Trade-off between stakeholders' goals in the home care nurse-to-patient assignment problem

Giuliana Carello [a], Ettore Lanzarone [b,*], Sara Mattia [c]

[a] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
[b] Istituto di Matematica Applicata e Tecnologie Informatiche (IMATI), Consiglio Nazionale delle Ricerche (CNR), Milan, Italy [c] Istituto di Analisi dei Sistemi ed Informatica (IASI), Consiglio Nazionale delle Ricerche (CNR), Rome, Italy

Three stakeholders are involved in health care services: patients, operators and service provider managers. They usually have conflicting needs: patients aim at obtaining good quality of service, operators require fair workloads, and managers try to reduce costs. Moreover, a fourth stakeholder, i.e., the contracting authority, pays for the service and fixes the requirements in terms of costs, quality of service and working conditions, to guarantee that the needs of the three actors are all taken into account and well balanced. Home care services represent a relevant example in which all of the different stakeholder perspectives should be included in the decision process, whereas they have been never compared nor considered together in the literature. We propose a set of mathematical models for the nurse-to-patient assignment problem in home care under continuity of care, inspired to multi-criteria optimization, to investigate the effect of each stakeholder's goal on the others and their interactions. The three stakeholder perspectives are modeled as alternative objective functions of an integer linear programming model, and a threshold method to include all of them is proposed. The approach is then tested on real-life instances, and both deterministic and uncertain patient demands are considered. For the considered setting, results show that the service provider can achieve good quality of service and fair workloads with limited cost increase. Thus, including the other perspectives in the decision making process and respecting the service authority's requirements is not as costly as expected.

Keywords:
Stakeholders' perspectives
Home care
Nurse-to-patient assignments
Continuity of care

1. Introduction

Health care services are provided by complex private and public organizations, and they affect society, citizens and national and regional budgets. Three stakeholders are usually involved in such services, i.e., *patients*, *service provider* and *operators*. Each one has his/her own point of view, goal and requests: patients are interested in a good quality of care; service provider managers must guarantee the service while keeping the operating costs low; operators require good working conditions and fair workload assignments. Unfortunately, these goals might be conflicting. Thus, although the real decision maker is the service provider, patients' and operators' perspectives cannot be neglected to pursue an overall good service quality. Satisfying the needs of the three actors is also the goal of a fourth stakeholder, i.e., the contracting authority, usually a public institution, which pays for the service and sets the requirements in terms of costs, quality of service and working conditions.

* Corresponding author.
E-mail addresses: giuliana.carello@polimi.it (G. Carello), ettore.lanzarone@cnr.it (E. Lanzarone), sara.mattia@iasi.cnr.it (S. Mattia).

We consider in this paper the case of the Home Care (HC) service. HC consists of delivering treatments and cares to patients at their home rather than in hospitals or other structures. HC avoids hospitalization costs and improves patients' quality of life, as they continue living in a familiar environment during the treatment [1]. Thus, HC services are nowadays gaining importance in many Western countries [2].

Several optimization problems arise in managing HC services and can be classified according to the planning level [3,4]. Long-term strategic planning involves decisions on districting and staff dimensioning, while short-term decisions focus on the daily activity planning, such as visits scheduling and operators routing. In addition, when the so-called *continuity of care* is preserved, mid-term planning includes stable assignments of operators to patients. Continuity of care is preserved when patients are assigned to only one nurse, named the *reference nurse*, who follows the patient's entire care pathway and preferably provides all of the visits during the whole treatment period. Although several aspects influence the quality of care, continuity of care is usually considered crucial in the literature, as it prevents patients from continuously developing new relations with new nurses and avoids potential loss of information among operators [5,6].

Published Journal Article available at: https://doi.org/10.1016/j.orhc.2017.12.002

This paper focuses on the *nurse-to-patient assignment problem* in HC, which asks to assign a set of patients to a set of reference nurses over a planning horizon divided into time periods [4,7]. We address an operational problem – see [3] for a description of the different planning levels – assuming fixed staff and fixed capacity for each nurse, i.e., the available working time without incurring overtime. Assignments are decided taking into account nurses' skills and patients' locations, and according to three different continuity of care requirements. On the one hand, this is an important added value because different requirements fairly describe the patients' point of view; on the other hand, the simultaneous presence of different requirements, even though common in the practice, is rarely included in the literature.

As for other management problems in health care, the nurse-to-patient assignment problem in HC must account for the different stakeholders' points of view (patients, service provider and nurses). Patients' continuity of care requirements must be satisfied, as high quality of service is pursued; limiting costs is crucial because HC is specifically designed to keep the care costs low while guaranteeing the service; nurses' activities should be accurately planned because they are particularly exposed to stress and burnout.[1] Indeed, to guarantee an ideal working environment, the workloads must be evenly distributed among the nurses, so that no operator is overloaded while others have reduced duties.

We first analyze the problem considering a single point of view at a time: each perspective is represented by a metrics, the problems are modeled as Integer Linear Programming (ILP) models that optimize only one of these metrics, and their solutions are compared to study the relations among the different perspectives. Then, to account for the contracting authority, we balance the stakeholders' perspectives by means of ILP models in which one metrics is optimized in the objective function and the others are accounted for as additional constraints (threshold method).

Although many approaches can be applied to deal with multi-criteria problems, the threshold method has been chosen because it models the actual stakeholders' roles in the decision making process. Indeed, the service is managed by only one of them (i.e., the service provider) under some regulations imposed by the National or Regional health care system and by the working contracts. Thus, the best way to naturally reproduce this situation is to consider the provider's point of view as the objective function and the needs of the other actors involved, together with the limits imposed by the contracting authority, as constraints. In this way, the provider is the decision maker but, at the same time, it takes into account the needs of the other actors by limiting to which extent their point of view can be stressed out.

Our analysis also includes uncertain patients' demands, which may cause high variability of the assigned workloads. Uncertainty is modeled using the cardinality-constrained approach [9,10].

Models have been tested on the historical data of one of the largest Italian public HC providers, which operates in the Northern Italy. Results confirm that the ILP models can handle real-life instances and, therefore, can be applied in practice for the management of real HC services (they always provide a solution in reasonable time). Moreover, in the considered instance, we found that the service provider can achieve good quality of service (patients' perspective) and fair workloads (nurses' perspective) with limited cost increase. This result, which was not *a priori* expected, has an important practical impact, as it shows that the service provider can include the other stakeholders' perspective (and thus respect the constraints from the contracting authority) without increasing costs.

[1] Burnout is a prolonged response to chronic job-related stressors, which brings workers to stress-related health problems and low career satisfaction [8].

Summing up, the contributions of the paper are: (i) to represent the three stakeholders' perspectives with three associated metrics; (ii) to investigate the effect and the relations of the three metrics through an approach that combines the metrics into one model that fits the decisional process; (iii) to verify that the approach can be applied in real life instances, to provide low cost solutions that also guarantee suitable working conditions for nurses and good quality of service for patients; (iv) to evaluate the benefit of accounting for uncertain patients' demands in this approach through the cardinality-constrained model. To pursue these goals, we exploited proven methodologies of the multi-criteria optimization, considering in particular the so-called threshold method.

Alternative objectives have been previously compared in HC. However, authors compared different ways of pursuing the goal of a given stakeholder, or global functions that include the different stakeholder's perspectives together. On the contrary, separately addressing the stakeholders' perspectives in HC has not been considered in the literature. To the best of our knowledge, an analysis related to this one in HC can be found only in [11], where equity and efficiency criteria are compared; in particular, the authors compared two alternative balancing objective functions via optimization and simulation, showing their impact on diverse indicators.

Looking at the HC nurse-to-patient assignment problem, a similar approach with different continuity of care requirements has been only considered in [12], where a penalty function that combines costs and reassignments of the reference nurses was considered. However, in that paper, the quality of the obtained solutions with respect to the single points of view was not investigated. Thus, the present work is the first attempt to investigate the three point of views in such problem and their relations.

The paper is structured as follows. Section 2 revises the literature about multi-criteria problems, particularly in health care management. Section 3 describes the addressed problem and gives the ILP formulations. Then, the experimental setting is presented in Section 4, and the obtained results are discussed in Section 5. The robust counterpart of the models, based on the cardinality-constrained approach, is presented in Section 6. Finally, the conclusions of the work are given in Section 7.

2. Related works

Many real-life problems involve multiple, possibly conflicting, objectives to be simultaneously optimized by a unique decision maker [13]. Such problems are referred to as *multi-objective* or *multi-criteria* optimization problems.

Let us consider a solution of a multi-criteria optimization problem. From the optimization perspective, the solution is said to be dominated if there exists another feasible solution with a better value for at least one of the objective functions while the others are not degraded. If a solution is non-dominated, i.e., it cannot be improved in any of the objectives without degrading at least one of the others, it is said to be *Pareto optimal*. The set of Pareto optimal solutions is called *Pareto frontier*. No solution in the Pareto front is better than the others in the multi-criteria framework, although the decision maker may prefer a solution depending on his/her own preferences. Multi-criteria decision problems are harder to solve than single-objective problems and several techniques have been applied to cope with them [14] and to compute the Pareto front [15,16].

