
ELSEVIER
Volume 236, Issue 2, 16 July 2014
ISSN 0377-2217
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
Available online at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/ejor

European Journal of Operational Research 236 (2014) 748–762

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

A cardinality-constrained robust model for the assignment problem in Home Care services

Giuliana Carello [a], Ettore Lanzarone [b,*]

[a] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, P.za Leonardo da Vinci 32, 20133 Milan, Italy
[b] Istituto di Matematica Applicata e Tecnologie Informatiche (IMATI), Consiglio Nazionale delle Ricerche (CNR), Via Bassini 15, 20133 Milan, Italy

ARTICLE INFO

Article history:
Received 19 March 2013
Accepted 3 January 2014
Available online 17 January 2014

Keywords:
Home Care
Robust optimization
Nurse-to-patient assignment
Continuity of care

ABSTRACT

Home Care includes medical, paramedical and social services which are delivered to patients at their domicile rather than in hospital. Managing human and material resources in Home Care services is a difficult task, as the provider has to deal with peculiar constraints (e.g., the continuity of care, which imposes that a patient is always cared for by the same nurse) and to manage the high variability of patients' demands. One of the main issues encountered in planning Home Care services under continuity of care requirement is the nurse-to-patient assignment. Despite the importance of this topic, the problem is only marginally addressed in the literature, where continuity of care is usually treated as a soft-constraint rather than as a hard one. Uncertainty is another relevant feature of nurse-to-patient assignment problem, and it is usually managed adopting stochastic programming or analytical policies. However, both these approaches proved to be limited, even if they improve the quality of the assignments upon those actually provided in practice. In this paper, we develop a cardinality-constrained robust assignment model, which allows exploiting the potentialities of a mathematical programming model without the necessity of generating scenarios. The developed model is tested on real-life instances related to a relevant Home Care provider operating in Italy, in order to evaluate its capability of reducing the costs related to nurses' overtimes.

1. Introduction

Randomness in data and parameters is a common feature of several optimization problems. It arises in many applications, such as telecommunications, where uncertainty is usually associated with traffic demands, and health care, where a high uncertainty is related to patients' conditions and demands.

Indeed, uncertainty is inherent in many health care optimization problems and cannot be neglected, as it may have a significant impact on quality and feasibility of problem solution. For instance, in emergency vehicle location problems, uncertainty is associated with the availability of ambulances (Brotcorne, Laporte, & Semet, 2003), whereas in planning and scheduling operating room theaters uncertainty is due to surgery durations (Denton, Miller, Balasubramanian, & Huschka, 2010) or care demands in different specialties (Holte & Mannino, 2013). Uncertainty also occurs in managing Home Care (HC) services.

HC providers must synchronize the use of resources at patient's domicile, while usually delivering the service to a large number of patients in a vast territory. Furthermore, random events may affect service delivery, undermine feasibility of plans, and cause a high variability in nurses' workloads and, consequently, in the cost of service. One of the most critical and frequent of such events is a sudden variation in the amount of service required by patients, which is in general highly variable. Hence, managing human resources in HC services is a difficult task, which is made more complex by uncertain patient demands.

Different approaches are usually applied to deal with uncertainty in health care problems, such as probabilistic models or stochastic optimization approaches.

Recently, an innovative robust optimization approach, named cardinality-constrained approach, has been proposed in Bertsimas and Sim (2004) and already applied in several fields, such as portfolio optimization or network design. It allows to account for a certain degree of uncertainty with a reasonable computational effort, providing a trade-off between computational time and robustness. In addition, it can be tuned to take into account the specific degree of risk the decision maker accepts. Although the approach seems to fit well to many health care problems, and in particular to the HC

* Corresponding author. Tel.: +39 02 2369 9531.
E-mail addresses: giuliana.carello@polimi.it (G. Carello), ettore.lanzarone@cnr.it (E. Lanzarone).

0377-2217/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2014.01.009

planning, to the best of our knowledge it has not been applied to many health care problems so far.[1]

In this paper we present an application of the above mentioned robust approach to the nurse-to-patient assignment problem in HC, which is an interesting and meaningful example of health care optimization problem. The approach can be easily applied to the problem and proved to be able to produce good quality solutions with a reasonable computational effort. Hence, it is also worth being tested on other health care problems.

1.1. Problem description

HC consists of delivering medical, paramedical and social services to patients at their domicile rather than in a hospital. On the one hand this leads to a significant improvement in the quality of life for patients, as they continue to live at their home in a familiar environment. On the other hand it leads to considerable cost savings for the health care system, as hospitalization costs are avoided. HC is a relevant and growing sector in western countries, due to population aging, increase in chronic pathologies, introduction of innovative technologies, and pressure of governments to contain health care costs.

HC services involve different professional figures. Patients are usually assisted by nurses and, according to the case, may require also other figures, such as physiotherapist, physician, or psychologist. In some cases, they may require specific facilities to be brought and managed at home.

Hence, many resources are involved in delivering HC services, including nurses and other categories of operators, support staff and material resources. HC providers must carefully optimize limited human resources. In addition some features, such as the continuity of care and the risk of incurring in operator burnout, make the HC resource planning different from the planning problems that occur in other services within the health care domain. Besides, as HC patients have to be cared after for weeks, middle and long-term planning are key issues in providing a good quality of service, while keeping balanced personnel workloads over the whole considered time horizon.

In this work, we focus on the nurse-to-patient assignment under continuity of care. Continuity of care means that an HC provider assigns only one nurse to each patient, named the *reference nurse*, and the assignment is kept for a long period. Continuity of care is an important factor in service provision in all health care facilities (Haggerty et al., 2003; Heffernan & Husni, 2009). Two main aspects are preserved when continuity of care is pursued. On the one hand, continuity of care preserves the quality of service perceived by patients, as patients receive care from the same operator and do not have to continuously develop new relationships with new operators. On the other hand, potential loss of information among operators is avoided. Both these aspects are extremely important in HC, where the patient is assisted for a long period, patient's family and domestic context are involved in providing care, and, for palliative care, very serious pathologies are treated. Hence, many HC providers pursue continuity of care. However, continuity of care constraint limits service flexibility and, despite of quality of care, some providers do not adopt reference nurses to increase operational efficiency. In general, in order to guarantee a good tradeoff between quality and flexibility, continuity of care should be preserved at least for critical patients (e.g., palliative patients) or patients with particular needs. In other words, patients should be classified according to their needs and continuity of care should be guaranteed for a subset of them.

Nurses are usually divided into districts depending on skills and territory. Patients are divided into classes, according to the type of service required (e.g., the main classification is between palliative and non-palliative patients), and nurses take care of patients they are skilled for. Furthermore, in large HC providers, nurses are divided into territorial groups and take care of patients who are resident in their territory. Districts are usually assumed to be independent in planning the assignments, i.e., each patient is cared for by nurses with a skill compatible to his/her pathology and working in his/her geographical area.

Each patient requires a number of visits per week from each professional figure, which is described by an uncertain parameter. Each operator has an amount of contracted working time per time period for providing visits. If the care volume exceeds the operator contract capacity, overtime has an extra cost that providers want to minimize. Hence, the assignment aims at reducing the visits that each operator provides above his/her contracted time.

1.2. State of the art

Home Care involves the management of several resources and must take into account many requirements and constraints. It is related to nurse rostering in hospitals (see Burke, De Causmaecker, Vanden Berghe, & Van Landeghem, 2004; Cheang, Li, Lim, & Rodrigues, 2003) as it deals with nurse management. However, HC management involves several issues which are not usually addressed in nurse rostering problems, such as the continuity of care (Haggerty et al., 2003) or the burnout risk (Borsani, Matta, Beschi, & Sommaruga, 2006), which make the HC nurse management peculiar. The main issues to be considered are the partitioning of a territory into districts, the dimensioning of human resources, the assignment of visits to operators (or patients to operators in case of continuity of care), the scheduling of nurses' duties and the routing optimization.

Literature about HC can be mainly divided into two groups: a first group deals with daily schedule of visits and routing of nurses, and a second group deals with staff planning and management in a mid-term and long-term perspective. The nurse-to-patient assignment under continuity of care analyzed in this paper is related to the mid-term management.

From a long-term point of view, the districting problem consists of grouping patients and nurses according to geographical and skill compatibility, usually aiming at balancing workload among different groups (Blais, Lapierre, & Laporte, 2003; Lahrichi, Lapierre, Hertz, Talib, & Bouvier, 2006).

Dimensioning of human resources consists of determining the number of operators, together with their skills, to meet patient demand in each part of the territory. Patient demand uncertainty may be taken into account (De Angelis, 1998). Funding has to be taken into account in dimensioning HC resources, as well (Busby & Carter, 2006).

The nurse-to-patient assignment problem consists of assigning personnel to visits in a fair way. Different features can be considered, such as the continuity of care and the uncertainty in patients' demands. It has been rarely studied as a stand-alone problem, i.e. not considering scheduling (Boldy & Howell, 1980), and, to the best of our knowledge, the assignment problem taking into account the continuity of care issue is only marginally addressed in literature (Borsani et al., 2006; Hertz & Lahrichi, 2009; Lanzarone, Matta, & Sahin, 2012). Besides, continuity of care is often considered as an objective rather than a strict requirement and, therefore, dealt with as a soft-constraint rather than as a hard one, see for instance Nickel, Schroder, and Steeg (2012).

[1] There are the only five papers (i.e., Banditori, Cappanera, & Visintin, 2013; Chan, Bortfeld, & Tsitsiklis, 2006; Denton et al., 2010; Holte & Mannino, 2013; Mannino, Nilssen, & Nordlander, 2012) found in September 2013 through a search on ISI web of knowledge and Scopus, referring to Health Care, among the papers citing Bertsimas and Sim (2004). However, none deals with HC management.

If continuity of care is not considered, the assignment problem turns out to be an assignment of operators to visits rather than patients; in this case the aim is to jointly optimize the assignment of operators to visits and the scheduling and routing problem (see Rasmussen, Justesen, Dohn, & Larsen, 2012; Trautsamwieser & Hirsch, 2011). However, this approach has two drawbacks. On the one hand, if the districts are given and have a limited extensions, as it usually is the case in Europe, the travel distances can be considered as constant and therefore the impact of scheduling and routing is not very significant. On the other hand, joint optimization requires significant computational effort, thus reducing the length of the considered time horizon and reducing the possibility of long-term workload balancing.

As mentioned, uncertainty inherently arises in HC due to unpredictable changes in patients' needs. It affects personnel workloads and the number of patients who can be treated. In Koeleman, Bhulai, and Van Meersbergen (2012) uncertainty is managed by representing the whole system as a Markov chain and developing admittance policies for patients. The nurse-to-patients assignment problem, in which both continuity of care and demand uncertainty are considered, has been rarely addressed in literature. The problem was tackled with stochastic programming on one hand (Lanzarone et al., 2012), and with analytical policies on the other (Lanzarone & Matta, 2012). However, both these approaches proved to be limited even if they improved the quality of the assignment upon those actually provided by the HC structures in practice. The stochastic programming approach is based on scenario generation and, due to the high number of patients and the associated demand variability, requires to include a very high number of scenarios. Only a limited number of them can be considered for a computationally acceptable solution and, therefore, a high expected value of perfect information (EVPI) and a low value of stochastic solution (VSS) are obtained (Lanzarone et al., 2012). Analytical policies are related to strict assumptions regarding the shape of workload probability density functions, the number of assignable patients (one patient at a time after deciding an ordering of new patients) and the number of periods in the planning horizon, which is equal to only one (Lanzarone & Matta, 2012).

With the cardinality-constrained approach, we aim at exploiting the potentialities of a linear programming model rather than an analytical approach, avoiding the necessity of generating scenarios.

2. Robust assignment model

We consider the problem of assigning a set of HC patients P to a set of nurses I over a time horizon T represented by a set of time slots.

