A Flexible Discrete Optimization Approach to the Physician Scheduling Problem

Renato Bruni*

"Sapienza" Università di Roma, Dip. di Ingegneria Informatica, Automatica e Gestionale, Via Ariosto, 25 - 00185 Roma, Italy

Paolo Detti

Università di Siena, Dip. di Ingegneria dell'Informazione e Scienze Matematiche, Via Roma, 56 - 53100 Siena, Italy

Abstract

Personnel scheduling deals with the attribution of a number of duty shifts to a number of workers respecting several types of requirements. In this work, the problem of scheduling physicians in health care departments is studied. This problem is NP-hard, and we propose a flexible Mixed Integer Linear Programming formulation that allows easy modifications for representing different situations and scenarios. This formulation can be solved to optimality by a standard Branch-and-Cut procedure even for very long planning horizons. A real-world case study is considered. A comparison of the solutions obtained by the proposed approach with the solutions currently adopted in the considered structure is presented. Results are very encouraging both from the schedule quality (e.g., workload balancing) and from the computational point of view.

Keywords: Staff Scheduling, Personnel Scheduling, Physicians, Rostering, Mixed Integer Linear Model.

1. Introduction

Personnel scheduling problems deal with the assignment of a number of tasks to a number of workers while respecting a number of restrictions that often make the problem harder than a simple assignment problem. Generally this attribution must be done in order to strictly satisfy service requirements and contractual agreements, and to maximize workers' preferences and/or minimize costs, where these objectives are applicable. These problems are very relevant in many working environments and assume different connotations in the different contexts, but typically constitute highly constrained optimization problems.

*Corresponding author
Email address: bruni@dis.uniroma1.it (Renato Bruni)

Consequently, in the last few decades they have been studied widely and several different approaches have been proposed (see for references [10, 26]). In the field of health care, personnel scheduling problems are particularly pervasive, because many services need to be assured on a continuous basis, twenty-four hours a day, seven days a week, and so any organization of the work must be based on duty rostering.

This work deals with the problem of scheduling physicians in the departments of a health care structure (e.g., hospital), where a schedule is an assignment of physicians to perform different *medical guard services* in those departments. Scheduling health care personnel is in general particularly challenging. However, the quality of these schedules is very important, because high performances in health care services could hardly be achieved when using low quality rosters, e.g., rosters causing unbalanced workloads or excessive stress in the workers. Performing periods of medical guard service in the hospitals is part of the physicians charges. Those periods usually last 10/12 hours, hence their good scheduling, taking into account suitable rest periods and any other additional requirement, is essential in improving the overall efficiency of the system.

The problem of personnel scheduling in health care has been addressed by several authors. However, most of the work has been focused on nurse scheduling (see, e.g., [7, 9, 16, 19]). The physician scheduling problem is different from nurse scheduling, since, in general, nurses work under collective agreement while physicians may have individual and *ad hoc* contractual duties. Another difference is in the objectives. In the nurse staff problem, both preference satisfaction and costs aspects must be considered. In the physician scheduling, only the satisfaction of the physician is typically taken into account, as physician retention is another issue often faced by hospital administrations. As a consequence, nurse scheduling is usually performed in two steps [8]. First, a schedule is generated to satisfy the collective agreement for the regular staff, minimizing staff shortages and surplus. Then, shortages are solved by using external personnel and overtime. The physician schedule is usually a one step process, where scheduling requirements must be joined with physician availability.

Most of works from the literature on physician scheduling deal with the problem of scheduling physicians in particular contexts, such as emergency rooms [1, 8, 20, 15] or operating rooms [21, 24]. Recently, in [17], the problem of generating a master schedule for the physicians in a hospital is addressed, by considering the full range of day-to-day activities of the physicians (including surgery, clinics, scopes, calls, administration). However, for large instances, this overall problem could not be solved exactly but only heuristically. In [23], a set covering model for scheduling physicians in a hospital is presented and tested on real-world instances, but only with a 2-week planning horizon. The same instances are solved heuristically in [6]. Despite the differences from nurse and physician scheduling, the mathematical formulations and the solution approaches used for solving the two problems may have points in common. Indeed, the solution approach employed in [2] for the nurse scheduling problem was slightly modified and successfully applied to the physician scheduling problem in [1].

In this work, a Mixed Integer Linear Programming formulation for the physician scheduling problem is presented. This model imposes satisfying all service requirements and contractual agreements, while trying to respect, as far as possible, workers' preferences. The proposed model is able to represent the various aspects of the problem generally considered in literature, and is also flexible, allowing easy modifications for representing different situations. This model can be solved to optimality by means of a standard Branch&Cut procedure for planning horizons even beyond the standard needs. The paper is organized as follows. In the rest of this Section, a brief discussion of the problem complexity is given. In Section 2, the general structure of the problem is formally defined. Section 3 describes the formulation proposed for the general problem, while Section 4 analyzes the capability of this model to represent the different aspects of the problem. Section 5, finally, reports computational results and comparison on real world instances.

1.1. Complexity Issues

Since the personnel scheduling problem generally is an assignment of shifts to physicians, in its very basic version it could be modeled as a standard assignment problem, and consequently solved in polynomial time [14]. ¿From the literature, easy cases of personnel scheduling have been modeled as *network flow* models and solved in polynomial time (e.g. [3, 27]). However, as soon as the problem gains more richness, more complex models are needed, as, for example, the *multicommodity* network flow models used in the case of railways [25], that are NP-hard [12].

In the case of physician scheduling, a number of additional requirements exists, that make the problem particularly difficult. One of the main complicating requirement is the fact that physicians need to have rest periods after any shift performed. Another complicating requirement is assigning shifts to doctors in order to balance the workload as much as possible. Those requirements imply using additional mathematical machinery for somehow "counting" the number of shifts. Indeed, Brucker et al. [4] show that the physician scheduling problem is NP-hard when the planning horizon is composed by $3n$ shifts, where n is the total number of physicians, and 3 not consecutive shifts must be assigned to each physician (by reduction from the *exact covering by 3-sets* (X3C) problem [14]). Recall that the problem *3-Satisfiability* (3SAT) reduces to X3C [14], and that the problem qSAT remains NP-complete for any $q \geq 3$. From this, it follows that *exact covering by p-sets* for $p \geq 3$ is NP-complete, too. Therefore, the physician scheduling problem is NP-hard also when the planning horizon is composed by pn shifts and p not consecutive shifts must be assigned to each physician, with $p \geq 3$.

