared among the aforementioned clinics, were also taken into account as hard constraints. On the contrary, other conflicting conditions such as particular preferences of physicians in addition to the ergonomic and fairness of assigned shifts among them were considered as soft constraints and their violations were minimized in the model. In order to better justify the significance of integrating physician scheduling problem (PSP) with clinic session scheduling in such health delivery centers, we measured the impact of adding hard constraints associated with clinic resources and administrative rules into the PSP model on a set of randomly generated benchmark instances.

The complexity for optimally solving the proposed PCSP model with a standard optimization software motivated us to develop an iterated variable neighborhood descent (IVND) algorithm to obtain high quality solutions in a reasonable time limit. The algorithm combines iterated local search (ILS) and variable neighborhood descent (VND) procedures, where the ILS generates a sequence of schedules generated by VND algorithm by using a perturbation strategy. Our computational results on the aforementioned test instances revealed the high quality of schedules provided by this algorithm in comparison with a standard optimization software.

The PCSP formulation and solution algorithm proposed in this article can be embedded into a decision support system to assist the polyclinic's management in a more efficient scheduling of different clinics with their designated physicians and in adjusting the schedule based on the updated data regarding the number of booked patients, number of clustered sessions, physicians days-off, etc. Another interesting extension of the current study would be to integrate the uncertainty inherent in the number of weekly arriving patients, emergency referrals, and the availability of physicians in different shifts in the PCSP model.

Acknowledgments

This research was partly funded by the Canadian Natural Sciences and Engineering Research Council under grants 418609-2012 and 402043-2011. This support is gratefully acknowledged. The authors thank four anonymous reviewers for their valuable comments on a previous version of the paper.

Chapter 3

A physicians planning framework for polyclinics under uncertainty

Abstract

In this paper, we present a comprehensive bi-level physicians planning framework for polyclinics under uncertainty. The first level focuses on clinic scheduling and capacity planning decisions, whereas the second level deals with physicians scheduling and operational adjustments decisions. In order to protect the generated schedules against demand uncertainty, the first level is modeled as an adjustable robust scheduling problem which is solved using an ad-hoc cutting plane algorithm. To cope with variability in patients' treatment times, we formulate the second level as a two-stage stochastic problem and use a sample average approximation scheme to obtain solutions with small optimality gaps. We use a Monte-Carlo simulation algorithm and data obtained from a university health center in Montreal, Canada, to demonstrate the benefits of our planning framework. In particular, we show that the schedule generated by our approach is superior in terms of total cost as compared with the one obtained from a single-level deterministic model.

3.1 Introduction

Manpower planning in service industries frequently follows a three-level procedure: planning, scheduling and allocation [40–42]. The first level deals with the operating policies of service centers. The second level specifies the personnel working shifts over the planning horizon. The last level focuses on workforce allocation, by taking into account the decisions made in previous levels, and making the required last-minute operational adjustments.

This paper introduces a planning framework for physicians in ambulatory polyclinics. Polyclinics are health care facilities consisting of multiple multidisciplinary, interdisciplinary and high throughput clinics that differ in terms of patient flow and treatment time and share limited resources. From a planning perspective, interdisciplinary clinics are the most critical ones given that they collaborate in the assessment and treatment of patients, which requires them to be scheduled simultaneously for a predetermined minimum number of shifts.

In polyclinic physicians planning, the first level corresponds to the *clinic scheduling and capacity planning problem* (CSCPP). The CSCPP encompasses long-term (strategic) decisions to determine the number of patients to admit, the total working hours of physicians, and the number of required examination rooms. In this level, clinic schedules are determined according to a weekly demand forecast. Nevertheless, over a long term planning horizon (e.g., one year) the weekly forecast can fluctuate within a given uncertain interval. Hence, hospital managers seek a robust strategic planning tool to maximize the number of patients who can be served on a weekly basis, even in the presence of worst-case (extremely high) demand scenarios, while determining the optimal personnel and resource requirements at each shift of the planning horizon. The aforementioned clinics' schedule along with the capacity plan are expected to be effective over a period of six months to a year.

The second level of the planning process involves the tactical assignment of physicians to the established shifts while taking into account the requirements of the shifts and physicians preferences. We refer to this problem as the *physician scheduling problem* (PSP). Generally speaking, PSPs consist of constructing work schedules for physicians in a planning horizon such that at each shift there are enough physicians to satisfy demand [3]. Similarly, in polyclinics physicians are usually scheduled on a weekly-basis once the number of patients scheduled for every shift, physicians availability, and preferences are known. It is worth noting that the outputs of the first level CSCPP act as an input to the second level PSP. Hospital administration can employ the PSP as a tactical planning tool to obtain physicians' work schedules at the beginning of each week such that all scheduled patients can be assessed, while taking into account capacity constraints and physicians' preferences.

Uncertain events disturb planned operations in health services. For instance, physicians' schedules are affected by the uncertainty in patients' treatment times. On the one hand, if the actual treatment times are longer than the estimated ones, the number of scheduled physicians might not be enough to serve all patients during the regular shifts. In this case, either patients would need to be rescheduled on another session or extra resources, such as on-call doctors or overtime shifts, would need to be deployed. On the other hand, if treatment times are shorter that the estimated ones, physicians would be idle which is undesirable for the administration. These scenarios lead to the third level of the physicians planning procedure, which is to perform last-minute operational adjustments to the planned schedules (i.e., prior to each shift). In this operational planning level, proper corrective (recourse) actions must be foreseen in order to minimize the average expected cost of schedules in the presence of uncertainty in treatment times.

Inspired by a real case study in an ambulatory cancer treatment polyclinic of the McGill University Health Centre (MUHC) in Montreal, Canada, we present a comprehensive bi-level physician planning framework. A distinguishing feature of this framework is that it takes into account the uncertainty in the number of arriving patients in the first level and the uncertainty in patients' treatment times in the second level. The decision levels in our approach are similar to the ones introduced by Zaerpour et al. [43] and follow the three-stage manpower planning process commonly used in service centers. Figure 3.1 summarizes the physicians planning procedure in polyclinics, and displays our planning framework.

At the first level, clinics' weekly sessions and their associated capacities are determined on a long-term cyclic weekly basis. Given the strategic nature of this decision level, the goal of polyclinic managers is to maximize the number of patients that can be served per week for each clinic under uncertain demand. That is, the managers seek a robust capacity plan such that the number of rejected patients under high-volume demand scenarios is

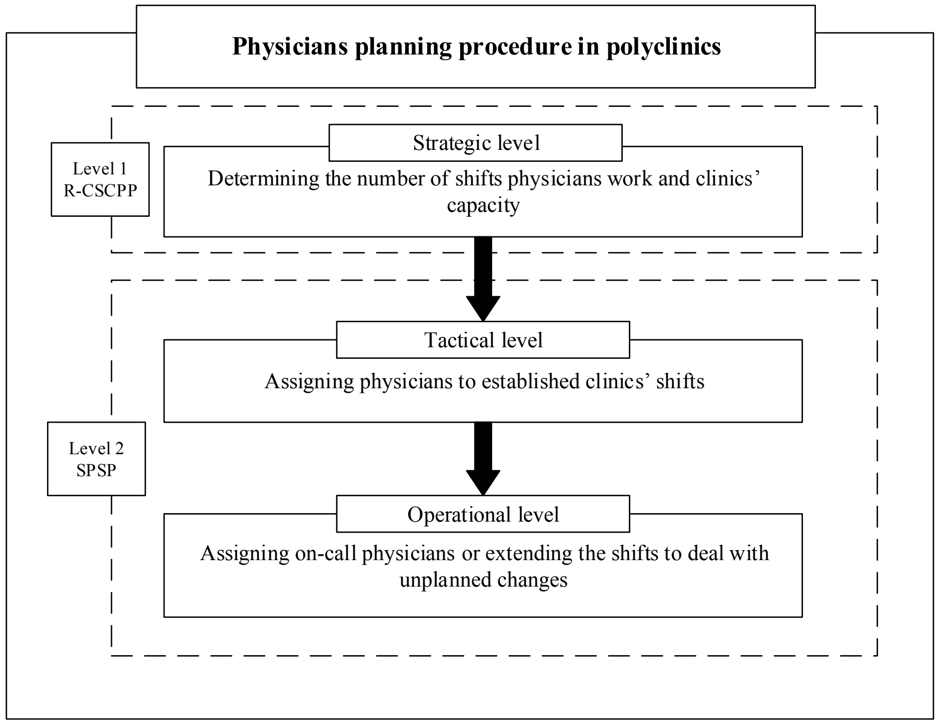

Figure 3.1: Physicians planning procedure in polyclinics

minimized. For this reason, we formulate this problem as an adjustable robust optimization problem and denote it as the *robust clinic scheduling and capacity planning problem* (R-CSCPP). The R-CSCPP provides decision-makers the possibility to control the level of conservatism of the plan by considering a budget of uncertainty. At the second level, we integrate the tactical decisions (i.e., weekly physician scheduling) with operational adjustments under uncertain treatment time within a single decision problem and refer to it as the *stochastic physician scheduling problem* (SPSP). A two-stage stochastic program is used to model in an integrated way these tactical and operational decisions. In particular, the physicians schedule corresponds to the first-stage decisions whereas the operational adjustments (i.e., calling on-call physicians and/or extending the shift duration) are the second-stage decisions (recourse actions) as a response to random treatment time scenarios.

The main contributions of our paper are as follows:

- To propose a comprehensive bi-level physician planning framework for polyclinics including strategic, tactical, and operational decisions while incorporating uncertainty in the amount of demand and treatment times.

- To present a mixed integer programming formulation (MIP) for the first level R-CSCPP in which a budget of uncertainty is used to control the level of conservatism. Given that this MIP requires an exponential number of constraints to model the demand uncertainty set, we develop an ad-hoc cutting plane algorithm for efficiently solving it.

- To present a two-stage integer stochastic programming formulation for the SPSP. The first-stage decisions focus on the assignment of physicians to regular or on-call duties in different shifts and the second-stage (recourse) actions consider calling on-call doctors and assessing patients during extended shift if treatment times are stretched-out. We use a sample average approximation (SAA) scheme to obtain feasible solutions for this problem with small statistical optimality gaps.

- To develop a Monte-Carlo simulation algorithm to demonstrate the importance and potential benefits of our planning framework using real data from a polyclinic in Montreal, Canada.

The remainder of this paper is structured as follows. Section 3.2 gives a concise review of the related literature on personnel scheduling problems, with an emphasis on applications in health care. In Section 3.3, we first present the CSCPP and its robust counterpart and describe an MIP and the cutting plane algorithm used for solving it. In Section 3.4, we introduce the deterministic PSP and stochastic SPSP and describe the SAA scheme used for solving the latter. Using the data provided by the university health center, in Section 3.5 we present the results of computational experiments to evaluate the performance of the proposed solution algorithms. We also compare the generated schedules with the results of a deterministic model that integrates clinics and physicians scheduling into a single level procedure. Finally, concluding remarks are given in Section 3.6.

3.2 Literature Review

Workforce allocation and personnel scheduling problems have been widely studied in the literature. Numerous applications arising in different service industries have been considered such as telephone operators, flight crews, bus drivers, physicians and nurses. We refer to Ernst et al. [13], Burke et al. [16], Van den Bergh et al. [14], and Erhard et al. [1], for review papers on these classes of decision problems.

Beaulieu et al. [17] propose the first MIP to model a PSP arising in the emergency room (ER) of a major hospital in Montreal. They divide scheduling constraints into two main categories: *compulsory* and *flexible*. Rousseau et al. [10] argue that a combination of constraint programming techniques with local search heuristics can be a promising generic methodology for solving a wide variety of PSPs. Carter and Lapierre [11] study physicians of ERs at six major hospitals in Canada and propose a tabu search algorithm for assigning physicians to work shifts. In Bard et al. [44], a three-phase methodology is proposed to generate monthly clinic assignments for internal medicine housestaff. Brunner [15], Brunner et al. [4], Brunner et al. [23], and Stolletz and Brunner [22] study flexible physicians scheduling problems in which the shifts have variable starting time and duration. Gunawan and Lau [26] and Van Huele and Vanhoucke [27] integrate scheduling physicians with other types of decision problems (i.e., duty and surgery scheduling) and consider resource capacity constraints (e.g., available operation rooms, available recovery beds). Tohidi et al. [45] propose an integrated clinic and physician scheduling problem in polyclinics in a deterministic context.

Some studies focus on a single level manpower planning (e.g., staffing or tactical planning), and consider the variability in daily demand or personnel's capacity. Bard and Purnomo [42, 46, 47] present a reactive and real-time scheduling approach to adjust midterm schedules on a daily-basis in response to fluctuations in demand. They apply their method as a recourse decision-making tool for the assignment of nurses to work shifts in a hospital in the U.S. Hur et al. [48] also investigate real-time work schedule adjustment decisions, and provide heuristics to solve the considered problem. Easton and Goodale [49] analyze different strategies (e.g., cross-trained workers, overtime, call-in employee, and temporary workers) for scheduling personnel with unplanned absenteeism in a real-time decision-making framework. Gross et al. [50] propose an MIP model for rescheduling physicians in a hospital as a response to unplanned absences of scheduled personnel. The goal is to minimize the changes in the initial roster while satisfying work regulations, qualification requirements, and physician preferences. Campbell [51] shows cross-trained workers are beneficial to service centers with multiple departments where the demand is subject to uncertainty. Wright and Mahar [52] investigate on centrally scheduling cross-trained nurses across multiple departments in two hospitals in the U.S. They determine the likelihood of violating the minimum nurse-to-patient ratios through queueing methods.

Other studies aim at integrating different levels of manpower planning, and analyzing the decisions to be taken at each level. Abernathy et al. [40] introduce a comprehensive three-level manpower planning procedure, which includes policy, staffing, and scheduling decisions, for service industries with demand fluctuations. They present an iterative solution methodology as well as a chance constraint-based method to solve this problem. Venkataraman and Brusco [53] propose an integrated staffing and scheduling action plan that allows recursion between the staffing and scheduling models and thus enables management to rapidly evaluate the impact of both staffing and scheduling policies. Wright et al. [54] merge nurse staffing and scheduling by developing a model that incorporates nurse-to-patient ratios and controls the amount of work given to each nurse. Wright and Bretthauer

[55] study two-phase scheduling and rescheduling models for cross-trained nurses in a hospital in the U.S. The authors propose a procedure that assigns the nurses to shifts over a mid-term planning horizon in the first phase. In the second phase, the schedules generated during the first phase are adjusted according to the demand at the beginning of each shift. Maenhout and Vanhoucke [56] integrate staffing and scheduling decisions in a nurse scheduling problem and demonstrate that the schedules generated by the integrated approach are preferable in terms of cost and personnel satisfaction. Ingels and Maenhout [57] propose a two-phase framework for tackling personnel scheduling problems under demand and employees' availability uncertainty. In the first phase, they assign employees to shifts as regular and reserved workforce. In the second phase, they simulate the random processes, and if needed, optimize the utilization of reserved workforce via an optimization model. They suggest a sequential solution method where the output of the first phase is inserted into the operational phase. In Zhong et al. [58], nurse staffing and scheduling are done in a two-stage heuristic algorithm that considers fairness objective on weekends and balanced workloads under different shift designs. Additionally, Chiaramonte and Caswell [59] propose an agent-based nurse rostering system that integrates nurse scheduling and rescheduling into a single framework. In their method, the agent system minimizes the difference between the initial roster and the rescheduled one using a local search algorithm.

A handful of studies integrate some stages of manpower planning into a two-stage stochastic programming model. Campbell [41] analyzes a staffing and scheduling problem for cross-trained workers in a service industry with multiple departments and formulates it as a two-stage stochastic program. The first-stage decisions are to assign workers to work tours, and the recourse actions are to assign cross-trained workers. Bard et al. [60] tackle the workforce planning at one of the U.S. post service distribution centers by using stochastic programming. The staffing decisions regarding full time and part time workers are made in the first stage, and the final allocations of workers to the shifts and decisions regarding the assignment of additional resources are made during the second-stage. Punnakitikashem et al. [61] formulate the workload assignment of nurses at a hospital in the U.S. as a two-stage stochastic program. The authors consider the first-stage decisions to be the assignment of patients to nurses. The recourse actions are considered as the amount of direct or indirect care performed by each nurse. Punnakitikashem et al. [62] extend this work to consider the nurse staffing problem along with the workload assignment problem. The authors introduce a two-stage stochastic program in which the staffing decisions are made in the first stage, and in the second-stage the duty assignments are executed with respect to outcomes of uncertain parameters. Zhu and Sherali [63] formulate a multi-category workforce planning problem with recruitment capacity constraints under demand uncertainty as a two-stage stochastic program. The staffing and allocation decisions are considered as first-stage decisions, and the workload assignment decisions are considered as recourse actions. Kim and Mehrotra [64] formulate the scheduling and rescheduling of nurses in polyclinics as a two-stage stochastic program, and propose a formulation which defines the convex hull of the second-stage MIP. El-Rifai et al. [65] propose a two-stage stochastic program for modeling a personnel scheduling problem in an emergency department with the goal of minimizing the expected patients wait time.

The above literature review indicates the paucity of research on incorporating uncertainty into physician scheduling problems. While the uncertainty has been only considered in some stages of workforce planning (e.g., strategic or tactical level) in service industries, to the best of our knowledge, no prior contribution exists that investigates the three stages of manpower planning considering uncertainty under a unifying decision framework. Therefore, this article aims to fill this void in the literature given the fact that adopting a deterministic approach could significantly affect the economic viability of the schedule.

3.3 The Clinic Scheduling and Capacity Planning Problem

As shown in Figure 1, the CSCPP deals with the strategic decisions in terms of clinics' weekly sessions in addition to their associated capacities, i.e., the number of patients to admit, treatment rooms and physicians' working hours. The goal is to maximize the service level (i.e., the number of patients that can be visited) under uncertain weekly demand for each clinic. We first formally define the deterministic CSCPP and provide an MIP formulation for it. We then present the robust R-CSCPP together with an MIP formulation and an exact solution algorithm for solving it.