Continuity of care is taken into account; indeed, different continuity of care requirements are considered, depending on the type of patient and on his/her requests. Hence, the set of patients P is partitioned into five subsets:

1. **Patients who require hard continuity of care (*C*)**
   - P_c^a: set of patients who require hard continuity of care – i.e., their reference nurse cannot be changed – and are already under treatment (and therefore assigned) at the beginning of the considered time horizon. As they cannot be reassigned, they keep their assignment as it was in the past.
   - P_c^n: set of patients who require hard continuity of care and start their treatment during the considered time horizon (therefore, they are not yet assigned). A subset of them begin the treatment at the first time slot. Furthermore, other patients may enter in each of the following time slots.
2. **Patients who require partial continuity of care (*PC*)**
   - P_{pc}^a: set of patients who require partial continuity of care – i.e. their reference nurse can be changed although it is preferable not to – and are already under treatment (and therefore assigned) at the beginning of the considered time horizon. They can be reassigned at the beginning of each time slot t, in order to balance workloads. However, each reassignment is penalized by a cost γ to keep the number of reassignments limited.
   - P_{pc}^n: set of patients who require partial continuity of care and start their treatment during the considered time horizon (therefore, they are not yet assigned). A subset of them begin the treatment at the first time slot, and other patients may enter in each of the following time slots. Similarly to patients in P_{pc}^a, they can be reassigned at the beginning of each time slot t with a reassignment cost γ.
3. **Patients who do not require continuity of care (*NC*)**
   - P_{nc}: set of patients who do not require continuity of care. They can be assigned to more than one nurse even in the same time slot t and their assignments can be changed from a time slot to another without reassignment costs.

The division in districts (usually based on territory and skills) is taken into account, i.e., each district is assigned to a subset of nurses. A parameter m_{ij} is given for each nurse $i \in I$ and patient $j \in P$, which is equal to 1 if i operates in the district of j, and 0 otherwise.

The amount of working time required by each patient $j \in P$ during each time slot $t \in T$ is an uncertain parameter r_{jt}, with expected value $\bar{r}_{jt}$ and maximum value $\bar{r}_{jt} + \hat{r}_{jt}$. Further, each nurse $i \in I$ has an amount of available working time υ_i in each time slot t. If the workload amount of a nurse exceeds the available time, overtime must be paid for.

An objective usually pursued in practice is the reduction of overtime. A possible approach to pursue this goal is to minimize first the highest overtime value, then the second highest, and so on. For an analytical description, overtime cost is modeled as a stepwise function, where highest per time unit overtime costs are associated with highest overtime values. For this purpose, a set of overtime levels L_i is defined for each nurse $i \in I$ and two parameters are given for each level $l \in L_i$: a threshold Δ_i^l and a cost per time unit c_l, where c_l is the cost for each overtime unit above $\upsilon_i + \sum_{k=1}^{l-1} \Delta_i^k$ and below $\upsilon_i + \sum_{k=1}^{l} \Delta_i^k$. As example, the cost for each overtime unit above υ_i and below a first threshold $\upsilon_i + \Delta_i^1$ is c_1, the cost for each overtime unit above $\upsilon_i + \Delta_i^1$ and below a second threshold $\upsilon_i + \Delta_i^1 + \Delta_i^2$ is c_2, and so on. Thresholds Δ_i^l may be different for different nurses, whereas costs c_l are usually the same and $c_l < c_{l+1}$ to get a monotonically increasing stepwise function.

A limit is also given to the total workload amount of each nurse $i \in I$, which cannot exceeds twice the value of υ_i. Hence, parameters Δ_i^l are defined in such a way that $\sum_{l \in L_i} \Delta_i^l = \upsilon_i, \ \forall i$. Although this limit is very high with respect to the contracted working time υ_i, indeed such limit allows to deal with the robustness by means of the cardinality-constrained approach, as described later in this section.

The assignment of patients requiring continuity of care is described by binary variables. A binary variable x_{ji} is defined for each patient $j \in P_c^a \cup P_c^n$ who requires hard continuity of care and each nurse $i \in I : x_{ji}$ is equal to 1 if j is assigned to i during the whole time horizon, and 0 otherwise. A binary variable ξ_{ji}^t is defined for each patient $j \in P_{pc}^a \cup P_{pc}^n$ who requires partial continuity of care, each nurse $i \in I$ and each time slot $t \in T : \xi_{ji}^t$ is equal to 1 if i is in the charge of j during t, and 0 otherwise. Furthermore, a binary variable y_j^t is introduced for each patient $j \in P_{pc}^a \cup P_{pc}^n$ who requires partial continuity of care and each time slot t. y_j^t is equal to 1 if the assignment of j is changed from $t-1$ to t, and 0 otherwise.

The assignments of patients $\in P_c^a \cup P_{pc}^a$ to reference nurses before the beginning of the considered time horizon are described by parameters $\tilde{x}_{ji}$: $\tilde{x}_{ji}$ is equal to 1 if j is initially assigned to i, and 0 otherwise.[2]

As for patients not requiring continuity of care, the fraction of the time needed by patient $j \in P_{nc}$ during time slot $t \in T$ which is provided by nurse $i \in I$ is represented by a continuous variable $\chi_{ji}^t \in [0, 1]$.

The overtime amount assigned to each nurse $i \in I$ during each time slot $t \in T$ is described by a continuous variable w_{it}^l for each level $l \in L_i$, which represents the overtime related to cost c_l.

Sets, parameters and variables of the model are summarized in Table 1.

The objective function, which aims at minimizing the overall overtime costs and the cost associated to reassignments, is written as follows:

$$\min \sum_{i \in I} \sum_{t \in T} \sum_{l \in L_i} c_l w_{it}^l + \gamma \sum_{j \in P_{pc}^a \cup P_{pc}^n} \sum_{t \in T} y_j^t. \tag{1}$$

Variables are subject to the following constraints:

$$\sum_{i \in I} m_{ij} x_{ji} = 1, \quad \forall j \in P_c^a \cup P_c^n \tag{2}$$

$$\sum_{i \in I} m_{ij} \xi_{ji}^t = 1, \quad \forall j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T \tag{3}$$

$$\sum_{i \in I} m_{ij} \chi_{ji}^t = 1, \quad \forall j \in P_{nc}, \quad t \in T \tag{4}$$

$$\sum_{j \in P_c^a \cup P_c^n} r_{jt} x_{ji} + \sum_{j \in P_{pc}^a \cup P_{pc}^n} r_{jt} \xi_{ji}^t + \sum_{j \in P_{nc}} r_{jt} \chi_{ji}^t \leqslant \upsilon_i + \sum_{l \in L_i} w_{it}^l, \quad \forall i \in I, \ t \in T \tag{5}$$

$$0 \leqslant w_{it}^l \leqslant \Delta_i^l, \quad \forall i \in I, \quad t \in T, \quad l \in L_i \tag{6}$$

$$x_{ji} \geqslant \tilde{x}_{ji}, \quad \forall i \in I, \quad j \in P_c^a \tag{7}$$

$$y_j^t \geqslant \xi_{ji}^t - \xi_{ji}^{t-1}, \quad \forall t \in T \setminus \{t_1\}, \quad j \in P_{pc}^a \cup P_{pc}^n, \quad i \in I \tag{8}$$

$$y_j^{t_1} \geqslant \xi_{ji}^{t_1} - \tilde{x}_{ji}, \quad \forall j \in P_{pc}^a, \quad i \in I \tag{9}$$

$$x_{ji} \in \{0, 1\}, \quad \forall j \in P_c^a \cup P_c^n, \quad i \in I \tag{10}$$

$$\xi_{ji}^t \in \{0, 1\}, \quad \forall j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T, \quad i \in I \tag{11}$$

$$\chi_{ji}^t \in [0, 1], \quad \forall j \in P_{nc}, \quad t \in T, \quad i \in I \tag{12}$$

$$y_j^t \in \{0, 1\}, \quad \forall j \in P_{nc}, \quad t \in T \tag{13}$$

Constraints (2)–(4) guarantee that patients are assigned to suitable nurses, taking into account district compatibility and the pertinent continuity of care requirement. Constraints (5) compute nurse workload (left hand side) and overtime, which is divided into the different cost levels (right hand side). Constraints (6) set the maximum overtime workload for each level. Constraints (7) guarantee that patients belonging to subset P_c^a do not change their assignment at the beginning of the considered time horizon. Finally, constraints (8) and (9) compute the number of reassignments. In particular, constraints (9) compute the number of reassignments at the initial time slot t_1 for patients in the subset P_{pc}^a.

As mentioned, we consider the cardinality-constrained approach (Bertsimas & Sim, 2004) to deal with uncertainty. Based on such approach, we introduce three cardinality parameters $\Gamma_c^i, \Gamma_{pc}^i$ and Γ_{nc}^i for each nurse $i \in I$. We also introduce three sets S_c^{it}, S_{pc}^{it} and S_{nc}^{it}, and three patients p_c^{it}, p_{pc}^{it} and p_{nc}^{it} for each nurse $i \in I$ and each time slot $t \in T$.

$S_c^{it} \subseteq P_c^a \cup P_c^n$ is the subset of patients requiring hard continuity of care and assigned to i, whose demand in time slot t is equal to the maximum treatment time (i.e., $\bar{r}_{jt} + \hat{r}_{jt}$). $S_{pc}^{it} \subseteq P_{pc}^a \cup P_{pc}^n$ and $S_{nc}^{it} \subseteq P_{nc}$ are analogously defined for patients requiring partial and no continuity of care, respectively. At most $\lfloor \Gamma_c^i \rfloor$ and $\lfloor \Gamma_{pc}^i \rfloor$ and $\lfloor \Gamma_{nc}^i \rfloor$ patients (with hard, partial and no continuity of care) are assumed to belong to these subsets, respectively.

[2] We could remove some variables x_{ji} and replace them with a fixed charge. However, variables x_{ji} are needed to deal with robustness.

Table 1
Sets, parameters and variables.

Sets	
T	Time horizon
P	Patients
P_c^n	Patients requiring hard continuity of care and already under treatment
P_c^a	Patients requiring hard continuity of care and not under treatment
P_{pc}^n	Patients requiring partial continuity of care and already under treatment
P_{pc}^a	Patients requiring partial continuity of care and not under treatment
P_{nc}	Patients requiring no continuity of care
I	Nurses
L_i	Overtime levels
Parameters	
r_{jt}	Demand of patient j in time slot t (expected value $\bar{r}_{jt}$, maximum increase $\hat{r}_{jt}$)
m_{ij}	Skill and territory based coverage parameter
υ_i	Nurse i's available working time per time slot
Δ_i^l	Threshold of overtime level l
c_l	Cost of overtime level l
$\tilde{x}_{ji}$	Patient j's initial assignment to nurse i
γ	Reassignment cost
Variables	
x_{ji}	Assignment variables for hard continuity of care patients
ξ_{ji}^t	Assignment variables for partial continuity of care patients
χ_{ji}^t	Assignment variables for no continuity of care patients
y_j^t	Reassignment variables
w_{it}^l	Nurse i's workload at time slot t related to overtime level l

Further, in case Γ_c^i and Γ_{pc}^i and Γ_{nc}^i are not integer, three patients are selected for each nurse i and each time slot t, whose demand charged to i is between $\bar{r}_{jt}$ and $\bar{r}_{jt} + \hat{r}_{jt}$. We denote with p_c^{it}, p_{pc}^{it} and p_{nc}^{it} such patients: p_c^{it} is a patient belonging to $P_c^a \cup P_c^n$ but not to S_c^{it}, p_{pc}^{it} is a patient belonging to $P_{pc}^a \cup P_{pc}^n$ but not to S_{pc}^{it}, and p_{nc}^{it} is a patient belonging to P_{nc} but not to S_{nc}^{it}.

The charged demand of patients not belonging to S_c^{it}, S_{pc}^{it} and S_{nc}^{it} and different from p_c^{it}, p_{pc}^{it} and p_{nc}^{it} is the expected treatment time (i.e., $\bar{r}_{jt}$).