Evidently, the physician scheduling problem has a *combinatorial optimization* structure, with the *ground set* being the set of every possible attribution of shifts to doctors, and each feasible solution is only a *subset* of this ground set producing a certain value for the selected objective (which can be, for example, a measure of the workload balancedness).

2. Definition and General Problem Structure

In this section, a formal definition of the Physician Scheduling problem is given. We consider the scheduling of physicians to perform different medical guard services in the departments of health care structures such as hospitals. Physicians, hereinafter called also *doctors*, have to guarantee their presence in a number of *departments*, so as to be ready to assist patients and solve the different issues that may arise. This means doctors must cover a number of time slots, or *shifts*, in those departments, according to a number of *rules*. Such rules may be different from case to case, but several general aspects are common to all cases. Basically, there are rules for ensuring *coverage* of the departments, sufficient *rest*, vacation time or other unavailabilities of the doctors, and workload *balancing* among the doctors. Balancing generally constitutes the most important preference aspect in practice, and indeed other preference aspects were not allowed in the analyzed real world case for avoiding disputes. Additional preferences could of course be allowed in specific cases, constituting additional elements in this structure. Therefore, a mathematical model of the problem should provide enough flexibility for introducing them. We now formalize the problem structure, giving also some concrete examples just for helping comprehension. Clearly, there are cases that may considerably differer from such examples, but still fall within the same general structure.

Shifts. The scheduling is generally planned by using a certain *planning horizon*, for instance 6 months or 1 year, and all the shifts in the planning horizon constitute a set $S = \{s_1, \ldots, s_m\}$. This horizon cannot be too short, since doctors need to know their duties with some advance, but not exceedingly long, because otherwise re-organizations, changes in the personnel, etc. would often nullify the rostering. Generally, we need to distinguish among:

1. Shifts located on regular working days, constituting a set $S_d \subset S$. They are for instance shifts lasting the 12 hours from 8 a.m. to 8 p.m. of the week days from Monday to Saturday.
2. Shifts located on regular nights, constituting a set $S_n \subset S$. They are for instance shifts lasting the 12 hours from 8 p.m. to 8 a.m. following the week days from Monday to Saturday (e.g. Monday night, in common terminology, is the night between Monday and Tuesday, even if Monday officially starts at 0.00 am of the night between Sunday and Monday).
3. Shifts located on Sundays, constituting a set $S_s \subset S$. They are for instance the 12 hours day shift and the 12 hours night shift located on Sundays.
4. Shifts located on holidays, constituting a set $S_{ho} \subset S$. They are the 12 hours day or night shifts that happen to be on a major holiday (Christmas, New Year's day, etc.).
5. Shifts located in weekends, constituting a set $S_{we} \subset S$. They are the 12 hours day or night shifts that happen to be on Saturdays or on Sundays.

Clearly, $S_d \cap S_n = \emptyset$, $S_d \cap S_s = \emptyset$, $S_n \cap S_s = \emptyset$, and $S_d \cup S_n \cup S_s = S$. On the contrary, S_{we} and S_{ho} may intersect the above S_d, S_n, S_s, and are needed for evaluating the weight of the workload of each doctor. Some type of shifts (nights, weekends, ...) are considered by the physicians unpleasant and more weighty than ordinary. To represent this, we define *inconvenient shifts* those belonging to S_n, S_{we}, S_{ho}. We denote by T the set of the types of shifts $\{d, n, s, ho, we\}$, so that the generic type of shift can be expressed as S_t with $t \in T$.

Coverage. Let the set of the different departments be $D = \{d_1, \ldots, d_q\}$; let the set of the different doctors be G. Each department d_h must be covered by a number c_{1h} of doctors (usually one, may depend on the department) during each regular day shift in S_d, while it has a number (generally $>> c_{1h}$) of doctors that can be on duty there during the day. Each doctor, on the basis of his/her competence, is able to cover only a specific department. In general, we may be interested in further distinguishing among the doctors that can cover department d_h: for example they may have different contractual duties even if similar competences. Therfore, we partition the set G of all doctors into *groups* $G_1, \ldots, G_t$ so that group G_k is homogeneous with regard to competence and contractual duties, and we have $G = G_1 \cup \ldots \cup G_t$ and $G_k \cap G_{k'} = \phi$ for $k \neq k'$. For each department d_h we have a set H_h of the indices of the groups of doctors that can be on duty there during the day. For example, if department d_1 can be covered by groups G_1 and G_2, we have $H_1 = \{1, 2\}$. Hence, the relationship between departments and groups is such that

$$H_1 \cup \cdots \cup H_q = \{1, \ldots, t\}.$$

We denote the elements of G_1 by the numbers $\{1, \ldots, n_1\}$, those of G_2 by the numbers $\{n_1 + 1, \ldots, n_1 + n_2\}$, so that in general

$$G_k = \{\sum_{p=1}^{k-1} n_p + 1, \ldots, \sum_{p=1}^{k} n_p\}.$$

The total number of doctors is $n = n_1 + \cdots + n_t$. During nights, sundays and holidays (S_n, S_s, S_{ho}, the above defined inconvenient shifts), the requirement for a presence is relaxed, and usually the whole set of departments is covered by only a number c_2 of doctors (for example just one). Clearly, the competence requirements must be relaxed as well.

Rest or unavailability. When assigning shifts to doctors, some contractual rules regulating rest periods, annual leave, exemptions must be respected. Those rules may vary from case to case, but should at least respect the applicable laws, see also the European Working Time Directive 2003/88/EC [11].

1. After one shift (for instance 12 hours) on duty, a doctor has to rest for at least a number r_1 of shifts (for instance 4 shifts, 48 hours, or sometimes even less, depending on the specific case). This very basic rest requirement is common to all cases of the problem.