In practice, solving a multi-objective problem consists of finding a solution that is *good enough* according to all the considered metrics [17]. Two main approaches are used. The first one combines the metrics into a single utility function to optimize (*value function method*). However, this method does not allow to directly control the value obtained for each metrics, and their relative weights within the utility function may be difficult to tune. The second approach uses one metrics as objective function, requiring at the same time that the values of the other metrics remain above or below given thresholds (*threshold method*). Reasonable thresholds are required to guarantee a feasible and good solution. Remarkably, under some assumptions, both approaches produce solutions belonging to the Pareto front. Multi-criteria optimization is becoming increasingly popular, and recent works can be found in the literature dealing with several application fields, e.g., manufacturing [18], real-time train scheduling [19], urban passenger transport systems [20], waste management [21], and water resources management [22].

2.1. Multi-criteria in health care

The number of health care applications where multi-criteria decisions are investigated is constantly increasing over time, in particular for management problems [23]. Such applications spread from the nurse rostering problem [24–27] to operating rooms scheduling [28,29] and hospital bed planning [30].

Recently, some works started to explicitly deal with different stakeholders' perspectives in the management of health care facilities. Two relevant examples deal with the operating room theaters: Cappanera et al. [31] addressed conflicting stakeholders' priorities in surgical scheduling by goal programming, while Marques and Captivo [32] proposed deterministic and robust approaches to evaluate the different stakeholders' perspectives in a surgical case assignment problem.

As for the short-term HC operations management (the daily schedule and routing of nurses) the quality of nurse schedules mainly depends on travel times, overtime, unscheduled tasks, nurses' and patients' preferences. In the literature, such criteria are often combined in a single objective function. Trautsamwieser and Hirsch [33] addressed the daily planning of HC services assigning nurses to routes and visits; the objective considers traveling time, unscheduled jobs, patients' dissatisfaction due to assignments to more than one nurse, and nurses' dissatisfaction due to overtime and long travel times. Rasmussen et al. [34] focused on the HC nurse routing and scheduling problem considering competence, time windows and working times; the objective function combines travel times, nurses' preferences and unassigned visits. Nickel et al. [35] considered a routing and scheduling problem where each visit must be assigned to a route; unassigned visits and lack of nurse–patient loyalty are penalized in the objective function, which also considers overtime costs and traveled distances. As mentioned, Cappanera et al. [11] compared equity and efficiency criteria by means of two alternative balancing objective functions, showing their impact on diverse indicators. More recently, Bowers et al. [36] analyzed the trade-off between continuity of care and travel times in a home midwifery service. Yalçındağ et al. [37] considered an objective function which includes both traveling times and utilization rates in an assignment and routing problem, while afterwards Yalçındağ et al. [38] proposed a pattern-based decomposition for this problem. Finally, Braekers et al. [39] mod-eled a routing and scheduling HC problem using a bi-objective optimization, where the two objectives are the minimization of travel costs and client inconveniences (in terms of penalties for not respecting the patients' preferences about nurses and visit-ing times); non-dominated solutions are produced using a multi-directional local search.

Focusing on the HC nurse-to patient assignment problem under continuity of care, Lanzarone et al. [4] considered a stochastic optimization model whose goal is to balance workloads under continuity of care. Lanzarone and Matta [7,40] developed analytical policies to reduce overtime penalty costs, i.e., a convex function of overtime that also balances the assigned workloads. Carello and Lanzarone [12] proposed a robust mathematical formulation where the objective function combines costs and reassignments of the reference nurse; as mentioned, it is the only paper, to the best of our knowledge, that includes the three types of continuity of care considered in this work.

All of the above-mentioned papers use the *value function method*, and none of them investigate the relations among the stakeholders' goals nor propose a method that can be used by the contracting authority to set minimal service requirements and working conditions. The *value function method* does not to allow to control the values of each single metrics and requires to tune the weights of the metrics within the objective function. On the contrary, as discussed in the introduction, the *threshold method* represents the most natural way to reproduce the decision process in many health care services, where the service provider is in charge of making decisions while considering the other stakeholders' needs and respecting the limits imposed by the contracting authority.

3. Problem and models

This section presents the addressed problem (Section 3.1) and the proposed ILP models. We first describe the common parts of all formulations in Section 3.2; then, we present the different stakeholders' metrics and the related objective functions in Sections 3.3–3.5. Finally, the adopted threshold method is presented in Sections 3.6 and 3.7. The common parts are similar to those reported in [12], whereas the different objectives and the additional constraints are defined afresh.

3.1. Problem description

We consider a set P of patients requiring care during a planning horizon of $|T|$ time periods. Each patient $j \in P$ requires a treatment duration r_{jt} in each time period $t \in T$. Patients are classified accord-ing to their continuity of care requirement: P_{hc} (*hard continuity of care*), P_{pc} (*partial continuity of care*) and P_{nc} (*no continuity of care*). A patient $j \in P_{hc}$ must be assigned to a single reference nurse over the entire time horizon; a patient $j \in P_{pc}$ must be assigned to a single nurse in each period $t \in T$, but the reference nurse may change from period to period; a patient $j \in P_{nc}$ has no restrictions and can be assigned to more than one nurse even in the same period, i.e., a nurse can provide only a part of the treatments required by j in t. Moreover, either patients are already under treatment before the beginning of the planning horizon or they start the treatment at the beginning of the planning horizon. Hence, subsets P_{hc} and P_{pc} are further divided into patients already under treatment (P^a_{hc} and P^a_{pc}, respectively) and new patients (P^n_{hc} and P^n_{pc}, respectively). The number of patients is fixed during the planning horizon. However, patients' characteristics may vary as for each patient j a specific treatment duration r_{jt} is considered for each period t. This also allows considering a variable number of patients even though set P and its subsets have fixed cardinalities. In fact, we can represent patients who exit from the service at time t^* by setting $r_{jt} = 0 \; \forall t > t^*$, and patients who enter in the service at time t^* by setting $r_{jt} = 0 \; \forall t < t^*$.

A set I of nurses is currently employed. Each nurse $i \in I$ has a working time v_i per time period according to the employment contract. However, nurse i may also work overtime (above v_i), provided that his/her total working time does not exceed λv_i in a single time period and $\eta |T| v_i$ over the whole time horizon. The overtime is not included in the salary and must be paid for. Finally, nurses have different skills and operate in different areas. Hence, each patient j can be treated only by nurses in $I_j \subseteq I$ who operate in the patient's neighborhood and have the required skills.

Table 1
Common sets, parameters and decision variables.

Sets	
T	Time horizon
P	Patients
$P_{hc} \subseteq P$	Patients with hard continuity
$P_{pc} \subseteq P$	Patients with partial continuity of care
$P_{nc} \subseteq P$	Patients with no continuity of care
$P_{hc}^a \subseteq P_{hc}$	Patients already under treatment with hard continuity of care
$P_{hc}^n \subseteq P_{hc}$	New patients with hard continuity of care
$P_{pc}^a \subseteq P_{pc}$	Patients already under treatment with partial continuity of care
$P_{pc}^n \subseteq P_{pc}$	New patients with partial continuity of care
I	Nurses
$I_j \subseteq I$	Nurses compatible with patient j
Parameters	
r_{jt}	Patient j's demand at time period t
v_i	Nurse i's available working time per time period
λ	Maximum workload ratio in each time period
η	Maximum workload ratio over the whole time horizon
$\tilde{x}_{ji}$	Assignment of patient $j \in P_{hc}^a$ to nurse i
Decision variables	
x_{ji}	Assignment of patient $j \in P_{hc}^n$ to nurse i (binary)
ξ_{ji}^t	Assignment of patient $j \in P_{pc}$ to nurse i at time period t (binary)
χ_{ji}^t	Fraction of r_{jt} for patient $j \in P_{nc}$ assigned to i (continuous non-negative)
w_{it}	Overall workload of nurse i at time period t (continuous non-negative)

3.2. Common parts of the formulations

A set of assignment variables for each type of continuity of care are introduced to model the problem. Binary variable $x_{ji} \in \{0, 1\}$ equals 1 if nurse i is in charge of patient $j \in P_{hc}^n$ during the entire time horizon, and 0 otherwise. Binary variable $\xi_{ji}^t \in \{0, 1\}$ equals 1 if nurse i is in charge of patient $j \in P_{pc}$ in time period t, and 0 otherwise. Continuous variable χ_{ji}^t $(0 \leq \chi_{ji}^t \leq 1)$ represents the fraction of r_{jt} for patient $j \in P_{nc}$ assigned to nurse i in time period t. As for patients in P_{hc}^a, they must keep the reference nurse assigned before the beginning of the planning horizon, as specified by parameter $\tilde{x}_{ji}$ that takes value 1 if patient $j \in P_{hc}^a$ is already assigned to nurse i, and 0 otherwise. Besides, a continuous non-negative variable w_{it} represents the overall workload (including overtime) of nurse i in time period t. Sets, parameters and decision variables of the problem are summarized in Table 1.