In this formulation, the approach proposed in Bertsimas and Sim (2004) is slightly modified. Whereas only one subset of parameters is selected for each constraint in Bertsimas and Sim (2004), in our model three subsets are selected for each constraint (i.e., for each nurse) as three disjoint set of patients may charge each nurse. It is worth noting that the computed solutions are feasible if at most $(\Gamma_c^i + \Gamma_{pc}^i + \Gamma_{nc}^i)|I|$ patients require their maximum number of visits simultaneously.

The robustness is taken into account considering the worst possible charge for each nurse i in each time slot t, given cardinalities $\Gamma_c^i, \Gamma_{pc}^i$ and Γ_{nc}^i. Hence, the model is modified in order to include such worst case in constraints (5):

$$\begin{aligned}
&\sum_{j \in P_c^a \cup P_c^n} \bar{r}_{jt} x_{ji} + \sum_{j \in P_{pc}^a \cup P_{pc}^n} \bar{r}_{jt} \xi_{ji}^t + \sum_{j \in P_{nc}} \bar{r}_{jt} \chi_{ji}^t \\
&+ \max_{\substack{S_c^{it} \cup \{p_c^{it}\} | S_c^{it} \subseteq P_c^a \cup P_c^n, \\ |S_c^{it}| = \lfloor \Gamma_c^i \rfloor, \, p_c^{it} \in P_c^a \cup P_c^n \setminus S_c^{it}}} \left\{ \sum_{j \in S_c^{it}} \hat{r}_{jt} x_{ji} + \left(\Gamma_c^i - \lfloor \Gamma_c^i \rfloor \right) \hat{r}_{p_c^{it} t} x_{p_c^{it} i} \right\} \\
&+ \max_{\substack{S_{pc}^{it} \cup \{p_{pc}^{it}\} | \\ S_{pc}^{it} \subseteq P_{pc}^a \cup P_{pc}^n, |S_{pc}^{it}| = \lfloor \Gamma_{pc}^i \rfloor, \\ p_{pc}^{it} \in P_{pc}^a \cup P_{pc}^n \setminus S_{pc}^{it}}} \left\{ \sum_{j \in S_{pc}^{it}} \hat{r}_{jt} \xi_{ji}^t + \left(\Gamma_{pc}^i - \lfloor \Gamma_{pc}^i \rfloor \right) \hat{r}_{p_{pc}^{it} t} \xi_{p_{pc}^{it} i}^t \right\} \\
&+ \max_{\substack{S_{nc}^{it} \cup \{p_{nc}^{it}\} | S_{nc}^{it} \subseteq P_{nc}, \\ |S_{nc}^{it}| = \lfloor \Gamma_{nc}^i \rfloor, \, p_{nc}^{it} \in P_{nc} \setminus S_{nc}^{it}}} \left\{ \sum_{j \in S_{nc}^{it}} \hat{r}_{jt} \chi_{ji}^t + \left(\Gamma_{nc}^i - \lfloor \Gamma_{nc}^i \rfloor \right) \hat{r}_{p_{nc}^{it} t} \chi_{p_{nc}^{it} i}^t \right\} \\
&\leqslant \upsilon_i + \sum_{l \in L_i} w_{it}^l, \quad \forall i \in I, \quad t \in T
\end{aligned} \tag{14}$$

Patients in subsets S_c^{it}, S_{pc}^{it} and S_{nc}^{it} and patients p_c^{it}, p_{pc}^{it} and p_{nc}^{it} may change in consecutive time slots. This means that the level of charged demand for patient j in a time slot t is independent from the value in the previous time slot $t-1$ (null covariance between the events). Even if unrealistic, this assumption increases the robustness of the obtained solution as the worst scenario is found out and the solution is guaranteed to be able to deal with it.

Let us denote with $\beta_c^{it}(x^*, \Gamma_c^i, t), \beta_{pc}^{it}(\xi^*, \Gamma_{pc}^i, t)$ and $\beta_{nc}^{it}(\chi^*, \Gamma_{nc}^i, t)$ the three maxima included in (14), which are associated with a given solution $\{x^*, \xi^*, \chi^*\}$. They are computed independently through three linear programming models:

$$\beta_c^{it}(x^*, \Gamma_c^i, t) = \max_{\substack{S_c^{it} \cup \{p_c^{it}\} | S_c^{it} \subseteq P_c^a \cup P_c^n, \\ |S_c^{it}| = \lfloor \Gamma_c^i \rfloor, \, p_c^{it} \in P_c^a \cup P_c^n \setminus S_c^{it}}} \left\{ \sum_{j \in S_c^{it}} \hat{r}_{jt} x_{ji}^* + \left(\Gamma_c^i - \lfloor \Gamma_c^i \rfloor \right) \hat{r}_{p_c^{it} t} x_{p_c^{it} i}^* \right\} \tag{15}$$

$$\beta_{pc}^{it}(\xi^*, \Gamma_{pc}^i, t) = \max_{\substack{S_{pc}^{it} \cup \{p_{pc}^{it}\} | \\ S_{pc}^{it} \subseteq P_{pc}^a \cup P_{pc}^n, |S_{pc}^{it}| = \lfloor \Gamma_{pc}^i \rfloor, \\ p_{pc}^{it} \in P_{pc}^a \cup P_{pc}^n \setminus S_{pc}^{it}}} \left\{ \sum_{j \in S_{pc}^{it}} \hat{r}_{jt} \xi_{ji}^{*t} + \left(\Gamma_{pc}^i - \lfloor \Gamma_{pc}^i \rfloor \right) \hat{r}_{p_{pc}^{it} t} \xi_{p_{pc}^{it} i}^{*t} \right\} \tag{16}$$

$$\beta_{nc}^{it}(\chi^*, \Gamma_{nc}^i, t) = \max_{\substack{S_{nc}^{it} \cup \{p_{nc}^{it}\} | S_{nc}^{it} \subseteq P_{nc}, \\ |S_{nc}^{it}| = \lfloor \Gamma_{nc}^i \rfloor, \, p_{nc}^{it} \in P_{nc} \setminus S_{nc}^{it}}} \left\{ \sum_{j \in S_{nc}^{it}} \hat{r}_{jt} \chi_{ji}^{*t} + \left(\Gamma_{nc}^i - \lfloor \Gamma_{nc}^i \rfloor \right) \hat{r}_{p_{nc}^{it} t} \chi_{p_{nc}^{it} i}^{*t} \right\} \tag{17}$$

As example, $\beta_c^{it}(x^*, \Gamma_c^i, t)$ is computed for each nurse $i \in I$ and each time slot $t \in T$ by solving the following linear programming problem:

$$(\mathcal{P}_c^{\beta it}) = \max \sum_{j \in P_c^a \cup P_c^n} \hat{r}_{jt} x_{ji}^* z_{ji}^t \tag{18}$$

$$\sum_{j \in P_c^a \cup P_c^n} z_{ji}^t \leqslant \Gamma_c^i \tag{19}$$

$$0 \leqslant z_{ji}^t \leqslant 1, \quad \forall j \in P_c^a \cup P_c^n \tag{20}$$

where i and t are fixed, and $z_{ji}^t \in [0,1]$ are continuous variables which represent the choice of elements in subset S_c^{it} and of p_c^{it}.[3]

Let us denote with ζ_{it}^c the dual variables associated to (19) and with π_{jit}^c the dual variables associated to $z_{ji}^t \leqslant 1$ (20). The dual problem is formulated as follows:

$$(\mathcal{D}_c^{\beta it}) = \min \quad \Gamma_c^i \zeta_{it}^c + \sum_{j \in P_c^a \cup P_c^n} \pi_{jit}^c \tag{21}$$

$$\zeta_{it}^c + \pi_{jit}^c \geqslant \hat{r}_{jt} x_{ji}^*, \quad \forall j \in P_c^a \cup P_c^n \tag{22}$$

$$\pi_{jit}^c \geqslant 0, \quad \forall j \in P_c^a \cup P_c^n \tag{23}$$

$$\zeta_{it}^c \geqslant 0 \tag{24}$$

Optimal values $(\mathcal{P}_c^{\beta it})$ and $(\mathcal{D}_c^{\beta it})$ coincide and, therefore, the forth addend of the left hand side of (14) can be replaced by $\Gamma_c^i \zeta_{it}^c + \sum_{j \in P_c^a \cup P_c^n} \pi_{jit}^c$ with the following variables and constraints added to the model:

$$\zeta_{it}^c + \pi_{jit}^c \geqslant \hat{r}_{jt} x_{ji}, \quad \forall i \in I, \quad j \in P_c^a \cup P_c^n, \quad t \in T$$

$$\zeta_{it}^c \geqslant 0, \quad \forall i \in I, \quad t \in T$$

$$\pi_{jit}^c \geqslant 0, \quad \forall i \in I, \quad j \in P_c^a \cup P_c^n, \quad t \in T$$

[3] Due to the structure of the model, at most one of variables z_{ji}^t is fractional in any optimal solution, if Γ_c^i is not integer, and represents p_c^{it}.

The same idea is applied to $\beta_{pc}^{it}(\xi^*, \Gamma_{pc}^i, t)$ and $\beta_{nc}^{it}(\chi^*, \Gamma_{nc}^i, t)$, deriving the primal problems $(\mathcal{P}_{pc}^{\beta it})$ and $(\mathcal{P}_{nc}^{\beta it})$ and the corresponding dual problems $(\mathcal{D}_{pc}^{\beta it})$ and $(\mathcal{D}_{nc}^{\beta it})$. As the three subsets of patients are disjoint, the worst case is obtained by maximizing the three primal problems independently.

Hence, the fifth addend of the left hand side of (14) is replaced by $\Gamma_{pc}^i \zeta_{it}^{pc} + \sum_{j \in P_{pc}^a \cup P_{pc}^n} \pi_{jit}^{pc}$ with the following variables and constraints added to the model:

$$\zeta_{it}^{pc} + \pi_{jit}^{pc} \geqslant \hat{r}_{jt} \xi_{ji}^t, \quad \forall i \in I, \quad j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T$$

$$\zeta_{it}^{pc} \geqslant 0, \quad \forall i \in I, \quad t \in T$$

$$\pi_{jit}^{pc} \geqslant 0, \quad \forall i \in I, \quad j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T$$

The sixth addend of the left hand side of (14) is replaced by $\Gamma_{nc}^i \zeta_{it}^{nc} + \sum_{j \in P_{nc}} \pi_{jit}^{nc}$ with the following variable and constraints added to the model:

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

$$\zeta_{it}^{nc} \geqslant 0, \quad \forall i \in I, \quad t \in T$$

$$\pi_{jit}^{nc} \geqslant 0, \quad \forall i \in I, \quad j \in P_{nc}, \quad t \in T$$

Summarizing, the robust cardinality-constrained model is defined by the objective function of (1) and the following set of constraints:

$$\sum_{i \in I} m_{ij} x_{ji} = 1, \quad \forall j \in P_c^a \cup P_c^n$$

$$\sum_{i \in I} m_{ij} \xi_{ji}^t = 1, \quad \forall j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T$$

$$\sum_{i \in I} m_{ij} \chi_{ji}^t = 1, \quad \forall j \in P_{nc}, \quad t \in T$$

$$\Gamma_c^i \zeta_{it}^c + \sum_{j \in P_c^a \cup P_c^n} \bar{r}_{jt} x_{ji} + \sum_{j \in P_c^a \cup P_c^n} \pi_{jit}^c + \Gamma_{pc}^i \zeta_{it}^{pc} + \sum_{j \in P_{pc}^a \cup P_{pc}^n} \bar{r}_{jt} \xi_{ji}^t + \sum_{j \in P_{pc}^a \cup P_{pc}^n} \pi_{jit}^{pc}$$
$$+ \Gamma_{nc}^i \zeta_{it}^{nc} + \sum_{j \in P_{nc}} \bar{r}_{jt} \chi_{ji}^t + \sum_{j \in P_{nc}} \pi_{jit}^{nc} \leqslant v_i + \sum_{l \in L_i} w_{it}^l, \quad \forall i \in I, \quad t \in T$$

$$\zeta_{it}^c + \pi_{jit}^c \geqslant \hat{r}_{jt} x_{ji}, \quad \forall i \in I, \quad j \in P_c^a \cup P_c^n, \quad t \in T$$

$$\zeta_{it}^{pc} + \pi_{jit}^{pc} \geqslant \hat{r}_{jt} \xi_{ji}^t, \quad \forall i \in I, \quad j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T$$