2. After a shift located in a weekend, a doctor should not receive another shift located in a weekend for at least a number r_2 of shifts (for instance 28 shifts, 14 days; in other words, anybody doing a weekend shift will have no weekend duty shifts in two weeks, but may have other shifts on regular working days).
3. Every year a number of days for annual leave must be granted. Usually they are about 30 days in an year, depending also on contracts and past usage, and generally at least 15 of these days can be taken as a consecutive holiday during the summer period. The exact periods are generally negotiated in advance by the doctors, and are assigned so as not to leave the departments completely uncovered.
4. Some categories of doctors may have contracts that do not oblige them to take some types of inconvenient shifts. For instance, some categories cannot have nights assigned, or cannot have Sundays.
5. Some doctors may have acceptable reasons for being considered unavailable for some type of shifts. For example, they may have priority administrative or teaching duties, or religious obligations. Such unavailability may be either permanent or limited to a period of time.
6. Finally, in some cases, there are categories of doctors (e.g. senior ones) that may perform shifts (or specific types of shifts), but only until a maximum admissible workload.

Balancing. Even if all shifts are covered and all the above requirements are respected, a physician scheduling is not acceptable if the amount of work is not fairly divided among all the doctors. This is a basic workers' preference condition, and generally implies two orders of requirements.

First, a requirement based on sheer numbers: as far as possible, all doctors in a group should receive the same number of shifts to do along all the planning horizon. Clearly, this constitutes for the doctors a main parameter in evaluating the quality of a physician scheduling

Second, the total "weight" of all the assignments received by each doctor in a group along all the planning horizon should be comparable. This means that the number of nights and shifts during weekends or holidays (inconvenient shifts) should be, as far as possible, *similar* for all doctors of the group. Clearly, this can be difficult to achieve, and there are many real-world situations in which one can prove that the number of nights, weekend and holiday shifts cannot be exactly the same, even for doctors of the same group. Nonetheless, this constitutes for the doctors another main parameter in evaluating the quality of the scheduling.

Following this general equity principles, several specific criteria for fairly dividing the workload among doctors can be devised, as explained in Section 3.

3. A Flexible Formulation of the Problem

We describe here the proposed Mixed Integer Linear formulation for the physician scheduling problem described in Section 2. This model includes all the constraints described in the following Subsection 3.1 for satisfying all service requirements and contractual agreements, while trying to respect, as far as possible, workers' preferences about workload balancing, as described in Subsection 3.2. A way of taking into account additional preference aspects, when they should be considered, is described in Subsection 3.3.

3.1. Satifying Hard Requirements

The following set of binary decision variables is used for assigning duty shifts to doctors.

$$x_{ij} = \begin{cases} 1 & \text{if doctor } i \text{ covers duty shift } j, \\ & \text{with } i \in G, j \in S \\ 0 & \text{otherwise} \end{cases}$$

By using such x_{ij} variables we can express all the above introduced rules as linear constraints. Coverage of all the day shifts, for department d_h, becomes the following set of constraints.

$$\sum_{i \in G_k : k \in H_h} x_{ij} = c_{1h} \quad \forall j \in S_d \tag{1}$$

Clearly, similar sets of constraints are needed for each department, i.e., for all $h = 1, \dots, q$. On the other hand, coverage of the inconvenient shifts (nights, sundays and holidays), for all departments, becomes the following set of constraints.

$$\sum_{i \in G} x_{ij} = c_2 \quad \forall j \in \{S_n \cup S_s \cup S_{ho}\} \tag{2}$$

Define the set $R(j, \rho_1) = \{r \in S \mid r = j, j+1, \dots, j+\rho_1\}$. Now, resting after each shift, for doctor i, becomes the following set of constraints.

$$\sum_{r \in R(j,\rho_1)} x_{ir} \le 1 \quad \forall j \in S \tag{3}$$

For instance, the very basic legal requirement of having at least 11 hours of rest every 24 hours [11] can be obtained for any $\rho_1 \ge 1$. Also, the requirement of a minimum weekly rest period of 24 uninterrupted hours every each 7-day period [11] can be obtained for example with $\rho_1 \ge 3$ (working 1 shift in a series of 4 guarantees at least 2 consecutive rest shifts).

By defining the set $W(j, \rho_2) = \{R(j, \rho_2) \cap S_{we}\}$, forbidding too near weekend shifts after each weekend shift, for doctor i, becomes the following set of constraints, where ρ_2 should be large enough to cover the desired number of weekend (e.g. $\rho_2 \ge 14$).

$$\sum_{r \in W(j,\rho_2)} x_{ir} \le 1 \quad \forall j \in S_{we} \tag{4}$$

The sets of rest constraints (3) and (4) are needed for each doctor, i.e., for all $i \in G$. Note that, in case shifts having different lengths are in use, we can compute different ρ_1 and ρ_2 for the different types of shifts. More precisely, we can associate a parameter $\rho_1(j)$ to each duty shift j in order to respect, for any different length of shift j and of the following shifts $j+1,\ldots,j+\rho_1(j)$, the granted number of hours after shift j. Similarly, we can associate a $\rho_2(j)$ to each duty shift j in order to respect in any case the granted number of free weekends after shift j. This computation can clearly be done offline.

Moreover, we define for the generic doctor i his or her unavailability set U_i, i.e. the set of shifts that cannot be covered by doctor i. This includes shifts belonging to the vacation time granted to doctor i (e.g., annual leaves), shifts during which i has other priority duties, shift types that are not contractually assignable to i, etc.

$$x_{ij} = 0 \quad \forall i \in G,\ \forall j \in U_i \tag{5}$$

Note that computation of the sets $\{U_i\}$ is done offline, taking into account also whether the unavailability is permanent or limited. The unavailability specification can also constitute a way of expressing a (hard) preference for a specific set of shifts, when this is allowed, by forbidding the complement of that set. Since sets $\{U_i\}$ contain the information about annual leave, a vacancies request that would leave a department uncovered would be detected because the constraints of type (5) would make model (11) infeasible. In this case, a post-infeasibility analysis is required, see e.g. [5] for a possible approach to the problem and references on the subject.

