Journal of the Operational Research Society manuscript No.
(will be inserted by the editor)

Integrating Physician and Clinic Scheduling in Ambulatory Polyclinics

Mohammad Tohidi · Masoumeh Kazemi Zanjani · Ivan Contreras

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 decent 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.

Keywords Healthcare · Physician scheduling · Integer programming · Heuristics

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 (Poissonnet and Véron, 2000). 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.

Physician scheduling problems (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 (Gendreau

Mohammad Tohidi · Masoumeh Kazemi Zanjani · Ivan Contreras
Concordia University and Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, Canada H3G 1M8
E-mail: m_tohidi@encs.concordia.ca, kazemi@encs.concordia.ca, icontrer@encs.concordia.ca

et al, 2006). Creating physicians' work schedules must be done by abiding to many rules and regulations. Most frequently, those rules are 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. Gendreau et al (2006) provide four categories of rules commonly considered in PSPs: supply and demand, workload, fairness, and ergonomic constraints. PSPs are combinatorial and finding a feasible solution that satisfies all the physicians' requests is challenging (Rousseau et al, 2002). The objective in PSPs is often to create physicians' work schedules according to their preferences and to improve the ergonomic aspects of the schedules. 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 (Carter and Lapierre, 2001; Bruni and Detti, 2014). 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 (Brunner et al, 2009).

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, pre-defined 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.

In this paper we present an integrated *physician and clinic scheduling problem* (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 mixed integer programing (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 decent 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 Iterated Variable Neighborhood Decent (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”.

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 (2004) and Van den Bergh et al (2013). 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 (Brunner, 2010). 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 (2016) and Burke et al (2004), respectively.

As Erhard et al (2016) 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 (2000) 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 (2006) 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 modeled as a multi-objective MIP model in which the objective function is a weighted function of the deviation variables. The weights are generated through pairwise comparison. Topaloglu (2009) 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.

Brunner et al (2009) 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 (2012) reformulate the problem investigated in (Brunner et al, 2009) as a set covering problem. The authors include some additional ergonomic and distribution constraints in their reformulation. Brunner (2010) propose an MIP for scheduling physicians with multiple experience levels for the same case study in (Brunner et al, 2009) 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 (2010) tackle the same problem as (Brunner et al, 2009) 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 (2002) 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 (2011) 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 (2001) 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 (2009) 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 (2014) 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 (2013) and Van Huele and Vanhoucke (2014) 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 (2013) 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 (2014) 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. The results also show that certain specific physicians' preferences have the most impact on the operational surgery schedule. Roland and Riane (2011) 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 (2010) 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. The authors also 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.

The physician and clinic scheduling problem

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.

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., Beaulieu et al, 2000; Burke et al, 2003, 2010; Gunawan and Lau, 2013; Roland and Riane, 2011). 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 together must be scheduled in the same shift for a minimum number of shifts.

- Utilization of the waiting room cannot fluctuate more than a certain level over the planning horizon. 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*' (Cayirli et al, 2006).
- 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.

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

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

The idea behind this objective is assigning blocks of consecutive days to physicians within a one-week planning horizon (Monday to Friday). 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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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) ensure that, for each clinic $c \in t$ in each cluster $t \in \boldsymbol{T}$, at least F^t shifts in a week are scheduled. 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 (13) are the standard integrality conditions on the decision variables.

A weighted sum approach

The weighted sum method is a classical approach when dealing with multiple objectives (Marler and Arora, 2010; Koski, 1985). 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 (Koski, 1985; Geoffrion, 1968). Studies with a posteriori articulation of preferences focus on providing the Pareto optimal set (Marler and Arora, 2010).

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

$$\sum_{k \in \boldsymbol{K}} x_{i(j+1)k} + \sum_{k \in \boldsymbol{K}} x_{i(j-1)k} - x_{ijk'} + n_{ij} + m_i \geq 0 \quad \forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I}_c,\ \forall j = 2 \ldots |\boldsymbol{J}|-1,\ \forall k' \in \boldsymbol{K} \tag{18}$$

$$\sum_{k \in \boldsymbol{K}} x_{i2k} - x_{i1k'} + n_{i1} + m_i \geq 0 \qquad \forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I}_c,\ \forall k' \in \boldsymbol{K} \tag{19}$$

$$\sum_{k \in \boldsymbol{K}} x_{i(|\boldsymbol{J}|-1)k} - x_{i|\boldsymbol{J}|k'} + n_{i|\boldsymbol{J}|} + m_i \geq 0 \qquad \forall c \in \boldsymbol{C},\ \forall i \in \boldsymbol{I}_c,\ \forall k' \in \boldsymbol{K} \tag{20}$$

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

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

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

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

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

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

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

We can now formulate the PCSP as the following MIP:

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

subject to (2) – (25).