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

$$0 \leqslant w_{it}^l \leqslant \Delta_i^l, \quad \forall i \in I, \quad t \in T, \quad l \in L_i$$

$$x_{ji} \geqslant \tilde{x}_{ji}, \quad \forall i \in I, \quad j \in P_c^a$$

$$y_j^t \geqslant \xi_{ji}^t - \xi_{ji}^{t-1}, \quad \forall t \in T \setminus \{t_1\}, \quad j \in P_{pc}^a \cup P_{pc}^n, \quad i \in I$$

$$y_j^{t_1} \geqslant \xi_{ji}^{t_1} - \tilde{x}_{ji}, \quad \forall j \in P_{pc}^a, \quad i \in I$$

$$x_{ji} \in \{0,1\}, \quad \forall j \in P_c^a \cup P_c^n, \quad i \in I$$

$$\xi_{ji}^t \in \{0,1\}, \quad \forall j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T, \quad i \in I$$

$$\chi_{ji}^t \in [0,1], \quad \forall j \in P_{nc}, \quad t \in T, \quad i \in I$$

$$y_j^t \in \{0,1\}, \quad \forall j \in P_{nc}, \quad t \in T$$

$$\zeta_{it}^c \geqslant 0, \quad \forall i \in I, \quad t \in T$$

$$\pi_{jit}^c \geqslant 0, \quad \forall i \in I, \quad j \in P_c^a \cup P_c^n, \quad t \in T$$

$$\zeta_{it}^{pc} \geqslant 0, \quad \forall i \in I, \quad t \in T$$

$$\pi_{jit}^{pc} \geqslant 0, \quad \forall i \in I, \quad j \in P_{pc}^a \cup P_{pc}^n, \quad t \in T$$

$$\zeta_{it}^{nc} \geqslant 0, \quad \forall i \in I, \quad t \in T$$

$$\pi_{jit}^{nc} \geqslant 0, \quad \forall i \in I, \quad j \in P_{nc}, \quad t \in T$$

The model guarantees that the optimal solution is feasible for any three subsets of patients with cardinality $\Gamma_c^i, \Gamma_{pc}^i$ and Γ_{nc}^i. As also the worst subset is considered among all of the possible subsets, the solution can face the worst scenario.

3. Computational tests

Computational tests have been run in order to evaluate the applicability of the proposed approach to the management of realistic HC services. Therefore, tests are run in order to evaluate the quality of solutions obtained within acceptable computational time, and the impact of such solutions when applied to realistic scenarios. Besides, we want to evaluate the impact of different degrees of robustness and to derive suggestions on how cardinality parameters have to be tuned to meet HC needs. The aim is to evaluate the performance of the robust assignment model with regard to reduction of overtime costs while preserving continuity of care where required. First, we tested our model on real life instances, in order to evaluate the impact of different levels of robustness and of other problem features, such as the required continuity of care. Then, we applied the obtained assignment on a set of sample paths.

In this section, we first describe the HC provider chosen for providing data, together with the reasons of the choice (Section 3.1), then we detail the experimental setup, describing instances and sample paths (Section 3.2). Finally, computational results are reported in Section 3.4.

3.1. Real case description

The considered HC provider is representative of a general class of providers in terms of organization and resource planning (Chahed, Dallery, Matta, & Sahin, 2006; Chahed, Marcon, Sahin, Feillet, & Dallery, 2009). It operates in the north of Italy, covering a region of about 800 km^2, with about 1000 patients assisted at the same time by about 50 nurses. It manages different professional figures; however, only the assignment of nurses is analyzed in this paper as they provide the largest number of visits, manage emergencies, meet short-term demand variations, and deal with highly uncertain workloads. Furthermore, continuity of care is usually pursued for nurses. This provider has been chosen for four reasons:

- The provider is one of the largest Italian public HC providers, with a complex human resource management.
- A part of the national standards for HC service was developed thanks to the experience conducted with the provider.
- The provider tries to pursue continuity of care, at least for critical patients, by preferably assigning each patient to only one reference nurse.
- The provider was already adopted to validate other assignment techniques (Lanzarone et al., 2012; Lanzarone & Matta, 2012).

Further, a patient stochastic model is available for this provider (Lanzarone, Matta, & Scaccabarozzi, 2010) to estimate the future patients' demands.

The provider covers three independent divisions, and the analysis is carried out for the largest one, which includes 22 nurses. The skills of nurses (nurses for palliative and for non-palliative care) and their territorial distribution are taken into account. Patients are divided into two classes (palliative care and non-palliative care patients) and three geographical areas are present in the analyzed division; hence, the division consists of six districts (one for each combination of skill and territory) and the assignments are planned considering the districts as independent. The features of the considered districts, e.g., the number of nurses and their working time, are reported in Table 2.

3.2. Experimental setup and instance description

Experiments are conducted with a rolling approach over a period of 26 weeks, referred to as the rolling weeks. The model is solved at the beginning of each rolling week, and 1 week time slots are considered with a planning horizon covering the considered week and the next 7 ones with a look ahead perspective (i.e., 8 time slots are in T, referred to as the horizon weeks). At each rolling week, the assignments of the first horizon week are kept, and the model is solved again for the next rolling week keeping the assignments of the previous rolling weeks. This is consistent with the policy of the analyzed HC provider, where assignments are mainly decided at the beginning of the week on a weekly basis. Finally, if a patient is discharged and readmitted during the 26 rolling weeks, he/she maintains the previously assigned reference nurse.

Data refers to the period from April to September 2008, which is the same time period studied in the other papers validated with this provider (Lanzarone et al., 2012; Lanzarone & Matta, 2012). The number of patients in the charge at each rolling week and their features are taken from the historical data of the provider (considering real arrivals of new patients and real discharges). In other words, the assignments are run taking into account the real mix of patients when the assignments are decided; if a patient exits during a rolling week he/she is consequently excluded from the following ones or, alternatively, if a patient enters during a rolling week he/she is considered for the next one.

In addition, stochastic demands r_{jt} of each considered patient j in each horizon week t (i.e., $\bar{r}_{jt}$ and $\hat{r}_{jt}$) are obtained from the stochastic model proposed in Lanzarone et al. (2010). The possible discharge of a patient during slot t^* of the planning horizon is represented by assuming $\bar{r}_{jt} = \hat{r}_{jt} = 0 \; \forall t > t^*$.

Concerning costs c_l and thresholds Δ_i^l, the current practice of the HC provider in the considered time period was not formalized by any cost function, even if the aim of the assignments was to reduce high overtimes. Thus, in our experiments, values are decided based on the salary provided for overtimes and generalizing among nurses. Ten levels are considered ($l = 1, \ldots, 10$), the costs are assumed as $c_l = l \; \forall l$, and each Δ_i^l is taken as the tenth part of the corresponding v_i ($\Delta_i^l = 0.1 v_i \;\; \forall i, l$), as different nurses may have different values of v_i.

Finally, as for reassignment penalty γ, a violation of care continuity is accepted if keeping the continuity would be too expensive. In the analyzed HC provider, the violation is accepted if keeping continuity would cause an overtime cost equal to or greater than 3 cost units. On the contrary, if the overtime cost due to the continuity of care satisfaction is equal to or lower than 2 cost units, continuity is guaranteed. Thus, an intermediate value $\gamma = 2.5$ is set.

Table 2
Districts of the analyzed division with the number of nurses and their weekly capacity.

Name of the district	Code of territory	Skill of the nurses	Number of nurses	Capacity in weekly hours for each nurse
NPA	A	Non-palliative	8	35, 40, 45, 50, 50, 50, 50, 50
PA	A	Palliative	3	20, 30, 30
NPB	B	Non-palliative	4	30, 35, 50, 50
PB	B	Palliative	1	35
NPC	C	Non-palliative	5	30, 35, 40, 50, 50
PC	C	Palliative	1	35

The assignments at the initial week (named rolling week 0) are computed considering all patients as newly admitted ones, i.e., without an assigned nurse. It is worth noting that instances at week 0 are more computationally expensive, as the number of admitted patients is significantly higher than in other weeks. Therefore, the initialization at rolling week 0 is obtained neglecting robustness (i.e., all patients require the expected demand $\bar{r}_{jt}$) and, consequently, constraints (5) are replaced by

$$\sum_{j\in P_c^a\cup P_c^n}\bar{r}_{jt}x_{ji}+\sum_{j\in P_{pc}^a\cup P_{pc}^n}\bar{r}_{jt}\xi_{ji}^t+\sum_{j\in P_{nc}}\bar{r}_{jt}\chi_{ji}^t\leqslant v_i+\sum_{l\in L_i}w_{it}^l,\quad \forall i\in I,\quad t\in T$$

In this way, we simulate the case in which, starting from a plan without robustness, we include robustness and observe how the quality of the assignments improves over time.

Different values of Γ_c^i,Γ_{pc}^i and Γ_{nc}^i, of maximum demands $\bar{r}_{jt}+\hat{r}_{jt}$, and different types of continuity of care are taken into account in the experiments.

As mentioned, patient demand distributions are estimated with the stochastic model proposed in Lanzarone et al. (2010). Indeed, expected demands $\bar{r}_{jt}$ and maximum demands $\bar{r}_{jt}+\hat{r}_{jt}$ for each patient and time slot are taken from the empirical distribution given by such stochastic model. However, these empirical distributions may have long positive tails associated with low probability, and it is useless to consider them while setting values $\hat{r}_{jt}$, since they may cause an excessive and unrealistic overload in the worst case. For this reason, we assume that maximum demands $\bar{r}_{jt}+\hat{r}_{jt}$ are taken in correspondence of a given quantile of the demand distribution, where this quantile must be close to 1. Thus, very unlikely scenarios are discarded and excessively conservative solutions are avoided. Two values are considered in the experiments, namely 0.9 and 0.8.

The continuity of care requirement is determined for each patient based on his/her characteristics. Patients belong to 14 different care profiles (CPs) for both the palliative and non-palliative cases Lanzarone et al. (2010). The type of continuity of care required is decided according to the CP at the moment in which the patient is first considered (i.e., the initial CP for patients entering from rolling week 1 to 25 and the current CP for patients at week 0), based on the idea that more complex CPs require a higher continuity of care.

One configuration is adopted for non-palliative patients. As for palliative patients, we know from the provider that hard continuity should be recommended, even if a partial continuity can be accepted in some cases. In addition, from the historical data, we know that partial continuity can be acceptable for about the 20% of palliative patients, even if we cannot associate this information with patients in the dataset. Hence, two different configurations are taken into account for palliative patients: either they all require hard continuity of care (optimal case in which the HC provider guarantees a hard continuity for all palliative treatments), or they require hard or partial continuity of care according to a random choice (to simulate the presence of a 20% of patients with partial continuity of care). In the latter case, denoted with "random 80% C and 20% PC", each palliative patient is randomly considered C (with probability 0.8) or PC (with probability 0.2) for the entire care duration. The adopted continuity of care requirements for the different CPs are reported in Table 3.[4]

Concerning the robustness, two values of cardinalities are considered: $\Gamma_c^i=\Gamma_{pc}^i=\Gamma_{nc}^i$ either equal to 1 or 2 $\forall i$.[5] Furthermore, in order to compare the solutions with their non-robust counterparts,

Table 3
Type of continuity of care (hard continuity (C), partial continuity (PC) and no continuity (NC)) associated to each CP (see (Lanzarone et al., 2010)).