Finally, if the specific doctor i has a limit on the maximum admissible workload, for example can do not more than a_t shifts of type t (i.e. belonging to the set S_t), this can be imposed with:

$$\sum_{j \in S_t} x_{ij} \leq a_t \tag{6}$$

These constraints should be considered for the specific $i \in G$ entitled to have this limit, and for the suitable $t \in T$. Note that, when no limits on the maximum admissible workload can be granted, doctors must cover shifts until all service requirements are satisfied, regardless to their total workload ("goal-oriented" point of view). On the contrary, when all doctors can specify a maximum admissible workload, this can make the problem infeasible ("worker-oriented" point of view). Clearly, when only some doctors can specify their maximum admissible workload, this may produce a heavier goal-oriented management of the others, so this opportunity should be used with care.

3.2. Pursuing Workload Balance

For balancing the overall workload, we use an upper bound to the number of shifts assigned to each doctor. Note that, since the cardinalities of the groups G_k may be different while the number of day shifts in S_h is the same for all groups, the required workload per doctor may be different from group to group. Therefore, we use a different real-valued nonnegative decision variable for each group:

$y_k \geq 0$ maximum number of shifts assigned to doctors in G_k

The workload balancing for the doctors of group G_k becomes the following set of constraints.

$$\sum_{j \in S} x_{ij} \leq y_k \quad \forall i \in G_k \tag{7}$$

Clearly, similar sets of constraints are needed for each group, i.e., for all $k = 1, \ldots, t$.

For balancing the workload of inconvenient shifts (nights, weekends, major holidays), there are essentially two alternatives. One alternative is upper bounding the number of the shifts assigned to each doctor differently for each single category of inconvenient shifts, therefore introducing, for the number of nights, the number of weekend shifts and the number of holiday shifts, different sets of variables similar to the y ones.

The second alternative is weighting the hardness of each type of inconvenient shift, and upper bounding only the total weight of the shifts that each doctor has to do in inconvenient shifts. The hardness of each type of inconvenient shift can be evaluated by using the following weights:

- w_n for regular week nights,
- w_{sa} for regular Saturday days,
- w_{sa_n} for regular Saturday nights,
- w_{su} for regular Sunday days,
- w_{su_n} for regular Sunday nights,
- w_{ho} for holidays (or also individual weights for each holiday).

The second alternative requires that a fair division of workload can be obtained not only by assigning the same number of each type of inconvenient shift to the doctors, but also by assigning, for instance, 10 nights and 3 Sundays to doctor a and 5 nights and 5 Sundays to doctor b, if the sum of the corresponding weights is the same. Note that the choice on how to model this balancing condition may appear rather trivial from the mathematical point of view, but has considerable impact on the life organization of the doctors.

The second alternative has several advantages on the first one. First, it allows more feasible solutions, because it is not limited to solutions having, for each type of inconvenient shift, the same number for all doctors, but accepts also solutions having different numbers of them but such that the sum of the weights of the inconvenient shifts assigned to each doctor is the same.

Second, it allows more flexibility, since different weights for the different holidays could be specified, and even *personalized* weights, if needed, could be specified: each doctor could give his/her weight values to each type of inconvenient shift (somebody could prefer doing nights instead of weekends, somebody the opposite, etc.). This would also be useful in dealing with the management of senior doctors.

Third, it can also be used for assigning, for each different type of inconvenient shift, the same number for doctors of the same group (as it is for the first alternative) by using *incomparable* weights for the different types of inconvenient shifts, i.e. such that the sum of any possible number of weights of one type can never reach the weight value of another type. For instance, if the total number of shifts is $s = 730$ (one year planning horizon with 2 shifts per day), when $w_n = 1$ and $w_{sa} = 1000$, the sum of any possible number of nights would never weight as much as one Saturday day, so they could never compensate each other and therefore the number of nights and the number of Saturday days should be the same for all doctors in the same group.

Fourth, it requires less variables in the model, with consequent computational advantages.

Therefore, we use another real-valued nonnegative decision variable for each group:

$z_k \geq 0$ maximum weight in inconvenient shifts assigned to doctors in G_k

The inconvenient shifts balancing for the doctors of group G_k becomes the following set of constraints, where w_j denotes the element, among the above defined weights, corresponding to the inconvenient shift j.

$$\sum_{j \in \{S_n \cup S_{we} \cup S_{ho}\}} w_j x_{ij} \leq z_k \quad \forall i \in G_k \tag{8}$$

Clearly, similar sets of constraints are needed for each group, i.e., for all $k = 1, \ldots, t$.

In order to make the balancing constraints work, we need to minimize the above maximum values y_k and z_k, obtaining a bi-objective structure with the following two objectives:

$$\min \sum_{k=1}^{t} y_k \tag{9}$$

$$\min \sum_{k=1}^{t} z_k \tag{10}$$

Using objective (9) corresponds to assigning the same number of shifts to the doctors of the same group; using objective (10) to assigning the same workload in inconvenient shifts to the doctors of the same group. Note that, when groups with different cardinalities correspond to the same department d_h, a Δ increase in the variables y_l and z_l upper bounding the larger group G_l has more coverage power (i.e., covers more shifts) than the same Δ increase in the variables y_s and z_s upper bounding the smaller group G_s. Therefore, an increase in y_l and in z_l would be preferred to an increase in y_s and in z_s during the search for the optimal solution. In case this behaviour wants to be avoided, the upper bounding variables $\{y_k\}$ and $\{z_k\}$ should receive weigths $\{n_k\}$ (the cardinalities of the groups $\{G_k\}$), so that an increase in y_l and in z_l would not be preferred anymore.