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 (Topaloglu, 2006, 2009). A detailed description of the AHP can be found in (Saaty, 1990). The author suggests that ranking the objectives should be done 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 a constraint 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. In our study, the comparison between different objectives are made according to information gathered through an interview with hospital staff as summarized in a pairwise comparison matrix in Table 1. In order to calculate the relative importance value, 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. According to Marler and Arora (2010), one should transform the objective function such that the objectives have similar ranges prior to setting the weights with the ranking method. However, when one is implementing a pairwise comparison method to set the weights, the objective functions are no longer required to be transformed/normalized.

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 relied mainly on finding 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 decent (VND). From now on, we refer to this hybrid algorithm as an iterated variable neighborhood decent (IVND) procedure. On

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

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 (Lourenço et al, 2010). 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 (Mladenović and Hansen, 1997). 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

```
A^0 = GenerateInitialSolution
A* = LocalSearch(A^0)
while termination condition has not been met do
    A^0 = Perturbation(A*, memory)
    A' = LocalSearch(A^0)
    A* = AcceptanceCriterion(A*,A',memory)
end while
```

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 k neighborhoods, $\mathcal{N}_1$,$\mathcal{N}_2$,...,$\mathcal{N}_k$. The algorithm starts from neighborhood $\mathcal{N}_1$ until a local optimal solution is found. After that, it switches to neighborhoods $\mathcal{N}_2$,...,$\mathcal{N}_k$, sequentially, until an improved solution is found. Each time the search improves the incumbent solution, the procedure restarts from neighborhood $\mathcal{N}_1$. 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 (2003) and Burke et al (2010) for the case of nurse rostering problems, which consider the exploration of neighborhoods in an increasing size order. Let $\boldsymbol{S} = \{1, \ldots, |\boldsymbol{J}||\boldsymbol{K}|\}$ be the set of shifts in the entire planning horizon (one week). In what follows, solutions 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 $k = 1, \ldots, |\boldsymbol{J}||\boldsymbol{K}|$, $\mathcal{N}_k$ consists of solutions which can be reached by swapping between any two physicians k consecutive shifts having the same starting shift s_r. That is,

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

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 1 illustrates possible swaps in $\mathcal{N}_3$ and $\mathcal{N}_{10}$. In particular, the figure shows an initial solution with three physicians and 10 shifts. A new solution in $\mathcal{N}_3(\boldsymbol{A})$ 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 solution $\boldsymbol{A}^{'}$, 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. Dashed lines depict the only possible swap between physician 1 and 3 in $\mathcal{N}_{10}(\boldsymbol{A})$, as in this neighborhood the entire weekly work schedules are exchanged between any two physicians.

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 a MIP. Let $\boldsymbol{Pt}$ 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 $\boldsymbol{Fe}$ as the set of elements whose values have changed during the VND

	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 1 VND neighborhood structures