Type of patient	CP	Type of continuity
Non-palliative	1	NC
Non-palliative	2	PC
Non-palliative	3	C
Non-palliative	4	C
Non-palliative	5	C
Palliative	6	C, or random 80% C and 20% PC
Palliative	7	C, or random 80% C and 20% PC
Palliative	8	C, or random 80% C and 20% PC
Non-palliative	9	PC
Non-palliative	10	PC
Non-palliative	12	PC
Non-palliative	13	C
Non-palliative	14	C
Non-palliative	15	NC

Table 4
Instances features.

Quantile for max	Type of continuity for palliative	Γ values	Robust solution	Non-robust solution
0.9	C	1	Conf. A	Conf. I
		2	Conf. B	Conf. I
	Random 80% C and 20% PC	1	Conf. C	Conf. J
		2	Conf. D	Conf. J
0.8	C	1	Conf. E	Conf. I
		2	Conf. F	Conf. I
	Random 80% C and 20% PC	1	Conf. G	Conf. J
		2	Conf. H	Conf. J

the case in which the robustness is always neglected at all rolling weeks from 0 to 25 is also considered for each parameter configuration. The different experimental settings generate a set of instances (Table 4), all involving the same set of patients and nurses. Each non-robust solution corresponds to 4 robust solutions; thus, the non-robust cases are reported several times in the table.

The dimension of the instances is given by the number of nurses and patients. Table 2 provides all the information about nurses: number, weekly working time, skill and territory. This configuration is constant over the 26 rolling weeks. With regard to patients, their numerousness and characteristics vary depending on the specific rolling week. Hence, to derive the dimension of the problem, the characteristics of patient mix have to be detailed for each rolling week and configuration. For this purpose, Table 5 gives the detailed cardinality of each set of patients.

3.3. Sample paths generation

In each configuration, assignments and workloads of the first horizon week t_1 of each run are taken, and the sequence of such assignments over the 25 rolling weeks is the planned solution.

The planned solutions are then applied in two different cases, as proposed in the other papers whose validation is conducted with this provider (Lanzarone et al., 2012; Lanzarone & Matta, 2012). They are either applied to a set of 10 simulated sample paths or, alternatively, to the real observed values of patients' demands from the historical data of the HC provider. The workload of each nurse over the rolling 25 weeks is determined by combining the provided assignments with patients' demands in the scenario (or with the real demands). For patients without continuity of care (NC), the per cent division among nurses is maintained while executing in the same week.

[4] A detailed description of the CPs can be found (Lanzarone et al., 2010). The division of the considered patients among the CPs can be found in the electronic companion of Lanzarone and Matta (2012), where the same instance is analyzed with a different technique.

[5] Integer values are taken for Γ_c^i,Γ_{pc}^i and Γ_{nc}^i; hence, p_c^{it},p_{pc}^{it} and p_{nc}^{it} are null.

Table 5
Number of patients in each subset for the robust configurations at each rolling week. The cardinality of the set P_{nc} is equal in all the configurations.

Rolling week	$\lvert P_{nc}\rvert$	Conf. A, B, E, F				Conf. C				Conf. D				Conf. G				Conf. H			
		$\lvert P_c^a\rvert$	$\lvert P_c^n\rvert$	$\lvert P_{pc}^a\rvert$	$\lvert P_{pc}^n\rvert$	$\lvert P_c^a\rvert$	$\lvert P_c^n\rvert$	$\lvert P_{pc}^a\rvert$	$\lvert P_{pc}^n\rvert$	$\lvert P_c^a\rvert$	$\lvert P_c^n\rvert$	$\lvert P_{pc}^a\rvert$	$\lvert P_{pc}^n\rvert$	$\lvert P_c^a\rvert$	$\lvert P_c^n\rvert$	$\lvert P_{pc}^a\rvert$	$\lvert P_{pc}^n\rvert$	$\lvert P_c^a\rvert$	$\lvert P_c^n\rvert$	$\lvert P_{pc}^a\rvert$	$\lvert P_{pc}^n\rvert$
1	153	302	21	110	3	296	17	116	7	294	21	118	3	294	17	118	7	296	21	116	3
2	150	315	14	110	4	305	13	120	5	307	12	118	6	303	13	122	5	309	13	116	5
3	153	319	19	107	3	308	19	118	3	309	19	117	3	307	19	119	3	312	17	114	5
4	151	318	7	109	2	310	7	117	2	311	7	116	2	308	6	119	3	311	7	116	2
5	149	311	7	110	1	303	7	118	1	305	7	116	1	300	7	121	1	306	6	115	1
6	151	308	14	107	3	300	13	115	4	302	14	113	3	296	14	119	3	302	14	113	3
7	156	311	15	108	1	303	12	116	4	306	15	113	1	299	15	120	1	306	15	113	1
8	158	315	15	107	1	303	14	119	1	310	14	112	1	305	15	117	1	310	14	112	1
9	161	318	12	106	2	307	12	117	2	313	11	111	3	309	12	115	2	312	12	112	2
10	148	315	8	105	1	305	8	115	1	311	8	109	1	306	8	114	1	311	8	109	1
11	148	308	9	109	1	299	9	118	1	303	9	114	1	299	9	118	1	303	9	114	1
12	154	311	5	106	1	302	5	115	1	306	5	111	1	302	4	115	2	306	5	111	1
13	156	305	13	102	4	297	13	110	4	302	12	105	5	295	13	112	4	301	13	106	4
14	159	301	9	104	2	294	8	111	3	297	9	108	2	291	9	114	2	297	9	108	2
15	159	308	10	106	3	300	9	114	4	303	9	111	4	298	10	116	3	303	10	111	3
16	154	306	13	103	3	297	10	112	6	301	13	108	3	296	13	113	3	301	12	108	4
17	151	307	11	105	9	298	10	114	10	302	11	110	9	297	11	115	9	301	11	111	9
18	154	300	15	113	1	290	14	123	2	297	15	116	1	291	15	122	1	295	15	118	1
19	154	313	14	110	5	302	14	121	5	310	14	113	5	304	14	119	5	309	14	114	5
20	158	318	7	111	1	308	7	121	1	315	7	114	1	310	7	119	1	314	6	115	2
21	157	319	7	111	3	310	7	120	3	316	7	114	3	311	7	119	3	314	7	116	3
22	160	316	9	112	1	308	9	120	1	312	5	116	4	308	9	120	1	311	7	117	2
23	157	312	11	110	2	303	9	119	4	304	11	118	2	304	9	118	4	306	11	116	2
24	159	312	10	109	4	302	10	119	4	305	10	116	4	303	10	118	4	306	10	115	4
25	162	315	3	116	2	306	3	125	2	309	3	122	2	307	3	124	2	309	1	122	4

The generation of sample paths consists of a mix between a Monte Carlo simulation from a semi-Markov model and the real historical data. The semi-Markov model is used for the majority of patients by modifying the approach proposed in the Markovian patient stochastic model (Lanzarone et al., 2010). The real conditions of patients are taken at the beginning of the care pathway (or at week 0 for patients already in the charge), and their requests and discharges are generated for all future weeks starting from these real conditions. Indeed, for each patient, a sample path is generated by sampling the sojourn duration in a CP, the related demand for visits at each week and the next CP. Only demands of long stay non-palliative patients are directly taken from their real historical values; these patients show a very low variability along with time, they do not represent a relevant uncertainty source, and the variability of the entire mix of patients is not reduced by this choice.

If an already assigned patient is present in a week of a sample path but not in the corresponding planning, he/she is considered in the execution by taking the last information available in the previous weeks (i.e., the last nurse assigned or, in case of NC, the last percentages).

3.4. Results analysis

This section is devoted to the analysis of results: first, results obtained by solving the optimization model are reported; then, results obtained by executing the assignments in the generated sample paths and with the real demands are detailed.

3.4.1. Solving the assignment problem

The model has been implemented with OPL 5.1 and solved with CPLEX; computational tests have been run on a PC equipped with CPU Intel Core i7 1.73 gigahertz and 6 gigabytes of installed RAM. A 5400 seconds Time Limit (TL) has been set.

First, Table 6 shows the computational time in seconds and the optimality gap per each rolling week and each robust configuration. Let UB be the value of the best solution found and LB the value of the best linear relaxation, the gap is defined as $\frac{UB-LB}{LB}$. Besides, their average and maximum values among the rolling weeks are given, together with the number of rolling weeks in which TL is reached. Results show that the needed computational time strongly increases while increasing the value of parameters $\Gamma_c^i, \Gamma_{pc}^i$ and Γ_{nc}^i, or the quantile chosen for the maximum demand. The different types of continuity of care chosen for palliative patients do not seem to influence the computational effort. TL is reached only in a few cases, in all of which $\Gamma_c^i = \Gamma_{pc}^i = \Gamma_{nc}^i = 2$. However, gaps are usually limited, always being below 5.2%. As for the non robust cases I and J, all the instances are solved to optimality in less than 3 seconds.

To evaluate the impact of TL value, the decrease of the optimality gap over time has been studied for instances that are not solved to optimality. The obtained evolution is reported in Fig. 1 for rolling week 6 of configuration B; it can be seen that increasing the computational time from 1.5 up to 10 hours only slightly improves the quality of the solution. Besides, the most relevant reduction of the gap is obtained within the first hour. Similar trends are obtained in the other cases that reached TL. Hence, for the analyzed instances and the required robustness levels, the chosen TL proves to be a good compromise between quality of solutions and practical applicability in real HC providers.

Table 7 shows the objective function and the number of reassignments for patients with partial continuity of care over the entire planning horizon. These values are expressed in terms of minimum, maximum and average values among the rolling weeks from 1 to 25; week 0 is excluded as it refers to the non-robust initialization.

The considered instances are always feasible, i.e., current workforce is able to serve the demand with the required level of robustness. Infeasibility might occur if robustness parameters are increased, and such infeasibility has to be intended as a warning about a too high robustness requirement for the current workforce.

Table 6
Computational time in seconds and optimality gap for the robust cases. Average CPU time is computed with TL = 5400 s.