3.3.1 Problem Definition and Formulation

Let $C, I, J,$ and K denote the sets of clinics, physicians, days and shifts per day, respectively. A shift in our study corresponds to a 4-hour period. We also assume (equal length) morning and afternoon shifts for each day of the week. The set of subsets of interdisciplinary clinics is denoted by T. Each clinic in subset $t \in T$ must be scheduled simultaneously with other clinics in the same subset for at least F_t shifts. The number of patients that can be assessed by a physician in clinic $c \in C$ in one shift is denoted as VPS_c. Let $\hat{H}_i$ and $\hat{Y}_c$ denote the minimum number of shifts that must be assigned to physician $i \in I$ and clinic $c \in C$, respectively. For every $c \in C$, PU_c is a measure of fairness that limits the total difference among the physicians in clinic c in terms of number of assigned shifts. For each $i \in I$, f_{1i} represents the cost of assigning physician i to a shift. Let f_3 be the cost of clinics' staff (non-physicians) per shift. Let R_c be the number of rooms required for non-physician staff in clinic $c \in C$ in a shift. We also assume that every physician needs one examination room when s/he works. Let f_2 denote the cost of using a room and NWV_c the number of patients per week visiting clinic $c \in C$. When demand is more than the planned capacity, we assume that patients will not be admitted and will thus have to be transferred to other hospitals, incurring a cost of f_4. The CSCPP consists of determining: *(i)* the set of shifts that each clinic is open during the week, *(ii)* a tentative work schedule for physicians affiliated with each clinic, *(iii)* the capacity of each clinic and the number of patients that can be assessed in each shift, and *(iv)* the number of rooms allocated to each shift, such that the demand, restrictions on interdisciplinary clinics, and physicians requirements are satisfied. The goal is to minimize the total cost of resources (physicians, rooms and clinics' staff) and the cost of rejected patients.

For each $j \in J$ and $k \in K$, we define clinic assignment variables y_{cjk} equal to 1 if and only if clinic $c \in C$ is assigned to shift k on day j. Similarly, for every $j \in J$ and $k \in K$, we define physician assignment variables x_{ijk} equal to 1 if and only if physician $i \in I$ is assigned to day j and shift k. For each $t \in T$, $j \in J$, $k \in K$, we define binary decision variables g_{tjk} equal to 1 if and only if all clinics in subset t are assigned to shift k on day j. For each clinic $c \in C$, let nd_c be an integer variable equal to the number of rejected patients. For every physician $i \in I$, h_i is an integer variable representing the total number of shifts assigned to physician i. Finally, r is an integer variable denoting number of required rooms per shift. Using these sets of variables, the CSCPP can be formulated as follows:

$$\text{minimize} \sum_{i\in I} f_{1i}h_i + f_2 r + \sum_{c\in C}\sum_{j\in J}\sum_{k\in K} f_3 y_{cjk} + \sum_{c\in C} f_4 nd_c \tag{27}$$

$$\text{subject to} \sum_{i\in I_c}\sum_{j\in J}\sum_{k\in K} VPS_c x_{ijk} + nd_c \geq NWV_c \quad c\in C \tag{28}$$

$$\sum_{i\in I} x_{ijk} + \sum_{c\in C} R_c y_{cjk} \leq r \quad j\in J, k\in K \tag{29}$$

$$\sum_{i\in I_c} x_{ijk} \geq y_{cjk} \quad c\in C, j\in J, k\in K \tag{30}$$

$$\sum_{c\in C|c\in t} y_{cjk} \geq |t| g_{tjk} \quad t\in T, j\in J, k\in K \tag{31}$$

$$\sum_{j\in J}\sum_{k\in K} g_{tjk} \geq F_t \quad t\in T \tag{32}$$

$$\sum_{j\in J}\sum_{k\in K} x_{ijk} \leq h_i \quad i\in I \tag{33}$$

$$\sum_{j\in J}\sum_{k\in K} x_{ijk} \geq \hat{H}_i \quad i\in I \tag{34}$$

$$\sum_{j\in J}\sum_{k\in K} y_{cjk} \geq \hat{Y}_c \quad c\in C \tag{35}$$

$$h_i - h_{i'} \leq PU_c \quad c\in C, i, i'\in I_c \tag{36}$$

$$\sum_{k\in K} x_{ijk} \leq 1 \quad i\in I, j\in J \tag{37}$$

$$x_{ijk},\ y_{cjk},\ g_{tjk} \in \{0,1\} \quad i\in I, j\in J, k\in K, c\in C, t\in T \tag{38}$$

$$h_i,\ r,\ nd_c \in \mathbb{Z}^+ \quad i\in I, c\in C. \tag{39}$$

The first three terms of the objective function are the total resource cost (i.e., cost of physicians, rooms, and clinics) and the last term is the total cost of rejected patients. Constraints (28) ensure that each clinic demand (i.e., visiting patients) is either served by physicians or rejected. Constraints (29) specify the number of required rooms in each shift.

Constraints (30) link the assignment of physicians with corresponding clinics. Constraints (31) and (32) guarantee that every clinic in each subset of interdisciplinary clinics is assigned with other clinics in the subset for the predefined number of shifts. In particular, constraints (31) force g_{tjk} to take value 1 if and only if all clinics in cluster t are assigned to shift k on day j, and constraints (32) ensures that all clinics in cluster t are scheduled simultaneously at least F_t shifts in a week. Constraints (33) record the total number of shifts that each physician works. Constraints (34) ensure that physicians are assigned to their minimum required shifts. Similarly, constraints (35) assure the minimum number of shifts that each clinic must be open. Constraints (36) guarantee a fair distribution of shifts among the physicians affiliated with the same clinic. More specifically, this condition restricts the difference in the number of assigned shifts between any two physicians of a clinic into a threshold value PU_c. Constraints (37) prevent physicians from working more than one shift per day. Finally, constraints (38) and (39) are the standard integrality and non-negativity constraints. We note that the physicians' schedule that is provided by the CSCPP can be considered as a rough estimate of the actual shift assignment that will be provided by the second-level SPSP.

3.3.2 A Robust Formulation for the CSCPP

In practice, the weekly demand NWV_c used in the CSCPP is usually not known in advance. At the strategic decision-making level, where it is not possible to have access to accurate information on the number of arriving patients to clinics on a weekly basis, estimating probability distributions associated with clinics' demand is not straightforward. Moreover, polyclinic managers are interested in finding robust cyclic schedules that could work well even in worst-case scenarios of increased demand. Therefore, we adopt a robust optimization approach in which weekly demands NWV_c are modeled with an interval of uncertainty. By estimating the upper and lower bound of the demand for each clinic, the uncertain demand set of clinic $c \in C$ is defined as:

$$\Omega_c = \left\{NWV_c \in \mathbb{Z}^+ : \underline{NWV_c} \leq NWV_c \leq \overline{NWV_c}\right\},$$

where $\mathbb{Z}^+$ is the set of positive integers and $\overline{NWV_c},\ \underline{NWV_c} \in \mathbb{Z}^+$. Given that it is very unlikely that all uncertain demand simultaneously achieve their upper bounds, we formulate the R-CSCPP such that the plan's cost is as small as possible when the demand takes its worst case scenario within a certain level of conservatism. In particular, we use a budget of uncertainty (Γ) similar to the one initially proposed by Bertsimas and Sim [66] to control the degree of conservatism desired during the decision-making process. Even though each clinic's demand may vary within the corresponding interval, we restrict the overall demand of all clinics to Γ by considering $\sum_{c \in C} NWV_c \leq \Gamma$. Let $NWV = (NWV_1, \ldots, NWV_{|C|})$ denote the vector of demands for each of the $|C|$ clinics. The demand uncertainty set Ω is defined as:

$$\Omega = \left\{NWV : NWV_c \in \Omega_c, \forall c \in C, \sum_{c \in C} NWV_c \leq \Gamma\right\},$$

where $\omega \in \Omega$ is a possible outcome of the uncertainty set. The R-CSCPP is formulated as follows:

minimize η

subject to (29)–(39)

$$\eta \geq \sum_{i \in I} f_{1i} h_i + f_2 r + \sum_{c \in C} \sum_{j \in J} \sum_{k \in K} f_3 y_{cjk}$$

$$+ \sum_{c \in C} f_4 nd_c^{\omega} \qquad \omega \in \Omega \tag{40}$$

$$\sum_{i \in I_c} \sum_{j \in J} \sum_{k \in K} VPS_c x_{ijk} + nd_c^{\omega} \geq NWV_c^{\omega} \quad c \in C,\ \omega \in \Omega \tag{41}$$

$$\eta,\ nd_c^{\omega} \in \mathbb{R}^+. \qquad c \in C, \omega \in \Omega \tag{42}$$

The R-CSCPP seeks a clinic schedule and capacity plan such that the cost is minimized under worst-case demand scenarios within the uncertainty set Ω. Constraints (40) formulate the total cost of each scenario, in which η is a decision variable that captures the maximum cost caused by the worst case scenario, and nd_c^{ω} is an integer variable equal to the number of rejected patients of clinic $c \in C$ in scenario $\omega \in \Omega$. Constraints (41) are the demand coverage constraint for each scenario $\omega \in \Omega$. Observe that although Ω is a finite set, its size grows exponentially in C. The R-CSCPP thus involves a huge number of variables nd_c^{ω}, and constraints (40) and (41) cannot be solved explicitly using a general purpose solver. In the next section, we describe an exact column-and-row generation algorithm that is used to efficiently solve this problem.

3.3.3 An Implementor/Adversary Algorithm for the R-CSCPP

In order to solve the proposed MIP formulation for the R-CSCPP, we use an iterative column-and-row generation procedure commonly referred to as the *implementor/adversary* algorithm [67]. This algorithm has been successfully applied to solve various robust optimization problems [see for instance, 68, 69]. It decomposes the original problem into two simpler ones: the *implementor* and the *adversary* subproblem. The main idea of the algorithm is to initially consider only a small subset of scenarios $\tilde{\Omega} \subset \Omega$ to define a *restricted* implementor problem (I-CSCPP), and to iteratively generate new scenarios via the adversary problem until an optimal solution to the original R-CSCPP is obtained. We note that I-CSCPP is a relaxation of the R-CSCPP given that it only considers the subset of constraints (40) and (41) associated with $\tilde{\Omega}$.

Let $\tilde{x} = (\tilde{h_i}, \tilde{r}, \tilde{x_{ijk}}, \tilde{y_{cjk}})$ be the optimal solution to the restricted implementor problem, and $\eta_{\tilde{x}}$ be the corresponding objective function value, which provides a valid lower bound (L) on the optimal solution value of R-CSCPP. For any $\omega \in \Omega$, let $f(\tilde{x}, \omega)$ be the associated cost of $\tilde{x}$, and $\tilde{nd}_c^\omega$ be the number of rejected patients in scenario ω. If $f(\tilde{x}, \omega) \leq \eta_{\tilde{x}}$, and $\tilde{nd}_c^\omega$ ensures the feasibility of constraints (41) for all $\omega \in \Omega$, then $\tilde{x}$ is the optimal solution to the original R-CSCPP with optimal solution value $\eta_{\tilde{x}}$. Otherwise, there exists at least one scenario $\omega \in \Omega \setminus \tilde{\Omega}$ that leads to a violated inequality of the type (40) and/or (41).

In order to prove optimality, or to determine which scenario leads to a violated inequality, we solve the adversary problem (i.e., separation problem) to find a scenario ω^* that maximizes the cost $f(\tilde{x}, \omega)$. In particular, for a given solution $\tilde{x}$ of the restricted implementor problem, the adversary problem can be stated as the following MIP:

$$\text{maximize} \sum_{c \in C} f_4 nd_c + \sum_{i \in I} f_{1i} \tilde{h_i} + f_2 \tilde{r} + \sum_{c \in C} \sum_{j \in J} \sum_{k \in K} f_3 \tilde{y_{cjk}} \tag{43}$$

$$\text{subject to} \quad \underline{NWV_c} \leq nwv_c \leq \overline{NWV_c} \qquad c \in C \tag{44}$$

$$\sum_{c \in C} nwv_c \leq \Gamma \tag{45}$$

$$nd_c \leq M z_c + nwv_c - \sum_{i \in I_c} \sum_{j \in J} \sum_{k \in K} VPS_c \tilde{x_{ijk}} \qquad c \in C \tag{46}$$

$$nd_c \leq M (1 - z_c) \qquad c \in C \tag{47}$$

$$z_c \in \{0, 1\} \qquad c \in C \tag{48}$$

$$nd_c \in \mathbb{R}^+ \qquad c \in C \tag{49}$$

$$nwv_c \in \mathbb{Z}^+, \qquad c \in C \tag{50}$$

where nwv_c are integer variables representing the demand for clinic $c \in C$, and M is a large constant. Constraints (44) and (45) ensure that for each $c \in C$, the clinic demand value nwv_c is within the uncertainty set Ω. For each $c \in C$, z_c are binary decision variables equal to 1 if the demand of clinic c is more than the capacity allocated according to schedule $\tilde{x}$ (i.e., $nwv_c - \sum_{i \in I_c} \sum_{j \in J} \sum_{k \in K} VPS_c \tilde{x_{ijk}} < 0$), and 0 otherwise. Constraints (46) and (47) ensure that for each $c \in C$, nd_c represents the number of rejected patients associated with current assignments of physicians $\tilde{x_{ijk}}$.

The optimal solution value of the adversary problem, denoted by $f(\tilde{x}, \omega^*)$, provides a valid upper bound (U) on the optimal solution value of R-CSCPP. If there is a new scenario ω^* that leads to a violated inequality, it is added to $\tilde{\Omega}$, and the updated restricted implementor problem is solved again. This iterative procedure is performed until the gap between the best upper and lower bounds reaches a predefined threshold value ϵ. The implementor/adversary algorithm is summarized in Algorithm 1.

Algorithm 3 Implementor/adversary algorithm for R-CSCPP

Initialization: $\tilde{\Omega} = \emptyset, L = -\infty, U = +\infty$
while $U - L > \epsilon$ **do**
  Solve the restricted implementor problem with $\tilde{\Omega}$ to obtain $\tilde{x}$ and L.
  $L \leftarrow \eta_{\tilde{x}}$
  Solve the adversary problem (43) - (50) to obtain ω^*
  if $U > f(\tilde{x}, \omega^*)$ **then**
    $U \leftarrow f(\tilde{x}, \omega^*)$
  $\tilde{\Omega} \leftarrow \tilde{\Omega} \cup \omega*$

3.4 The Physician Scheduling Problem

In this section, we address the tactical and operational levels of our physician planning framework summarized in Figure 3.1. At these levels, the physicians are assigned to different shifts over a one-week planning horizon to satisfy clinics' demand while taking into account the uncertainty in patient treatment times, physician preferences, as well as fairness and ergonomic criteria. We note that the polyclinic capacity plan derived from the first level of the planning framework acts as the input to the second-level physician scheduling problem. In what follows, we first provide a formal definition and an MIP formulation for the deterministic variant of the PSP. We then show how we incorporate the uncertainty in patient treatment times into the PSP and formulate it as a two-stage integer stochastic program. Finally, we present an SAA scheme to generate feasible solutions and obtain a statistical estimation of their optimality gap.

3.4.1 Problem Definition and Formulation

Consider the sets and parameters previously defined for the CSCPP. In addition, for each $j \in J$ and $k \in K$, let D_{cjk} denote the number of patients visiting clinic $c \in C$, and R_{jk} the number of available examination rooms. For every physician $i \in I$, H_i denotes the maximum number of on-duty shifts. Using the optimal solution $(\tilde{h_i}, \tilde{r}, \tilde{x_{ijk}}, \tilde{y_{cjk}})$ of the

R-CSCPP, the above parameters can be defined as follows:

$$D_{cjk} = \sum_{i \in I_c} VPS_{cjk} \tilde{x_{ijk}} \quad c \in C, j \in J, k \in K \tag{51}$$

$$R_{jk} = \tilde{r} - \sum_{c \in C} R_c \tilde{y_{cjk}} \quad j \in J, k \in K \tag{52}$$

$$H_i = \tilde{h_i} \quad i \in I. \tag{53}$$

Let NJ_i be the set of days that physician $i \in I$ is not available to work. For each $i \in I$, let TK_i denote the percentage of total working shifts that physician i must be assigned to the shifts that s/he prefers to work in. To reflect physicians preferences for working in a specific shift, for each $i \in I$, $j \in J$ and $k \in K$, we define PR_{ijk} equal to 1 if physician i prefers to work on day j, shift k, and 0 otherwise. The PSP consists of assigning physicians to a set of shifts on a weekly basis, such that the demand per shift, examination room capacity, and physicians preferences are satisfied while minimizing the total physicians' cost. Physician assignment variables x_{ijk} are defined as in the CSCPP. The PSP can be formulated as the following MIP:

$$
\begin{aligned}
\text{minimize} & \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} f_{1i}x_{ijk} && && (54)\\
\text{subject to} & \sum_{i\in I_c} VPS_{cjk}x_{ijk} = D_{cjk} && c\in C,\ j\in J,\ k\in K && (55)\\
& \sum_{i\in I} x_{ijk} \le R_{jk} && j\in J,\ k\in K && (56)\\
& \sum_{j\in J}\sum_{k\in K}(x_{ijk} - x_{i'jk}) \le PU_c && c\in C,\ i,i'\in I_c && (57)\\
& \sum_{k\in K} x_{ijk} = 0 && i\in I,\ j\in NJ_i && (58)\\
& \sum_{j\in J}\sum_{k\in K} x_{ijk} \le H_i && i\in I && (59)\\
& \sum_{j\in J}\sum_{k\in K}(TK_i x_{ijk} - PR_{ijk}x_{ijk}) \le 0 && i\in I && (60)\\
& \sum_{k\in K} x_{ijk} \le 1 && i\in I,\ j\in J && (61)\\
& x_{ijk}\in\{0,1\} && i\in I,\ j\in J,\ k\in K. && (62)
\end{aligned}
$$

The objective function represents the total cost of physicians. Constraints (55) ensure that an adequate number of physicians is assigned to each shift to provide service to the scheduled patients. Constraints (56) restrict the number of physicians in each shift to the number of available examination rooms. Constraints (57) restrict the difference in the number of assigned shifts between any two physicians of each clinic. Constraints (58) prevent assigning physicians to the shifts that they are not available to work. Constraints (59) limit the number of shifts that a physician can be on-duty by considering her/his maximum workload. Constraints (60) ensure that a certain percentage of physicians' workload are assigned according to their preferences. Similarly, constraints (61) prohibit physicians from working more than one shift per day. Lastly, constraints (62) are the standard integrality constraints.