This bi-objective can be scalarized by giving weights b_1 and b_2 to the two objectives, or by using one of them as constraint and the other as unique objective. By choosing the first option, the overall Mixed Integer Linear model is the following, where we have, in order: objective function given by (9) and (10); coverage constraints (1) and (2); rest constraints (3) and (4); annual leave/unavailability constraints (5); maximum workload constraints (6); workload balance constraints (7) and (8).

$$\left\{\begin{array}{lll}
\min & b_1 \sum_{k=1}^{t} y_k + b_2 \sum_{k=1}^{t} z_k & \\
s.t. & \sum_{i \in G_k : k \in H_h} x_{ij} = c_{1h} & \forall j \in S_d,\ \forall h \in \{1,\dots,q\} \\
 & \sum_{i \in G} x_{ij} = c_2 & \forall j \in \{S_n \cup S_s \cup S_{ho}\} \\
 & \sum_{r \in R(j,\rho_1)} x_{ir} \leq 1 & \forall j \in S,\ \forall i \in G \\
 & \sum_{r \in W(j,\rho_2)} x_{ir} \leq 1 & \forall j \in S_{we},\ \forall i \in G \\
 & x_{ij} = 0 & \forall i \in G,\ \forall j \in U_i \\
 & \sum_{j \in S_t} x_{ij} \leq a_t & i \in G, t \in T \\
 & \sum_{j \in S} x_{ij} \leq y_k & \forall i \in G_k,\ \forall k \in \{1,\dots,t\} \\
 & \sum_{j \in \{S_n \cup S_{we} \cup S_{ho}\}} w_j x_{ij} \leq z_k & \forall i \in G_k,\ \forall k \in \{1,\dots,t\} \\
 & x_{ij} \in \{0,1\} & \forall i \in G,\ \forall j \in S \\
 & y_k \in \mathbb{R}_+ & \forall k \in \{1,\dots,t\} \\
 & z_k \in \mathbb{R}_+ & \forall k \in \{1,\dots,t\}
\end{array}\right. \tag{11}$$

Note that this model has no specific structure except for linearity, and this generality allows for easy modifications in the cases when additional requirements (e.g. different workers' preferences) should be taken into account.

3.3. Additional Preferences

We remarked that, in many practical cases, additional preferences are not allowed, for avoiding disputes. Nonetheless, there may be cases when they should be considered, and we assume them expressed by means of the following *preference values* for each $i \in G$ and $j \in S$, with 1 meaning maximum preference.

$$p_{ij} \in [0,1] \quad \text{preference of doctor } i \text{ for shift } j$$

A first option in these cases would be considering an additional term in the objective of (11) with the following structure

$$\sum_{i \in G} \sum_{j \in S} (1 - p_{ij}) x_{ij}$$

Clearly, the smaller its value, the more the shifts assignment corresponding to the values of the x_{ij} variables satisfies the higher preference values (the number of x_{ij} variables at 1 in any feasible solution is the same, but each of them is multiplied by a value $(1 - p_{ij})$ representing the "unsatisfaction" that each 1 causes). However, this structure would consider only the aggregate preference satisfaction, and the preferences of the different doctors may be satisfied at very different levels. In other words, we would respect as much as possible the overall preferences, but there may be some doctors whose preferences are respected at a very low degree.

For balancing also the amount of the preferences that are respected, we use the following set of real-valued nonnegative decision variables.

$v_k \geq 0$ maximum preference unsatisfaction for the doctors in G_k

The unsatisfaction level of doctor $i \in G_k$ would be $\sum_{j \in S}(1 - p_{ij}) x_{ij}$. Therefore, for each group, i.e., for all $k \in \{1, \ldots, t\}$, we need the following set of constraints

$$\sum_{j \in S} (1 - p_{ij}) x_{ij} \leq v_k \quad \forall i \in G_k \tag{12}$$

and the term added in the objective function of (11) becomes

$$b_3 \sum_{k=1}^{q} v_k. \tag{13}$$

where parameter b_3 should be carefully set in order to weight the relative importance of the preference satisfaction compared to that of balancing both shifts and the workload in inconvenient shifts (first two terms of the objective in (11)).

4. Comparison to Other Models from the Literature

In this section, we analyze the capability of our model to represent practical requirements in comparison with that of the other approaches proposed in the literature. Gendreau et al. [15] distinguish the constraints of the physician scheduling problem (for emergency rooms management) on the basis of their logical meaning, and present a quite exhaustive classification of them by using the following four categories:

- supply and demand constraints,
- workload constraints,

- fairness constraints,
- ergonomic constraints.

The supply and demand constraints deal with the availabilities of the physicians and the requirements of the emergency rooms. The workload constraints deal with the workload (number of hours or number of shifts) that is assigned to physicians during a given planning period. These constraints impose that each physician should be assigned an amount of work that lies within a specified interval, and they also limit the number of shifts of the same type. The fairness constraints control the distribution of different kinds of shifts during the whole planning period, assuring in particular that the inconvenient shifts are fairly distributed. The ergonomic constraints, finally, model rules ensuring a certain level of quality for the schedules produced, such as controlling the number of consecutive night shifts, assigning suitable rest periods after each duty shift, allowing some free weekends after a duty shift on a weekend, etc. The main contribution of this work is the above analysis, and no mathematical model is explicitly presented. Physicians' preferences are also briefly analyzed.

A classification of constraints in categories is proposed also by Beaulieu et al. [1]. The considered classes are similar to the previous ones, but they also distinguish the constraints into compulsory rules (e.g., rules that must absolutely be enforced) and flexible rules (e.g., rules that can be violated). Using the class names given above [15], the supply and demand constraints, the workload constraints, as well as the ergonomic constraints arising from contractual agreements (e.g., regulating rest periods, annual leave, vacations, exemptions) are classified as compulsory rules. The flexible rules include the ergonomic constraints which aim at improving the quality of the schedule and the fairness constraints. In this case, physicians' preferences are not explicitly considered.

Following a different approach, Gunawan and Lau [17] propose a number of mathematical models, mainly focused on the physicians' preferences satisfaction: the goal is to generate a schedule taking into account the ideal schedule of each physician. The above works could be considered to represent collectively all the aspects of the physician scheduling problem: the constraints proposed in the other works on this subject do not move away from this scheme and often cover only a subset of it.