Week	Conf. A		Conf. B		Conf. C		Conf. D		Conf. E		Conf. F		Conf. G		Conf. H	
	CPU time	gap %	CPU time	gap %	CPU time	gap %	CPU time	gap %	CPU time	gap %	CPU time	gap %	CPU time	gap %	CPU time	gap %
1	134.82	0.00	**TL**	2.68	87.94	0.00	**TL**	2.06	16.21	0.00	4477.20	0.00	42.84	0.00	2177.76	0.00
2	85.32	0.00	**TL**	4.32	256.31	0.00	**TL**	5.18	40.51	0.00	**TL**	1.26	20.61	0.00	**TL**	0.92
3	13.18	0.00	1183.93	0.00	37.16	0.00	1871.99	0.00	5.29	0.00	90.04	0.00	7.74	0.00	114.57	0.00
4	5.99	0.00	416.79	0.00	10.75	0.00	234.24	0.00	4.99	0.00	38.36	0.00	6.13	0.00	46.16	0.00
5	36.04	0.00	970.21	0.00	8.25	0.00	1997.32	0.00	3.87	0.00	287.90	0.00	4.29	0.00	36.18	0.00
6	94.16	0.00	**TL**	3.42	13.49	0.00	**TL**	2.36	12.54	0.00	92.87	0.00	5.27	0.00	280.99	0.00
7	28.08	0.00	**TL**	1.12	34.37	0.00	**TL**	0.05	14.59	0.00	626.19	0.00	14.56	0.00	541.71	0.00
8	16.65	0.00	**TL**	0.94	18.83	0.00	**TL**	1.00	10.84	0.00	414.04	0.00	7.21	0.00	1568.49	0.00
9	19.00	0.00	**TL**	1.17	25.24	0.00	133.52	0.00	19.36	0.00	53.03	0.00	5.91	0.00	209.90	0.00
10	6.33	0.00	700.78	0.00	5.85	0.00	314.28	0.00	3.98	0.00	52.20	0.00	4.09	0.00	67.22	0.00
11	5.60	0.00	1424.85	0.00	4.98	0.00	110.67	0.00	10.28	0.00	27.38	0.00	4.52	0.00	69.23	0.00
12	6.04	0.00	133.83	0.00	5.37	0.00	147.59	0.00	3.81	0.00	26.69	0.00	4.24	0.00	78.78	0.00
13	24.23	0.00	48.20	0.00	18.83	0.00	126.70	0.00	11.53	0.00	9.73	0.00	8.95	0.00	52.64	0.00
14	24.90	0.00	59.68	0.00	18.97	0.00	89.19	0.00	3.24	0.00	14.91	0.00	5.02	0.00	52.15	0.00
15	4.85	0.00	6.51	0.00	6.10	0.00	6.83	0.00	4.32	0.00	5.37	0.00	4.32	0.00	4.91	0.00
16	8.97	0.00	7.80	0.00	6.85	0.00	7.72	0.00	3.15	0.00	7.60	0.00	4.68	0.00	4.23	0.00
17	10.94	0.00	101.55	0.00	8.55	0.00	15.85	0.00	6.38	0.00	13.85	0.00	6.16	0.00	7.13	0.00
18	5.82	0.00	101.40	0.00	4.70	0.00	10.62	0.00	3.48	0.00	10.16	0.00	4.21	0.00	3.78	0.00
19	12.90	0.00	1111.42	0.00	73.32	0.00	1197.21	0.00	18.19	0.00	111.68	0.00	6.52	0.00	888.45	0.00
20	13.65	0.00	94.56	0.00	37.11	0.00	345.62	0.00	12.25	0.00	401.72	0.00	9.84	0.00	1140.90	0.00
21	12.23	0.00	19.69	0.00	16.33	0.00	57.11	0.00	11.93	0.00	62.85	0.00	9.47	0.00	59.24	0.00
22	9.24	0.00	53.35	0.00	18.00	0.00	136.05	0.00	10.81	0.00	164.16	0.00	6.05	0.00	550.00	0.00
23	9.36	0.00	18.66	0.00	5.26	0.00	40.26	0.00	4.48	0.00	38.00	0.00	5.20	0.00	49.90	0.00
24	17.25	0.00	95.99	0.00	53.82	0.00	126.42	0.00	36.39	0.00	92.17	0.00	5.54	0.00	82.78	0.00
25	18.83	0.00	**TL**	2.16	101.49	0.00	4884.73	0.00	40.30	0.00	976.83	0.00	5.76	0.00	1049.65	0.00
Average	24.97	0.00	1773.97	0.63	35.11	0.00	1554.16	0.43	12.51	0.00	539.80	0.05	8.37	0.00	581.47	0.04
Maximum	134.82	0.00	**TL**	4.32	256.31	0.00	**TL**	5.18	40.51	0.00	**TL**	1.26	42.84	0.00	**TL**	0.92
Number of TL	0		7		0		5		0		1		0		1	

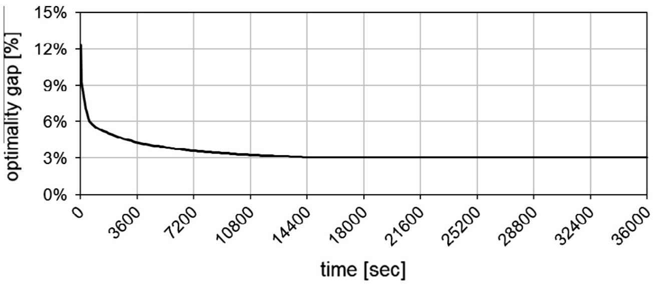

Fig. 1. Optimality gap versus computational time in seconds for the rolling week 6 of configuration B.

Table 7
Objective function and number of reassignments.

Conf.	Objective function			Number of reassignments		
	Min	Max	Average	Min	Max	Average
A	58.8	156.9	98.9	0	3	0.5
B	186.3	478.6	299.1	0	9	1.4
C	58.1	193.2	102.4	0	2	0.5
D	160.1	523.1	298.6	0	7	1.6
E	32.0	96.8	62.1	0	2	0.4
F	91.1	225.4	143.4	0	5	1.0
G	31.2	112.1	61.6	0	2	0.2
H	86.1	237.2	144.0	0	3	0.9
I	3.9	27.8	13.6	0	1	0.1
J	3.9	27.8	13.4	0	1	0.0

Table 8
Ratio $\frac{\bar{r}_{jt_1}+\tilde{r}_{jt_1}}{\bar{r}_{jt_1}}$ in rolling week 10 for the cases of maximum demand equal to quantile 0.8 and to quantile 0.9 of the corresponding density function, mediated among the patients of each district.

District	Quantile 0.8	Quantile 0.9
NPA	1.738	2.242
PA	1.483	1.951
NPB	1.744	2.198
PB	1.510	1.952
NPC	1.711	2.245
PC	1.499	1.927

It is a choice of the HC provider how to face such situation, whether by increasing the workforce or accepting lower levels of robustness.[6]

The objective function increases with the values of Γ^i_c, Γ^i_{pc}, and Γ^i_{nc} both in terms of overtime cost and number of reassignments, as the demand of the worst scenario increases and more robust solutions are selected. Furthermore, the objective function is higher if quantile of maximum demand is equal to 0.9, as maximum demand of each patient increases as well. Also the number of reassignments increases with quantile, but the variation is limited in all configurations. Obviously, value of the objective function is significantly increased with respect to the non-robust counterpart.[7]

The ratio between maximum and expected demands in the first week of the planning horizon $\left(\text{i.e.,} \frac{\bar{r}_{jt_1}+\tilde{r}_{jt_1}}{\bar{r}_{jt_1}}\right)$ is taken as index of patient demand increase. This is reported in Table 8 for rolling week 10, as example, in two cases: maximum demand corresponding to quantile 0.8 or to 0.9 of the corresponding distribution (Lanzarone et al., 2010). In the following weeks of planning horizon, the ratio tends to increase, because a descending trend is observed for the expected value, which is only marginally replicated by the maximum value.

Table 9
Overtime cost from the objective function and overtime expected cost with the expected demands (minimum, maximum and average value among the rolling weeks 1–25).

Conf.	Overtime cost			Overtime expected cost		
	Min	Max	Average	Min	Max	Average
A	58.8	151.8	97.7	5.0	30.0	15.7
B	181.3	478.6	295.5	7.3	45.6	20.8
C	55.6	188.2	101.2	5.4	34.7	16.1
D	160.1	520.6	294.6	6.4	40.8	19.3
E	32.0	96.8	61.1	4.6	29.5	15.1
F	88.6	220.6	141.0	6.2	33.7	18.7
G	31.2	107.1	61.1	4.6	29.9	15.3
H	86.1	237.2	141.8	6.9	36.5	17.9
I	3.9	27.8	13.4	3.9	27.8	13.4
J	3.9	27.8	13.3	3.9	27.8	13.3

The overtime cost is significantly affected by the degree of robustness of the solution, as maximum demands of patients belonging to S^{it}_c, S^{it}_{pc} and S^{it}_{nc} have an impact on the overall workload. However, the question arises on how a robust solution behaves if no patient requires the maximum amount of care. For evaluating the expected behavior of solutions, overtime costs of the obtained assignments are recomputed with respect to the expected demands $\bar{r}_{jt}$ for all patients. These values are reported in terms of minimum, maximum and average values among the rolling weeks from 1 to 25 in Table 9.

It can be seen that robustness determines an increase in the overtime expected costs, as expected. However, it is worth noting that, when considering expected demands, the robust assignment is not significantly penalized with respect to the optimal non-robust counterpart. Indeed, overtime expected costs are always lower than twice the non-robust case.

Objective function, overtime cost, overtime expected cost and cost related to the reassignments (i.e., the number of reassignments multiplied by γ) over the rolling weeks are finally shown in Fig. 2 for robust configurations A and E and non-robust configuration I. Very similar trends are also observed in all other configurations. Indeed, peaks in the objective function are in the same rolling week, independent of the specific configuration; this means that they are related to patient mix rather than determined by the solution. Furthermore, for robust configurations, higher objective functions are observed in rolling weeks 1–3, which usually decrease as weeks go by, because a non robust assignment is computed for the initial rolling week 0 and the solution gradually improves.

3.4.2. Execution of the assignments

Each obtained solution is applied to 10 generated sample paths and to the real historical patients' demands.

The quality of solutions is analyzed in each district in terms of two indicators: mean overtime cost per nurse and fairness (i.e., the range between minimum and maximum utilization). Obviously, these indicators are extracted in districts where more than one nurse is present (i.e., districts NPA, PA, NPB and NPC).

The mean overtime cost per nurse is computed at each rolling week as the ratio between the total cost of the provider (computed according to the same levels, $c_l = l$ and Δ^l_i applied in the planning) and the number of nurses. The same is repeated considering each district separately. As for the fairness, utilization of a nurse in a

[6] It is worth noting that, should the non-robust counterpart be infeasible for a given set of patients, the only option is to increase the workforce.

[7] It is worth noting that non-robust solutions are always optimal, whereas the optimality of robust solutions is not always guaranteed as a time TL limit is set.

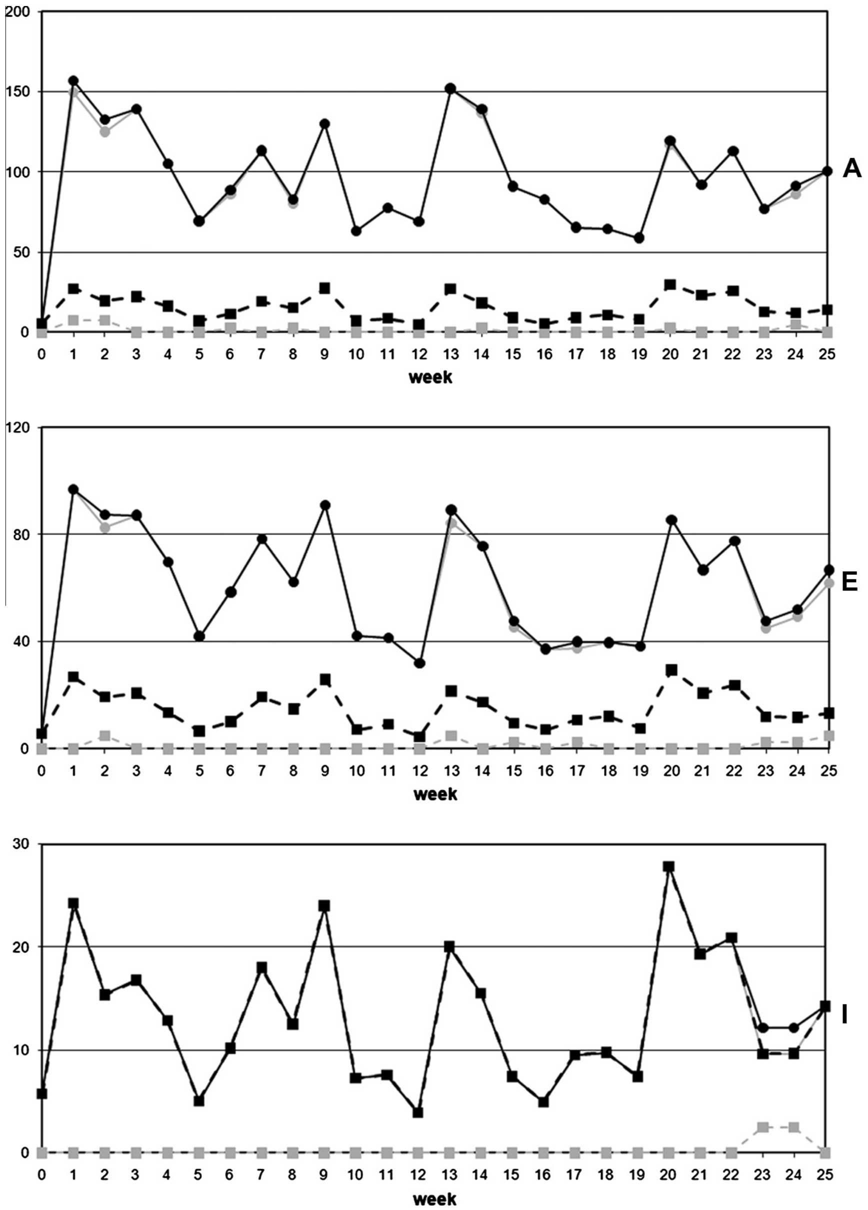

Fig. 2. Values assumed by objective function (continuous black line with round indicators), overtime cost (continuous gray line with round indicators), overtime expected cost (dotted black line with square indicators) and cost related to the reassignments (dotted gray line with square indicators) over the rolling weeks in configurations A, E and I.