3.4.2 Dealing with Patient Treatment Time Uncertainty

The PSP described in the previous section assumes that the number of patients that a physician can assess per shift (VPS_{cjk}) is known and deterministic. However, in practice the treatment time of each patient changes depending on their condition and acuity, thus the values VPS_{cjk} are uncertain. In this section, we assume that treatment times are stochastic and can be represented as a finite set of scenarios with known probability. This is a realistic assumption due to the fact that the patients that visit the polyclinics under discussion can actually be classified into a rather small number of classes of patients, each associated with a different condition. As a consequence, the probability distribution of treatment times and the percentage of patients of each class can be realistically estimated. In this context, we assume that the physicians schedule must be determined at the beginning of the planning horizon (week) without full knowledge of the list of patients that must be served during each shift. In other words, the treatment times are not revealed to the decision maker at this stage. The patients' treatment times, on the contrary, will be revealed only few hours before the beginning of a shift. At this point, the decision maker would have a much more accurate estimate on the treatment times by referring to the initial clinical assessment of each patient, which is done prior to the patients' appointments with the doctors. Given that physicians have already been assigned to shifts, depending on the patient mix assigned to a particular shift, it may be possible that not all scheduled patients can be assessed during the regular duration of the shift. In order to ensure service is provided to all schedule patients, three types of corrective (recourse) actions with associated costs are considered: *i)* assign an extra number of physicians from on-call physicians, *ii)* extend shift duration (i.e. over-time), and *iii)* request additional examination rooms. It should be noted that the third recourse action is a consequence of the first one in the sense that additional physicians require more resources. On the other hand, the initial schedule might lead to an idle time for physicians due to low treatment times required for patients scheduled during a shift. The under-utilization of physicians would then incur in non-essential service costs for hospital's management.

Given the above assumptions, the stochastic PSP can be formulated as a multi-stage stochastic program with recourse where each stage is associated with a shift of the planning horizon. The first-stage decisions (those made at the beginning of the week) revolve around assigning physicians to either regular or on-call duties over all shifts during the planning horizon. On the contrary, calling on-call physicians to work during a shift, assigning physicians to extended (over-time) shifts, and requesting extra examination rooms are the decisions that are made as corrective actions at the beginning of each shift (stage) for different scenarios that depend on patients' treatment times. Similar to the majority of personnel scheduling problems (see e.g., [60]), we note that the recourse decisions made at the beginning of each stage (shift) do not affect the decisions in subsequent stages. That is, the non-anticipativity condition (NAC) should only be satisfied for the first-stage decisions. Therefore, the multi-stage stochastic program can be transformed into a two-stage stochastic program with recourse, denoted by SPSP.

Consider the sets and parameters defined for the deterministic PSP. We denote Ξ as the set of random scenarios, where $\Xi = \{\xi_1, \ldots, \xi_{|\Xi|}\}$, and P^{ξ} denotes the probability of scenario $\xi \in \Xi$. Let VPS^{ξ}_{cjk} represent the number of patients that can be visited in clinic $c \in C$ during shift $j \in J$ by physician $k \in K$ under scenario $\xi \in \Xi$. Similar to assigning on-duty physicians, for each $i \in I$, $j \in J$, $k \in K$, we define $x^{'}_{ijk}$ equal to 1, if and only if physician i is assigned on day j, shift k, as on-call. Let f_5 be the cost of assigning a physician as on-call per shift. As mentioned, if the number of on-duty physicians is not enough for a certain scenario during a shift, on-call physicians are called to work, and additional examination rooms are prepared to accommodate them. For each $i \in I$, $j \in J$, $k \in K$, $\xi \in \Xi$, we define s^{ξ}_{ijk} equal to 1, if and only if on-call physician i is called to work on day j, shift k, under scenario ξ. Also, for each $j \in J$, $k \in K$, $\xi \in \Xi$, the decision variable er^{ξ}_{jk} represents the number of additional examination rooms prepared for shift k on day j, under scenario ξ. f_6 and f_7 denote the costs for calling an on-call physician to work and preparing an additional examination room, respectively.

For each $c \in C$, $j \in J$, $k \in K$, $\xi \in \Xi$, the recourse decision ot^{ξ}_{cjk} represents the number of patients visiting clinic c in shift k on day j who are processed during overtime shift, under scenario ξ, and f_8 is the corresponding cost. To measure the under-utilization of on-duty physicians under some scenarios, we define the recourse decision oa^{ξ}_{cjk} for each $c \in C$, $j \in J$, $k \in K$, $\xi \in \Xi$ to represent the proportion of the total available physicians' time that is idle in clinic c, shift k, day j, and scenario ξ. f_9 is the cost of idle physicians. The objective of the SPSP is to minimize the cost of on-duty and on-call physicians, plus the expected cost of recourse decisions over all treatment time scenarios. The SPSP can be formulated as follows:

$$\begin{aligned} \text{minimize} \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} (f_{1i} x_{ijk} + f_5 x'_{ijk}) + \sum_{\xi \in \Xi} P^{\xi} \Bigg[\sum_{i \in I} \sum_{j \in J} \sum_{k \in K} f_6 s^{\xi}_{ijk} \\ + \sum_{j \in J} \sum_{k \in K} f_7 er^{\xi}_{jk} + \sum_{c \in C} \sum_{j \in J} \sum_{k \in K} (f_8 ot^{\xi}_{cjk} + f_9 oa^{\xi}_{cjk}) \Bigg] \end{aligned} \tag{63}$$

subject to (59) – (62)

$$\sum_{i \in I_c} VPS^{\xi}_{cjk}(x_{ijk} + s^{\xi}_{ijk}) + ot^{\xi}_{cjk} - VPS^{\xi}_{cjk} oa^{\xi}_{cjk} = D_{cjk} \qquad c \in C, j \in J, k \in K, \xi \in \Xi \tag{64}$$

$$\sum_{i \in I} (x_{ijk} + s^{\xi}_{ijk}) - er^{\xi}_{jk} \leq R_{jk} \qquad j \in J, k \in K, \xi \in \Xi \tag{65}$$

$$\sum_{j \in J} \sum_{k \in K} (x_{ijk} + x'_{ijk} - x_{i'jk} - x'_{i'jk}) \leq PU_c \quad c \in C, i, i' \in I_c \tag{66}$$

$$\sum_{k \in K} (x_{ijk} + x^{'}_{ijk}) \leq 0 \qquad i \in I, j \in NJ_i \tag{67}$$

$$x_{ijk} + x^{'}_{ijk} \leq 1 \qquad i \in I, j \in J, k \in K \tag{68}$$

$$\sum_{k \in K} (x_{ijk} + s^{\xi}_{ijk}) \leq 1 \qquad i \in I, j \in J, \xi \in \Xi \tag{69}$$

$$s^{\xi}_{ijk} \leq x^{'}_{ijk} \qquad i \in I, j \in J, k \in K, \xi \in \Xi \tag{70}$$

$$x^{'}_{ijk},\ s^{\xi}_{ijk} \in \{0, 1\} \qquad i \in I, j \in J, k \in K, \xi \in \Xi \tag{71}$$

$$ot^{\xi}_{cjk},\ er^{\xi}_{jk} \in \mathbb{Z}^{+} \qquad c \in C, j \in J, k \in K, \xi \in \Xi \tag{72}$$

$$oa^{\xi}_{cjk} \in \mathbb{R}^{+}. \qquad c \in C, j \in J, k \in K, \xi \in \Xi \tag{73}$$

The first two terms of the objective are the total cost of assigning physicians as on-duty and on-call, respectively. The last four terms are the expected cost of calling on-call physicians to work, preparing additional examination rooms, serving patients in overtime shifts, and physicians' idle time over all scenarios. Constraints (64) specify the number of patients served by on-duty or on-call physicians, the number of patients assessed during overtime, and the amount of physicians' idle time during each shift for all scenarios. Constraints (65) are the examination rooms capacity constraints. Constraints (66) ensure fair assignments of on-duty and on-call shifts among the physicians of each clinic. Constraints (67) forbid the on-call or on-duty assignments to the shifts that physicians are not available to work in. Constraints (68) prevent simultaneous assignments of physicians as on-call and on-duty at any given shift. Constraints (69) guarantee that each physician works no more than one shift per day. Constraints (70) assure that physicians can be called to work in a shift only if they have been already assigned as on-call to that shift. Finally, constraints (71) $-$ (73) are the standard integrality and non-negativity constraints.

3.4.3 Sample Average Approximation

We now present an algorithm for solving the SPSP that incorporates a Monte-Carlo sampling technique, known as the SAA scheme [70–73], with a general purpose MIP solver. Broadly speaking, the SAA scheme generates a random sample and approximates the expected value function by the corresponding sample average function. The associated deterministic sample average optimization problem is then solved to obtain a solution of the SPSP, and the procedure is repeated. The SAA scheme not only generates high quality solutions when solving the sample average problems, but is also able to produce a statistical estimation of their optimality gap.

The main difficulty in solving the SPSP is the number of variables and constraints needed to explicitly consider its very large scenario set Ξ. In the SAA scheme, a random sample $N = \{\xi_1, \ldots, \xi_{|N|}\}$ of scenarios from the original set Ξ is generated, and the second stage expectation

$$\sum_{\xi \in \Xi} P^{\xi} \left[\sum_{i \in I} \sum_{j \in J} \sum_{k \in K} f_6 s^{\xi}_{ijk} + \sum_{j \in J} \sum_{k \in K} f_7 er^{\xi}_{jk} + \sum_{c \in C} \sum_{j \in J} \sum_{k \in K} (f_8 ot^{\xi}_{cjk} + f_9 oa^{\xi}_{cjk}) \right],$$

is approximated by the sample average function

$$\frac{1}{|N|}\sum_{n\in N}\left[\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} f_6 s^n_{ijk} + \sum_{j\in J}\sum_{k\in K} f_7 er^n_{jk} + \sum_{c\in C}\sum_{j\in J}\sum_{k\in K}(f_8 ot^n_{cjk} + f_9 oa^n_{cjk})\right],$$

where the index associated with the specific scenario $\xi \in \Xi$ is replaced by the index associated with the sample scenario $n \in N$. Therefore, after replacing ξ by n in constraints (64)-(73), the original two-stage stochastic program SPSP is approximated by the SAA problem

$$\begin{aligned}
&\text{minimize} \sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\left(f_{1i}x_{ijk} + f_5 x^{'}_{ijk}\right)\\
&+\frac{1}{|N|}\sum_{n\in N}\left[\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} f_6 s^n_{ijk} + \sum_{j\in J}\sum_{k\in K} f_7 er^n_{jk} + \sum_{c\in C}\sum_{j\in J}\sum_{k\in K}(f_8 ot^n_{cjk} + f_9 oa^n_{cjk})\right]\\
&\text{subject to} (59) - (62), (64) - (73).
\end{aligned}$$

Let $\tilde{x}_N$ denote the corresponding optimal solution of the above SAA problem and z_N its objective value. It can be shown that under mild regularity conditions, $\tilde{x}_N$ and z_N converge with probability one to their true counterparts, as the sample size $|N|$ increases. Moreover, $\tilde{x}_N$ converges to an optimal solution of the original problem with probability approaching one exponentially fast [72]. It is also possible to estimate the sample size $|N|$ needed to generate and $\epsilon-$ optimal solution to the original problem with a probability at least equal to $1 - \alpha$. However, it is known that the sample size estimate is too conservative for practical applications. Therefore, instead of solving one large-scale SAA problem, the SAA algorithm involves in the generation of a set of M independent samples of solutions for the corresponding smaller SAA problems. Using the optimal solution values $z^1_N, \ldots, z^{|M|}_N$ from these $|M|$ SAA problems, we compute statistical lower and upper bounds for the optimal solution value of the original problem. We now describe our implementation of the overall

SAA procedure. Later, in Section 3.5.2 we will show how to select a sample size capable of producing tight and accurate statistical bounds.

1. Generate a set $M = \left\{N_1, \ldots, N_{|M|}\right\}$ of independent samples, each of size $|N|$, i.e., $\xi_j^i, \ldots, \xi_j^{|N|}$ for $j \in M$. For each sample N_j solve the corresponding SAA problem using a general purpose MIP solver. Let z_{N_j} and $\tilde{x}_{N_j}$, $j \in M$, be the corresponding optimal objective value and an optimal solution, respectively.

2. Compute the average of all optimal solution values from the SAA problems and their variance:
$$\mu_M^N = \frac{1}{|M|} \sum_{j \in M} z_{N_j},$$
$$\sigma^2_{\mu_M^N} = \frac{1}{(|M|-1)|M|} \sum_{j \in M} \left(z_{N_j} - \mu_M^N\right)^2.$$
It is known that the average μ_M^N provides a statistical lower bound for the optimal value of the original problem SPSP [71], i.e. $E\left[\mu_M^N\right] \leq z^*$, and $\sigma^2_{\mu_M^N}$ is an estimate of the variance of this estimator.

3. Choose a feasible solution $\tilde{x}$ of the original problem, for instance, one of the previously obtained solutions $\tilde{x}_{N_j}$. Using this solution, it is possible to estimate the optimal solution value z^* of the original problem SPSP as follows:

$$\begin{aligned}
z_{N'}(\tilde{x}) = \min & \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} (f_{1i} \tilde{x_{ijk}} + f_5 \tilde{x_{ijk}}') \\
& + \frac{1}{|N'|} \sum_{n \in N'} \left[\sum_{j \in J} \sum_{k \in K} (\sum_{i \in I} f_6 s_{ijk}^n + f_7 er_{jk}^n) \right. \\
& \left. \qquad + \sum_{c \in C} \sum_{j \in J} \sum_{k \in K} (f_8 ot_{cjk}^n + f_9 oa_{cjk}^n) \right] \\
& \text{subject to} (59) - (62), (64) - (73),
\end{aligned}$$

where $\xi^i, \ldots, \xi^{N'}$ is a sample of size N' generated independently of the samples employed in the SAA problems, and ξ is replaced by n in constraints (59)-(62), (64)-(73). Given that the first-stage variables are fixed, one can take $|N'|$ much larger than the sample size $|N|$ used in the SAA problems. Note that $z_{N'}(\tilde{x})$ is an estimate on the upper bound on the optimal solution value z^* of the original problem, i.e. $E[z_{N'}(\tilde{x})] \geq z^*$. The variance of this estimate can be computed as

$$\sigma^2_{N'}(\tilde{x}) = \frac{1}{(|N'|-1)|N'|} \sum_{n \in N'} \Bigg[\sum_{i\in I} \sum_{j\in J} \sum_{k\in K} (f_1 \tilde{x_{ijk}} + f_5 \tilde{x_{ijk}}') + \sum_{j\in J} \sum_{k\in K} (\sum_{i\in I} f_6 s^n_{ijk} + f_7 er^n_{jk}) + \sum_{c\in C} \sum_{j\in J} \sum_{k\in K} (f_8 ot^n_{cjk} + f_9 oa^n_{cjk}) - \tilde{Z}_{N'}(\tilde{x}) \Bigg]^2 .$$

According to Verweij et al. [73], one should take $\tilde{x}^*$ as the best solution among $\tilde{x}^1_N, \ldots, \tilde{x}^M_N$ candidate solutions, that is:

$$\tilde{x}^* \in \arg\min \left\{ z_{N'}(\tilde{x}) : \tilde{x} \in \{\tilde{x}^1_N, \ldots, \tilde{x}^M_N\} \right\}.$$

4. Compute an estimate of the % optimality gap of solution $\hat{z}$ by using the lower and upper bound estimates on the optimal solution value of the original problem SPSP obtained in Steps 2 and 3:

$$\text{GAP}(\tilde{x}^*) = \frac{\left(z_{N'}(\tilde{x}^*) - \mu^N_M\right)}{z_{N'}(\tilde{x}^*)} \times 100.$$

3.5 Computational Results

This section presents the results of computational experiments obtained using data from a real case study in a polyclinic in Montreal, Canada. The computational experiments consist of three parts. We first evaluate the convergence and solution time of the Implementor/Adversary (I/A) algorithm developed to solve the R-CSCPP for the real-case problem. We then analyze the performance of the proposed SAA algorithm for solving the SPSP. Finally, using a Monte Carlo simulation, we compare the cost of schedules obtained from our framework with the cost of the *expected value problem* (EVP) in which uncertain parameters are replaced with their expected values. All experiments were run on an HP server with an Intel(R) Xeon(R) CPU E5-2687W v3 processor running at 3.10 GHz and 512 GB of RAM under a Linux environment. All algorithms were coded in C++, and the SAA problems were solved using *Concert Technology* of CPLEX 12.7.0.

The studied polyclinic operates five days a week and two shifts each day. Table 3.1 shows the information of the polyclinic, in which the first two columns are the clinics' identification number and their discipline. The third column contains the number of physicians in each clinic. The fourth and fifth columns provide lower and upper bounds on the number of visiting patients to each clinic, respectively. Finally, the last three columns contain the minimum, average, and maximum number of patients that can be assessed by one physician in one shift, respectively. This information was obtained from historical data.

Table 3.1: Polyclinic information

ID	Clinic	# Physicians	Demand		# Patients/Physician/Shift		
			$\underline{NWV_c}$	$\overline{NWV_c}$	$\underline{VPS_c}$	VPS_c	$\overline{VPS_c}$
1	Breast	19	90	270	2	4	13
2	Urology	15	115	345	3	7	21
3	Hematology	21	95	285	2	4	12
4	Gynecology	7	66	197	7	10	20
5	Hepatology	10	110	330	5	7	12
6	Lung	19	85	255	2	4	10
7	Musculoskeletal	6	35	105	2	5	18
8	Melanoma	9	200	600	12	18	24
9	Upper GI	10	10	30	1	2	4
10	Pain	7	8	23	1	2	3
11	Cancer Rehab.	6	13	38	1	2	3
12	Colorectal	12	75	225	2	4	8
13	Brain	6	15	45	1	2	12

Table 3.2 contains the information regarding interdisciplinary clinics. There are five clusters of interdisciplinary clinics that must be simultaneously scheduled for a given number of shifts. The first two columns show the clusters' identification number and the interdisciplinary clinics in each cluster. The third column contains the number of shifts that each cluster must be scheduled simultaneously during a week.

Table 3.2: Interdisciplinary clinics

Clusters	Interdisciplinary groups	Frequency
1	(6) & (10)	4
2	(6) & (11)	4
3	(12) & (5)	4
4	(2) & (10)	3
5	(6) & (8)	1

Without loss of generality, we assume all costs to be rational numbers and to be set according to their relative priority levels. Based on the results of a survey conducted at the polyclinic, service-level-related costs have the highest priority for the administration, followed by the cost of physicians that are higher than the cost of the other resources. Tables 3.3 and 3.4 provide the considered cost values in our experiments for the R-CSCPP and the SPSP, respectively. As can be seen, we assume the costs of rejecting and rescheduling a patient (f_4 and f_8) to be significantly higher than the costs of resources and physicians, as the polyclinic administration would like to process as many patients as possible during regular shifts. Moreover, we assume equal costs for all physicians.