The model proposed in our work is indeed able to represent most of the typologies of constraints and rules introduced above, including issues on the physicians' preferences. Supply and demand constraints correspond to the coverage requirements reported in Section 2 and are modeled by constraints (1) and (2). Observe, in this regard, that our model is able to take into account the requirements of different departments at the same time, while the mathematical models presented in [1] and [15] can not be used when the requirements of more than one department must be considered. The workload constraints are modeled by means of conditions (6). The compulsory ergonomic constraints are expressed by conditions (3), (4) and (5).

All the flexible rules are properly modeled by considering the bounding variables (y_k, z_k, v_k) and suitable terms in our objective. More precisely, the fairly distribution of the regular and inconvenient shifts are controlled by the constraints (7) and (8) and by the two terms $b_1 \sum_{k=1}^{q} y_k$ and $b_2 \sum_{k=1}^{q} z_k$ of the objective function. Finally, physicians' preferences are taken into account by adding the extra term (13) in the objective function. Observe that, the importance of physicians' preferences with respect to the other objectives can be tuned by properly setting parameters b_1, b_2 and b_3. Note also that the model proposed in our work allows more flexibility than other models. For example, the amount of rest after work may be changed by simply modifying constants ρ_1 and ρ_2 in constraints (3) and (4) , while in other models the entire structure of the constraints would need to be changed.

5. Performance Analysis on a Real-World Case Study

The described formulation has been particularized to the case of the Hematology Center of the "Umberto I" University Hospital in Rome, Italy, that is part of the Department of Cellular Biotechnologies and Hematology of "Sapienza" University of Rome. In this real context, four groups of doctors exist, composed of 24, 1, 5 and 2 units, i.e., $G_1 = \{1, \dots, 24\}$, $G_2 = \{25\}$, $G_3 = \{26, \dots, 30\}$ and $G_4 = \{31, 32\}$ and two departments d_1 and d_2. Night shifts are not contractually assignable to doctors in G_2 and G_4. Group G_2 covers the same department than group G_1, so the two should be considered together for the daily covering constraints. The same happens for groups G_3 and G_4. Hence, $H_1 = \{1, 2\}$ and $H_2 = \{3, 4\}$. The parameter values are: $\rho_1 = 4$, $\rho_2 = 28$, $c_{11} = c_{12} = 1$, $c_2 = 1$. The management of the above Hematology Center use to compute the schedule by hand, by applying a two-phase procedure roughly described in the following.

During the first phase, the shifts of the set S_d (containing regular working days only) are assigned to physicians of each set G_k in a *round robin* way (following, for example, alphabetic order). In particular, for each department d_h, the schedule is built by trying to assign each shift to exactly c_{1h} physicians of G_k, with $k \in H_h$, and by assuring that a rest period of ρ_1 shifts exists between any two shifts assigned to the same physician.

During the second phase, the shifts in the sets S_n, S_{we}, S_{ho} (i.e., the inconvenient shifts) are considered and assigned to all the groups of physician that are charged with those kinds of shifts, in order to respect all the rest of the constraints. This should be obtained by initially assigning them using another round robin procedure, assuring now that a rest period of ρ_2 shifts exists between any two shifts in S_{we} assigned to the same physician. After this, exchanges and new insertions are performed, in order to yield a solution respecting all the constraints on resting and being as balanced as possible with respect both to regular and inconvenient shifts. Clearly, the criteria for doing this last exchange and insertion operations are not algorithmically expressed, and can be also quite subjective.

On the other hand, the proposed model (11) has been implemented by using AMPL modeling language [13] and solved by means of ILOG Cplex 11.2 solver

	n	m	k	y	$\sigma(y)$	z	$\sigma(z)$
By hand (~ 4 hours)	32	240	4	$\begin{pmatrix} 11 \\ 5 \\ 20 \\ 16 \end{pmatrix}$	$\begin{pmatrix} 0.71 \\ 0.00 \\ 0.93 \\ 0.81 \end{pmatrix}$	$\begin{pmatrix} 25 \\ 7 \\ 25 \\ 14 \end{pmatrix}$	$\begin{pmatrix} 4.07 \\ 0.00 \\ 2.42 \\ 12.83 \end{pmatrix}$
Solving model (11) (2 sec.)	32	240	4	$\begin{pmatrix} 9 \\ 8 \\ 17 \\ 17 \end{pmatrix}$	$\begin{pmatrix} 0.38 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix}$	$\begin{pmatrix} 23 \\ 15 \\ 23 \\ 23 \end{pmatrix}$	$\begin{pmatrix} 2.40 \\ 0.00 \\ 2.11 \\ 0.0 \end{pmatrix}$

Table 1: Comparison with the current solution approach of the case study

[18] running on an Intel Core 2 processor PC at 2.4 GHz with 2Gb of RAM. The results were compared to those of the described *by hand* approach. The planning horizon for this comparison was necessarily short (4 months), because this is generally the maximum size that the hospital management can solve by hand with reasonable effort.

Table 1 contains the aggregate result of this comparison. In particular, it reports a description of the instance: number of doctors n; number of shifts in the planning horizon m; number of groups k; followed by an analysis of the workload balancing of the schedules obtained with the two approaches. In order to do this, we a consider the value of the y variables for each group (maximum number of shifts in the group); an evaluation, for each group, of the imbalance in the number of shifts, denoted as $\sigma(y)$ and explained below; the value of the z variables for each group (maximum weight in inconvenient shifts in the group); an evaluation, for each group, of the imbalance in the weight of inconvenient shifts, denoted $\sigma(z)$ and explained below.

For evaluating the $k-$th group imbalance in the number of shifts, we use the *standard deviation* $\sigma_k(y)$ of the number of shifts for group k. By denoting with t_i the number of shifts assigned to each doctor i of group k and with θ_k the average of such values, we have:

$$\sigma_k(y) = \sqrt{\frac{1}{n_k}\sum_{i=1}^{n_k}(t_i - \theta_k)^2}$$

Components $\sigma_k(y)$ for $k = 1, \ldots, q$ constitute the vector $\sigma(y)$. Values $\sigma_k(z)$ are computed analogously by considering the weight in inconvenient shifts assigned to each doctor of group k, and constitute the vector $\sigma(z)$. The weight in inconvenient shifts is computed by using weights $w_n = 3$, $w_{sa} = 2$, $w_{sa\_n} = 4$,

$w_{su} = 5$, $w_{su\_n} = 4$, $w_{ho} = 6$. Those values were determined by gathering the opinions of the physicians of the mentioned structure.