Table 10
Minimum, maximum and average values among the rolling weeks from 1 to 25 of mean overtime cost per nurse: sum of districts NPA, PA, NPB and NCP.

Conf.	Overtime cost: Districts NPA + PA + NPB + NPC					
	Sample paths			Real execution		
	Min	Max	Average	Min	Max	Average
A	0.00	10.56	2.01	0.19	17.19	4.47
B	0.00	10.12	2.00	0.36	13.22	4.29
C	0.00	9.22	1.58	0.25	13.26	4.10
D	0.00	9.25	1.43	0.03	11.51	3.48
E	0.00	10.73	2.14	0.35	18.18	4.71
F	0.00	10.16	1.77	0.27	12.00	4.03
G	0.00	8.35	1.81	0.21	10.99	3.70
H	0.00	12.44	1.68	0.30	16.51	3.82
I	0.00	13.35	2.17	0.64	16.12	5.81
J	0.01	11.79	2.30	0.92	19.16	5.94

week is computed as the ratio between the workload (i.e., the number of weekly provided visits) and the availability v_i. Then, the fairness of a district in a rolling week is the difference between maximum and minimum utilization among nurses.

These two indicators are directly taken for the execution with the historical demands, whereas for sample paths the reported and analyzed indicator is the average at each rolling week among the paths. Hence, for each configuration and district, results are the list of average costs and fairnesses over the rolling weeks in two cases: executed with the historical demands or averaged among the paths. We remark that objective functions and planned costs reported in Section 3.4.1 refer to the entire planning horizon, equal to 8 weeks, whereas for the execution only the first horizon week is extracted from each rolling week.

Average costs are synthetically reported in terms of minimum, maximum and average values among the rolling weeks from 1 to 25 in Table 10.

Table 11
Minimum, maximum and average values among the rolling weeks from 1 to 25 of mean overtime cost per nurse and fairness in districts NPA, PA, NPB and NCP.

Conf.	Sample paths						Real execution					
	Overtime cost			Fairness			Overtime cost			Fairness		
	Min	Max	Average	Min	Max	Average	Min	Max	Average	Min	Max	Average
District NPA												
A	0.00	6.18	1.01	0.16	0.73	0.43	0.16	20.31	4.62	0.51	1.27	0.75
B	0.00	4.00	0.65	0.18	0.65	0.39	0.00	12.64	3.50	0.33	1.01	0.68
C	0.00	3.87	0.73	0.24	0.78	0.42	0.16	5.72	1.75	0.33	0.77	0.51
D	0.00	5.07	0.64	0.30	0.80	0.47	0.00	7.47	1.83	0.37	0.88	0.56
E	0.00	6.52	1.13	0.31	0.82	0.51	0.61	19.21	3.71	0.41	0.98	0.65
F	0.00	4.04	0.72	0.20	0.68	0.41	0.25	10.13	3.67	0.46	0.91	0.67
G	0.00	5.33	1.02	0.19	0.84	0.54	0.31	7.57	2.46	0.25	0.82	0.56
H	0.00	5.83	0.57	0.15	0.74	0.37	0.00	6.95	1.45	0.32	0.76	0.49
I	0.00	11.80	1.83	0.43	0.96	0.67	1.24	15.89	7.00	0.61	1.18	0.95
J	0.02	10.06	2.46	0.45	1.07	0.76	1.91	18.77	6.79	0.65	1.27	0.91
District PA												
A	0.00	0.29	0.01	0.10	0.62	0.30	0.00	0.15	0.01	0.02	0.68	0.29
B	0.00	2.93	0.13	0.08	0.60	0.36	0.00	1.40	0.06	0.09	0.75	0.33
C	0.00	0.84	0.05	0.07	0.54	0.24	0.00	0.67	0.05	0.10	0.58	0.34
D	0.00	0.90	0.07	0.14	0.64	0.34	0.00	0.33	0.03	0.04	0.58	0.28
E	0.00	0.87	0.11	0.02	0.72	0.38	0.00	2.00	0.25	0.02	0.90	0.52
F	0.00	0.17	0.01	0.03	0.39	0.23	0.00	0.52	0.02	0.10	0.68	0.36
G	0.00	0.40	0.02	0.06	0.60	0.26	0.00	2.41	0.18	0.10	0.93	0.39
H	0.00	0.12	0.01	0.08	0.57	0.32	0.00	0.90	0.10	0.02	0.70	0.33
I	0.00	1.13	0.09	0.08	0.74	0.39	0.00	4.00	0.51	0.20	1.02	0.55
J	0.00	7.38	0.63	0.06	0.96	0.44	0.00	12.09	0.87	0.03	1.29	0.50
District NPB												
A	0.00	20.31	4.00	0.24	0.92	0.48	0.66	27.62	9.70	0.06	0.58	0.31
B	0.00	24.07	4.05	0.21	0.98	0.48	1.81	27.80	10.98	0.05	0.71	0.38
C	0.00	25.15	3.93	0.19	0.68	0.40	0.94	37.85	12.76	0.19	0.77	0.40
D	0.00	24.82	3.93	0.08	0.93	0.41	0.14	33.22	10.95	0.02	0.68	0.34
E	0.00	20.21	3.81	0.17	0.86	0.43	0.56	26.47	9.64	0.07	0.61	0.30
F	0.00	22.31	3.92	0.12	0.99	0.41	0.86	30.07	9.70	0.14	0.56	0.33
G	0.00	19.17	4.17	0.15	0.77	0.45	0.41	26.52	10.09	0.10	0.68	0.31
H	0.00	31.51	4.92	0.20	0.81	0.43	1.48	41.79	11.33	0.07	0.75	0.37
I	0.00	21.11	3.22	0.06	0.56	0.36	0.75	30.25	10.30	0.13	0.67	0.37
J	0.00	22.72	3.52	0.10	0.79	0.35	0.77	35.39	11.33	0.17	0.92	0.37
District NPC												
A	0.00	15.93	3.22	0.20	0.64	0.42	0.00	14.09	2.72	0.17	0.82	0.46
B	0.00	13.05	3.65	0.31	0.67	0.49	0.00	9.56	2.73	0.25	0.73	0.45
C	0.00	10.06	1.96	0.08	0.57	0.35	0.00	13.20	3.37	0.09	0.86	0.43
D	0.00	8.47	1.52	0.14	0.64	0.34	0.00	7.33	2.20	0.20	0.63	0.40
E	0.00	15.80	3.63	0.20	0.79	0.46	0.00	19.60	5.06	0.15	1.18	0.58
F	0.00	16.21	2.78	0.21	0.59	0.43	0.00	7.42	2.49	0.21	0.63	0.45
G	0.00	9.31	2.24	0.12	0.83	0.38	0.00	9.20	2.71	0.20	0.68	0.43
H	0.00	15.15	1.86	0.16	0.71	0.32	0.00	20.94	3.84	0.22	0.84	0.48
I	0.00	16.94	3.11	0.16	0.78	0.45	0.00	12.46	3.50	0.24	0.92	0.50
J	0.00	8.46	2.06	0.15	0.80	0.37	0.00	11.05	3.32	0.19	0.86	0.51

Synthetic results show that robust solutions perform better than their non-robust counterparts. Average overtime costs are always lower in robust solutions for both sample paths and real data. Concerning maximum costs, they are always lower in robust solutions but in three cases, i.e., configuration H for sample paths and configurations A and E for real data. Thus, robustness improves solutions allowing cost savings. To give an idea of the obtained cost saving, we can assume that one unit of cost corresponds to about 15 euros. Considering that the four districts include 20 nurses (see Table 2), that the observed period refers to 25 rolling weeks, and that all overtimes belong to the first level, each cost reduction of one unit corresponds to a global saving of 7500 euros in the period. As example, comparing solution C and the corresponding non robust solution J, a global cost saving of 13800 euros is expected with the real execution and a global cost saving of 5400 euros is expected in the average among the paths.

In the analyzed instance, the two adopted values of $\Gamma^i_c = \Gamma^i_{pc} = \Gamma^i_{nc}$ show similar quality of solutions with respect to the average overtime costs for both sample paths and real data. Concerning the maximum value, a certain benefit of $\Gamma^i_c = \Gamma^i_{pc} = \Gamma^i_{nc} = 2$ is observed in case of historical data. Higher values of cardinality parameters, although requiring higher computational time, seem to guarantee better solutions, in terms of behavior in the worst case. However, as computational times significantly increase with $\Gamma^i_c = \Gamma^i_{pc} = \Gamma^i_{nc}$, we can affirm that the adopted values are reliable for the considered instance and robustness level is not to be increased. Besides, we recall that the provided solutions are feasible even if at most $(\Gamma^i_c + \Gamma^i_{pc} + \Gamma^i_{nc})|I|$ patients require their maximum demand $\bar{r}_{jt} + \hat{r}_{jt}$. Thus, increasing too much the cardinality parameters would produce over-conservative solutions. On the contrary, if a relevant benefit was observed passing from $\Gamma^i_c = \Gamma^i_{pc} = \Gamma^i_{nc} = 1$ to $\Gamma^i_c = \Gamma^i_{pc} = \Gamma^i_{nc} = 2$ in an instance, higher value of robustness would be worth investigating.

Results about overtime costs and fairness are also detailed for each district: minimum, maximum and average values among the rolling weeks from 1 to 25 are reported in Table 11. Their trends over the weeks are plotted in Fig. 3 for the sample paths, for robust configurations A and E and non-robust configuration I;