Table 3.3: Cost values of R-CSCPP

Parameter	Value
f_1	5
f_2	1
f_3	1
f_4	50

Table 3.4: Cost values of SPSP

Parameter	Value
f_1	5
f_5	2.5
f_6	10
f_7	7.5
f_8	50
f_9	15

3.5.1 Performance of the I/A on the R-CSCPP

We now present the results of the experiments for assessing the performance of the I/A algorithm. In particular, R-CSCPP is solved using different values of Γ. We note that Γ can take values in the interval $\left[\sum_{c\in C} \underline{NWV_c}, \sum_{c\in C} \overline{NWV_c}\right] = [917, 2748]$. However, since a physician serves multiple patients in each shift (i.e., VPS_c), and the demand constraint is not an equality constraint, the total number of served patients is not necessarily equal to Γ. In our experiments, $\Gamma = 1667$ represents the case in which all clinics' demand are simultaneously at their upper bounds.

The results for $\Gamma = 1067$ are given in Figure 3.2. In this case, our I/A algorithm converges after 173 seconds and 20 iterations. For each iteration, the lower bound returned by the implementor (lower dotted line) and the upper bound returned by the adversary (upper solid line) are plotted and their associated values given in the left $y-$axis. The dash line with circle marks shows the optimality gap (i.e., percent deviation of implementor from the best known solution of the adversary problem) on the right $y-$axis. As can be seen, the algorithm requires few iterations to converge (we observed a similar behavior under different Γ values).

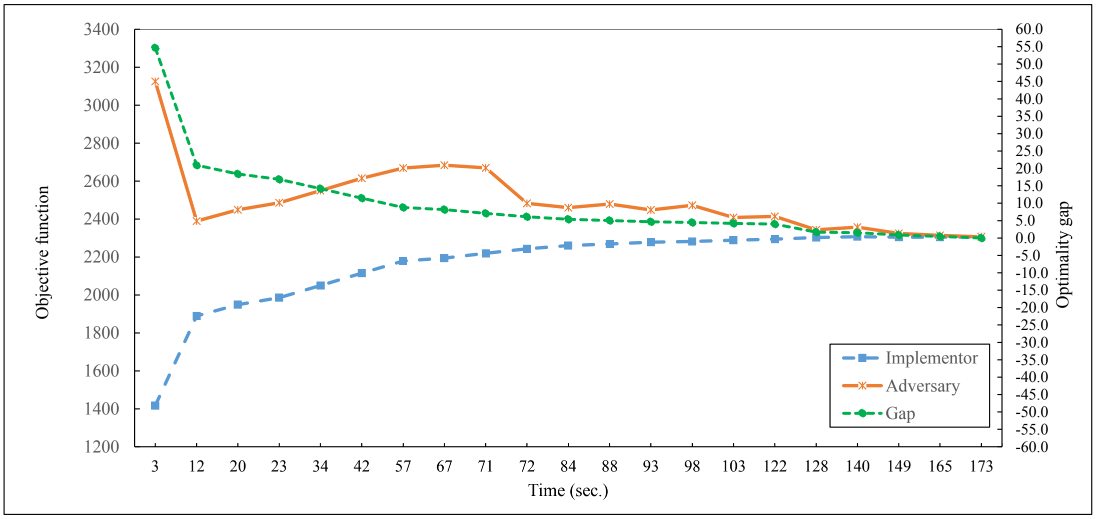

Figure 3.2: Convergence profile of I/A algorithm for $\Gamma = 1067$

Table 3.5 shows the percentage of maximum demand (worst-case scenario) for the clinics that can be satisfied under the optimal solution obtained by the R-CSCPP when considering different values of budget of uncertainty (Γ). The first column gives the value of Γ associated with each instance. The second column provides the running time of the algorithm (in seconds) to terminate. The rest of the columns give the percentage of the worst-case demand served in each clinic, starting from the breast clinic (c_1) to the brain clinic (c_{13}). As can be seen in Table 3.5, the algorithm terminates with an optimal solution in less than 200 seconds in all considered instances. In these experiments, in each instance we included the variables and constraints generated during the solution of the instances with smaller Γ values in the implementor problem. As the row $\Gamma = 1667$ shows, 100 percent of the worst case demand in all clinics is already served. Therefore, any Γ value $1667 \leq \Gamma \leq 2780$ will provide the same solution as $\Gamma = 1667$.

Table 3.5: % of worst-case demand level served at the optimal solution of the R-CSCPP for different Γ values

Γ	Time	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}
917	3	34	34	33	35	34	34	33	35	33	33	35	33	35
967	21	34	34	33	35	34	34	38	35	33	33	35	33	35
1017	107	38	38	38	40	38	39	48	35	40	33	35	39	35
1067	173	66	58	64	75	62	69	90	47	60	50	55	72	48
1117	57	99	90	99	100	94	98	100	65	100	92	95	98	96
1167	31	99	98	99	100	100	100	100	74	100	100	95	98	96
1217	20	99	98	99	100	100	100	100	82	100	100	95	98	96
1267	32	99	98	99	100	100	100	100	91	100	100	95	98	100
1317	35	99	98	99	100	100	100	100	100	100	100	95	98	100
1367	25	100	100	99	100	100	100	100	100	100	100	95	98	100
1417	7	100	100	99	100	100	100	100	100	100	100	95	98	100
1467	6	100	100	99	100	100	100	100	100	100	100	95	98	100
1517	25	100	100	99	100	100	100	100	100	100	100	95	100	100
1567	22	100	100	100	100	100	100	100	100	100	100	95	100	100
1667	5	100	100	100	100	100	100	100	100	100	100	100	100	100

3.5.2 Performance of the SAA Scheme for the SPSP

Table 3.6 summarizes the computational results of the SAA algorithm for the SPSP. The realization of the uncertain parameters corresponding to the number of patients assessed by a physician during a shift for each scenario ξ are generated according to a truncated Poisson distribution with a mean value equal to VPS_c, for each $c \in C$. We recall that our bi-level framework starts by solving the R-CSCPP with a given budget of uncertainty Γ. Then, it proceeds to solving the SPSP. For each combination of Γ and a number of sample scenarios N, the rows of Table 3.6 provide the following information: $\bar{Z}_N$, $\sigma_{\bar{Z}_N}$, $\tilde{Z}_{N'}(\tilde{x})$, $\sigma_{\tilde{Z}_{N'}(\tilde{x})}$, % GAP, and CPU time. For these experiments, we consider $M = 40$ and $N' = 2,000$.

As can be seen from Table 3.6, high quality solutions can be obtained in all considered instances by using a relatively small sample size. In particular over all the instances, a sample size of $N = 30$ can provide solutions within 2% optimality gap. It is also noteworthy that standard deviations of statistical lower and upper bounds are small enough, which suggests that the values of M and N' are sufficiently large. As expected, the CPU time increases as the sample size increases, taking almost 24 hours for a sample size of $N = 60$. Observe that the total CPU time for solving all the instances with a sample size of $N = 30$ is no more than three and half hours.

Table 3.6: Results of the SAA algorithm with $M = 40$ and $N' = 2,000$

Γ	Ind.	N							
		1	5	10	20	30	40	50	60
917	$\bar{Z}_N$	3818.62	4447.86	4551.53	4573.13	4614.51	4635.68	4642.20	4647.52
	$\sigma_{\bar{Z}_N}$	138.93	63.35	35.82	17.60	21.28	15.68	13.78	12.44
	$\tilde{Z}_{N'}(\tilde{x})$	11677.80	6053.71	4837.10	4681.55	4677.83	4672.75	4645.12	4642.61
	$\sigma_{\tilde{Z}_{N'}(\tilde{x})}$	50.70	32.98	14.72	13.89	14.09	13.88	13.97	13.57
	GAP	67.30	26.53	5.90	2.32	1.35	0.79	0.06	-0.11
	CPU	274.37	475.33	4162.30	9812.14	11055.88	25280.52	59168.13	84156.51
1067	$\bar{Z}_N$	7521.90	8414.72	8848.81	8793.63	8865.92	8891.89	8890.51	8900.37
	$\sigma_{\bar{Z}_N}$	377.74	113.12	68.65	44.72	35.51	36.58	28.18	21.48
	$\tilde{Z}_{N'}(\tilde{x})$	20857.40	10355.80	9146.95	8948.85	8947.39	8934.56	8903.35	8915.84
	$\sigma_{\tilde{Z}_{N'}(\tilde{x})}$	75.31	47.73	29.85	29.14	29.27	29.05	28.98	28.50
	GAP	63.94	18.74	3.26	1.73	0.91	0.48	0.14	0.17
	CPU	280.00	1058.00	3122.27	4583.59	9672.30	13190.48	28403.95	63560.44
1167	$\bar{Z}_N$	10610.00	11014.70	11637.70	11718.30	11786.20	11801.78	11804.80	11788.80
	$\sigma_{\bar{Z}_N}$	432.94	241.72	103.67	69.92	64.60	51.89	45.56	36.00
	$\tilde{Z}_{N'}(\tilde{x})$	22360.90	13735.20	12100.60	11913.60	11882.50	11836.80	11806.00	11793.80
	$\sigma_{\tilde{Z}_{N'}(\tilde{x})}$	90.22	70.45	44.85	42.49	43.50	43.81	43.36	45.67
	GAP	52.55	19.81	3.83	1.64	0.81	0.30	0.01	0.04
	CPU	269.46	593.03	4371.88	6143.82	9264.97	15477.57	21402.85	40532.60
1467	$\bar{Z}_N$	10472.00	11392.70	12041.50	12088.74	12068.50	12085.90	12158.70	12174.89
	$\sigma_{\bar{Z}_N}$	542.15	202.76	89.53	71.13	53.61	53.88	43.53	41.07
	$\tilde{Z}_{N'}(\tilde{x})$	24354.10	13266.30	12348.50	12234.30	12240.30	12239.90	12197.80	12180.40
	$\sigma_{\tilde{Z}_{N'}(\tilde{x})}$	88.39	61.85	46.17	46.42	45.16	45.20	46.00	45.36
	GAP	57.00	14.12	2.49	1.19	1.40	1.26	0.32	0.05
	CPU	269.65	582.08	5344.20	7482.70	8612.77	11942.59	15724.38	20116.55
1667	$\bar{Z}_N$	10528.70	11466.40	12442.00	12567.60	12579.21	12591.60	12533.60	12610.00
	$\sigma_{\bar{Z}_N}$	400.60	157.42	131.86	62.31	62.10	48.19	43.89	46.25
	$\tilde{Z}_{N'}(\tilde{x})$	23860.80	14012.70	12914.60	12681.90	12683.40	12676.50	12632.00	12635.50
	$\sigma_{\tilde{Z}_{N'}(\tilde{x})}$	90.27	68.92	48.62	47.88	47.50	47.69	48.36	47.51
	GAP	55.87	18.17	3.66	0.90	0.82	0.67	0.78	0.20
	CPU	272.27	541.83	4494.05	6239.83	10121.72	17450.84	24467.91	24806.62

Figure 3.3 visualizes the results of the table and plots the percent deviation gap between the statistical lower bounds and upper bounds and CPU times over different sample sizes. The left $y-$axis provides the optimality gap for the solid lines, and the right $y-$axis shows the CPU time for the series plotted by the dash lines. As the figure shows, the solutions obtained with $N = 5$ can be at most 27% away from its lower bound. As we increase the sample size, both the lower bounds and upper bounds improve (i.e., former increase and latter decrease). Nevertheless, the improvement in the gap for sample sizes greater than $N = 30$ is only marginal. We can also observe a significant increase in the CPU time for sample sizes greater than $N = 30$. For these reasons, during the rest of the computational experiments we use a sample size of $N = 30$.

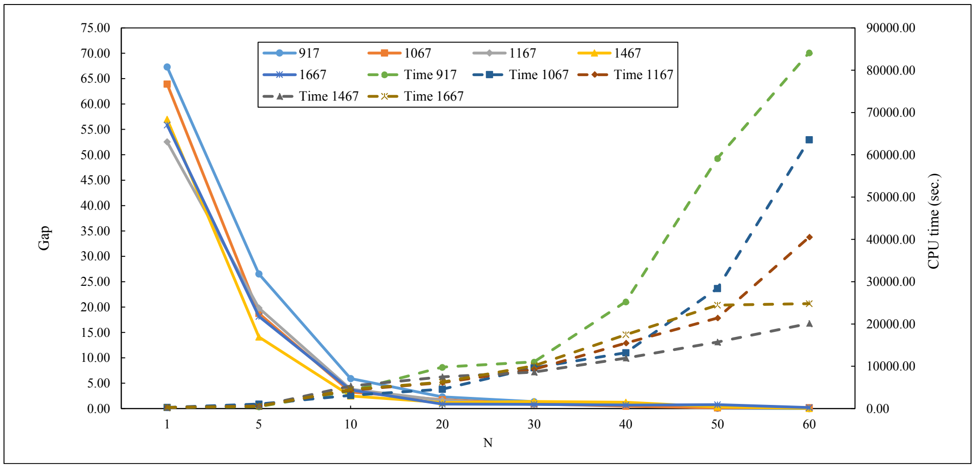

Figure 3.3: Percent deviation gap and CPU time in convergence of the SAA method

3.5.3 The Value of Stochastic Schedule

We now demonstrate the superiority of the stochastic solution over the deterministic schedule by comparing the results of the single level deterministic EVP with our bi-level framework through a Monte Carlo simulation. In the absence of uncertainty, the CSCPP and the PSP can be integrated into a single-level problem. However, in our approach, we first solve the R-CSCPP with a certain budget of uncertainty Γ. After that, we plug the obtained clinics and resource plans into the SPSP and develop physicians' work schedules. Throughout this section, we refer to the bi-level framework as *sequential approach*. Figure 3.4 summarizes the simulation procedure.

As can be seen, in the sequential approach, the R-CSCPP is first solved (with a specific Γ value), then its solution (D_{cjk}, R_{jk}, H_i) is given as an input to the SPSP. The SPSP is then solved with a sample size of $N = 30$, and the physicians' work schedules $(\bar{x_{ijk}}, \bar{x_{ijk}}')$ are generated. In order to make the decisions regarding utilizing the available on-call physicians, assigning additional examination rooms in each shift, and to calculate the total cost of processed patients during over-time shifts as well as physicians' idle time, we formulate a mathematical program (SIM) that receives physicians works schedules and the number

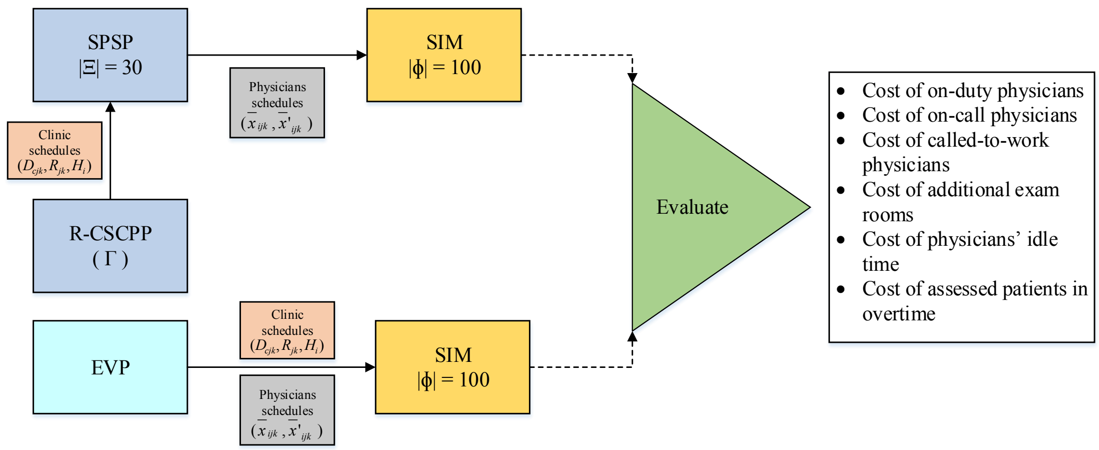

Figure 3.4: Simulation procedure

of available rooms in different shifts as inputs. In this problem, the number of weekly arriving patients as well as patients' treatment times can vary simultaneously in each replication of the simulation. Hence, the generated physicians' work schedules are plugged into the SIM, and then it is optimized for a set of $\Phi = \{\phi_1, \ldots, \phi_{100}\}$ replications. In the deterministic approach, physicians' work schedules are developed in one step and plugged into the SIM model.

For a given replication ϕ, the SIM problem is formulated as follows:

$$
\begin{aligned}
&\text{minimize} \sum_{i\in I}\sum_{j\in J}\sum_{k\in K}(f_1\bar{x_{ijk}} + f_5\bar{x_{ijk}}' + f_6 s_{ijk}) \\
&+\sum_{j\in J}\sum_{k\in K}\left\{f_7 er_{jk} + \sum_{c\in C}(f_8 ot_{cjk} + f_9 oa_{cjk})\right\} && && (74)\\
&\text{subject to} \sum_{i\in I_c} VPS^{\phi}_{cjk}(\bar{x_{ijk}} + s_{ijk}) + \sum_{j\in J}\sum_{k\in K} nd_{cjk} \\
&\qquad -\sum_{j\in J}\sum_{k\in K} VPS^{\phi}_{cjk} oa_{cjk} = NWV^{\phi}_c && c\in C && (75)\\
&\qquad \sum_{i\in I}(\bar{x_{ijk}} + s_{ijk}) - er_{jk} \le R_{jk} && j\in J,\ k\in K && (76)\\
&\qquad \sum_{k\in K}(\bar{x_{ijk}} + s_{ijk}) \le 1 && i\in I, j\in J && (77)\\
&\qquad s_{ijk} \le \bar{x_{ijk}}' && i\in I, j\in J, k\in K && (78)\\
&\qquad s_{ijk} \in \{0,1\} && i\in I, j\in J, k\in K && (79)\\
&\qquad ot_{cjk}\ er_{jk} \in \mathbb{Z}^+ && c\in C, j\in J, k\in K && (80)\\
&\qquad oa_{cjk} \in \mathbb{R}^+ && c\in C, j\in J, k\in K && (81)
\end{aligned}
$$

As can be seen in Figure 3.4, physicians' work schedules ($\bar{x_{ijk}}, \bar{x_{ijk}}'$) as well as the number of available examination rooms (R_{jk}) are the inputs of the problem. In order to incorporate the randomness in the number of weekly arriving patients, the demand constraint (75) is formulated as weekly demand for each clinic. Values of the random parameters (VPS^{ϕ}_{cjk}, NWV^{ϕ}_c) are generated from the distributions provided in Table 3.7.