The common constraints are listed below:

$$\sum_{i \in I_j} x_{ji} = 1, \qquad \forall j \in P_{hc}^n \tag{1}$$

$$\sum_{i \in I_j} \xi_{ji}^t = 1, \qquad \forall j \in P_{pc}, t \in T \tag{2}$$

$$\sum_{i \in I_j} \chi_{ji}^t = 1, \qquad \forall j \in P_{nc}, t \in T \tag{3}$$

$$w_{it} \leq \lambda v_i, \qquad \forall i \in I, t \in T \tag{4}$$

$$\sum_{t \in T} w_{it} \leq \eta |T| v_i, \qquad \forall i \in I \tag{5}$$

$$\sum_{j \in P_{hc}^a} r_{jt} \tilde{x}_{ji} + \sum_{j \in P_{hc}^n} r_{jt} x_{ji} + \sum_{j \in P_{pc}} r_{jt} \xi_{ji}^t + \sum_{j \in P_{nc}} r_{jt} \chi_{ji}^t = w_{it}, \qquad \forall i \in I, t \in T \tag{6}$$

Constraints (1)–(3) force to assign each patient to one or more suitable nurses in each time period, according to the required continuity of care. Constraints (4) and (5) guarantee that the maximum allowed workload is not exceeded. Constraints (6) compute the workload of each nurse in each time period.

3.3. The service provider's perspective

The service provider aims at minimizing the overtime costs, where the overtime cost per time unit of nurse i is denoted by c_i. A continuous non-negative variable ω_{it}^o represents the overtime of nurse i in time period t and is computed through the following constraints:

$$\omega_{it}^o \geq w_{it} - v_i \qquad \forall i \in I, t \in T \tag{7}$$

The *cost metrics* of the provider is then the overall overtime cost, which is computed as $\sum_{t \in T} \sum_{i \in I} c_i \omega_{it}^o$. The provider's objective function asks for minimizing the cost metrics:

$$\min \sum_{t \in T} \sum_{i \in I} c_i \omega_{it}^o \tag{8}$$

We refer to the problem (1)–(8) as the *Provider Problem* (PrP).

We remark here that we are dealing with an operational problem, in which staffing decisions are not considered, and that the considered time horizon is too short to include changes in the staff. Moreover, as usual in real life services, staff is not overdimensioned and the available capacity often results in a tight constraint, making the overtime crucial for the provider.

3.4. The nurses' perspective

Nurses should not be overloaded to avoid burnout and other stress-related problems; thus, the service provider's perspective is also related to nurses' requirements.

However, nurses ask for something more, and it is important to prevent them from developing the feeling of being unequally treated, i.e., to avoid situations in which one nurse works at a high percentage of his/her capacity and another one at a low percentage.

Thus, the nurses' utilization rates should be limited and evenly balanced, where the utilization rate of nurse i in period t is the ratio between the workload w_{it} and the available working time v_i. Accordingly, in the nurses' objective (referred to as *fairness metrics*) a weighted sum of the maximum utilization rate and of the difference between the maximum and the minimum utilization rates is minimized. The second term of the objective function prevents from having a limited but not evenly distributed utilization rates, i.e., some nurses are far more utilized than others. The maximum and minimum utilization rates are represented by continuous non-negative variables z_{max} and z_{min}, respectively, whose value is set through the following constraints:

$$z_{max} \geq \frac{w_{it}}{v_i} \qquad \forall i \in I, t \in T \tag{9}$$

$$z_{min} \leq \frac{w_{it}}{v_i} \qquad \forall i \in I, t \in T \tag{10}$$

The *fairness metrics* is then given by $\alpha_n z_{max} + (z_{max} - z_{min})$, where α_n is the relative weight of the first term. A high value of α_n mainly accounts for the maximum utilization, whereas a low value for the imbalances among nurses. The nurses' objective function is thus the following:

$$\min \ \alpha_n z_{max} + (z_{max} - z_{min}) \tag{11}$$

We refer to the model (1)–(6), (9)–(11) as the *Nurse Problem* (NuP).

Fairness is not guaranteed by (4) and (5), nor automatically ensured by the optimal solution of the service provider's perspective. Moreover, although a provider may be interested in guaranteeing good working conditions, fairness is not its main goal when it does not affect provider's costs or even increases costs. Thus, fairness has to be explicitly required when aimed.

3.5. The patients' perspective

The quality of care received by patients is influenced by several aspects, which are difficult to separate and capture with a mathematical function. Most of them are also related to the nurses' working conditions, thus involving the service provider's and the nurses' perspectives.

However, continuity of care is described in the literature to be a key point in guaranteeing a good level of care; thus, continuity of care is the major point to consider when focusing on the patients' perspective. Continuity is mandatory for patients in P_{hc}, whereas it is not required for patients in P_{nc}. Thus, the goal is to guarantee the continuity of care for the largest number of patients in P_{pc} (*reassignment metrics*).

Binary variable $y_{jt} \in \{0, \ 1\}$ represents the change in the reference nurse between two consecutive time periods for patient $j \in P_{pc}$. It takes value 1 if two different nurses are in charge of patient $j \in P_{pc}$ at time periods $t-1$ and t. In addition, a contin-uous non-negative variable s represents the maximum number of reassignments for a single patient over the whole planning horizon. Values of y_{jt} and s are set through the following constraints:

$$y_j^t \geq \xi_{ji}^t - \xi_{ji}^{t-1} \qquad \forall j \in P_{pc}, i \in I, t \in T \setminus \{t = 1\} \tag{12}$$

$$y_j^1 \geq \xi_{ji}^1 - \tilde{\xi}_{ji} \qquad \forall j \in P_{pc}^a, i \in I \tag{13}$$

$$s \geq \sum_{t \in T} y_{jt} \qquad \forall j \in P_{pc} \tag{14}$$

Parameter $\tilde{\xi}_{ji}$ represents the last assignment of patient $j \in P_{pc}^a$ before the beginning of the planning horizon.

We aim at minimizing the total number of reassignments, provided that no patients are particularly penalized, i.e., the reassignments must be fairly distributed among the patients. Thus, the *reassignment metrics* consists of a weighted sum of the total number of reassignments and of the highest number of reassignments for a single patient, i.e., $\alpha_p |P_{pc}||T|s + \sum_{j \in P_{pc}} \sum_{t \in T} y_j^t$, where α_p is the relative weight of the first term. The patients' objective function is therefore:

$$\min \ \alpha_p |P_{pc}||T|s + \sum_{j \in P_{pc}} \sum_{t \in T} y_j^t \tag{15}$$

Model (1)–(6), (12)–(15) is called the *Patient Problem* (PaP).

3.6. Threshold method: Adding points of view as constraints

According to the threshold method for multi-criteria optimization, we add a set of constraints to each model, to limit the values of the other metrics, thus obtaining the corresponding *constrained* model. This naturally replicates the role of the contracting authority, which imposes limits to guarantee the needs of all stakeholders. Moreover, as these needs are related, this harmonizes all of the requirements to improve the overall level of the service for all actors involved.

When minimizing the *fairness* or the *reassignment metrics*, costs are limited by adding the following constraint:

$$\sum_{t \in T} \sum_{i \in I} c_i \omega_{it}^o \leq \bar{c} \tag{16}$$

where $\bar{c}$ is the maximum acceptable overtime cost. Of course, (16) also requires constraints (7) to compute ω_{it}^o.

When minimizing the *cost* or the *reassignment metrics*, the nurses' point of view can be accounted for by limiting the difference in the utilization rates between any two nurses and the utilization rates themselves, coherently with the two terms of (11):

$$\frac{w_{ht}}{v_h} - \frac{w_{kt}}{v_k} \leq \delta_1 \quad h, k \in I, t \in T \tag{17}$$

$$\frac{w_{it}}{v_i} \leq \delta_2 \quad i \in I, t \in T \tag{18}$$

where the δ_1 and δ_2 are suitable threshold values.

Finally, when minimizing the *cost* or the *fairness metrics*, we limit the individual number of reassignments as follows, where q is the given threshold.

$$\sum_{t \in T} y_{jt} \leq q \quad j \in P_{pc} \tag{19}$$

We then consider the following problems:

- *Constrained Provider Problem* (CPrP): PrP with additional parameters δ_1, δ_2 and q, and additional constraints (17)–(19);
- *Constrained Nurse Problem* (CNuP): NuP with additional parameters $\bar{c}$ and q, and additional constraints (7), (16), (19);
- *Constrained Patient Problem* (CPaP): PaP with additional parameters $\bar{c}$, δ_1 and δ_2, and additional constraints (7), (16)–(18).

3.7. Setting and refreshing the thresholds

Thresholds represent the minimal values that a particular stakeholder may consider satisfactory for its own metrics. Each time the problem is solved, either the value of each threshold can be provided by the stakeholder itself (first case) or it can be computed from the current conditions of the provider (second case). In the latter case, to get the maximum utility for the stakeholder, the threshold can be set at the optimum of the single objective problem, provided that a certain degradation may be necessary to guarantee the feasibility of the constrained problem. Considering that the assignments are refreshed at a fixed frequency or every time one or more new patients enter the service, thresholds are fixed in the first case, while they need to be updated in the second case.