procedure in the current iteration. During the perturbation phase, a percentage of elements of the current solution $\boldsymbol{A}$ are selected randomly and if they do not belong to set $\boldsymbol{Fe}$, they will be added to set $\boldsymbol{Pt}$. 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
        end if
    end while
    if F(Ā) < F(A*) then
        A* ← Ā
    else
        nit++
    end if
    if nit > NIL then
        Ā ← A*
        nit ← 0
    end if
    Update Pt and R
    Solve PCSP-R with (26) to obtain a new trial solution Ā
end while
return A*
```

Computational results

In this section, we present the results of computational experiments we have run in order to compare and to 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. We use a traditional (deterministic) branch-and-bound solution algorithm with all CPLEX parameters set to their default values.

Table 2 Characteristics of benchmark instances

Class	#Days	#Shifts	# Clinics	# Physicians	# Clusters
A	5	2	5	15-50	3
B	5	2	10	60-100	5
C	5	2	13	133	5

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 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 waiting room capacity is defined as

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

and the number of examination rooms as

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

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. In our experiments, we considered two cases for every instance: objectives with equal weights and weights determined by the AHP method. We thus considered 60 instances in total.

Table 3 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

The impact of clinic scheduling constraints on physicians work schedules

In this section we analyze the impact of clinics' and administration's constraints on physicians' work schedules. To this end, 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 3 shows the clinic's constraints considered in this experiment.

The analysis was conducted on all considered instances and the obtained results are given in Figure 2. We use two different sets of right-hand-side (RHS) values of the aforementioned constraints: the nominal case and one in which the RHS is increased by 20%. The horizontal axis represents results obtained when adding each of the constraints individually to the basic PSP model, while the vertical axis shows the percent increase in the objective function value of PSP when each set of constraints is added to PSP.

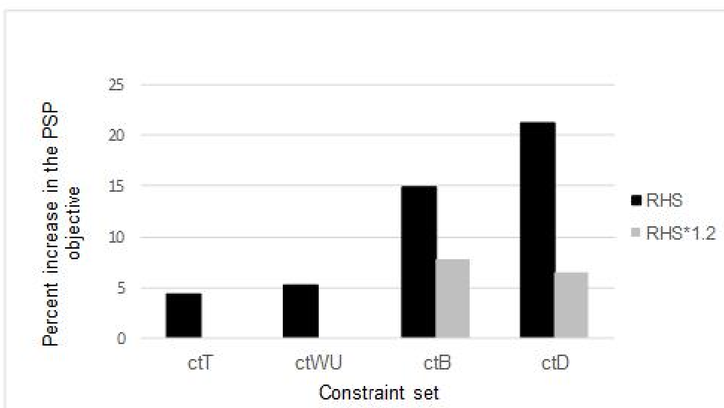

Figure 2 Analysis of adding clinic scheduling constraints to the PSP

The results given in Figure 2 show that incorporating the clinics' requirements in the physician scheduling problem 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. The simultaneous scheduling of clusters of clinics and the fluctuation level in waiting room utilization seem to have a lower impact on the physician's preferences.

The observed impact of clinic constraints on the physicians' work schedule provides insights to the hospital administrator in terms determining what are the resources that should be prioritized. Increasing the capacity of the waiting and examination rooms or limiting the number of patients to admit are the most critical choices. In other words, increasing the available resources for a fixed weekly demand provides more flexibility in assigning physicians to shifts and improves the quality of their work schedules.

Analyzing the computational performance of the PCSP formulation and IVNS 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 IVNS algorithm. Table 4 summarizes the values of the parameters of the IVND used in these experiments. These values were fine-tuned to provide the best overall results for our algorithm for the considered instances.

Table 4 The values of parameters of the IVND

Class	*perc%*	TL (sec.)	*repair_time_limit* (sec.)	NIL
A	$0.90 * \|\boldsymbol{I}\|\|\boldsymbol{J}\|\boldsymbol{K}\|$	10	5	10
B	$0.70 * \|\boldsymbol{I}\|\|\boldsymbol{J}\|\boldsymbol{K}\|$	50-100	30-50	10
C	$0.70 * \|\boldsymbol{I}\|\|\boldsymbol{J}\|\boldsymbol{K}\|$	100	100	10

Table 5 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 digit presents the instance number within its class, and the second one implies whether the objectives of the problem have equal weights (1) or AHP-based weights (2). 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 $100 * |Incumbent - Optimum|/|Optimum|$.