Table 3.7: Distributions of the random parameters

Parameter	Distribution	Mean	Min.	Max.
NWV^{ϕ}_c	Uniform	-	$\underline{NWV_c}$	$\overline{NWV_c}$
VPS^{ϕ}_{cjk}	Truncated Poisson	VPS_c	$\underline{VPS_c}$	$\overline{VPS_c}$

Figure 3.5 provides the simulation results. The $y-$axis corresponds to the operational

cost calculated by the simulation model, and the $x-$axis gives the replication number. As can be observed, in 99 out of 100 replications the cost of the deterministic approach is higher than the cost of the sequential method with $\Gamma \geq 1067$. Throughout the replications, we observe an intense fluctuation in the cost of the deterministic approach with a standard deviation of 4858 versus 452 for the sequential approach with $\Gamma = 1067$. Moreover, the worst case cost of the deterministic method can be as high as 30000 (replication 14), whereas the worst case cost of the sequential approach is no more than 15000 (replication 59) for any budget of uncertainty. The significant superiority of the sequential method over the deterministic approach, regardless of the considered uncertainty budget, implies the effectiveness of the recourse actions against variability in patients' treatment times in the SPSP. We also note that when Γ increases from its lower bound (i.e., 917), the cost decreases remarkably. The decline in the cost by increasing the uncertainty budget confirms the robustness of the R-CSCPP.

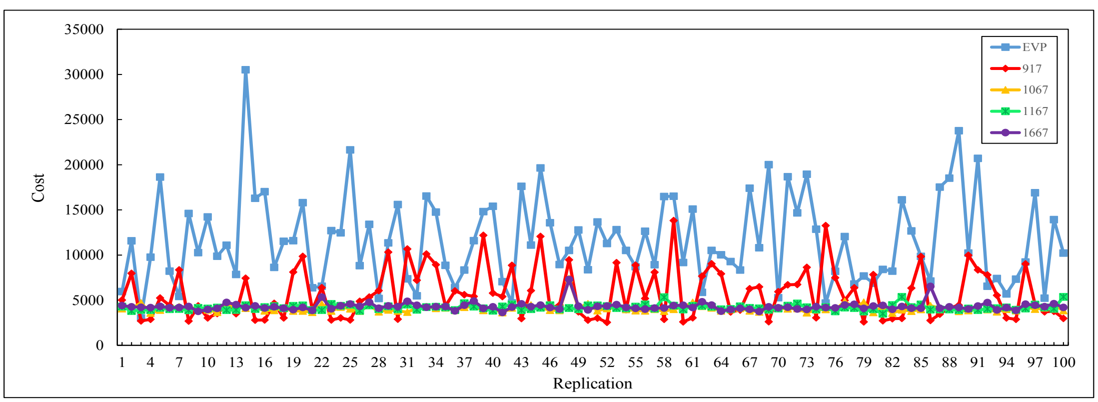

Figure 3.5: Simulation results

Figure 3.6 gives the average operational cost of the sequential approach with different uncertainty budgets along with the average cost of the EVP. The cost of the sequential approach is plotted with respect to the left $y-$axis whereas the cost of the deterministic approach is given with respect to the right $y-$axis. The dash lines correspond to the percent deviation of the sequential approach from the deterministic one. As can be observed, the average cost of the sequential approach is significantly lower than the average cost of the deterministic one over all simulation replications. Additionally, the decrease of the operational cost when increasing the uncertainty budget is more evident. Moreover, an interesting observation is the trend of the average cost throughout the uncertainty budget values. It can be noted that the decrease in the average cost stops at a certain level of Γ (1067), and it rises from $\Gamma = 1067$ to $\Gamma = 1667$. The turnabout in the average cost is the result of over-protecting the schedules against demand uncertainty. Recall that the R-CSCPP provides as an output the weekly number of patients that must be served by each clinic, which in turn, becomes the input to the SPSP. Under a high budget of uncertainty, more clinic sessions and working hours could be scheduled for physicians to maximize the number of patients that could be scheduled during regular shifts. On the contrary, the actual number of patients who require appointments and hence scheduled during each week (generated in the simulator) might be significantly smaller than the number obtained by the R-CSCPP under a high uncertainty budget. This, in turn, would lead to an increased number of idle hours for physicians initially scheduled by the SPSP under the aforementioned uncertainty budget. In other words, assigning extra resources to shifts in order to hedge against the demand uncertainty may lead to an increase in physicians' idle time. It may also increase the number of on-duty physicians, which reduces the number of on-call physicians and results in less flexibility toward patients' treatment times variability.

3.6 Conclusion

This paper introduced a framework for planning physicians in polyclinics under uncertainty. The procedure addresses the problem at the strategic, tactical, and operational planning levels. In the strategic level, we proposed an adjustable robust approach that plans

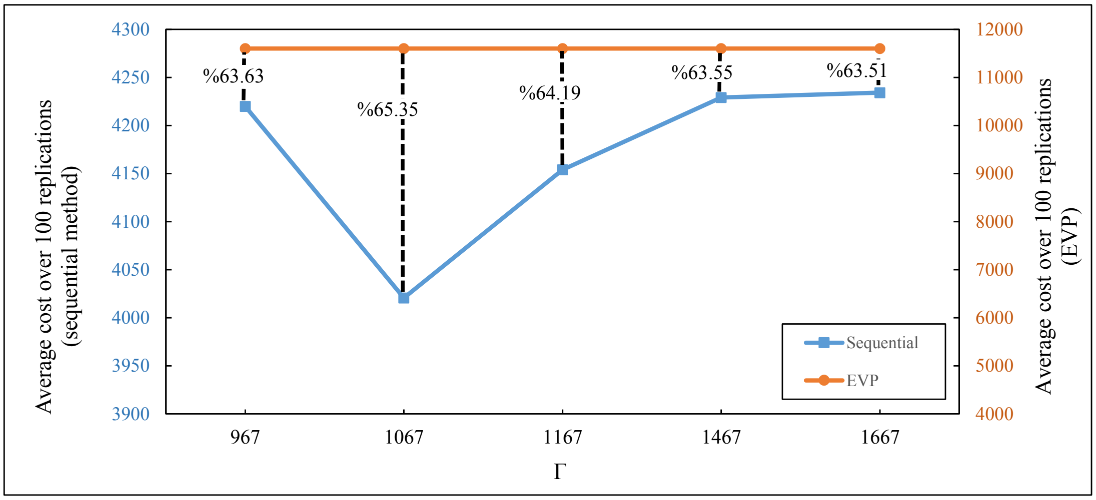

Figure 3.6: Average cost for different Γ values

clinics work schedules and assigns required capacity to each shift. In particular, the model provides the decision maker with the option of protecting the plans against the uncertainty in the number of arriving patients to the polyclinic. The robust problem was solved with an implementor/adversary algorithm, which can prove optimality in reasonable CPU times. We also combined tactical physicians scheduling with operational rescheduling decisions into a two-stage stochastic program that incorporates the uncertainty in patients' treatment times. In the first stage, physicians are assigned as on-call or on-duty, and in the second-stage, the on-call physicians are called to work if needed. Since the variety of patients' treatment times in each clinic results in a fairly large number of scenarios, we used a sample average approximation scheme to obtain high quality solutions by considering only a sub-set of scenarios. The results from the computational experiments with the data provided by a polyclinic in a university health center in Montreal confirmed the efficiency of the proposed framework. Furthermore, we investigated the impact of including uncertainty in our solution framework by using a Monte Carlo simulation. We compared our framework to its deterministic counterpart and demonstrated that the additional capacity included in the plan as a consequence of the use of a robust model, and the corrective (recourse) decisions implemented by the stochastic model resulted in schedules that have significantly lower costs than those generated by the deterministic one.

Acknowledgments

This research was partly funded by the Canadian Natural Sciences and Engineering Research Council under grants 418609-2012 and 402043-2011. This support is gratefully acknowledged.

Chapter 4

Simulation optimization for physicians scheduling in polyclinics

Abstract

Reducing wait-times and resource expenses are essential factors to enhance service quality and financial viability of health delivery systems. This paper presents a simulation-optimization (SO) modeling approach for physicians scheduling in an outpatient polyclinic under uncertain arrival pattern and treatment time of patients. We provide a methodology that combines discrete-event simulation with an optimization search routine to minimize patient wait-time and physician overtime subject to several scheduling/resource restrictions. Our main goal is to investigate the impact of physicians work schedules on patient wait-time in such systems. Our experimental results on a real case study in Canada clearly indicate the significant impact of adopting the proposed SO framework for physician scheduling on reducing the aforementioned key performance measures. Further, the optimization search routine developed in our SO model outperforms the existing search routines embedded into commercial SO software packages in terms of solution quality and CPU time.

4.1 Introduction

Outpatient polyclinics reduce the burden in hospitals and help bridge the gap between primary and secondary care. They provide a better structure for physicians of different disciplines to work together and enable patients with chronic and complex conditions to visit multiple clinics at the same place during the same visit. Patient wait-time is one of the essential elements in outpatients care service quality [74]. Studies have indicated that every aspect of patient experience such as confidence in the care provider and perceived quality of care are correlated negatively with longer wait-time [75]. According to Erhard et al. [1], planning efficient schedules for physicians in outpatient clinics is essential in reducing expenses and patients wait-time.

Physician scheduling problems (PSPs) consist of creating work schedules for physicians in a predetermined planning horizon such that in every given work shift enough physicians are available to process all scheduled patients [3]. Physicians' work schedule must abide to several regulations and constraints (e.g., physicians preferences and contract terms). Such constraints are more binding in the context of polyclinics where the resources such as treatment rooms and medical staff are shared among different clinics that differ in terms of patient flow and throughput. These problems are usually formulated as mixed-integer programs incorporating several families of conflicting constraints (see e.g., Tohidi et al. [45]). Further, the physicians' schedule is prone to uncertain events such as patients' unpunctuality/no-shows and variable treatment times. Such disruptions might lead to increased patient wait-time and/or extended shifts (overtime) for physicians.

A comprehensive literature review on PSPs is provided in Erhard et al. [1]. According to this article and references therein, the majority of existing approaches investigate this problem in a deterministic context[see for instance, 12, 25, 45]. On the contrary, PSPs under uncertain patient demand and treatment time have only been investigated in few articles. Robust optimization and stochastic programming are the main approaches exploited to incorporate uncertainty into decision models proposed for scheduling physicians [see, 64, 76]. Nevertheless, the objective function of the aforementioned models is to minimize the expected cost that the hospital/polyclinic incurs as a consequence of using extended shifts (over-time) and extra number of (on-call) physicians or nurses in order to visit all scheduled patients. To the best of our knowledge, none of the existing models explicitly considers patient wait-time as part of the objective function. This might be due to the fact that the analytic estimation of this performance measure is nontrivial in complex queuing networks such as the case in outpatient polyclinics.

Given that in recent years hospitals have been forced to reduce the expenses while improving service quality, the PSP should be addressed as a multi-objective optimization problem, where the decision-makers should aim for the simultaneous minimization of resource costs and patient wait-time. To fill this important void in the literature, we propose a simulation-optimization (SO) modeling approach for physician scheduling that aims to minimize the aforementioned objectives under uncertain arrival pattern and treatment times in an outpatient polyclinic. To the best of our knowledge, this is the first time that the impact of physician schedules on patient wait-time is explicitly incorporated into a physician scheduling framework.

SO is the amalgamation of a simulation model with an optimization algorithm to obtain the optimal configuration to a system where an analytic form of the objective function and/or constraints is not known and can only be evaluated through simulation [77]. The idea is to avoid simulating exhaustively all possible system configurations in order to find the optimal one. While SO approaches have been extensively exploited for resource planning (e.g., staffing) in health care delivery systems [see e.g., 78–81], their implementation in a highly-constrained optimization problem similar to the PSP under investigation is less trivial. In the context of resource planning problems, the decision maker is looking for the optimal number of nurses, physicians, and treatment rooms under a budget limitation. Hence, generating a high quality initial solution in the SO framework is quite straightforward due to the relatively small number of decision variables and constraints. On the contrary, the PSP incorporates a large number of binary decision variables (assignment of physicians to work shifts) and several classes of conflicting constraints (e.g., preferences, fairness and ergonomic constraints, work contracts, and resources limitations and interactions). Hence, generating an initial feasible solution in the absence of an analytic objective function requires solving a feasibility problem for which it is not possible to verify the quality of the obtained solution apriori. The latter could significantly impact the performance of the optimization search routine (e.g., a heuristic or a meta-heuristic) embedded within the SO framework. As a consequence, the existing SO softwares that combine simulated annealing (SA) or other meta-heuristics within a discrete-event simulation (DES) commercial software (e.g., OptQuest in Arena) are less efficient in the context of this study. By the same token, meta model-based SO approaches [see, for instance, 82] are also less efficient in the presence of a large number of decision variables. In particular, constructing the approximation of the simulation model involves investigating the relationship between a very large sample of inputs (i.e., feasible schedules) and their simulation output.

To alleviate the shortcomings of the existing SO software packages, our second contribution relies on the integration of a mathematical optimizer, based on an enhanced SA algorithm, with a simulation kernel into a common programming language (i.e., C++). In particular, our SO framework aims to improve the quality of the incumbent solution in the SA algorithm via implementing two enhancement strategies. The first strategy is to employ a non-linear mixed-integer-program to approximate the PSP in order to generate a high quality initial feasible solution. This is achieved via modeling the polyclinic under investigation as a D/M/1 queue and approximating the expected patient wait-time with an analytical (nonlinear) function. The second strategy relies on introducing a diversification mechanism within the SA algorithm in order to avoid getting trapped in local optimal solutions. We also conduct a set of numerical experiments in the context of an outpatient polyclinic in a university hospital in Montreal (Canada) with the goal of investigating the impact of the proposed schedule on improving the two key performance measures as well as the aforementioned enhanced optimization search routine on the quality of converged solution.

The remainder of the paper is organized as follows. Section 4.2 is a review of relevant literature to the PSP and SO. Section 4.3 presents the formal description of the polyclinic under investigation. Section 4.4 describes the proposed SO framework. Section 4.5 contains the results of our computational experiments, and it is followed by our conclusions in Section 4.6.

4.2 Literature Review

DES techniques have been extensively applied to a broad range of health care applications including patients appointment scheduling and admissions policies as well as resource planning (e.g., personnel staffing) in outpatient facilities. Numerous studies use DES to assess the impact of staff schedules on different key performance measures in health delivery systems. Dittus et al. [83] focus on residents work schedules and predict the effects of alternative work schedules on the sleep and activity profile of the residents using a DES model. Additionally, Evans et al. [84] examine the significance of nurses and physicians work schedules on patients' length-of-stay (LOS) in an ER. Similarly, Rossetti et al. [85] evaluate the impact of ER physicians' schedule on patient throughput and resource utilization. Swisher et al. [86] investigate the impact of various compositions of medical staff, registration windows, and clinic's available space on the profit of the clinic as well as patients and staff satisfaction. Spry and Lawley [87] analyze the impact of staffing levels and alternative work schedules on the average time of medication delivery to the patients in a hospital pharmacy. Al-Najjar and Ali [88] estimate the required number of nurses and doctors in two ERs by the aid of a DES model. In Abo-Hamad and Arisha [89], seven scenarios based on adding resources and a patients admission policy are investigated through a DES model with the objective to minimize patients' LOS and maximize resources and layout efficiency in an ER. Using a DES model, Oh et al. [90] prove that an LOS of under 3 hours can be achieved via reconfiguration of staff levels and resource allocations in an ER.

Several papers attempt to optimize the performance measures of health care systems by combining simulation with optimization models. Centeno et al. [78] determine the number of required staff under demand and service time uncertainty using a simulation model. They exploit the simulation results as the input to a mathematical model that creates staff's shift-based schedules. De Angelis et al. [79] use simulation to estimate the stochastic objective function of a mathematical programming model that determines the configuration of required servers in a medical center. In Oddoye et al. [81], multiple patient-related performance measures (e.g., LOS and patient wait-time) are improved through the optimization of resource levels by the aid of DES in a medical center. Zeinali et al. [82] propose a simulation-based meta model to optimize the number of required resources (e.g., doctors, nurses, beds) for improving patient wait-time in an ER.

SO frameworks that incorporate heuristic search algorithms into the simulation models have been used for optimizing resource configurations in health care systems. Swisher and Jacobson [91] use an object-oriented simulation model for determining optimal staffing and resource configuration in a family practice clinic. They combine several objectives related to clinic's profitability and patient satisfaction into a single objective and choose the best configuration by using a rank and selection (R&S) method. Additionally, Yeh and Lin [92] address nurse staffing in an ER and use a genetic algorithm embedded in a commercial solver to obtain a near-optimal solution. In Ahmed and Alkhamis [80], SO is used to determine the appropriate number of staff in an ER with the goal of improving patients throughput. Cabrera et al. [93] and Cabrera et al. [94] use Agent-Based Simulation and exhaustive search to determine ER staff configuration in a university hospital. More recently, Ozcan et al. [95] propose a SO modeling approach to obtain the optimal number of required beds in addition to operation room block times and duration in a hospital.

To the best of our knowledge, none of the existing SO models proposed in health care applications address a PSP involving a large number of possible schedules that should comply with several families of constraints. On the contrary, they are mainly focused on staffing and resource configuration decisions. Furthermore, the majority of these models are implemented by using commercial simulation packages. Nonetheless, the limitations of the aforementioned packages in efficiently solving highly-constrained optimization models similar to the ones associated with PSPs, motivated us to code our DES model and the heuristic search procedure in a common programming language.

4.3 System Description

The Polyclinic under investigation is an ambulatory outpatient cancer center that provides several hospital services such as X-rays, biopsy, minor surgeries and other outpatient treatments. The center operates over 2 four-hour shifts (morning and afternoon) every day from Monday to Friday. It consists of 145 physicians and 13 cancer clinics, namely breast, urology, hematology, gynecology, hepatobiliary, lung, musculoskeletal, melanoma, upper gastrointestinal, pain, cancer rehabilitation, colorectal and brain metastases. One waiting room and a limited number of examination rooms are shared among all clinics. Moreover, some clinics such as lung and cancer rehabilitation are interdisciplinary and must be scheduled simultaneously for a certain number of shifts.