Updating could be ideally performed each time the problem is solved. In such layout, solving the multi-criteria optimization always requires solving two single objective models in advance, to get the values of the thresholds for the two metrics not included in the objective function. Moreover, in case the set thresholds give infeasibility for the constrained problem, the multi-criteria optimization has to be repeated while increasing the thresholds.

Table 2
Patients per week.

Week	$\|P^a\|$	$\|P^n\|$	Week	$\|P^a\|$	$\|P^n\|$	Week	$\|P^a\|$	$\|P^n\|$
1	560	29	10	568	8	19	575	21
2	571	22	11	559	16	20	582	13
3	571	30	12	563	14	21	582	15
4	573	14	13	557	23	22	581	16
5	568	9	14	557	18	23	576	16
6	559	24	15	568	18	24	575	19
7	566	25	16	557	22	25	590	8
8	575	20	17	557	26			
9	579	20	18	564	19			

Alternatively, if the provider conditions are quite stable, the values of the thresholds can be provided based on the past experience, to avoid the additional computational times every time the problem is solved. Thus, the thresholds are kept constant over a mid-term horizon, and refreshed at a fixed frequency or in the presence of a certain deviation secondary to a variation in the provider working conditions (different number of patients, composition of the patient mix, …). We address the reader to [15] for details about the threshold method.

4. Experimental setting

Tests are run on real-life instances, in which the planning is performed on a weekly basis. To test the approach over a long time period, we consider a rolling horizon framework, in which the assignments computed in a week feed the next week instance as input (i.e., they provide the values for parameters $\tilde{x}_{ji}$ and $\tilde{\xi}_{ji}$). Indeed, we consider the historical data of a large HC provider operating in the Northern Italy [4,7,12,40] over 26 weeks (period April 2008–September 2008). The initial week (named week 0) is used for initializing the assignments, while the assignments are rolling updated in the remaining weeks 1–25.

The considered HC provider is in charge of about 1000 patients who are cared after by about 50 nurses. It is divided into three divisions, and the number of patients in the charge together with their features are taken from the historical data of the largest division. The patient set P and its subsets (cardinalities and patients' characteristics) are then updated at each rolling week based on the actual historical data at the corresponding week. The number of patients at each week is reported in Table 2.

Patients are classified into 14 different care profiles (11 profiles for non-palliative patients and 3 for palliative ones) and demands r_{jt} are estimated from current patients' conditions through the stochastic model of Lanzarone et al. [41], i.e., r_{jt} are obtained as the expected value of an estimated probability density function.

Nurses and patients of the considered division are divided into 6 independent districts based on territory and skill (palliative or non-palliative). We consider a separate management of districts, in which each nurse operates in only one district. Table 3 reports the number of nurses, their available working times v_i and the percentage of patients for each district.

Two continuity of care scenarios are considered: (i) scenario *Partial* in which all of the patients require partial continuity of care; (ii) scenario *Mix* which includes all types of continuity (hard, partial, none). Continuity requirements in the *Mix* scenario depend on the patient's care profile. As for non-palliative patients, intensive care profiles require a harder type of continuity; as for palliative patients, they are randomly assigned to hard and partial continuity of care requirement with probability 0.8 and 0.2, respectively. The same division is kept in all tested *Mix* cases. On average, the 54% of the patients ask for hard continuity of care, the 20% for partial continuity, and the 26% do not ask for any continuity.

With a look ahead perspective, we consider a planning horizon of $|T| = 8$ weeks for each instance, which corresponds to as many future weeks starting from the current one. Parameters λ and η are set equal to 1.8 and 1.4, respectively, and both α_n and α_p are set equal to 10, i.e., we give more importance to the maximum utilization rate (in the *fairness metrics*) and to the individual number of reassignments (in the *reassignment metrics*). Finally, we assume that the overtime cost c_i is equal to 1 for all nurses.

Concerning the thresholds for the multi-criteria approach, we work in the second case presented in Section 3.7 as, based on the available historical data, we may assume that the service conditions are rather stable over the considered rolling weeks. Thus, we consider that the provider sets the threshold values for the metrics based on its past experience, and that such thresholds are kept constant over the weeks.

To emulate such behavior in the experiments, we treat the first 7 weeks after the initialization (weeks 1–7) as the past experience of the provider, and we use them to set the thresholds for the following weeks (weeks 8–25). We run each unconstrained model up to week 7, and we take the maximum of the metrics over the weeks 1–7 as initial thresholds. Then, for each constrained model, we check the feasibility of the metrics constraints with these thresholds; in case of unfeasibility, we iteratively increase the threshold values by 10% until a feasible solution is found up to week 7. The obtained thresholds are then applied to the following weeks 8–25. If these thresholds are too tight for a week, they are increased only for this unfeasible week (iteratively by 10% until the feasibility is reached); if the infeasibility occurs in several weeks, this suggests an updating of the thresholds.

Models have been implemented in the CPLEX STUDIO IDE environment, and solved with CPLEX 12.6.1 on a PC equipped with processor Intel i7-6700MQ at 2.60G Hz and with 16 Gb of RAM. A time limit of 5400 s, a tree memory limit of 3 Gb and an acceptable optimality gap equal to the 0.05% have been imposed. As we are considering a weekly based planning, one hour and a half of computation is reasonable.

5. Results

This section analyzes the computational results, aiming at:

(i) Investigating the relations among the three unconstrained problems, so as to verify whether a single objective can produce good results also for the other two metrics. Hence, the three unconstrained problems are compared in Section 5.1.
(ii) Evaluating the impact of the thresholds and the possibility of providing a low cost solution while guaranteeing suitable working conditions for nurses and good quality of service for patients (thus satisfying the contracting authority). Hence, the behavior of the constrained problems is analyzed with respect to all of the metrics in Section 5.2.
(iii) Evaluating the computational effort required to solve the models, either with or without the additional constraints, to show that the proposed approach can be used in practice. Results are reported in Section 5.3.
(iv) Evaluating the benefit of planning the assignments accounting for the uncertain patient demands through a cardinality-constrained robust model. Results are separately reported in Section 6.

It is worth pointing out that, due to the rolling approach, the different models are not compared on the very same instances at each week. Indeed, the previously computed assignments are kept to define parameters $\tilde{x}_{ji}$ and $\tilde{\xi}_{ji}$; thus, solving different models in a week leads to different instances in the successive weeks. However, as we are interested in the long-term behavior of each model, the rolling approach follows the impact of a model over the time.

Table 3
Number of nurses, v_i and percentage of patients in weeks 1–25 for each district. NP* districts include the non-palliative patients, whereas P* districts the palliative ones.

District	Number of nurses	Nurse weekly working time v_i	Percentage of patients
NP1	8	35, 40, 45, 50, 50, 50, 50, 50	39%
P1	3	20, 30, 30	6%
NP2	4	30, 35, 50, 50	22%
P2	1	35	3%
NP3	5	30, 35, 40, 50, 50	26%
P3	1	35	4%

Table 4
Objective function values for the unconstrained models: average value among the weeks together with the minimum and maximum values in brackets; weeks 1–7 (a) and weeks 8–25 (b).

Scenario	PrP	NuP	PaP	
Partial	13.60 (5.05–19.03)	11.62 (10.99–13.10)	20264 (0–47442)	(a)
Mix	13.60 (5.05–19.03)	11.63 (10.99–13.13)	0 (0–0)	(a)
Hard	16.75 (6.70–23.59)	12.64 (11.57–13.65)		(a)
Scenario	PrP	NuP	PaP	
Partial	12.25 (3.88–25.25)	11.43 (10.94–12.02)	20863 (0–47921)	(b)
Mix	12.36 (3.88–25.25)	11.44 (10.97–12.01)	533 (0–9601)	(b)
Hard	15.29 (6.29–28.71)	12.07 (11.43–12.75)		(b)

5.1. *Comparison among the unconstrained problems*

As first step of the analysis, we report in Table 4 the values of the objective function obtained by the three unconstrained problems for both weeks 1–7, used for setting the thresholds, and the actually tested weeks 8–25. Only in this first analysis, to highlight the impact of the continuity of care, we also report the results of a third, less realistic, scenario, i.e., the *Hard* scenario in which all patients require hard continuity of care. Of course, the PaP is not reported for the *Hard* scenario, as the continuity is always respected.

In the initial 7 weeks, the PrP and the NuP have the same objective function value for the *Partial* and the *Mix* scenarios. Instead, the PaP objective function is significantly higher for the *Partial* scenario, both because the number of patients who require partial continuity is higher in the *Partial* scenario and because in the *Mix* scenario patients in P_{nc} allow reassigning nurses without penalties; in fact, in the *Mix* scenario, there are no reassignments in all of the weeks. Fairness does not change much from week to week both for the *Partial* and the *Mix* scenarios. As for the *Hard* scenario, the objective function values are higher than for the other scenarios: hard continuity of care reduces the flexibility of the service, thus increasing the overtime costs for the PrP and slightly increasing the fairness metrics for the NuP. On the contrary, in the *Mix* scenario, patients requiring hard continuity are somehow balanced by those requiring partial or no continuity, who give flexibility to the system.