The results of Table 5 show that the proposed formulation is capable of finding the optimal solution in all considered instances in less than 10 seconds. The IVND is also capable of finding the optimal solution for all instances in less than 25 seconds, expect for one instance which takes about 4 minutes. Also, the results of the last column indicate that for all small instances (except for 4.1) the optimal solution is obtained in the first iteration of IVND algorithm. It is worth to note that class A of instances are considered to demonstrate the IVND heuristic is capable to obtain the optimal solution for easier test instances.

Table 6 and 7 show the results for classes B and C instances. In both tables, the results are grouped in three categories: CPLEX, IVND, and VND. More specifically, for CPLEX results, *%D1H* provides the

Table 5 Comparison of the IVND and solver on class A instances

Instance	CPLEX Time (sec.)	IVND Time (sec.)	VND %Dev.
1.1	<1	<1	0.00
1.2	<1	<1	0.00
2.1	<1	<1	0.00
2.2	3	5	0.00
3.1	3	<1	0.00
3.2	3	<1	0.00
4.1	9	231	2.50
4.2	2	20	0.00
5.1	9	25	0.00
5.2	10	20	0.00
6.1	5	5	0.00
6.2	5	4	0.00
7.1	10	15	0.00
7.2	7	15	0.00
8.1	3	<1	0.00
8.2	3	<1	0.00
9.1	<1	<1	0.00
9.2	<1	<1	0.00
10.1	7	15	0.00
10.2	7	10	0.00

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 in 6 hour time limit 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. *%G2H* and *%G6H* represent the optimal gap of the best solution (UB) found by the solver within 2 hours and 6 hours, respectively, calculated according to $|UB - LB|/|UB|{*}100$, where (LB) is the best valid lower bound obtained by the solver in the corresponding time limit. *%D1H′* and *%D2H′* are the percent deviation of the best solution, found by the IVND in 1 and 2 hours, respectively, from the incumbent solution for the corresponding instance. They are calculated similarly to *%D1H* and *%D2H* according to the bounds obtained by the IVND algorithm. *%BG* is the optimal gap of the best solution found by the IVND. It is calculated according to $|Incumbent - LB|/|Incumbent|{*}100$, where (LB) is the best valid lower bound obtained by CPLEX after 6 hours. It is worth noting that *%BG* demonstrates the quality of the solutions found by the IVND algorithm as compared to CPLEX. *%D* provides the percent deviation of the best solution obtained in the first iteration of the IVND from the incumbent solution of the corresponding instance. This information is particularly useful to evaluate the impact of using destroy/repair strategies within the IVND algorithm. Finally, *Time* shows the CPU time required for completing the first iteration of the IVND. The last row in Table 6 and 7 presents the average values of each column. Note that an N.A. entry indicates that either CPLEX was not able to find a feasible solution within the time limit or it consumed all the available memory before terminating.

Table 6 Comparison of the IVND and CPLEX on class B instances

Inst.	CPLEX				IVND			VND	
	%D1H	%D2H	%G2H	%G6H	%D1H′	%D2H′	%BG	%D	Time
1.1	0.58	0.58	4.30	N.A	0.00	0.00	3.69	1.16	204
1.2	0.31	0.11	2.28	N.A	0.00	0.00	1.87	0.69	200
2.1	N.A	N.A	N.A	N.A	0.00	0.00	3.17	6.21	145
2.2	0.68	0.39	2.05	1.77	0.00	0.00	1.62	7.85	120
3.1	2.36	2.36	6.70	6.61	0.00	0.00	4.10	14.17	210
3.2	0.15	0.15	2.92	2.70	0.12	0.00	2.54	7.10	203
4.1	N.A	N.A	N.A	N.A	0.00	0.00	7.01	6.60	201
4.2	N.A	N.A	N.A	3.11	0.00	0.00	3.02	11.76	205
5.1	29.03	9.68	16.67	8.89	0.00	0.00	5.38	48.39	203
5.2	N.A	8.90	11.02	1.27	0.00	0.00	1.14	28.07	260
6.1	51.94	6.20	11.61	8.69	0.00	0.00	4.48	10.85	151
6.2	16.95	0.84	3.34	2.84	0.00	0.00	2.46	8.73	152
7.1	3.68	1.47	4.7	3.08	0.00	0.00	3.08	9.56	117
7.2	1.08	0.06	1.52	1.21	0.00	0.00	1.21	9.45	119
8.1	2.13	2.13	4.7	N.A	0.00	0.00	2.48	9.93	117
8.2	0.93	0.93	2.74	N.A	0.00	0.00	1.79	8.31	117
9.1	28.28	2.07	5.09	4.18	0.00	0.00	2.74	7.59	121
9.2	3.12	0.18	1.96	1.38	0.00	0.00	1.38	9.82	121
10.1	5.30	1.99	5.82	5.71	0.00	0.00	3.60	11.26	202
10.2	3.18	0.52	3.34	N.A	0.15	0.00	2.80	8.03	187
Ave.	9.36	2.27	5.33	3.95	0.01	0.00	2.98	11.28	165.25

From the results provided in *%D1H′* and *%D1H* columns of Tables 6 and 7, 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 0.01 and 0.04 compared to 9.36 and 17.10, 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 0.00 for class B instances and 0.002 for class C instances compared to 2.27 and 1.95, respectively. Contrasting the average values of *%BG* and *%G6H* reveals that in all the instances of class B and almost all the instances (except instance 1.2 and 8.2) of class C, the solutions provided by the IVND in 2 hours are significantly better than the ones obtained by CPLEX even after 6 hours. The averages of column *%BG* in both Tables 6 and 7 indicate that the average optimal gap of the solutions provided by the IVND in 2 hours is at most 2.08 percent. From the results of *%D* in both Tables 6 and 7, it is also remarkable that running the VND iteratively significantly improves the solution obtained by a single iteration of VND in all