All patients coming to the center are referred by general practitioners or specialists and have an appointment. They follow a process as depicted in Figure 4.1. Although the process flow is not identical in all clinics, the figure presents the general flow which is similar among all. The process begins with arrival of the patient to the polyclinic and ends when the patient is released. The arriving patient goes to the receptionist who registers patient's information and collects his/her referral note. Afterwards, the patient waits in the waiting room to visit the doctor at his/her scheduled appointment time. The patient is assessed by the doctor in an examination room, where s/he will decide if the patient needs further tests such as X-ray, biopsy, or consultation with other physicians and a second assessment. Consequently, according to the required process, patients can be divided into two categories. Category 1 patients are those who need further processing after physicians assessment. These patients go to the waiting room and wait for the next process they require. We categorize the next processes for Category 1 patients as: Test 1, Test 2, and second assessment. Test 1 are those diagnostic tests that require long processing times such as blood tests, biopsy, and etc. Test 2 are imaging tests such as X-ray which usually require less processing times than Test 1. Second assessments are for patients with interdisciplinary diagnosis; after the first assessment, the physician usually consults with other physicians in the same interdisciplinary team regarding patient's condition, and the patient is informed with an accurate diagnosis in the second assessment session. Category 2 refers to the patients who do not need further diagnostics, or directly are referred for a treatment in future sessions.

Currently, the extended duration of diagnostics and interdisciplinary consultations along with the unbalanced number of patients scheduled on different sessions cause long patient wait-times and inevitable overtime for physicians. Hence, the polyclinic is looking for the optimal scheduling of physicians in addition to the number of scheduled patients in each session such that a significant improvement in terms of both key performance measures is achieved.

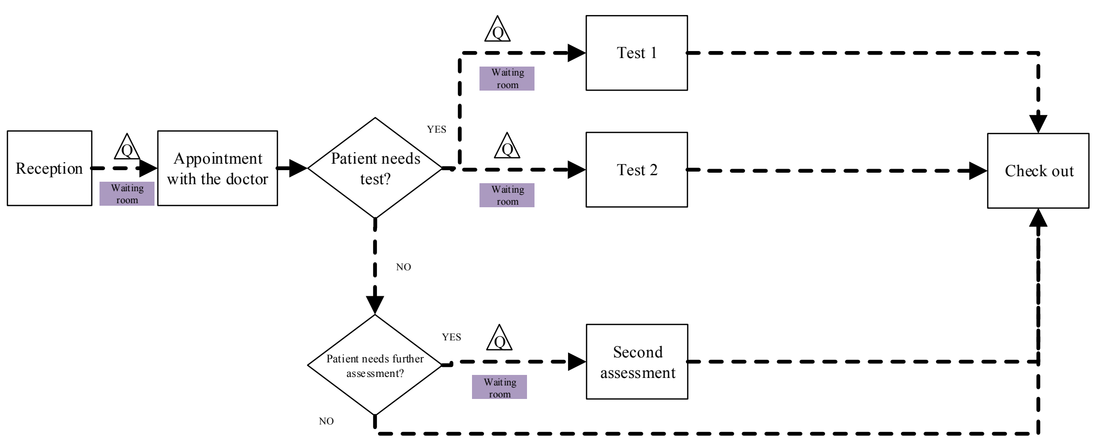

Figure 4.1: Polyclinic process flow chart

4.4 Simulation Optimization Modeling Approach

As it was mentioned earlier, by the aid of SO, the decision maker combines simulation with an optimization search procedure to find the optimal system settings of a stochastic system. Figure 4.2 depicts our proposed SO framework that entails two main components, i.e., the DES model *(SIM)* and the optimization search routine *(SAC)*. The latter is composed of a SA meta-heuristic algorithm, a restricted physicians scheduling mathematical program ($P^{''}$), and a physician scheduling approximation model ($P^{'}$). This procedure starts by generating an initial feasible solution (schedule) by the aid of $P^{'}$. Starting from this initial solution, the SA algorithm discovers its neighborhood and generates a new candidate solution. A diversification strategy is also incorporated within the SA algorithm by the aid of restricted model $P^{''}$. It is worth mentioning that the objective function of each candidate solution generated in the *SAC* search routine is estimated by the aid of the DES simulation model.

In what follows, each component of the SO framework is explained in details. We first start by describing the DES model corresponding to the polyclinic under investigation along with the experiments conducted to validate this model. Afterwards, we describe the optimization search routine of the framework, where we start by providing the mathematical formulation of our PSP. This is followed by the description of the approximation of the PSP model that is able to generate a high quality initial solution. Finally, the neighborhood structure of the search in addition to the diversification strategy adopted within the SA algorithm are presented.

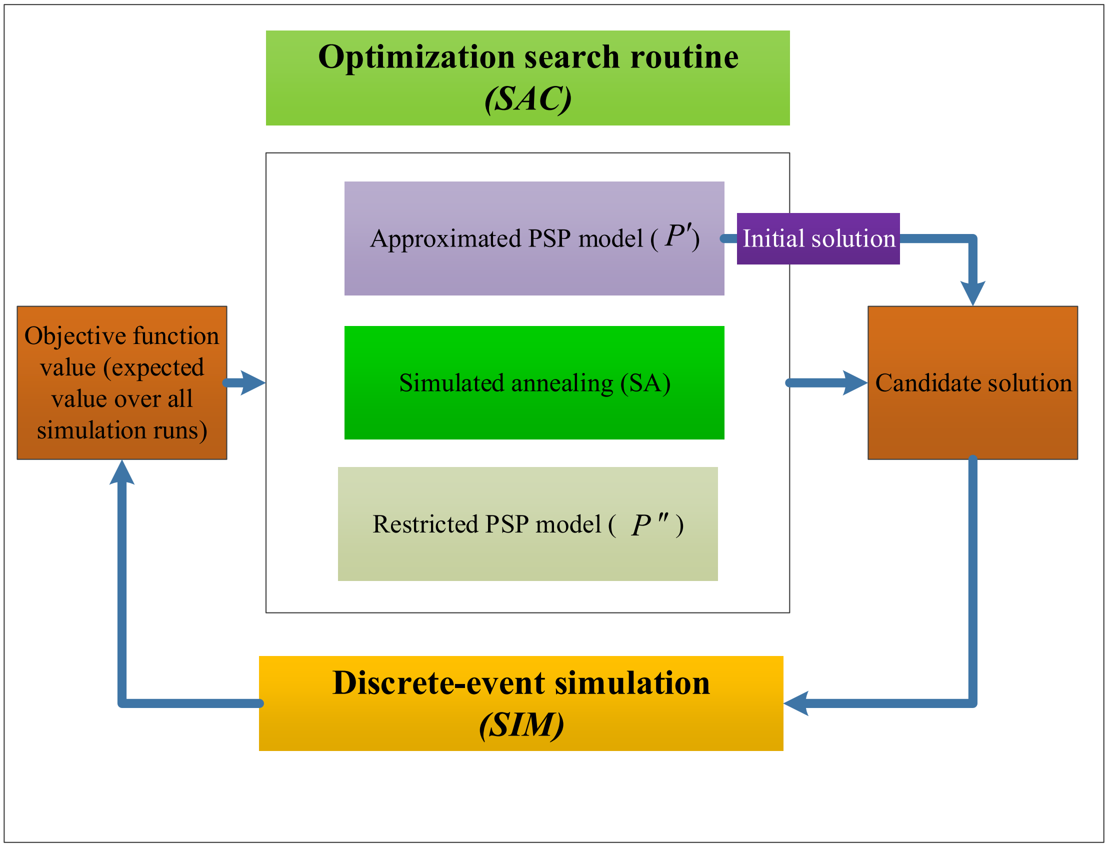

Figure 4.2: Simulation optimization framework

4.4.1 Discrete Event Simulation Model

A survey was conducted in the polyclinic to study the patients arrival pattern, physicians' assessment times, and duration of diagnostic tests. Afterwards, collected data were fitted into different probability distributions in order to find the best fit. Consequently, patients' arrival times (in minutes) follow a Uniform distribution $U(t-10, t+10)$, where t denotes the initial patient's appointment time. This distribution models patients' unpunctuality. Further, we observed that appointments are scheduled according to the rule in Bailey [96], where two appointments are booked at the beginning of the shift, and successive appointments are then booked at the intervals equal to the mean assessment time. Table 4.1 summarizes the probability distributions of the service times corresponding to different stages of the process flow depicted in Figure 4.1. These distributions were also validated by the physicians, nurses and administration. In this table, the first column represents different clinics starting from the breast clinic (c_1) to the brain clinic (c_{13}). Whereas, the other columns contain the parameters of probability distributions associated with different stages of the process flow in the polyclinic.

Table 4.1: Service time distribution at each stage of the process

Stage	Reception	Doctor	Test 1	Test 2	Second Assessment
Distribution	Uniform	Exponential	Uniform	Uniform	Uniform
Clinic 1	[5 10]	50	[30 50]	[5 10]	[5 10]
Clinic 2	[5 10]	34	[5 10]	-	[5 10]
Clinic 3	[5 10]	51	[5 10]	[5 10]	[5 10]
Clinic 4	[5 10]	22	[20 30]	[5 10]	[5 10]
Clinic 5	[5 10]	32	[20 30]	-	[5 10]
Clinic 6	[5 10]	52	[30 50]	-	[5 10]
Clinic 7	[5 10]	44	[30 50]	[5 10]	[5 10]
Clinic 8	[5 10]	13	[10 30]	-	[5 10]
Clinic 9	[5 10]	100	[30 50]	[5 10]	[5 10]
Clinic 10	[5 10]	120	-	-	-
Clinic 11	[5 10]	109	-	-	[5 10]
Clinic 12	[5 10]	51	[30 50]	[5 10]	[5 10]
Clinic 13	[5 10]	96	[60 90]	[10 15]	[5 10]

Our simulation model is structured based on a *resource view* approach, where the system is simulated as a queuing network constituting of a set of communicating resources. We instantiate the resource classes as the process stages (i.e., reception, doctor, imaging, biopsy, treatment) and specify their associated events as waiting to process a patient, processing a patient, and referring the patient to the next server. Moreover, *clients* are the patients who travel through the resource network and hold various attributes such as process route and priority rank.

4.4.1.1 Validating The Discrete Event Simulation Model

In our study, we conducted multiple audits, reviews, and object-flow testing by tracing patients of different clinics through reception, doctors, different clinical tests, and observed their throughput. We also performed a turning test by presenting several outputs (e.g., maximum number of patients in the waiting room (MW)) from the simulation model and historical data to the hospital's administration. The administration were not able to tell the difference between the results, hence this test provided a measure of validity for the simulation model. To further confirm the turning test, we conducted a t-test on the same Key Performance Measure (KPI), (MW), based on the results of the system simulation over 14 days. The results are provided in Table 4.2, where $\bar{X}, \mu, s, df, t, \alpha$ and $C.V.$ represent the mean value of MW, population mean (obtained from the historical data), sample standard deviation, degree of freedom, t-statistic, level of significance, and critical value. Similar to the previous test, the results confirm that the simulated mean value of MW is sufficiently close to its historical mean.

Table 4.2: One-sample t-test for maximum number of patients waiting in the waiting room

KPI	$\bar{X}$	μ	s	df	t	α	$C.V.$
MW	90.744	90	2.468948	19	1.3135	0.05	2.093

Finally, we carried out a sensitivity analysis by running the simulation model with various input parameters. Table 4.3 contains the results of the sensitivity analysis on the total number of patients assigned to each physician. In particular, for a given solution, the first column presents a certain percentage of assigned patients plugged into the simulation model. The rest of the columns present the value of different KPIs. The results show a positive correlation between the number of scheduled patients and overall polyclinic utilization. When there are few appointments, physicians' overtime and patient wait-time are minimal.

Table 4.3: Sensitivity analysis on inputs of the simulation model

Percentage of assigned patients	MW	Total patient waiting time (min)	Total physicians overtime (min)
20	34	1968.94	0.160642
50	47.5	28759	233.96
70	68.45	54962.3	1133.9
100	96.23	110229	7470.46

As for the number of replications, the average patient wait-time was analyzed. Figure 4.3 plots the average total patient wait-time over the number of replications, ranging between 1 and 150. The trend on the graph shows that 100 replications are adequate to reach stability and steady state condition. The results of the above validation tests confirm that the model provides an accurate estimation of patient wait-time and physicians overtime.

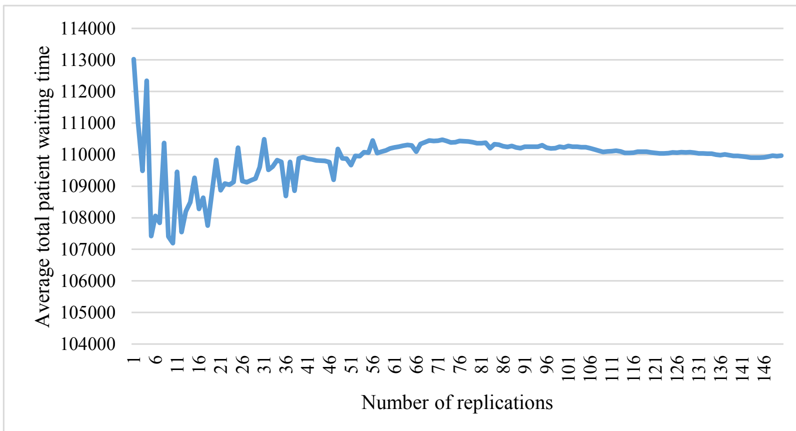

Figure 4.3: Average patient wait-time for alternative numbers of replications

4.4.2 The Physician Scheduling Problem

In our PSP, we are looking for the assignment of physicians affiliated with different clinics to the shifts as well as the number of patients to assign to physicians in each shift such that all the scheduled patients are visited. Let C, I, J, and K denote the sets of clinics, physicians, days and shifts per day, respectively. The set of subsets of interdisciplinary clinics is denoted by T. Each clinic in subset $t \in T$ must be scheduled simultaneously with other clinics in the subset for at least F_t shifts. R represents the total number of available examination rooms in the polyclinic, and we also assume that every physician needs one room when he/she is on-duty. Let NWV_c be the number of patients scheduled in clinic c over the planning horizon, and we assume that all patients must be assessed. As mentioned earlier, several classes of constraints must be taken into account when scheduling physicians in polyclinics. The following parameters are defined to formulate such limitations. Let H_i denote the number of shifts that physician i must work according to his/her contract. NJ_i is the set of days that physician i is not available. P_i represent the minimum number of patients that must be assigned to physician i in the planning period which is defined according to doctors' workload and target revenue. Additionally, we define M_i as the maximum number of patients that can be assigned to physician i in any given shift. The value of M_i is determined either according to physician's preference or based on mean assessment times and shift duration. In order to reflect physicians' preferences for working during a specific shift, we define PR_{ijk} equal to 1, if physician i prefers to work on day j, shift k, 0 otherwise. Moreover, TK_i is the percentage of total working shifts that physician i must be assigned to the shifts that he/she prefers to work in.

For each $c \in C$, $j \in J$, $k \in K$, we define clinic assignment variables y_{cjk} equal to 1 if and only if clinic c is assigned to shift k on day j. Similarly, for every $i \in I, j \in J, k \in K$, we define physician assignment variables x_{ijk} equal to 1 if and only if physician i is assigned to day j, shift k. Furthermore, for every $i \in I, j \in J, k \in K$, variables xp_{ijk} are the number of patients assigned to physician i on day j, shift k. Finally, for all $t \in T$, $j \in J$, $k \in K$, g_{tjk} is an auxiliary binary decision variable that equals to 1 if and only if all clinics in subset t are assigned to shift k on day j.

The optimization model considered in this study aims to minimize: (i) total patient wait-time, and (ii) total physicians' overtime, subject to clinics' and physicians' requirements and resource capacity. As it was mentioned earlier, the problem includes uncertainty in patient arrivals and treatment times. Therefore the aforementioned objectives are stochastic functions for which an analytic approximation is not straightforward. Hence, we denote patient wait-time, and physicians' overtime, by $f_1(xp)$ and $f_2(xp)$, respectively. The PSP can be formulated as follows:

(P) Minimize $f_1(xp), f_2(xp)$

subject to:
$$\sum_{i\in I_c}\sum_{j\in J}\sum_{k\in k} xp_{ijk} = NWV_c \qquad c\in C \tag{82}$$

$$\sum_{i\in I} x_{ijk} \le R \qquad j\in J, k\in K \tag{83}$$

$$\sum_{i\in I_c} x_{ijk} \ge y_{cjk} \qquad c\in C, j\in J, k\in K \tag{84}$$

$$\sum_{c\in C|c\in t} y_{cjk} \ge |t| g_{tjk} \qquad t\in T, j\in J, k\in K \tag{85}$$

$$\sum_{j\in J}\sum_{k\in K} g_{tjk} \ge F^t \qquad t\in T \tag{86}$$

$$\sum_{j\in J}\sum_{k\in K} x_{ijk} = H_i \qquad i\in I \tag{87}$$

$$\sum_{j\in J}\sum_{k\in K} xp_{ijk} \ge P_i \qquad i\in I \tag{88}$$

$$xp_{ijk} \le M_i x_{ijk} \qquad i\in I, j\in J, k\in K \tag{89}$$

$$xp_{ijk} \ge x_{ijk} \qquad i\in I, j\in J, k\in K \tag{90}$$

$$\sum_{k\in K} x_{ijk} \le 1 \qquad i\in I, j\in J \tag{91}$$

$$\sum_{j\in J}\sum_{k\in K} (PR_{ijk}x_{ijk} - TK_i x_{ijk}) \ge 0 \quad i\in I \tag{92}$$

$$\sum_{k\in K} x_{ijk} = 0 \qquad i\in I, j\in NJ_i \tag{93}$$

$$x_{ijk},\ y_{cjk},\ g_{tjk} \in \{0,1\} \qquad c\in C, i\in I, j\in J, k\in K, t\in T$$

$$xp_{ijk} \in \mathbb{Z}^+ \qquad i\in I, j\in J, k\in K \tag{94}$$

Constraints (82) ensure that all patients visiting different clinics are assessed. Constraints (83) limit the number of on-duty physicians in each shift to the number of available rooms. Constraints (84) link the assignment of physicians with corresponding clinics. Constraints (85) - (86) guarantee that every clinic in each subset of interdisciplinary clinics is assigned with other clinics in the subset for at least the predefined number of shifts. In particular, constraints (85) force g_{tjk} to one when all clinics in cluster t are scheduled simultaneously on day j, shift k, and constraints (86) ensure that the simultaneous assignments happen at least F_t times during the planning horizon. Constraints (87) ensure that physicians are assigned to their pre-specified number of shifts. Constraints (88) make sure that every physician receives at least the minimum number of patients that they require. Constraints (89) restrict the number of assigned patients to a physician in any given shift to the maximum allowable number. Constraints (90) stipulate the assignment of physicians to the number of patients that they assess. Constraints (91) prevent physicians from working more than one shift per day. Constraints (92) guarantee that a certain percentage of physicians' working shifts are assigned according to their preferences. Constraints (93) prevent the assignment of physicians to the shifts that they are not available. Lastly, constraints (94) are the standard integrality and non-negativity constraints.