Similar remarks also hold for the actually tested weeks 8–25. Indeed, the PrP and NuP objective functions do not change much in the *Partial* and *Mix* scenarios, while they both increase in the *Hard* one. The number of reassignments is significantly lower for the *Mix* scenario when compared to the *Partial* one, although not always null as in weeks 1–7 (there is one reassignment in week 24 in the *Mix* scenario). Comparing these weeks with the initial ones, we may note that the average value of the PrP objective function is lower than in the first weeks, while the minimum–maximum interval is larger. Instead, the NuP objective function is a little bit lower (average, minimum and maximum values) than in the first weeks. Finally, the PaP objective increases in weeks 8–25.

We show now the relations among the different perspectives, by analyzing the value of the three metrics when solving each unconstrained problem. Results are reported in Table 5, where columns are associated with models and rows with metrics; thus, the best possible value of each metrics can be found on the diagonal.

Similar relations among the metrics are found in both scenarios and both in weeks 1–7 and 8–25. The PrP guarantees reasonably good fairness, as the increase of the average fairness metrics is between the 10% and the 16% for any scenario and considered set of weeks. Instead, the opposite is not true, as overtime costs may significantly increase when solving the NuP (between 1.3 and 7.21 times as far as the average value is concerned; up to about 14 times for the maximum value). Yet, overtime and fairness are somehow related. On the contrary, the patient perspective conflicts with the others: in fact, when solving the PrP or the NuP, the number of reassignments dramatically increases, and the PaP produces the highest values of overtime costs and fairness metrics.

Results show that we cannot rely on an unconstrained model to guarantee good values for each metrics. In fact, even the PrP, which produces not only the best overtime cost but also good fairness values, is not able to produce the optimal value of fairness and generates many reassignments. Thus, we resort to the threshold-constrained models described in Section 3.6 to obtain good quality solutions with respect to all of the metrics.

5.2. *Including the threshold constraints*

Here we analyze the results obtained solving the constrained problems, where threshold values are computed with the procedure described at the end of Section 4 using the results obtained in weeks 1–7 as thresholds for weeks 8–25.

Initial tested values are $\bar{c} = 19.03$, $\delta_2 = 1.22$ and $q = 1$ for both the *Partial* and the *Mix* scenarios, while $\delta_1 = 0.86$ in the *Partial* scenario and $\delta_1 = 0.88$ in the *Mix* scenario. Although q never reaches value 1 in the *Mix* scenario, nevertheless we set $q = 1$ in both cases so as to provide some flexibility. After the required increments to get the feasibility of the constrained problems in weeks 1–7, the obtained thresholds are as follows (10% of increase with respect to the initial values):

- CPrP: $\delta_2 = 1.34$ and $q = 1$ in both scenarios; $\delta_1 = 0.97$ and $\delta_1 = 0.95$ in the *Mix* and *Partial* scenario, respectively.
- CNuP: $\bar{c} = 20.93$ and $q = 1$ in both scenarios.
- CPaP: $\delta_2 = 1.34$ and $\bar{c} = 20.93$ in both scenarios; $\delta_1 = 0.97$ and $\delta_1 = 0.95$ in the *Mix* and *Partial* scenario, respectively.

Results are reported in Table 6, where rows are associated with metrics and columns are associated with problems.

Let us first note that the adopted thresholds (derived from weeks 1–7) always allow to find a feasible solution for the constrained problems but in two weeks, namely week 9 and 20, thus showing that the service demand is stable enough and that historical data can be used to provide suitable thresholds for a long

Table 5
Metrics of the unconstrained models: average value among the weeks together with the minimum and maximum values in brackets. Weeks 1–7 for scenario *Partial* (a); weeks 8–25 for scenario *Partial* (b); weeks 1–7 for scenario *Mix* (c); weeks 8–25 for scenario *Mix* (d).

Metrics	PrP	NuP	PaP	
Overtime	13.60 (5.05–19.03)	64.07 (8.20–281.86)	530.89 (433.18–645.70)	(a)
Fairness	12.83 (11.95–13.53)	11.62 (10.99–13.10)	19.48 (19.06–19.73)	
Reassignment	380151 (372589–388109)	380123 (372578–388007)	20264 (0–47442)	
Metrics	PrP	NuP	PaP	
Overtime	12.25 (3.88–25.25)	28.16 (8.40–96.48)	379.26 (309.97–475.82)	(b)
Fairness	13.03 (11.51–15.76)	11.43 (10.94–12.02)	19.23 (17.74–19.66)	
Reassignment	378975 (370412–386847)	379400 (370921–386834)	20863 (0–47921)	
Metrics	PrP	NuP	PaP	
Overtime	13.60 (5.05–19.03)	111.66 (43.66–255.63)	407.79 (330.64–554.01)	(c)
Fairness	13.24 (11.60–14.24)	11.63 (10.99–13.13)	19.00 (16.96–19.80)	
Reassignment	69878 (65891–77788)	71134 (65886–77149)	0 (0–0)	
Metrics	PrP	NuP	PaP	
Overtime	12.36 (3.88–25.25)	83.69 (31.75–158.16)	214.70 (109.63–413.48)	(d)
Fairness	13.26 (11.58–15.09)	11.44 (10.97–12.01)	16.99 (15.95–19.17)	
Reassignment	70484 (63656–77840)	70376 (63573–77073) 1	533 (0–9601)	

Table 6
Metrics of the constrained models: average value over weeks 8–25 together with the minimum and maximum values in brackets. Number of weeks in which the adopted thresholds make the problem infeasible and have been recomputed. Scenario *Partial* (a) and scenario *Mix* (b).

Metrics	CPrP	CNuP	CPaP	
Overtime	12.25 (3.88–25.25)	13.78 (3.88–26.07)	16.97 (8.57–25.25)	(a)
Fairness	12.71 (11.47–14.34)	11.43 (10.94–12.02)	12.19 (11.76–12.85)	
Reassignment	47478 (46449–48419)	47513 (46450–48412)	18312 (0–47921)	
Weeks 8–25 with threshold recomputation	0	2	2	
Metrics	CPrP	CNuP	CPaP	
Overtime	12.25 (3.88–25.25)	14.02 (3.88–26.64)	19.38 (8.72–26.64)	(b)
Fairness	13.06 (11.97–14.70)	11.43 (10.93–12.01)	13.47 (12.52–16.98)	
Reassignment	9421 (8866–10067)	9435 (8863–10063)	1587 (0–9762)	
Weeks 8–25 with threshold recomputation	0	2	2	

time horizon (up to 6 months) with few critical weeks in which the thresholds must be recomputed.

Two main behaviors are *a priori* possible when solving the constrained models: either there are equivalent optima and the additional constraints select among them, or the metrics are conflicting and the additional constraints worsen the objective function value with respect to the unconstrained optimum. Our results show that adding constraints (16)–(19) with suitable thresholds allows selecting, among a set of almost equivalent optimal solutions with respect to a single point of view, a solution that provides a good value also for the other two metrics. In fact, by comparing the results with those in Table 5, we observe that in almost all cases the objective function does not get worse when the values of other metrics are limited, while the other metrics are significantly improved.

Let us first consider the *Partial* scenario and the CPrP and CNuP problems. On the one hand, their objective does not change when adding the threshold constraints (average, minimum and maximum values). On the other hand, as for the CPrP, the fairness metrics improves with respect to the PrP, and the reassignment metrics dramatically decreases at the same time. Similarly, when solving the CNuP, the overtime cost is almost halved with respect to NuP and the reassignment metrics is reduced by almost one order of magnitude. As for the CPaP in the *Partial* scenario, not only the fairness metrics and the overtime cost are reduced (up to about 20 times as far as the average overtime cost is concerned) but also the average reassignment metrics is improved. This happens because the alternative assignments that meet the threshold constraints in one week result in a different (and possibly better) input for the following weeks, leading to solutions that may be even better than the unconstrained ones. The behavior is confirmed in the *Mix* scenario: the CPrP and the CNuP improve all of the metrics with respect to the PrP and the NuP, respectively, including that optimized in the objective. Instead, the CPaP only improves overtime and fairness metrics with respect to the PaP, whereas the reassignment metrics increases.

From these results, we also note that overtime cost and fairness are somehow related, and that optimizing the cost yields to low values of the fairness metrics. Indeed, the PrP provides a good value of the fairness metrics even without the additional constraints and the fairness metrics value is only marginally reduced in the CPrP. On the contrary, as mentioned, a reduction is observed in the overtime cost metrics passing from NuP to CNuP. Further, the number of reassignments may significantly increase if not constrained; hence, a great reduction can be obtained by adding the corresponding constraints.

The selection among equivalent optimal solutions was not expected, especially for the reassignment metrics, which seems conflicting with the others. Such behavior is mainly driven by the ratio between the demands r_{jt} and the working times v_i. In fact, if r_{jt} values are small with respect to the working times v_i, it is possible to *adjust* the assignments so as to guarantee good quality of care at limited costs. On the contrary, higher ratios r_{jt} over v_i mean that workloads are less adjustable and, thus, one of the two metrics must be worsened in order to favor the other.