Table 7 Comparison of the IVND and CPLEX on class C instances

Inst.	CPLEX				IVND			VND	
	%D1H	%D2H	%G2H	%G6H	%D1H′	%D2H′	%BG	%D	Time
1.1	N.A	N.A	N.A	N.A	0.00	0.00	3.38	1.96	209
1.2	4.18	0.74	3.60	2.67	0.36	0.01	2.68	0.63	185
2.1	N.A	N.A	N.A	N.A	0.00	0.00	0.58	2.02	123
2.2	10.97	0.40	1.08	0.72	0.00	0.00	0.61	2.97	139
3.1	3.55	1.37	2.75	2.12	0.00	0.00	1.29	1.91	111
3.2	0.75	0.21	1.00	0.78	0.09	0.00	0.75	3.03	130
4.1	1.14	0.57	2.72	2.50	0.00	0.00	1.92	1.42	159
4.2	1.28	0.23	1.94	1.81	0.00	0.00	1.58	1.83	138
5.1	1.39	0.56	2.05	1.42	0.00	0.00	1.42	1.11	206
5.2	0.10	0.10	0.81	0.74	0.00	0.00	0.67	1.09	132
6.1	N.A	N.A	N.A	N.A	0.00	0.00	1.15	2.27	206
6.2	22.29	6.30	7.25	2.97	0.13	0.00	0.43	3.48	171
7.1	24.50	1.44	2.12	N.A	0.00	0.00	0.64	2.88	135
7.2	0.44	0.00	0.53	N.A	0.04	0.00	0.49	3.04	129
8.1	87.29	16.10	20.48	N.A	0.00	0.00	1.01	1.98	123
8.2	51.82	0.86	0.86	0.33	0.11	0.04	0.37	1.75	130
9.1	N.A	N.A	N.A	N.A	0.00	0.00	1.93	1.58	205
9.2	N.A	0.35	1.58	1.38	0.00	0.00	1.09	2.25	134
10.1	46.09	1.89	3.16	1.05	0.00	0.00	1.05	0.81	145
10.2	0.70	0.13	0.75	0.68	0.03	0.00	0.58	2.29	127
Ave.	17.10	1.95	3.29	1.47	0.04	0.00	1.18	2.02	151.85

the instances. Comparing the average values of columns *%G2H* and *%G6H* of Table 6 with those of Table 7, it can be concluded that class B instances are more difficult than class C instances. This can be partially explained by the fact that class C instances were generated in order to represent the real case. However, class B instances are designed to be as difficult as possible.

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 polyclinics 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.

References

Beaulieu H, Ferland JA, Gendron B, Michelon P (2000) A mathematical programm g approach for scheduling physicians in the emergency room. Health Care Management Science 3(3):193–200

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

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

Brunner JO (2010) Flexible shift planning in the service industry: the case of physicians in hospitals, vol 640. Springer Berlin Heidelberg, Berlin

Brunner JO, Bard JF, Kolisch R (2009) Flexible shift scheduling of physicians. Health Care Management Science 12(3):285–305

Brunner JO, Bard JF, Kolisch R (2010) Midterm scheduling of physicians with flexible shifts using branch and price. IIE Transactions 43(2):84–109

Burke E, De Causmaecker P, Petrovic S, Berghe GV (2003) Variable neighborhood search for nurse rostering problems. In: Metaheuristics: computer decision-making, Springer, New York, pp 153–172

Burke EK, De Causmaecker P, Berghe GV, Van Landeghem H (2004) The state of the art of nurse rostering. Journal of Scheduling 7(6):441–499

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

Carter MW, Lapierre SD (2001) Scheduling emergency room physicians. Health Care Management Science 4(4):347–360

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

Erhard M, Schoenfelder J, Fgener A, Brunner JO (2016) State of the art in physician scheduling. URL https://ssrn.com/abstract=2813360

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

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

Geoffrion AM (1968) Proper efficiency and the theory of vector maximization. Journal of Mathematical Analysis and Applications 22(3):618 – 630

Gunawan A, Lau HC (2013) Master physician scheduling problem. Journal of the Operational Research Society 64(3):410–425

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

Lourenço HR, Martin OC, Stützle T (2010) Iterated local search: Framework and applications. In: Handbook of Metaheuristics, Springer, New York, pp 363–397

Marler RT, Arora JS (2010) The weighted sum method for multi-objective optimization: new insights. Structural and Multidisciplinary Optimization 41(6):853–862

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

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

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

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

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

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

Saaty TL (1990) How to make a decision: the analytic hierarchy process. European Journal of Operational Research 48(1):9–26

Stolletz R, Brunner JO (2012) Fair optimization of fortnightly physician schedules with flexible shifts. European Journal of Operational Research 219(3):622–629

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

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

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

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