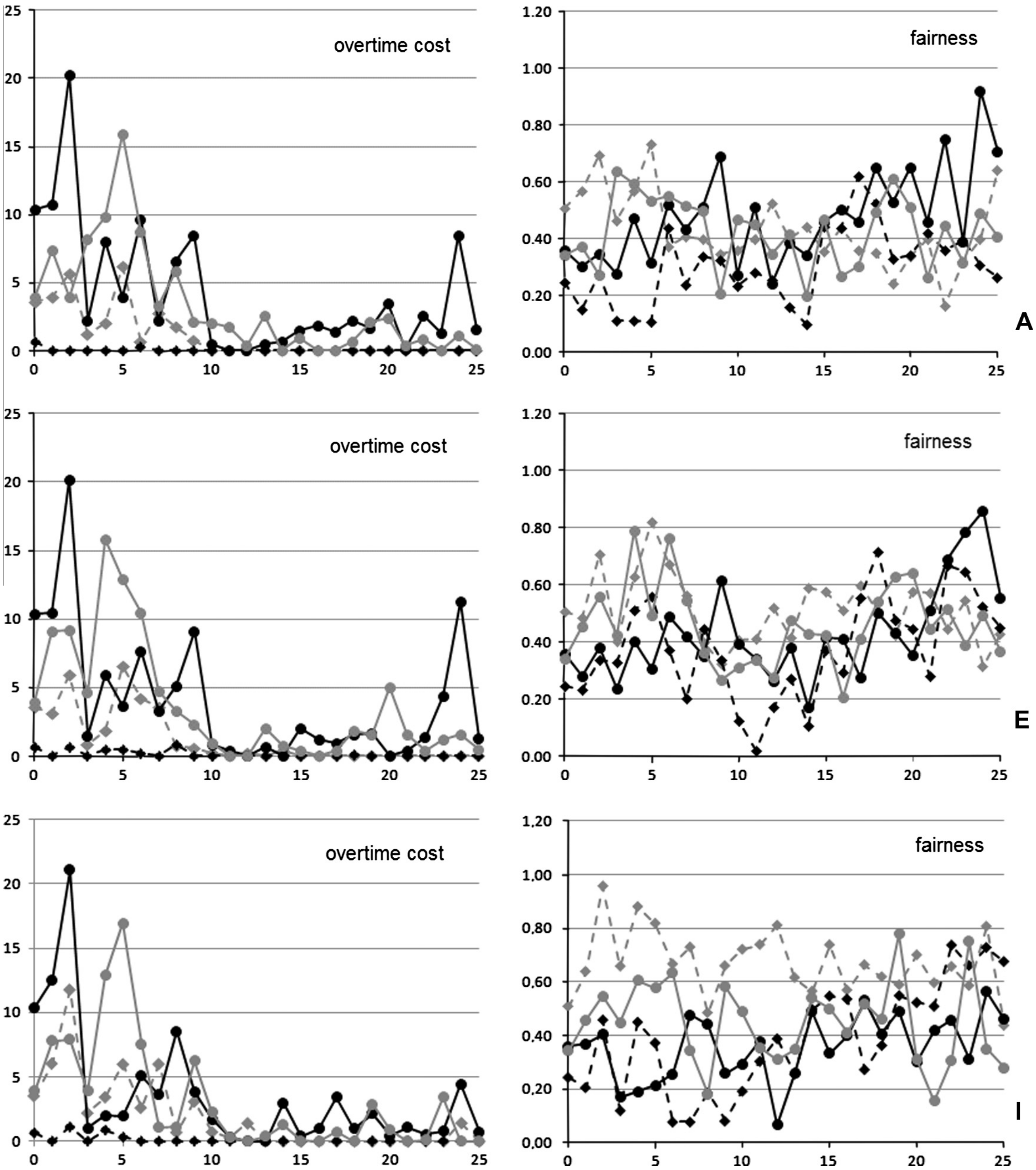

Fig. 3. Overtime cost per nurse and fairness over the rolling weeks in districts NPA (dotted gray line with rhomboid indicators), PA (dotted black line with rhomboid indicators), NPB (continuous black line with round indicators) and NPC (continuous gray line with round indicators) in the case of the sample paths.

similar trends are obtained in the other configurations. Plots refer to the average among the sample paths; similar behaviors are obtained with the real demands, with slightly higher oscillations from week to week.

The main benefit of robustness in terms of overtime costs is observed in districts NPA and PA both for average and maximum values. A lower benefit is observed in district NPC and hardly any benefit in district NPB. It seems that higher benefits are obtained for critical patients with higher demands (i.e., palliative patients) and for higher numbers of nurses. Indeed, district NPA includes 8 nurses, NPC 5 nurses and NPB 4 nurses.

Fairness shows a similar behavior. Although fairness is not considered in the objective function and costs are null for workloads below υ_i, the model balances workloads among overloaded nurses (with workload above υ_i) as $c_l < c_{l+1}\ \forall l \in L_i$. In the robust configurations, where workloads are increased due to patients in S_c^{it}, S_{pc}^{it} and S_{nc}^{it}, workloads are likely to be above υ_i and thus well balanced. This is the reason why, even in our instances, better fairness (i.e., lower ranges) is observed in the correspondence of lower costs.

The trends among the rolling weeks (Fig. 3) show higher oscillations in fairness, as it is not considered in the objective function. Furthermore, similar values of fairness are in general observed over the weeks. Overtime costs show lower oscillations, with similar behaviors among configurations; hence, they are related to patient mix rather than to the approach. In particular, the high overtime cost at rolling week 24 for NPB is always present in all robust configurations. As expected, given the planned solutions, costs are higher in the first rolling weeks, whereas they get lower along with them. Finally, it can be seen that higher costs are usually related to lower fairness (i.e., higher ranges); as example, the peak at rolling week 24 is also present in the plots of fairness.

The proposed approach produces better assignments than those applied by planners. Indeed, planners did not use any quantitative management tools. They aimed at guaranteeing that continuity of care and district division are satisfied. However, due to the poor quality of the assignments, they were forced to violate the constraints in order to avoid very high workloads. Such violations may help in reducing the burnout risks, but cause a bad quality of the service provided to patients. As the necessity of violating these constraints can be associated with low quality plans, violations would be avoided if good plans are defined from the beginning. Our results have shown that the assignments computed by the proposed robust model are compatible with both the satisfaction of the constraints and acceptable costs and workloads, differently from the historical condition of workloads. Hence, such constraints can be considered as hard ones.

Finally, a direct comparison with the other approaches applied to this problem (Lanzarone et al., 2012; Lanzarone & Matta, 2012) is not directly possible. Indeed, the proposed model takes into account features and requirements which are not considered in previous approaches. As example, a simplified model is considered in Lanzarone et al. (2012); in fact, the stochastic programming approach is computationally demanding, and it is not adequate for managing a full model that includes the three types of continuity of care requirement.

4. Discussions and conclusions

In this paper, we apply the robust cardinality-constrained approach proposed in Bertsimas and Sim (2004) to a health care application, namely, the nurse-to-patient assignment in HC services under continuity of care requirement. Nurse-to-patient assignment in HC is a relevant yet not very studied problem in HC management, and it significantly suffers from parameters uncertainty.

According to the cardinality-constrained approach, a deterministic assignment model is modified to take into account uncertainty in patients' demands without the necessity of assuming probability distributions or deriving scenarios.

The proposed model has been tested on both historical data of patients' demands and a set of generated sample paths, and better performances of the robust model in terms of overtime costs and fairness in nurses' workloads with respect to the non-robust counterpart have been highlighted. Moreover, computational times needed to solve the model with a weekly frequency, although sometimes not to optimality, are reasonable.

Considering the real historical patients' demands of the analyzed period, results show that, if the assignments proposed by the robust model had been applied to the demands of that period, they would have allowed the provider to obtain acceptable workloads and costs together with the full respect of the constraints (adequate continuity of care levels and division among districts). If compared to the other methodologies applied to this problem (Lanzarone et al., 2012; Lanzarone & Matta, 2012), the cardinality-constrained approach is able to solve the problem in less computational time (whereas including stochasticity with scenario generation of the stochastic programming requires a huge computational time) and with fewer assumptions on demand variability (whereas the analytical approach based on stochastic ordering requires to introduce several assumptions on the shape of density functions).

The cardinality-constrained robust approach is promising, as it suites well to many health care applications, and therefore its potentiality in health care field is worth being further investigated.

However, the approach has two main limits, which must be addressed. From the computational point of view, results show that there are some instances (i.e., some rolling weeks) that cannot be solved to optimality within the chosen time limit. Besides, the robust model is too computationally expensive to be applied to the initial assignment at rolling week 0, and it is likely to be computationally expensive for instances with a number of patients or nurses higher than those considered from rolling week 1 to 25. Hence, also alternative approaches are worth being investigated to deal with larger instances. In Bienstock (2007) the implementor-adversary approach is proposed: in the implementor step, the optimization problem is solved taking into account a subset of robust constraints, whereas in the adversary step new cuts are generated, which represent the worst possible realizations with respect to the current solution. In our problem, the adversary step should generate the worst possible realization of patients' demands for the current assignments. Further, other robustness features, besides the subsets with given cardinality, could be taken into account in generating the cuts. Concerning demand uncertainty, patients' demands are allowed to take only two values in the proposed approach, i.e., expected and maximum values. This may cause the solutions to be excessively conservative, as unlikely values may be taken into account as maximum demands. In addition, in the real practice, patients' demands may assume different values between the expected and the maximum ones. Further, some patients may have a demand lower than the expected value. Such limits could be overcome by introducing different levels of demand for each patient, as proposed in Mattia (2012) or Bienstock (2007), rather than only the two levels considered in this work. As future work, we will apply these approaches to the considered problem.

References

Banditori, C., Cappanera, P., & Visintin, F. (2013). A combined optimization–simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics, 24*(2), 155–186.

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

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

Blais, M., Lapierre, S. D., & Laporte, G. (2003). Solving a home-care districting problem in an urban setting. *Journal of Operational Research Society, 54*(11), 1141–1147.

Boldy, D., & Howell, N. (1980). The geographical allocation of community care resources – A case study. *Journal of Operational Research Society, 31*, 123–129.

Borsani, V., Matta, A., Beschi, G., & Sommaruga, F. (2006). A home care scheduling model for human resources. In *Proceedings of ICSSSM, Troyes, France* (pp. 449–454).

Brotcorne, L., Laporte, G., & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research, 147*, 451–463.

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.

Busby, C. R., & Carter, M. W. (2006). A decision tool for negotiating home care funding levels in Ontario. *Home Health Care Service Quaterly, 25*(3/4), 91–106.

Chahed, S., Dallery, Y., Matta, A., & Sahin, E. (2006). Operations management related activities in home health care structures. In *Proceedings of the INformation COntrol problems in Manufacturing conference (INCOM)* (vol. 3, pp. 641–646).

Chahed, S., Marcon, E., Sahin, E., Feillet, D., & Dallery, Y. (2009). Exploring new operational research opportunities within the home care context: The chemotherapy at home. *Health Care Management Science, 12*(2), 179–191.

Chan, T. C. Y., Bortfeld, T., & Tsitsiklis, J. N. (2006). A robust approach to IMRT optimization. *Physics in Medicine and Biology, 51*(10), 2567–2583.

Cheang, B., Li, H., Lim, A., & Rodrigues, B. (2003). Nurse fostering problems – A bibliographic survey. *European Journal of Operational Research, 151*, 447–460.

De Angelis, V. (1998). Planning home assistance for AIDS patients in the city of Rome, Italy. *Interfaces, 28*(3), 75–83.

Denton, B. T., Miller, A. J., Balasubramanian, H. J., & Huschka, T. R. (2010). Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research, 58*, 802–816.

Haggerty, J., Reid, R., Freeman, G., Starfield, B., Adair, C., & McKendry, R. (2003). Continuity of care: A multidisciplinary review. *British Medical Journal, 327*, 1219–1221.

Heffernan, J. F., & Husni, M. (2009). Continuity of care coordination, health needs and social deprivation. *The Psychiatrist, 33*, 132–134.

Hertz, A., & Lahrichi, N. (2009). A patient assignment algorithm for home care services. *Journal of Operational Research Society, 60*(4), 481–495.

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

Koeleman, P. M., Bhulai, S., & Van Meersbergen, M. (2012). Optimal patient and personnel scheduling policies for care-at-home service facilities. *European Journal of Operational Research, 219*, 557–563.

Lahrichi, N., Lapierre, S. D., Hertz, A., Talib, A., & Bouvier, L. (2006). Analysis of a territorial approach to the delivery of nursing home care services based on historical data. *Journal of Medical Systems, 30*(4), 283–291.

Lanzarone, E., & Matta, A. (2012). A cost assignment policy for home care patients. *Flexible Service and Manufacturing Journal, 24*(4), 465–495.

Lanzarone, E., Matta, A., & Sahin, E. (2012). Operations management applied to home care services: The problem of assigning human resources to patients. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 42*(6), 1346–1363.

Lanzarone, E., Matta, A., & Scaccabarozzi, G. (2010). A patient stochastic model to support human resource panning in home care. *Production Planning and Control, 21*(1), 3–25.

Mannino, C., Nilssen, E. J., & Nordlander, T. E. (2012). A pattern based, robust approach to cyclic master surgery scheduling. *Journal of Scheduling, 15*(5), 553–563.

Mattia, S. (2012). *Robust optimization with multiple intervals*. Technical Report IASI-CNR (2012), number 12-07.

Nickel, S., Schroder, M., & Steeg, J. (2012). Mid-term and short-term planning support for home health care services. *European Journal of Operational Research, 219*, 574–587.

Rasmussen, M. S., Justesen, T., Dohn, A., & Larsen, J. (2012). The home care crew scheduling problem: Preference-based visit clustering and temporal dependencies. *European Journal of Operational Research, 219*, 598–610.

Trautsamwieser, A., & Hirsch, P. (2011). Optimization of daily scheduling for home health care services. *Journal of Applied Operational Research, 3*(3), 124–136.