In the considered dataset, the expected r_{jt} at week $t = 1$ is on average 2.6 h, and it never rises above 6.5 h; moreover, these values decrease while t increases (see [41] for detailed values). Moreover, in the patient mix obtained from the real data of the considered weeks, the fraction of patients requiring the maximum treatment time (a subset of the palliative patients) is small, being about the 10%, while about the 70% of patients require less than 3 h.

More in general, high treatment times are not common in HC. In fact, patients requiring higher service times are usually addressed to other services (e.g., hospitals, nursing houses), or they hire a full time caregiver. Thus, our instance can be considered general enough, and the results provide indications that are also valid for other HC providers. In particular, they are highly impacting from the application point of view, as they mean that the provider can meet the other stakeholders' requirements without increasing the costs.

As a last analysis, let us compare the results with those obtained by the model proposed in [12] on a scenario similar to the *Mix* case in terms of number and characteristics of the patients. Both models include three types of continuity of care requirements, but the objective function considered in [12] is a weighted sum of two terms. The first term, aiming at both reducing overtime costs and balancing workloads, is a piecewise linear approximation of a quadratic function, where the overtime cost per time unit increases with the increasing overtime; the second term, aiming at minimizing the number of reassignments, is the sum of the reassignments. The weight of the second term with respect to the first one has been set so that one reassignment is as expensive as 2.5 h of overtime in the first piece. Also that model provides, in general, good performances for the three metrics. The quadratic penalty cost function keeps the overtimes limited and also limits the overtime differences among nurses, as it spreads the overtime among the nurses rather than assigning all of it to only one or few nurses; moreover, the high relative weight of the second term keeps the number of reassignments low. However, that model does not separate the effects of the different points of view, nor allows to evaluate the relations among the metrics or to tune the limits of the individual metrics. On the contrary, the threshold method here proposed can show how close or far each metrics is from its optimal value. Finally, we highlight that good performances were obtained with the model in [12] because the three individual metrics are somehow related, as shown before in this section; while the threshold method is able to analyze this trade-off and to deal with other cases in which the points of view may be conflicting, the other approach does not.

Let us conclude this analysis discussing the impact of keeping constant thresholds using the maximum value as starting point for computing them. It may be argued that keeping the same thresholds along the whole time horizon may excessively worsen the metrics in some weeks. However, the worsening is not dramatic in the tested instances. Let us consider the CPrP, which is more likely to be solved in the real practice; results in Table 5 show that, as far as the fairness metrics is concerned, the maximum value among the weeks is about the 20% higher than the smallest one; concerning the continuity of care metrics, the increase of the maximum value among the weeks with respect to the lowest one seems to be significant but, indeed, it represents a small number of reassignments. Anyway, even computing the thresholds starting with the maximum values and keeping them constant, the solution is good enough for both metrics. It proves, in the considered HC provider, that the demand is stationary enough to assume constant thresholds (as also confirmed by the small number of weeks in which the thresholds must be recomputed — see Tables 6(a) and 6(b), last row) and that the method is able to produce good results in practice.

5.3. Computational time

Computational times are analyzed to assess whether the proposed approach can be applied in practice for a weekly planning. The approach requires a threshold-setting phase, where first the unconstrained problems and then the constrained ones are solved on historical data, and the actual weekly assignment planning.

Results for the threshold setting phase are reported in Table 7, where the average, the minimum and the maximum computational times among weeks 1–7 are given, together with the sum over the weeks, for each unconstrained and constrained problem. Results for weeks 8–25 are reported in Table 8, where the average, the minimum and the maximum computational times among the weeks are reported for each constrained problem.

As for the unconstrained problems, results show that solving the problems on weeks 1–7 in the *Partial* scenario is not computationally challenging: the PrP requires 6.63 s on average and less than 1 min even in the worst case; NuP is slightly more time consuming on average (7.35 s) but it does not require more than 20 s; the PaP is the least time consuming, as it requires 4.14 s on average and never more than 8 s. Solving the unconstrained models in the *Mix* scenario is even less time consuming, as it always requires less than 1 s. Summing up, the *Partial* scenario is the most time consuming, requiring about 10 times the computational time of the *Mix* one, since no assignment is forced (subset P^a_{hc} is empty).

Including the metrics constraints increases the computational time. The computational time needed to solve the *Partial* scenario significantly increases for the CPrP and the CNuP with respect to the unconstrained versions, while CPaP increases only slightly. The *Mix* scenario, i.e., the most realistic one, suffers less from the additional constraints: in fact, whatever the considered case is, the constrained problem always requires about 1 s and its computational time never rises above 1.54 s.

Table 7
CPU times[s] for weeks 1–7, unconstrained and constrained problems.

	Partial				*Mix*			
	avg	min	max	sum weeks 1–7	avg	min	max	sum weeks 1–7
PrP	6.63	1.12	39.16	46.42	0.58	0.52	0.62	4.05
NuP	7.35	1.53	18.27	51.45	0.67	0.61	0.76	4.68
PaP	4.14	2.52	7.96	28.98	0.71	0.69	0.74	4.97
CPrP	19.65	13.64	37.61	137.53	1.09	1.01	1.17	7.65
CNuP	57.52	6.61	96.27	402.64	1.16	1.06	1.30	8.12
CPaP	5.93	4.38	8.93	41.53	1.00	0.94	1.09	7.01

Table 8
CPU times [s] for weeks 8–25, constrained problems.

	Partial			Mix		
	avg	min	max	avg	min	max
CPrP	21.63	12.20	47.33	1.10	0.97	1.40
CNuP	89.64	14.43	162.83	1.08	0.86	1.54
CPaP	5.29	4.33	6.98	0.93	0.84	1.23

As for the threshold setting phase, it requires to first solve the unconstrained problems in weeks 1–7, and then the constrained versions, possibly increasing the thresholds if an unfeasibility occurs in a week; in this case, the constrained problems must be solved more than once. In our experiments, thresholds are increased only once, by the 10%. Considering the worst case (unfeasibility at week 7), each constrained problem is solved twice per week. Thus, being the time for the feasibility check negligible, the computational effort for the overall thresholds setting procedure includes that to solve 7 times the unconstrained problem and to solve 14 times the constrained one, besides the time for the corresponding 3 initializations at week 0, which correspond to less than 15 min. Considering that the thresholds setting procedure is occasionally performed (in our experiments the same thresholds are kept for 17 weeks and they do not need to be updated but in two), the computational effort is highly acceptable and compatible with the real life application.

As for the weekly planning, results in Table 8 show that, even in the most time consuming case (the CNuP solved in the *Partial* scenario), the computational time for solving a constrained problem in a week never rises above 3 min. Moreover, solving the constrained problems in the *Mix* scenario, which is the most likely to occur in practice, is much faster and never rises above 1 s.

Finally, we remark that, even if the thresholds have to be increased in a week, being the time consumed for the feasibility check negligible, the computational times in Table 8 are valid and the weekly planning requires a certainly acceptable effort.

6. Robustness and executed solutions

So far, treatment durations r_{jt} have been assumed to be deterministic parameters. Instead, in real-life cases, r_{jt} are uncertain parameters that depend on patient's conditions [41]. To tackle this uncertainty, we propose robust counterparts of the developed models, based on the cardinality-constrained approach [9], and we analyze the behavior of the solutions when applied to real-life scenarios.

In the robust cardinality-constrained formulation, parameters r_{jt} are assumed to take values in an interval between a nominal value $\bar{r}_{jt}$ and a maximum value $\bar{r}_{jt} + \hat{r}_{jt}$. For each nurse i and time period t, a subset of patients of given cardinality for each continuity of care requirement (Γ_{hc} patients in P_{hc}, Γ_{pc} patients in P_{pc} and Γ_{nc} patients in P_{nc}, respectively) ask for the maximum treatment duration, while the others for the nominal one. The approach selects the worst possible case, thus guaranteeing that the obtained solution remains feasible for every other choice of the subsets with the given cardinality. In the following, we briefly describe how to formulate the robust counterpart; we refer to [12] for a detailed description of the formulation.