4.4.3 Optimization Search Routine

Given that model *P* has a stochastic objective function, it can not be solved by common commercial solvers such as CPLEX. Hence, we propose an enhanced SA algorithm *(SAC)*, to efficiently solve this problem. It is noteworthy that for some classes of optimization problems, such as PSPs, the quality of incumbent solutions of metaheuristics heavily rely on the initial solution. In the absence of analytic objective functions, one common practice is to solve a feasibility problem in order to obtain an initial feasible solution. Alternatively, we also approximate the stochastic objective function by a non-linear function and formulate an approximate model, denoted as *P′*, to generate a high quality initial solution. Our second enhancement strategy revolves around incorporating a diversification procedure within the SA algorithm. We next provide the details of different elements of the aforementioned optimization search routine.

4.4.3.1 Generating An Initial Feasible Solution

In order to generate a high quality initial solution to *SAC* algorithm, we solve model *P* with an objective function that approximates the wait-time. Intuitively, we assume patient wait-time and physicians overtime are two correctional KPIs in the system; hence we focus on approximating wait-time, as the only objective function by the aid of queuing theory. More precisely, when ignoring patients' unpunctuality, patients' visits to the system can be modeled as a queue D/M/1 given that patients arrive at the beginning of intervals $[n\beta, (n+1)\beta] \quad n = 0, 1, 2, \ldots$, where β is the length of intervals. In D/M/1 queues, the average wait-time of the customer who arrives in interval n, when n tends to infinity, is estimated as follows [97]:

$$E_n = (1/\mu).\sigma/(1-\sigma), \tag{95}$$

where μ^{-1} is the average service time, and σ is obtained from equation (96).

$$\sigma = e^{-\mu\beta(1-\sigma)} \tag{96}$$

Assuming that $\mu\beta\sigma$ is small enough, we can approximate $e^{\mu\beta\sigma}$ using following equation [98].

$$e^{\mu\beta\sigma} = 1 + \mu\beta\sigma \tag{97}$$

Finally, by substituting μ with M_i and β with $\frac{4\times 60}{xp_{ijk}}$, the approximation of average wait-time can be formulated as follows:

$$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\frac{1}{M_i\times(e^{\frac{M_i\times 4\times 60}{xp_{ijk}}}-\frac{4\times 60}{xp_{ijk}}-1)} \tag{98}$$

We call problem *P* with the above approximated objective function as P' and solve it by using the CP optimizer of CPLEX to obtain an initial feasible solution to *SAC* algorithm.

4.4.3.2 Enhanced Simulated Annealing Algorithm

We use SA as our local search algorithm within our SO framework. SA is an optimization paradigm based on the structural properties of physical materials that are melted down and then cooled in a controlled manner. The technique is designed in such as way to avoid getting trapped in a local optimal solution. At each iteration, a neighbor of the current solution is generated and either accepted as the new solution or rejected. When the temperature is high at the beginning, the acceptance process is almost random. As time progresses and the temperature cools down, it becomes increasingly dependent on the solution quality [99].

In our SA algorithm, we use a neighborhood that seeks to modify the number of assigned patients to each physician in each shift. Let $\boldsymbol{S}=\{1,\ldots,|\boldsymbol{J}||\boldsymbol{K}|\}$ be the set of shifts in the one-week planning horizon. Each candidate solution can be denoted by $|\boldsymbol{I}|\times|\boldsymbol{S}|$ matrix $\boldsymbol{A}$, where a_{is} is the number of patients assigned to physician $i\in\boldsymbol{I}$ in shift $s\in\boldsymbol{S}$. The neighborhood in our SA algorithm, denoted as $N(\boldsymbol{A})$, consists of swapping two shifts for each physician. $N(\boldsymbol{A})$ is formally defined as follows:

$$N(\boldsymbol{A})=\left\{\boldsymbol{A}^{'}:\exists s_1,s_2\in S,s_1\neq s_2,\text{ and } i\in I,a^{'}_{is_1}=a_{is_2},a^{'}_{is_2}=a_{is_1}\right\}.$$

The merit of the above neighborhood is that it maintains the feasibility of constraints (82) and (87)-(90), as in our search we aim to explore only feasible solutions. The objective value of each candidate solution is calculated by the aid of the *SIM* model. More precisely, for each candidate solution *A*, *f(A)* is calculated as the average wait-time and physicians overtime over all replications of the simulation.

In order to diversify the search we track the number of iterations that the best solution has not been improved. If the number of non-improved iterations exceeds a threshold value (*maxnit*), we generate a solution different from the best known solution and replace it with the current solution in the search. In other words, we use the information in the best known solution and define the following constraints:

$$\sum_{i \in I_c} xp_{ijk} - d_{cjk} \leq UB_{cjk} \quad \forall c \in C, j \in J, k \in K, \tag{99}$$

where d_{cjk} measures the positive deviation, and UB_{cjk} is calculated as follows:

$$\sum_{i \in I_c} xp^*_{ijk} = UB_{cjk} \quad \forall c \in C, j \in J, k \in K, \tag{100}$$

and xp^*_{ijk} denotes the number of patients assigned to each physician in each shift in the best known solution. Afterwards, we obtain a new solution that minimizes the maximum deviation from constraint (99) by solving the following optimization model ($P^{''}$):

($P^{''}$) Minimize Z

subject to: $Z \geq d_{cjk} \quad \forall c \in C, j \in J, k \in K$ (101)

(82)-(94) & (99)

It is noteworthy that $P^{''}$ is a restricted version of model P by considering a new objective function and two sets of constraints (99) and (101). We denote the optimal solution of $P^{''}$ as $A^{''}$. Algorithm 1 summarizes our SO algorithm.

Algorithm 4 Simulation optimization algorithm

```
Solve $P'$ to obtain initial solution $A$
Insert $A$ into $SIM$
$A^* \leftarrow A$
$f^* \leftarrow f(A)$
Set an initial temperature $T$ and a reduction factor $0 < r < 1$
while Not yet frozen do
    $count \leftarrow 0$
    $nit \leftarrow 0$
    while count$< L$ do
        Pick a random neighbor $A' \in N(A)$
        Insert $A'$ into $SIM$
        $\Delta = f(A') - f(A)$
        if $\Delta \leq 0$ then
            $A \leftarrow A'$
            if $f(A') < f^*$ then
                $A^* \leftarrow A'$
                $nit \leftarrow 0$
                $f^* \leftarrow f(A')$
        else
            Set $A \leftarrow A'$ with probability $e^{\frac{-\Delta}{T}}$
            $nit + +$
        if $nit > maxnit$ then
            Solve $P''$
            $nit \leftarrow 0$
            Set $A \leftarrow A''$
        $count + +$
    $T \leftarrow rT$
return $A^*$ and $f^*$
```

4.5 Numerical Results

In this section, we present the results of the computational experiments conducted in the context of an outpatient cancer treatment polyclinic in order to analyze the performance of the SO modeling approach proposed for our PSP. We first analyze the value of adopting an SO approach for physicians scheduling as compared with a deterministic mathematical programming approach. We also study the impact of adopting a multi-objective approach on the optimal value of each performance measure. Finally, we evaluate the performance of proposed optimization search routine *(SAC)* embedded into our SO framework both in terms of solution quality and CPU time. All experiments were run on a Dell station with an Intel(R) Core(TM) CPU i7-4790 processor at 3.60 GHz and 16 GB of RAM under Windows 7 environment. The formulations and algorithms were coded in C++, and the associated MIPs were solved using the Concert Technology of CPLEX 12.7.0. The simulation model was coded in DESP-C++, a DES package based on C++ [100], an open source object-oriented simulation kernel that provides classes for managing and ordering simulation events.

4.5.1 The Value Of SO Modeling Approach

In this section, we show the value of adopting a SO approach for our PSP in comparison with deterministic mathematical programming that is common in practice. To this end, we solve a deterministic PSP similar to the one in Tohidi et al. [45], where the objective function of the problem is merely formulated based on physicians' preferences, while abiding by the same constraints as the one in Model P. We denote the optimal solution of this model as *det*. Afterwards, this solution along with the converged solution of SO algorithm (denoted as SO) are plugged into *SIM*, and the expected patient wait-time and physician overtime are calculated for each solution in our PSP. Figure 4.4 summarizes the comparison between these KPIs (in minutes) obtained after simulating the schedules corresponding to *det* and *SO*. These results clearly confirm the importance of including average wait-time and overtime into the objective function of the PSP given that over 19% and 47% improvement can be achieved, respectively, in each KPI as compared with a deterministic physician scheduling approach where the analytic form of these KPIs cannot be formulated in the objective function.

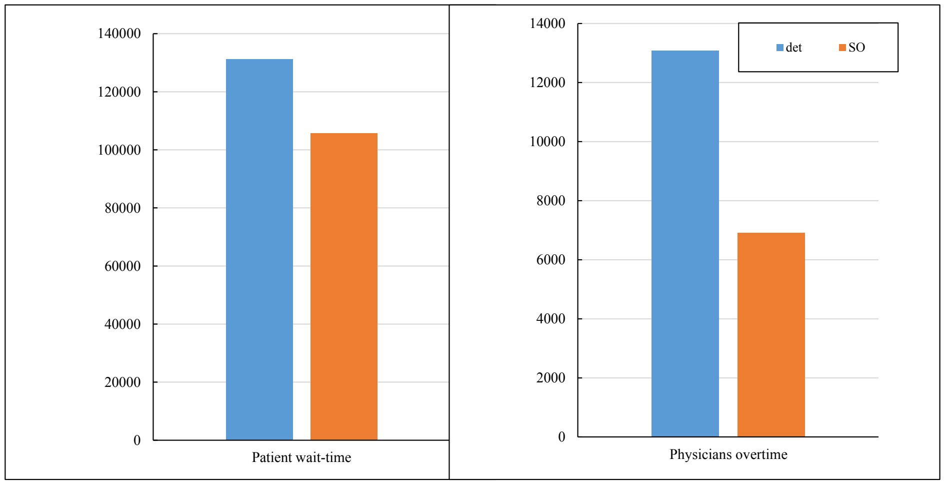

Figure 4.4: Average patient wait-time and physician overtime

4.5.2 Impact Of Multi-objective Approach On Each Performance Measure

In this section, we evaluate the impact of considering two objectives on the value of each via solving the problem with one objective at a time. Table 4.4 contains the value of each objective in the best known solution when solving the problem considering only one objective function versus considering the summation of both objectives. As can be observed in this table, when we consider physicians overtime as the objective function, the value of both objectives in the best solution deteriorate compared to patient wait-time or summation of both objectives as the objective function. Total patient wait-time improves significantly when it is the only objective. However, physicians overtime is very close to its value when physician overtime is the only objective. This gives us an interesting insight that by minimizing patient wait-time in our PSP, physician overtime will be implicitly minimized. Additionally, this confirms our initial assumption on considering a positive correlation between the two KPIs and considering patient wait-time as the only objective function while generating the initial feasible solution (model P'). Finally, the results suggest that considering a multi-objective approach is appropriate for obtaining high quality solutions with respect to both KPIs as compared with formulating the problem as a single-objective model.

Table 4.4: Value of each objective in the best solution considering different objective functions

Objective function	Physician overtime	Patient wait-time
Physician overtime	6979	108088
Patient wait-time	6969	105363
Physician overtime + Patient wait-time	6909	105724

4.5.3 Performance Of The Optimization Search Routine

In this section, we compare the performance of the optimization search routine *(SAC)* with a conventional SA algorithm (denoted by *SA*) and an exhaustive local search (denoted by LS) algorithm. In both *LS* and *SA*, we provide the initial solution by solving model P with an objective function similar to the one adopted in deterministic PSP models (see, e.g., Tohidi et al. [45]). In *SA*, we do not include the diversification strategy exploited in *SAC*. In *LS*, we use the same kind of neighborhood as in the *SA*, and we explore all feasible neighborhoods, starting from the first to the last physician.

Table 4.5 presents the values of parameters adopted in *SAC*; these values are chosen according to extensive computational experiments.

Table 4.5: The value of parameters in *SAC*

Parameter	Value
T	5000
r	0.9
$count$	200
$maxnit$	100

Figure 4.5 plots the CPU time versus the best know solution value for each of the above

mentioned search algorithms. By comparing the results of *LS* and *SA*, it can be concluded that *SA* outperforms *LS*, as in the first 600 seconds of the search, it finds a solution that is 5% better than the one found by *LS*. Further, it can be observed that our proposed approximation model $P^{'}$ is very efficient in finding a high quality initial feasible solution and significantly accelerating the search process. This initial solution is 8% better than the one obtained by considering a deterministic objective function in model P and 17% better than solving the problem without any objective. Finally, the objective function of the best solution obtained by *SAC* procedure is also 8% lower than the one obtained by a standard *SA* algorithm with a reduced CPU time of 500 sec.

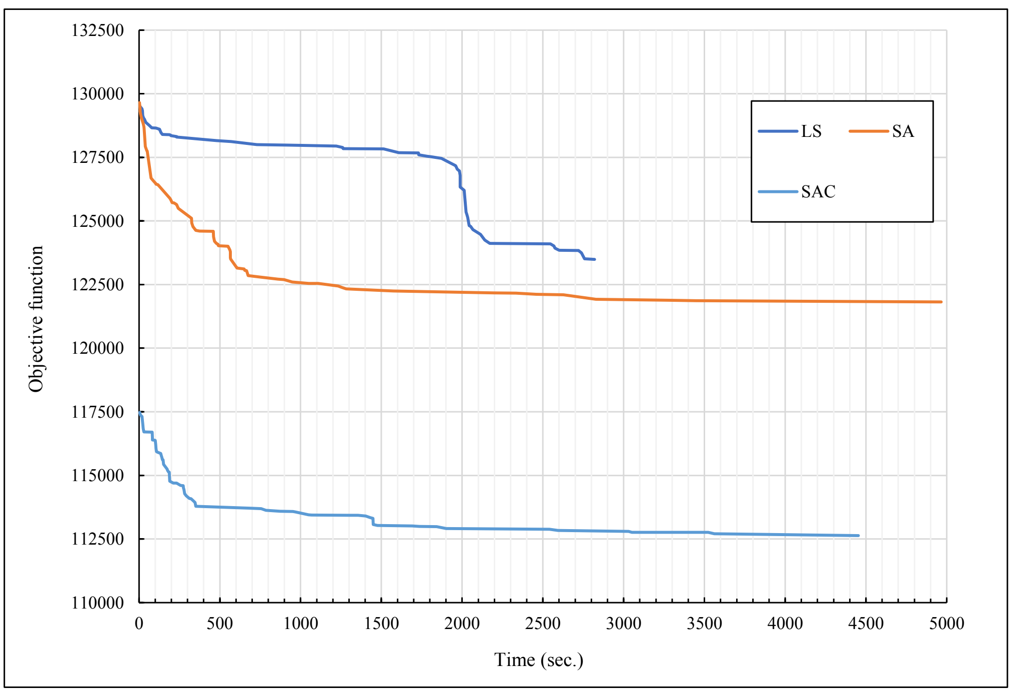

Figure 4.5: Comparison of solution methodologies

4.6 Conclusion

This paper combines discrete even simulation with an optimization search procedure to improve physicians work schedules and patients satisfaction in an outpatient polyclinic. To the best of our knowledge, the procedure is the first attempt in the literature to explicitly incorporate the minimization of patient wait-time in a physician scheduling problem while dealing with uncertain arrival patterns and treatment times. Given that commercial SO software packages are not efficient in solving a complex PSP similar to the one under investigation, we implemented a DES kernel integrated into an enhanced SA algorithm to solve this problem. In particular, modeling the system as a D/M/1 queue and approximating the stochastic wait-time as an analytic (non-linear) function leads to a high quality initial feasible solution within the search procedure. This approximation along with a diversification strategy embedded within SA algorithm significantly improves the quality of converged solution and the CPU time. Furthermore, our numerical results based upon the data provided by a polyclinic in a university health center in Montreal, Canada, confirmed the significance of incorporating stochastic functions such as patient wait-time and physician overtime into physician scheduling problems. More specifically, the results indicate that adopting a SO approach to address this problem can simultaneously improve patient experience via reducing wait-time and operational costs of the clinic in terms of extended shift duration for physicians.

Our future work will investigate other efficient ways of combining mathematical optimizers and different metaheuristics with simulation software packages in solving this type of highly-constrained real world problem. In particular, other approaches or approximation methods can be exploited to generate a high quality initial solution. Another interesting area of research on this topic is to strengthen the interaction between the simulation and optimization modules. Our next effort must be dedicated to extract more information from the simulation model to guide the local search and optimization search routine.

Acknowledgments

This research was partly funded by the Canadian Natural Sciences and Engineering Research Council under grants 418609-2012 and 402043-2011. This support is gratefully acknowledged.

Chapter 5

Conclusion and Future Work

5.1 Concluding Remarks

This thesis addressed physicians scheduling in the context of polyclinics, distinguished by clinics' requirements and resource availability. It accounted for several types of problems plausible in practice associated with the scheduling of physicians in outpatients polyclinics. Inspired by a real-life case example, i.e., the outpatient polyclinic of MUHC, we investigated the performance of the solution approaches on instances with realistic sizes. It was shown that the problem instances can be solved within reasonable amount of times utilizing the proposed solution schemes.

In the second chapter, we proposed a multi-objective mixed-integer programming model for integrating physician and clinic scheduling problems. In addition to the common physicians scheduling constraints, the limited capacity of resources, such as waiting and treatment rooms, and clinics requirements were also taken into account. Moreover, in order to better justify the significance of integrating physician scheduling problem with clinic session scheduling in polyclinics, we measured the impact of adding constraints associated with clinic resources and administrative rules into the physician scheduling model. We also developed an iterated variable neighborhood descent algorithm to obtain high quality solutions in a reasonable time limit. The algorithm combines iterated local search and variable neighborhood descent procedures. Our computational results on the aforementioned test instances revealed the high quality of schedules provided by this algorithm in comparison with a standard optimization software.

In the third chapter, we introduced a framework for planning physicians in polyclinics under uncertainty. The procedure addresses the problem at the strategic, tactical, and operational planning levels. In the strategic level, we proposed an adjustable robust approach that plans clinics work schedules and assigns required capacity to each shift. The robust problem was solved with an implementor/adversary algorithm, which can prove optimality in reasonable CPU times. We also combined tactical physicians scheduling with operational rescheduling decisions into a two-stage stochastic program that incorporates the uncertainty in patients' treatment times. Since the variety of patients' treatment times in each clinic results in a fairly large number of scenarios, we used a sample average approximation scheme to obtain high quality solutions by considering only a sub-set of scenarios. Furthermore, we investigated the impact of including uncertainty in our solution framework by using Monte Carlo simulation. We compared our framework to its deterministic counterpart and demonstrated that our framework produces significantly lower costs.