This analysis shows that the solution obtained by solving model (11) is more balanced ($\sigma(y)$ and $\sigma(z)$ have smaller values). Consequently, the maximum number of shifts and the maximum weight in inconvenient shifts, for each group, is lower for all the groups that could be fully utilized with the *by hand* approach. On the contrary, for groups that could not be fully utilized with the *by hand* approach because of their limitations in inconvenient shifts (e.g. G_2), the above values are higher. The two above aspects are therefore both in favor of the solution obtained by solving model (11). Note, moreover, that, by solving model (11), one can easily test increasing values of ρ_1 and ρ_2, exploring so the overall feasibility of more comfortable shift distributions. Clearly, there is no match in the analysis of the solution times required by the two approaches. Hence, the solution obtained by solving model (11) appear superior under every respect.

We also report more detailed information on the two solutions. Tables 2 and 3 give, for each doctor i, the number of shifts belonging to S_d (# regular); the number of shifts belonging to S_n (# night); the number of day shifts located in weekends (# we day); the number of night shifts located in weekends (# we night); the number of shifts located on holidays (# holiday); the total number of shifts (# tot). Observe that the solution by hand ranges from 6 to 11 shifts and from 18 to 20 shifts for each doctor in G_1 and G_3 respectively, while the solution found by model (11) ranges from 8 to 9 shifts and exactly 17 shifts for each doctor in G_1 and G_3 respectively, confirming what observed for Table 1.

Note, finally, that a perfect workload balancing could be unobtainable when considering short planning horizons (just as an example, when there is a small number of holiday shifts in the horizon, it could be impossible to assign them to each doctor). On the contrary, the longer the planning horizon, the better workload balancing can be achieved. For this reason, we consider other 4 instances of the described real-world problem (called Inst{1,2,3,4}), obtained by using, in the above described case, a planning horizon of respectively 6, 12, 18, 24 months. However, planning horizon longer than 1 year are considered mainly for computational reasons.

For every instance, Table 4 reports again practical aspects of the instance and an analysis of the workload balancing of the schedule obtained by solving that instance. This analysis shows that the workload balancing is in many cases perfect, since all members of the group have exactly the same workload (the standard deviation is 0), and in the other cases is anyway satisfactory. Note that the workload of the members of a group can be very different from that of another group, but balancing among different groups was not required and could be, in general, not perfectly obtainable, as observed in Section 3. However, if some balancing among different groups should also be taken into account, easy modifications of the proposed model, for example introducing additional or alternative upper bounding variables similar to the mentioned y and z, could provide this additional functionality.

Table 5, on the other hand, reports, for the same instances, some computational analysis: the overall number of simplex iterations needed to solve

i	# regular	# night	# we day	# we night	# holiday	tot
1	2	3	1	1	1	8
2	4	4	1	0	0	9
3	4	4	2	0	1	11
4	3	2	2	1	1	9
5	3	2	2	2	0	9
6	4	3	0	2	1	10
7	4	3	1	0	1	9
8	3	4	2	0	0	9
9	4	4	0	1	0	9
10	4	3	1	2	0	10
11	3	2	1	2	0	8
12	3	2	1	0	0	6
13	4	3	0	1	0	8
14	4	3	0	2	0	9
15	2	2	1	1	1	7
16	3	3	1	1	1	9
17	4	3	1	0	0	8
18	3	3	2	1	0	9
19	3	2	2	1	0	8
20	4	3	0	1	0	8
21	3	4	1	1	0	9
22	2	3	1	2	0	8
23	3	3	0	2	0	8
24	4	4	0	1	0	9
25	3	0	2	0	0	5
26	11	2	2	2	1	18
27	12	2	3	1	1	19
28	13	2	2	1	1	19
29	12	2	4	2	0	20
30	11	3	4	2	0	20
31	12	0	4	0	0	16
32	12	0	2	0	0	14

Table 2: Analysis of the solution obtained by hand

i	# regular	# night	# we day	# we night	# holiday	tot
1	3	3	1	1	1	9
2	3	3	1	1	0	8
3	3	3	1	1	1	9
4	3	3	1	1	0	8
5	3	3	1	2	1	9
6	2	4	2	0	0	8
7	3	3	2	1	0	9
8	4	3	1	1	0	9
9	4	2	1	1	1	9
10	4	2	1	2	0	9
11	3	3	1	1	1	9
12	4	2	1	2	0	9
13	3	3	2	1	0	9
14	3	3	2	1	0	9
15	4	3	1	1	0	9
16	3	4	1	0	1	9
17	3	4	2	0	0	9
18	3	3	1	1	1	9
19	3	3	1	1	1	9
20	4	3	1	1	0	9
21	3	3	1	1	1	9
22	3	2	1	2	0	8
23	4	2	1	2	0	9
24	3	3	1	1	1	9
25	5	0	3	0	0	8
26	12	2	1	2	0	17
27	11	3	2	1	0	17
28	11	3	3	0	0	17
29	12	2	1	2	0	17
30	11	3	2	1	0	17
31	13	0	4	0	0	17
32	13	0	4	0	0	17