Reformulation is based on duality, and a pair of variables for each type of continuity of care, ζ_{it}^{hc}, π_{jit}^{hc}, ζ_{it}^{pc}, π_{jit}^{pc}, ζ_{it}^{nc}, π_{jit}^{nc} are introduced to account for the dual amounts. Robust counterparts of the models are reformulated by replacing constraints (6) with the following sets of constraints:

$$\sum_{j\in P_{hc}} \bar{r}_{jt} x_{ji} + \Gamma_{hc}^{i} \zeta_{it}^{hc} + \sum_{j\in P_{hc}} \pi_{jit}^{hc} + \sum_{j\in P_{pc}} \bar{r}_{jt} \xi_{ji}^{t} + \Gamma_{pc}^{i} \zeta_{it}^{pc} + \sum_{j\in P_{pc}} \pi_{jit}^{pc} + \sum_{j\in P_{nc}} \bar{r}_{jt} \chi_{ji}^{t} + \Gamma_{nc}^{i} \zeta_{it}^{nc} + \sum_{j\in P_{nc}} \pi_{jit}^{nc} \le w_{it}, \qquad \forall i \in I, t \in T \tag{20}$$

$$\zeta_{it}^{c} + \pi_{jit}^{c} \ge \hat{r}_{jt} x_{ji}, \qquad \forall i \in I, j \in P_{hc}, t \in T \tag{21}$$

$$\zeta_{it}^{pc} + \pi_{jit}^{pc} \ge \hat{r}_{jt} \xi_{ji}^{t}, \qquad \forall i \in I, j \in P_{pc}, t \in T \tag{22}$$

$$\zeta_{it}^{nc} + \pi_{jit}^{nc} \ge \hat{r}_{jt} \chi_{ji}^{t}, \qquad \forall i \in I, j \in P_{nc}, t \in T \tag{23}$$

where variables w_{it} now represent the robust workloads in the worst case.

Robust increased workloads are used for computing the cost metrics. As for the fairness metrics, robust increased workloads are taken when computing z_{max} for the maximum utilization rate, i.e., the first term in (11). However, the difference between the maximum and minimum rates (the second term) is computed based on the nominal workload values without robustness to ensure the applicability of the approach.

In the analyses, we set each cardinality Γ equal to 2 and the maximum values $\hat{r}_{jt}$ at the 80% quantile of the probability density function from the patient stochastic model of Lanzarone et al.[41]. The initial assignment are computed solving the deterministic model on week 0; then, the robust model is solved both in weeks 1–7, when setting the thresholds, and in weeks 8–25. Parameters λ and η keep the already adopted values for the PrP and the NuP (and the corresponding CPrP and the CNuP), which are large enough to permit managing the worst case. As for the PaP and the corresponding CPaP, λ and η are increased by 0.1 in some weeks to guarantee feasibility, due to the increased workload in the robust approach.

For the constrained models, the thresholds are defined starting from the unconstrained robust models, considering the weeks 1–7 as described in Section 3.6. Initial tested values are $\bar{c} = 178.60$, $\delta_1 = 0.90$, $\delta_2 = 1.45$ and $q = 1$ (according to what mentioned above, δ_1 considers the robust workloads while δ_2 the determinis-tic ones). The thresholds have been increased by the 10%, i.e., the final adopted values are: $\bar{c} = 196.46$, $\delta_1 = 0.99$, $\delta_2 = 1.60$ and $q = 1$. Moreover, no additional increase has been required in weeks 8–25 in any experiment.

Due to the worst-case selection, the objective function of the robust solution has not a practical meaning. The impact of robustness

Table 9
Execution in 10 sample paths of the overtime per nurse and the utilization rate, for the constrained models in the *Mix* case over weeks 8–25.

Model	Metrics	Deterministic		Robust		District
		avg	mix–max	avg	mix–max	
CPrP	overtime	1.24	0.00–2.04	0.61	0.00–1.75	NP1
		0.26	0.00–2.79	0.13	0.00–1.52	P1
		1.02	0.00–4.03	2.23	0.00–6.16	NP2
		1.59	0.00–5.10	1.74	0.00–3.91	NP3
	z_{max}	1.19	0.99–1.41	1.11	0.95–1.35	NP1
		0.88	0.54–1.28	0.80	0.43–1.15	P1
		1.08	0.96–1.28	1.24	0.98–1.48	NP2
		1.14	0.93–1.51	1.19	1.00–1.47	NP3
	z_{min}	0.38	0.27–0.58	0.54	0.40–0.76	NP1
		0.25	0.08–0.66	0.35	0.14–0.66	P1
		0.75	0.57–0.93	0.69	0.51–0.91	NP2
		0.57	0.29–0.81	0.68	0.56–0.85	NP3
CNuP	overtime	0.81	0.00–3.43	0.61	0.00–1.60	NP1
		0.09	0.00–1.42	0.02	0.00–0.38	P1
		0.73	0.00–3.83	1.26	0.00–5.31	NP2
		1.43	0.00–4.97	1.22	0.00–3.79	NP3
	z_{max}	1.12	0.89–1.35	1.08	0.88–1.24	NP1
		0.75	0.47–1.21	0.69	0.46–1.04	P1
		1.05	0.92–1.29	1.10	0.91–1.31	NP2
		1.16	0.97–1.52	1.10	0.88–1.33	NP3
	z_{min}	0.48	0.33–0.64	0.50	0.36–0.70	NP1
		0.43	0.25–0.73	0.44	0.28–0.77	P1
		0.78	0.65–0.93	0.74	0.64–0.82	NP2
		0.60	0.48–0.76	0.63	0.35–0.84	NP3
CPaP	overtime	0.54	0.00–1.87	0.48	0.00–2.86	NP1
		0.38	0.00–2.99	0.10	0.00–0.95	P1
		1.24	0.00–5.37	3.36	0.00–6.78	NP2
		1.71	0.00–4.47	1.63	0.00–5.21	NP3
	z_{max}	1.06	0.89–1.18	1.05	0.86–1.39	NP1
		0.92	0.49–1.30	0.72	0.45–1.10	P1
		1.13	0.96–1.72	1.33	0.98–1.65	NP2
		1.19	0.92–1.52	1.15	0.99–1.40	NP3
	z_{min}	0.38	0.26–0.51	0.49	0.37–0.68	NP1
		0.27	0.14–0.58	0.43	0.29–0.65	P1
		0.76	0.59–0.98	0.65	0.49–0.78	NP2
		0.54	0.38–0.69	0.61	0.48–0.84	NP3

is then evaluated by applying both the deterministic solutions and their robust counterparts to 10 sample paths of patient demands, i.e., realizations of uncertain patient demands, as in [12].

The *Mix* case is considered, which is the most realistic in practice. We evaluate the behavior in terms of the overtime cost and of the minimum and maximum utilization rates z_{min} and z_{max} in the scenarios over the weeks. Reassignments are not evaluated, as they do not change from the planned solution to the executions.

Results obtained by applying the solutions of the constrained models to the 10 sample paths are reported in Table 9, where the overtime costs, the maximum and the minimum utilization rates are given for both the deterministic and the robust case in terms of average, minimum and maximum value over weeks 8–25. They are separately presented for the four districts with more than one nurse (namely NP1, P1, NP2, NP3). P1 includes palliative patients, while NP1, NP2 and NP3 non-palliative ones (Table 3).

We first analyze the executed solutions of the deterministic models. Results show that the constrained models allow preserving good results while executing the solutions. As for the overtime cost, the best average values are obtained by the CNuP in three out of four districts, while the CPaP provides the best average value for NP1 district. It may seem counter-intuitive; however, it is worth noting that in general the values obtained by different objectives are very similar. The CNuP provides the best values for the fairness metrics, i.e., the lowest z_{max} in two out of four districts and the highest z_{min} in all districts, while the CPaP provides the lowest

Table 10
CPU times in seconds for the constrained deterministic and robust models for scenario *Mix* over weeks 8–25 (TL denotes the time limit).

	Deterministic			Robust		
	avg	min	max	avg	min	max
CPrC	1.10	0.97	1.40	819.24	8.46	TL
CNuP	1.08	0.86	1.54	24.26	4.03	66.35
CPaP	0.93	0.84	1.23	6.04	5.05	7.78

(best) z_{max} in NP1. Summing up, the CNuP provides slightly better results for both the overtime and the fairness metrics.

Then, we analyze the impact of the robustness by comparing the behavior of the deterministic and the robust executed solutions. Adding robustness slightly reduces the average overtime, which decreases in two out of four districts with the CPrP, and in three out of four with the CNuP and CPaP solutions. A particular behavior is observed in district NP2, where the robust solution provides higher overtime costs whatever the considered model is. Robust solutions also provide a lower maximum overtime for CPrP and CNuP in all district but in NP2. Thus, robustness proves to be effective and provide a reliable solution in terms of overtime when facing demand realizations, especially for CPrP and CNuP.

Robustness is demanding from the computational effort point of view. We compare the CPU time required to solve the deterministic and the robust counterparts of the constrained models in Table 10. Computational times significantly increase, from few seconds up to the one hour and a half time limit. However, the three models react in a different way to robustness. The highest increase of the average computational time is found for the CPrP (from 1.10 s to about 15 min), whereas the lowest increase is found for the CPaP (from about 1 s to about 6 s). However, the required CPU time is always compatible with the threshold setting procedure and with a weekly planning. Gaps when the time limit is reached over weeks 8–25 (twice only in the robust CPrC) are below the 0.35%.

7. Conclusions

In this paper we consider the mid-term nurse-to-patient assignment problem in HC service planning under different continuity of care requirements. Three stakeholders' perspectives are considered: the patients' perspective, which asks for meeting continuity of care requirements; the nurses' perspective, which asks for good levels of fairness; and the provider's perspective, which asks for reducing costs. We discuss a threshold method to evaluate the relations and the trade-offs among the perspectives, formalized into *metrics*, and to support the decision process while accounting for the contracting authority as well, which forces to guarantee a good service according to the perspectives of all stakeholders.