Finally, in the fourth chapter, we combined discrete-event simulation modeling with an optimization search procedure to improve physicians work schedules and patients satisfaction. We focused on scheduling physicians under uncertain patients' arrival patterns and treatment times. Moreover, we developed a discrete-event simulation kernel integrated with an enhanced simulated annealing algorithm to solve this problem. In addition to that, we used a model of D/M/1 queue type and approximated the stochastic wait-time as an analytic non-linear function that gave us a high quality initial feasible solution. Furthermore, our numerical results confirmed the significance of incorporating stochastic functions such as patient wait-time and physician overtime into physician scheduling problem.

5.2 Future Research Directions

Immediate extensions of this thesis can revolve around the following directions.

- Developing a decomposition-based exact algorithm for the PCSP.
- Considering a variant of the proposed models accounting for uncertainty in unavailability of physicians, where the decisions made at each stage might affect the decisions in subsequent stages.
- Developing a multi-stage stochastic programming model for the variant mentioned above.
- Investigating on the impact physician scheduling on various appointment systems.
- Another interesting research direction is to develop MIP approximation models that can estimate the patients wait-time and physicians overtime.
- Developing efficient solution algorithms for the variants mentioned above would be another promising area of research.

Bibliography

[1] M. Erhard, J. Schoenfelder, A. Fügener, J. Brunner, State of the art in physician scheduling, European Journal of Operational Research 265 (2018) 1–18.

[2] M. Pinedo, Planning and scheduling in manufacturing and services, Springer, 2005.

[3] M. Gendreau, J. Ferland, B. Gendron, N. Hail, B. Jaumard, S. Lapierre, G. Pesant, P. Soriano, Physician scheduling in emergency rooms, in: Practice and Theory of Automated Timetabling VI, Springer, 53–66, 2006.

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

[5] J. P. Leigh, R. L. Kravitz, M. Schembri, S. J. Samuels, S. Mobley, Physician career satisfaction across specialties, Archives of internal medicine 162 (14) (2002) 1577–1584.

[6] D. Grembowski, D. Paschane, P. Diehr, W. Katon, D. Martin, D. L. Patrick, Managed care, physician job satisfaction, and the quality of primary care, Journal of General Internal Medicine 20 (3) (2005) 271–277.

[7] M. António Ferreira Rodrigues Nogueira dos Santos, H. Kurt Olof Eriksson, Insights into physician scheduling: a case study of public hospital departments in Sweden, International journal of health care quality assurance 27 (2) (2014) 76–90.

[8] J. T. Kolstad, A. E. Kowalski, The impact of health care reform on hospital and preventive care: evidence from Massachusetts, Tech. Rep., National Bureau of Economic Research, 2010.

[9] C. M. Poissonnet, M. Véron, Health effects of work schedules in healthcare professions, Journal of Clinical Nursing 9 (1) (2000) 13–23.

[10] L.-M. Rousseau, G. Pesant, M. Gendreau, A general approach to the physician rostering problem, Annals of Operations Research 115 (1-4) (2002) 193–205.

[11] M. W. Carter, S. D. Lapierre, Scheduling emergency room physicians, Health Care Management Science 4 (4) (2001) 347–360.

[12] R. Bruni, P. Detti, A flexible discrete optimization approach to the physician scheduling problem, Operations Research for Health Care 3 (4) (2014) 191–199.

[13] A. T. Ernst, H. Jiang, M. Krishnamoorthy, D. Sier, Staff scheduling and rostering: A review of applications, methods and models, European Journal of Operational Research 153 (1) (2004) 3–27.

[14] J. Van den Bergh, J. Beliën, P. De Bruecker, E. Demeulemeester, L. De Boeck, Personnel scheduling: A literature review, European Journal of Operational Research 226 (3) (2013) 367–385.

[15] J. O. Brunner, Flexible shift planning in the service industry: the case of physicians in hospitals, vol. 640, Springer Science & Business Media, 2010.

[16] E. K. Burke, P. De Causmaecker, G. V. Berghe, H. Van Landeghem, The state of the art of nurse rostering, Journal of Scheduling 7 (6) (2004) 441–499.

[17] H. Beaulieu, J. A. Ferland, B. Gendron, P. Michelon, A mathematical programming

approach for scheduling physicians in the emergency room, Health Care Management Science 3 (3) (2000) 193–200.

[18] S. Topaloglu, A shift scheduling model for employees with different seniority levels and an application in healthcare, European Journal of Operational Research 198 (3) (2009) 943–957.

[19] S. Topaloglu, A multi-objective programming model for scheduling emergency medicine residents, Computers & Industrial Engineering 51 (3) (2006) 375–388.

[20] W. T. Song, M. Chih, A. E. Bair, Improving the efficiency of physical examination services, Journal of medical systems 34 (4) (2010) 579–590.

[21] A. Raouf, M. Ben-Daya, Outpatient clinic's staffing: a case study, International Journal of Health Care Quality Assurance 10 (6) (1997) 229–232.

[22] R. Stolletz, J. O. Brunner, Fair optimization of fortnightly physician schedules with flexible shifts, European Journal of Operational Research 219 (3) (2012) 622–629.

[23] J. O. Brunner, J. F. Bard, R. Kolisch, Midterm scheduling of physicians with flexible shifts using branch and price, IIE Transactions 43 (2) (2010) 84–109.

[24] S. Topaloglu, I. Ozkarahan, A constraint programming-based solution approach for medical resident scheduling problems, Computers & Operations Research 38 (1) (2011) 246–255.

[25] J. Puente, A. Gómez, I. Fernández, P. Priore, Medical doctor rostering problem in a hospital emergency department by means of genetic algorithms, Computers & Industrial Engineering 56 (4) (2009) 1232–1242.

[26] A. Gunawan, H. C. Lau, Master physician scheduling problem, Journal of the Operational Research Society 64 (3) (2013) 410–425.

[27] C. Van Huele, M. Vanhoucke, Analysis of the integration of the physician rostering problem and the surgery scheduling problem, Journal of Medical Systems 38 (6) (2014) 1–16.

[28] B. Roland, F. Riane, Integrating surgeons' preferences in the operating theatre planning, European Journal of Industrial Engineering 5 (2) (2011) 232–250.

[29] B. Roland, C. Di Martinelly, F. Riane, Y. Pochet, Scheduling an operating theatre under human resource constraints, Computers & Industrial Engineering 58 (2) (2010) 212–220.

[30] E. Burke, P. De Causmaecker, S. Petrovic, G. V. Berghe, Variable neighborhood search for nurse rostering problems, in: Metaheuristics: computer decision-making, Springer, 153–172, 2003.

[31] E. K. Burke, J. Li, R. Qu, A hybrid model of integer programming and variable neighbourhood search for highly-constrained nurse rostering problems, European Journal of Operational Research 203 (2) (2010) 484–493.

[32] T. Cayirli, E. Veral, H. Rosen, Designing appointment scheduling systems for ambulatory care services, Health Care Management Science 9 (1) (2006) 47–58.

[33] R. T. Marler, J. S. Arora, The weighted sum method for multi-objective optimization: new insights, Structural and Multidisciplinary Optimization 41 (6) (2010) 853–862.

[34] J. Koski, Defectiveness of weighting method in multicriterion optimization of structures, Communications in Applied Numerical Methods 1 (6) (1985) 333–337.

[35] A. M. Geoffrion, Proper efficiency and the theory of vector maximization, Tech. Rep., DTIC Document, 1967.

[36] T. L. Saaty, How to make a decision: the analytic hierarchy process, European Journal of Operational Research 48 (1) (1990) 9–26.

[37] O. Grodzevich, O. Romanko, Normalization and Other Topics in Multiobjective Optimization, in: In Proceedings of the Fields MITACS Industrial Problems Workshop, Toronto, 2006.

[38] H. R. Lourenço, O. C. Martin, T. Stützle, Iterated local search: Framework and applications, in: Handbook of Metaheuristics, Springer, 363–397, 2010.

[39] N. Mladenović, P. Hansen, Variable neighborhood search, Computers & Operations Research 24 (11) (1997) 1097–1100.

[40] W. J. Abernathy, N. Baloff, J. C. Hershey, S. Wandel, A three-stage manpower planning and scheduling model—-a service-sector example, Operations Research 21 (3) (1973) 693–711.

[41] G. M. Campbell, A two-stage stochastic program for scheduling and allocating cross-trained workers, Journal of the Operational Research Society 62 (6) (2011) 1038–1047.

[42] J. F. Bard, H. W. Purnomo, Short-term nurse scheduling in response to daily fluctuations in supply and demand, Health Care Management Science 8 (4) (2005) 315–324.

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

[44] J. F. Bard, Z. Shu, L. Leykum, Monthly clinic assignments for internal medicine housestaff, IIE Transactions on Healthcare Systems Engineering 3 (4) (2013) 207–239.

[45] M. Tohidi, M. Kazemi, I. Contreras, Integrated physicians and clinic scheduling in polyclinics, Journal of the Operational Research Society 000 (2018) 1–15.

[46] J. F. Bard, H. W. Purnomo, Real-time scheduling for nurses in response to demand fluctuations and personnel shortages, in: Proceedings of the 5th international conference on the practice and theory of automated timetabling, 67–87, 2004.

[47] J. F. Bard, H. W. Purnomo, Hospital-wide reactive scheduling of nurses with preference considerations, IIE Transactions 37 (7) (2005) 589–608.

[48] D. Hur, V. A. Mabert, K. M. Bretthauer, Real-Time Work Schedule Adjustment Decisions: An Investigation and Evaluation, Production and Operations Management 13 (4) (2004) 322–339.

[49] F. F. Easton, J. C. Goodale, Schedule recovery: Unplanned absences in service operations, Decision Sciences 36 (3) (2005) 459–488.

[50] C. N. Gross, A. Fügener, J. O. Brunner, Online rescheduling of physicians in hospitals, Flexible Services and Manufacturing Journal (2016) 1–33.

[51] G. M. Campbell, Cross-utilization of workers whose capabilities differ, Management Science 45 (5) (1999) 722–732.

[52] P. D. Wright, S. Mahar, Centralized nurse scheduling to simultaneously improve schedule cost and nurse satisfaction, Omega 41 (6) (2013) 1042–1052.

[53] R. Venkataraman, M. Brusco, An integrated analysis of nurse staffing and scheduling policies, Omega 24 (1) (1996) 57–71.

[54] P. D. Wright, K. M. Bretthauer, M. J. Côté, Reexamining the nurse scheduling problem: Staffing ratios and nursing shortages, Decision Sciences 37 (1) (2006) 39–70.

[55] P. D. Wright, K. M. Bretthauer, Strategies for addressing the nursing shortage: Coordinated decision making and workforce flexibility, Decision Sciences 41 (2) (2010) 373–401.

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

[57] J. Ingels, B. Maenhout, The impact of reserve duties on the robustness of a personnel shift roster: An empirical investigation, Computers & Operations Research 61 (2015) 153–169.

[58] X. Zhong, J. Zhang, X. Zhang, A two-stage heuristic algorithm for the nurse scheduling problem with fairness objective on weekend workload under different shift designs, IISE Transactions on Healthcare Systems Engineering 7 (4) (2017) 224–235.

[59] M. Chiaramonte, D. Caswell, Rerostering of nurses with intelligent agents and iterated local search, IIE Transactions on Healthcare Systems Engineering 6 (4) (2016) 213–222.

[60] J. F. Bard, D. P. Morton, Y. M. Wang, Workforce planning at USPS mail processing and distribution centers using stochastic optimization, Annals of Operations Research 155 (1) (2007) 51–78.

[61] P. Punnakitikashem, J. M. Rosenberger, D. B. Behan, Stochastic programming for nurse assignment, Computational Optimization and Applications 40 (3) (2008) 321–349.

[62] P. Punnakitikashem, J. M. Rosenberber, D. F. Buckley-Behan, A stochastic programming approach for integrated nurse staffing and assignment, IIE Transactions 45 (10) (2013) 1059–1076.

[63] X. Zhu, H. D. Sherali, Two-stage workforce planning under demand fluctuations and uncertainty, Journal of the Operational Research Society 60 (1) (2009) 94–103.

[64] K. Kim, S. Mehrotra, A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management, Operations Research 63 (6) (2015) 1431–1451.

[65] O. El-Rifai, T. Garaix, V. Augusto, X. Xie, A stochastic optimization model for shift scheduling in emergency departments, Health Care Management Science 18(3) (2015) 289–302.

[66] D. Bertsimas, M. Sim, The price of robustness, Operations Research 52 (1) (2004) 35–53.

[67] D. Bienstock, Histogram models for robust portfolio optimization, Journal of Computational Finance 11 (1) (2007) 1–64.

[68] M. Holte, C. Mannino, The implementor/adversary algorithm for the cyclic and robust scheduling problem in health-care, European Journal of Operational Research 226 (3) (2013) 551–559.

[69] J. Tang, Y. Wang, An adjustable robust optimisation method for elective and emergency surgery capacity allocation with demand uncertainty, International Journal of Production Research 53 (24) (2015) 7317–7328.

[70] A. Shapiro, T. Homem-de Mello, On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs, SIAM journal on optimization 11 (1) (2000) 70–86.

[71] W.-K. Mak, D. P. Morton, R. K. Wood, Monte Carlo bounding techniques for determining solution quality in stochastic programs, Operations Research Letters 24 (1) (1999) 47–56.

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

[73] B. Verweij, S. Ahmed, A. J. Kleywegt, G. Nemhauser, A. Shapiro, The sample average approximation method applied to stochastic routing problems: a computational study, Computational Optimization and Applications 24 (2) (2003) 289–333.

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

[75] C. Bleustein, D. B. Rothschild, A. Valen, E. Valatis, L. Schweitzer, R. Jones, Wait times, patient satisfaction scores, and the perception of care, The American journal of managed care 20 (5) (2014) 393–400.

[76] M. Tohidi, M. Kazemi, I. Contreras, A Physicians planning framework for polyclinics under uncertainty, URL `https://users.encs.concordia.ca/~icontrer/web/tohidi2017.pdf`, 2018.

[77] M. C. Fu, Optimization for simulation: Theory vs. practice, INFORMS Journal on Computing 14 (3) (2002) 192–215.

[78] M. Centeno, R. Giachetti, R. Linn, A. Ismail, A simulation-ilp based tool for scheduling ER staff, in: Proceedings of the 2003 conference on Winter simulation, Winter Simulation Conference, 1930–1938, 2003.

[79] V. De Angelis, G. Felici, P. Impelluso, Integrating simulation and optimisation in health care centre management, European Journal of Operational Research 150 (1) (2003) 101–114.

[80] M. A. Ahmed, T. M. Alkhamis, Simulation optimization for an emergency department healthcare unit in Kuwait, European Journal of Operational Research 198 (3) (2009) 936–942.

[81] J. P. Oddoye, D. F. Jones, M. Tamiz, P. Schmidt, Combining simulation and goal programming for healthcare planning in a medical assessment unit, European Journal of Operational Research 193 (1) (2009) 250–261.

[82] F. Zeinali, M. Mahootchi, M. M. Sepehri, Resource planning in the emergency departments: A simulation-based metamodeling approach, Simulation Modelling Practice and Theory 53 (2015) 123–138.

[83] R. S. Dittus, R. W. Klein, D. J. DeBrota, M. A. Dame, J. F. Fitzgerald, Medical resident work schedules: Design and evaluation by stimulation modeling, Management Science 42 (6) (1996) 891–906.

[84] G. W. Evans, E. Unger, T. B. Gor, A simulation model for evaluating personnel schedules in a hospital emergency department, in: Simulation Conference, 1996. Proceedings. Winter, IEEE, 1205–1209, 1996.

[85] M. D. Rossetti, G. F. Trzcinski, S. A. Syverud, Emergency department simulation and determination of optimal attending physician staffing schedules, in: Simulation Conference Proceedings, 1999 Winter, vol. 2, IEEE, 1532–1540, 1999.

[86] J. R. Swisher, S. H. Jacobson, J. B. Jun, O. Balci, Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation, Computers & Operations Research 28 (2) (2001) 105–125.

[87] C. W. Spry, M. A. Lawley, Evaluating hospital pharmacy staffing and work scheduling using simulation, in: Proceedings of the 37th conference on Winter simulation, Winter Simulation Conference, 2256–2263, 2005.

[88] S. M. Al-Najjar, S. H. Ali, Staffing and scheduling emergency rooms in two public hospitals: a case study, International Journal of Business Administration 2 (2) (2011) 137.

[89] W. Abo-Hamad, A. Arisha, Simulation-based framework to improve patient experience in an emergency department, European Journal of Operational Research 224 (1) (2013) 154–166.

[90] C. Oh, A. M. Novotny, P. L. Carter, R. K. Ready, D. D. Campbell, M. C. Leckie, Use of a simulation-based decision support tool to improve emergency department throughput, Operations Research for Health Care 9 (2016) 29–39.

[91] J. R. Swisher, S. H. Jacobson, Evaluating the design of a family practice healthcare clinic using discrete-event simulation, Health Care Management Science 5 (2) (2002) 75–88.

[92] J.-Y. Yeh, W.-S. Lin, Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department, Expert Systems with Applications 32 (4) (2007) 1073–1083.

[93] E. Cabrera, M. Taboada, M. L. Iglesias, F. Epelde, E. Luque, Optimization of healthcare emergency departments by agent-based simulation, Procedia Computer Science 4 (2011) 1880–1889.

[94] E. Cabrera, M. Taboada, M. L. Iglesias, F. Epelde, E. Luque, Simulation optimization for healthcare emergency departments, Procedia Computer Science 9 (2012) 1464–1473.

[95] Y. A. Ozcan, E. Tànfani, A. Testi, Improving the performance of surgery-based clinical pathways: a simulation-optimization approach, Health Care Management Science 20 (1) (2017) 1–15.

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

[97] B. Jansson, Choosing a good appointment system—a study of queues of the type (D, M, 1), Operations Research 14 (2) (1966) 292–312.

[98] M. Abramowitz, I. A. Stegun, Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55, Courier Corporation, 1965.

[99] M. Gendreau, J.-Y. Potvin, Handbook of metaheuristics, vol. 2, Springer, 2010.

[100] J. Darmont, DESP-C++: a discrete-event simulation package for C++, Software: Practice and Experience 30 (1) (2000) 37–60.