Table 3: Analysis of the solution solving model (11)

instance	n	m	k	y	$\sigma(y)$	z	$\sigma(z)$
Inst1	32	362 (6 months)	4	$\begin{pmatrix} 15 \\ 15 \\ 22 \\ 22 \end{pmatrix}$	$\begin{pmatrix} 0.81 \\ 0.00 \\ 0.49 \\ 0.50 \end{pmatrix}$	$\begin{pmatrix} 33 \\ 33 \\ 8 \\ 8 \end{pmatrix}$	$\begin{pmatrix} 0.37 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix}$
Inst2	32	730 (12 months)	4	$\begin{pmatrix} 30 \\ 29 \\ 44 \\ 44 \end{pmatrix}$	$\begin{pmatrix} 0.81 \\ 0.00 \\ 0.40 \\ 0.00 \end{pmatrix}$	$\begin{pmatrix} 67 \\ 64 \\ 16 \\ 16 \end{pmatrix}$	$\begin{pmatrix} 0.90 \\ 0.00 \\ 0.40 \\ 0.00 \end{pmatrix}$
Inst3	32	1092 (18 months)	4	$\begin{pmatrix} 44 \\ 44 \\ 65 \\ 65 \end{pmatrix}$	$\begin{pmatrix} 0.69 \\ 0.00 \\ 0.00 \\ 0.50 \end{pmatrix}$	$\begin{pmatrix} 100 \\ 100 \\ 22 \\ 22 \end{pmatrix}$	$\begin{pmatrix} 0.63 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix}$
Inst4	32	1460 (24 months)	4	$\begin{pmatrix} 59 \\ 59 \\ 87 \\ 87 \end{pmatrix}$	$\begin{pmatrix} 0.99 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix}$	$\begin{pmatrix} 133 \\ 133 \\ 30 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 0.44 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix}$

Table 4: Practical aspects and workload imbalance in the solution of the case study

instance	simplex iter.	br. nodes	cuts added	time
Inst1	20,710	380	1	39 s.
Inst2	31,556	539	45	150 s.
Inst3	15,203	200	3	162 s.
Inst4	158,216	1,968	50	1,009 s.

Table 5: Computational aspects of the solution of the case study

the linear relaxations; the total number of branching nodes enumerated by the branch-and-cut procedure; the total number of cuts added by the branch-and-cut procedure; running time in seconds. Observe that the proposed model allows to compute in reasonable times the solution for planning horizons that are even beyond the practical needs (usually no more than 1 year). Observe, moreover, that the number of branching nodes is in general quite small for the different instances, while the number of simplex iterations is large. Therefore, the proposed formulation (11) appears quite a strong one.

6. Conclusions

We formalized the general aspects of the physician scheduling problem and analyzed its computational complexity. Then, we proposed a Mixed Integer Linear Programming formulation for this problem. This model imposes satisfying all service requirements and contractual agreements (including rest periods and annual leave), while trying to respect, as far as possible, workers' preferences, with particular attention on workload balancing. The proposed model is able to represent the various aspects of the problem generally considered in previous literature on the subject. In any case, the generality of the proposed model allows for easy modifications in the cases when additional requirements should be taken into account. This model has been used for solving, on several planning horizons, a real-world case of physician scheduling in some of the departments of one of the biggest Italian university hospitals. Results have been analyzed both form the schedule quality and from the computational point of view, and demonstrate the effectiveness of the proposed optimization approach, also in comparison with the solution approach currently adopted in the health care structure considered as case-study.

References

[1] Beaulieu H, Ferland JA, Gendron B, Michelon P (2000) A mathematical programming approach for scheduling physicians in the emergency room. Health Care Management Science 3: 139–200

[2] Berrada I, Ferland JA, Michelon P (1996) A multi-objective approach to nurse scheduling with both hard and soft constraints. Socio-Economic Planning Sciences 30: 183–193

[3] Brucker P, Qu R (2012) Network flow models for intraday personnel scheduling problems. Annals of Operations Research, in press, doi: 10.1007/s10479-012-1234-y

[4] Brucker P, Qu R, Burke E K (2011) Personnel scheduling: Models and complexity. European Journal of Operational Research 210(3): 467–473

[5] Bruni R, Bianchi G (2012) A Formal Procedure for Finding Contradictions into a Set of Rules. Applied Mathematical Sciences 6(126): 6253-6271

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

[7] Burke E K, De Causmaecker P, Vanden Berghe G, Van Landeghem H (2004) The State of The Art of Nurse Rostering. Journal of Scheduling 7: 441–499

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

[9] Cheang B, HLi A Lim, Rodrigues, B (2003) Nurse Rostering problems - a bibliographic survey. European Journal of Operational Research 151: 447–460

[10] Ernst AT, Jiang H, Krishnamoorthy M, Owens B, Sier D (2004) An Annotated Bibliography of Personnel Scheduling and Rostering. Annals of Operations Research 127: 211–144

[11] Directive 2003/88/EC of the European Parliament and of the Council of 4 November 2003 concerning certain aspects of the organisation of working time.

[12] Even S, Itai A, Shamir A (1976) On the Complexity of Timetable and Multicommodity Flow Problems. SIAM Journal on Computing 5(4): 691–703

[13] Fourer R, Gay DM, Kernighan BW (2002) AMPL: A Modeling Language for Mathematical Programming. Duxbury Press

[14] Garey MR, Johnson DS (1979) Computers and Intractability: A Guide to the Theory of NP-completeness. WH Freeman, San Francisco

[15] Gendreau M, Ferland J, Gendron B, Hail N, Jaumard B, Lapierre S, Pesant G, Soriano P (2007) Physician scheduling in emergency rooms. In: PATAT06 Lecture Notes in Computer Science 3867: 53–66

[16] Glass C, Knight R (2010) The nurse rostering problem: A critical appraisal of the problem structure. European Journal of Operational Research 202(2): 379–389

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

[18] IBM (2009) Ilog Cplex 121 Reference Manual. International Business Machines Corporation, NY

[19] M'Hallah R and Alkhabbaz A (2013) Scheduling of nurses: A case study of a Kuwaiti health care unit. Operations Research for Health Care, 2(1-2):1–19

[20] Puente J, Gomez A, Fernandez 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

[21] 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

[22] Rousseau LM, Pesant G, Gendreau M (2000) A Hybrid Algorithm to Solve a Physician Rostering Problem. In Second Workshop on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems, Paderborn, Germany

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

[24] Testi A, Tanfani E, Torre G (2007) A three-phase approach for operating theatre schedules. Health Care Management Science 10(2): 163–172

[25] Vaidyanathan B, Jha KC, Ahuja RK (2007) Multicommodity network flow approach to the railroad crew-scheduling problem. IBM Journal of Research and Development 51(3): 325–344

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

[27] Yan S, Tu YP (2002) A network model for airline cabin crew scheduling. European Journal of Operational Research 140: 531–540