We compared the different formulations on real-life data from a large HC provider operating in the Northern Italy, and both certain and uncertain patient demands were considered in the experiments through deterministic model formulations and their robust counterparts.

We found that, for the considered instances, the overtime and the fairness metrics are somehow related, while the reassignment metrics is conflicting, and that cheap solutions are also fair, while the opposite is not always true. Moreover, results show that fairly good values of the fairness and the reassignment metrics can be obtained with a minimal increase of the cost metrics. This was not *a priori* expected and has an important practical impact. In fact, although the real decision maker is the service provider, nurses' and patients' perspectives can be accounted without increasing costs, thus pursuing an overall good quality of service for all the stakeholders, as required by the contracting authority.

Results also prove that the considered formulations can solve real-life instances in reasonable times, thus ensuring the practical applicability of the approach.

As discussed above, these results can be extended to several other HC providers. In fact, the main characteristics of our dataset (i.e., low ratio between patient's demand and nurse's working time, and non over-dimensioned staff) are common to the majority of HC providers. Future work will be done in extending the analysis to other simulated data that do not fit with these characteristics. Even far from the real HC cases, the outcomes from these instances could provide additional insight to the approach.

References

[1] A. Matta, S. Chahed, E. Sahin, Y. Dallery, Modelling home care organisations from an operations management perspective, Flex. Serv. Manuf. J. 26 (2014) 295–319.

[2] Medicare Payment Advisory Commission (MEDPAC), Home health services, in: Report to the Congress: Medicare payment policy, 2011 (Chapter 8).

[3] E. Lanzarone, A. Matta, The nurse-to-patient assignment problem in home care services, in: Testi, Tànfani (Eds.), Advanced Decision Making Methods Applied To Health Care, Springer 2012, 2012a, pp. 121–139.

[4] E. Lanzarone, A. Matta, E. Sahin, Operations management applied to home care services: The problem of assigning human resources to patients, IEEE Trans. Syst. Man Cybern. A 42 (2012) 1346–1363.

[5] J. Haggerty, R. Reid, G. Freeman, B. Starfield, C. Adair, R. McKendry, Continuity of care: A multidisciplinary review, Br. Med. J. 327 (2003) 1219–1221.

[6] C.A. Woodward, J. Abelson, S. Tedfort, B. Hutchison, What is important to continuity in home care? Perspectives of key stakeholders, Soc. Sci. Med. 58 (2004) 177–192.

[7] E. Lanzarone, A. Matta, Robust nurse-to-patient assignment in home care services to minimize overtimes under continuity of care, Oper. Res. Health Care 3 (2014) 48–58.

[8] C. Cordes, T. Dougherty, A review and an integration of research on job burnout, Acad. Manag. Rev. 18 (1993) 621–656.

[9] D. Bertsimas, M. Sim, The price of robustness, Oper. Res. 52 (2004) 35–53.

[10] B. Addis, G. Carello, A. Grosso, E. Lanzarone, S. Mattia, E. Tànfani, Handling uncertainty in health care management using the cardinality-constrained approach: Advantages and remarks, Oper. Res. Health Care 4 (2015) 1–4.

[11] P. Cappanera, M.G. Scutellà, F. Visintin, Home care services delivery: Equity versus efficiency in optimization models, in: Proceedings of the International Conference on Health Care Systems Engineering, HCSE2013, in: Springer Proceedings in Mathematics and Statistics, vol. 61, 2014, pp. 1–13.

[12] G. Carello, E. Lanzarone, A cardinality-constrained robust model for the assignment problem in home care services, European J. Oper. Res. 236 (2014) 748–762.

[13] M.M.S. Sodhi, Conceptualising social responsibility in operations via stakeholder resource-based view, Prod. Oper. Manage. 24 (2015) 1375–1389.

[14] C. Zopounidis, P.M. Pardalos (Eds.), Handbook of Multicriteria Analysis, Springer, 2010.

[15] K. Miettinen, Nonlinear multiobjective optimization, in: International Series in Operations Research and Management Science, vol. 12, Kluwer Academic Publishers, Dordrecht, 1999.

[16] M. Ehrgott, Multicriteria Optimization, Springer, Berlin, 2005.

[17] W. Stadler (Ed.), Multicriteria optimization in engineering and in the sciences, in: Mathematical Concepts and Methods in Science and Engineering, vol. 37, Plenum Press, New York, ISBN: 0-306-42743-5, 1988.

[18] L.A. Ocampo, E.E. Clark, A sustainable manufacturing strategy decision framework in the context of multi-criteria decision-making, Jordan J. Mech. Ind. Eng. 9 (2015) 177–186.

[19] M. Samà, C. Meloni, A. D'Ariano, F. Corman, A multi-criteria decision support methodology for real-time train scheduling, J. Rail Transp. Plan. Manag. 5 (2015) 146–162.

[20] J.C. Pérez, M.H. Carrillo, J.R. Montoya-Torres, Multi-criteria approaches for urban passenger transport systems: A literature review, Ann. Oper. Res. 226 (2015) 69–87.

[21] A. Soltani, K. Hewage, B. Reza, R. Sadiq, Multiple stakeholders in multi-criteria decision-making in the context of municipal solid waste management: A review, Waste Manag. 35 (2015) 318–328.

[22] B. Asl Rousta, S. Araghinejad, Development of a multi criteria decision making tool for a water resources decision support system, Water Res. Manag. 29 (2015) 5713–5727.

[23] V. Diaby, K. Campbell, R. Goeree, Multi-criteria decision analysis (MCDA) in health care: A bibliometric analysis, Oper. Res. Health Care 2 (2013) 20–24. I.

[24] Berrada, J.A. Ferland, P. Michelon, A multi-objective approach to nurse scheduling with both hard and soft constraints, Socio-Econ. Plan. Sci. 30 (1996) 183–193.

[25] F. Bellanti, G. Carello, F. Della Croce, R. Tadei, A greedy-based neighborhood search approach to a nurse rostering problem, European J. Oper. Res. 153 (2004) 28–40.

[26] J.F. Bard, H.W. Purnomo, Preference scheduling for nurses using column generation, European J. Oper. Res. 164 (2005) 510–534.

[27] D. Parr, J.M. Thompson, Solving the multi-objective nurse scheduling problem with a weighted cost function, Ann. Oper. Res. 155 (2007) 279–288.

[28] S.N. Ogulata, R. Erol, A hierarchical multiple criteria mathematical programming approach for scheduling general surgery operations in large hospitals, J. Med. Syst. 27 (2003) 259–270.

[29] S. Gul, B.T. Denton, J.W. Fowler, T. Huschka, Bi-criteria scheduling of surgical services for an outpatient procedure center, Prod. Oper. Manage. 20 (2011) 406–417.

[30] F. Ben Abdelaziz, M. Masmoudi, A multiobjective stochastic program for hospital bed planning, J. Oper. Res. Soc. 63 (2012) 530–538.

[31] P. Cappanera, F. Visintin, C. Banditori, Addressing conflicting stakeholders' priorities in surgical scheduling by goal programming, Flex. Serv. Manuf. J. (2017). http://dx.doi.org/10.1007/s10696-016-9255-5.

[32] I. Marques, M.E. Captivo, Different stakeholders' perspectives for a surgical case assignment problem: Deterministic and robust approaches, European J. Oper. Res. 261 (2017) 260–278.

[33] A. Trautsamwieser, P. Hirsch, Optimization of daily scheduling for home health care services, J. Appl. Oper. Res. 3 (2011) 124–136.

[34] M.S. Rasmussen, T. Justesen, A. Dohn, J. Larsen, The home care crew scheduling problem: Preference-based visit clustering and temporal dependencies, European J. Oper. Res. 219 (2012) 598–610.

[35] S. Nickel, M. Schroder, J. Steeg, Mid-term and short-term planning support for home health care services, European J. Oper. Res. 219 (2012) 574–587.

[36] J. Bowers, H. Cheyne, G. Mould, M. Page, Continuity of care in community midwifery, Health Care Manag. Sci. 18 (2014) 195–204.

[37] S. Yalçındağ, A. Matta, E. Şahin, J.G. Shanthikumar, A two-stage approach for solving assignment and routing problems in home health care services, in: Springer Proceedings in Mathematics and Statistics, vol. 61, 2014, pp. 47–59.

[38] S. Yalçındağ, P. Cappanera, M.G. Scutella, E. Şahin, A. Matta, Pattern-based decompositions for human resource planning in home health care services, Comput. Oper. Res. 73 (2016) 12–26.

[39] K. Braekers, R.F. Hartl, S.N. Parragh, F. Tricoire, A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience, European J. Oper. Res. 248 (2016) 428–443.

[40] E. Lanzarone, A. Matta, A cost assignment policy for home care patients, Flex. Serv. Manuf. J. 24 (2012b) 465–495.

[41] E. Lanzarone, A. Matta, G. Scaccabarozzi, A patient stochastic model to support human resource planning in home care, Prod. Plan. Control 21 (2010) 3–25.