

Robust Nurse-to-Patient Assignment in Home Care Services to Minimize Overtimes under Continuity of Care

Ettore Lanzarone¹ and Andrea Matta²

¹*Istituto di Matematica Applicata e Tecnologie Informatiche (IMATI),
Consiglio Nazionale delle Ricerche (CNR), Via Bassini 15, 20133 Milano, Italy,
ettore.lanzarone@cnr.it (corresponding author).*

²*Dipartimento di Meccanica, Politecnico di Milano,
Via La Masa 1, 20156 Milano, Italy,
andrea.matta@polimi.it*

Abstract

Home Care (HC) providers are complex organizations that manage a large number of patients, different categories of operators, support staff and material resources in a context affected by high variability. Hence, robust resource planning is crucial for operations in HC organizations, in order to avoid process inefficiencies, treatment delays, and low quality of service. Under continuity of care, one of the main issues in HC planning is the assignment of a reference nurse to each assisted patient, because this decision has an impact on the workload assigned to the nurse for the entire patient's length of stay. In this paper, we derive an analytical structural policy for solving the nurse-to-patient assignment problem in the HC context under continuity of care. This policy accounts for the randomness that is related to the demands from patients already assigned to nurses and to the demands from new patients who need assignments. The policy is compared to other previously developed approaches, and applied to a relevant real case.

Keywords:

Home Care, Nurse-to-Patient Assignment, Stochastic Demands, Continuity of Care

1. Introduction

1.1. Background

Home Care (HC) consists of delivering medical, paramedical and social services to patients at their domicile rather than in hospital. HC leads to a significant increase in the quality of life for patients, who are assisted at home, and to considerable cost savings for

the entire health care system (Davies and Dale, 2003; Chevreul et al., 2005; Comondore et al., 2009). HC is a relevant sector of the health care domain in western countries: in the US, about 3.3 million beneficiaries received HC services from more than 11,400 agencies in 2009, and Medicare spent 19 billion dollars on HC (Medicare Payment Advisory Commission MEDPAC, 2011). Moreover, these numbers are continuously growing because of the aging of the population, the increase in chronic pathologies, the introduction of innovative technologies, and the continuous pressure of governments to contain health care costs (Chevreul et al., 2005; American Geriatrics Society AGS, 2006).

Many resources are involved in HC service delivery, including different categories of operators (nurses, physicians, physiotherapists, social assistants and psychologists), support staff and material resources. Managing these resources is a difficult task because the HC provider has a large number of assisted patients, must synchronize the use of the resources at each patient's home, and delivers the service to an often vast territory (Comondore et al., 2009; Chahed et al., 2009; Matta et al., 2012). Therefore, robust resource planning is crucial for operating in HC organizations, to avoid process inefficiencies, treatment delays, and low quality of the provided service.

In addition, random events affect the service delivery and undermine the feasibility of plans, e.g., variations in patients' conditions, resource unavailabilities and longer durations of operator transfers in the territory. The most critical and frequent events are sudden variations in patients' conditions, which make the service demand highly uncertain and the resource planning more complex. As example, Lanzarone et al. (2010) show that the coefficient of variation (i.e., the ratio between the standard deviation and the expected value) of the weekly demand from HC patients ranges from 0.39 for medium-high care intensity patients to 1.29 for low care intensity patients. Finally, the existence of some constraints, such as the continuity of care and the risk of incurring a burnout of operators (Cordes and Dougherty, 1993), makes the HC resource planning different from the planning problems encountered in other production and service systems, also within the health care domain.

In the HC context, continuity of care means that only one operator of each category is assigned to a patient, named the *reference* or *principal* operator, who is not changed for a long period, usually a semester (Borsani et al., 2006), and preferably provides all of the visits required by the patient to the operator's category. Continuity is considered an important quality indicator of the HC service, because the potential loss of information among operators is avoided, and the patient receives care from the same operator rather

than having to continuously develop new relationships (Haggerty et al., 2003).

Despite the complexity required for planning, in the majority of HC providers, human resource planning is not supported by proper skills, methodologies and tools that are needed for managing the logistics and the organizational activities of care delivery. Hence, the possibility of implementing adequate planning models and tools for the HC context could improve the robustness of plans with a limited investment of support staff sustained by the HC providers.

The main issues of HC resource planning are the partitioning of a territory into a given number of districts, the human resource dimensioning, the assignment of visits to operators, and the scheduling and routing determinations (Yalcindag et al., 2011; Matta et al., 2012). These issues involve three planning levels:

- *Districting and resource dimensioning*: HC patients are grouped into categories, depending on the type of required service. Typically, the main distinction is between palliative and non palliative patients; in line, each nurse is characterized by a skill, i.e., the set of patient classes he/she can handle. This planning level consists of dividing the territory served by the provider into regions and the patients of each region into groups based on the skill requested to the nurses. Each group represents a district, and a certain number of skilled nurses are assigned to the districts to satisfy the demand of their patients.
- *Operator assignment*: once allocated to a district, operators are chosen to provide each visit based on different criteria (e.g., the time of the day, the specific service requested during the visit, etc). Under continuity of care, patients (and not single visits) are assigned to the operators and the assignments are kept along with the time.
- *Scheduling, routing and control*: this is the definition of the weekly plans with the sequence of visits for each operator, taking into account the planned assignments. This level also includes the daily control of the activities to respond to unavailabilities or unexpected variations in service demand.

In this paper, we focus on the operator assignment under continuity of care, taking into account the nurses.

1.2. The nurse-to-patient assignment problem in Home Care

Patients need to be assisted by different categories of operators: they are always in charge of the nurses and, depending on the case, they may involve one or more figures such as a physiotherapist, physician, or psychologist. Operators of each category are divided into districts depending on their skills and territory. Assignments are usually planned considering the districts as independent, meaning that each patient is cared for by an operator with a skill that is compatible to his/her pathology and who works in his/her geographical area. However, in the real practice of HC providers, after the assignments are set, an operator may also be assigned to care for patients who do not belong to his/her district, to compensate for the infeasibilities of scheduling and routing, or to partially compensate for workload unbalancing among districts (Comondore et al., 2009; Yalcindag et al., 2011; Lanzarone and Matta, 2012).

A key issue for the assignments is the continuity of care, particularly for nurses. A large number of HC providers pursue the continuity of care. However, some other HC organizations do not adopt the concept of reference operator to increase the operational efficiency, and each visit to a patient is provided by any appropriate operator who has sufficient available capacity in the required period of time. In this way, at each planning period, no constraint deriving from previous assignments has to be included and no engaged workload of operators related to previous assignments has to be managed.

Nurse-to-patient assignment under continuity of care consists of assigning each newly admitted patient to his/her reference nurse, chosen among the compatible ones (i.e., belonging to the new patient's district). The goals pursued by HC providers can be different depending on the provider. A widespread goal is the minimization of the overtime incurred by operators. This minimization is highly important mainly for two reasons: the provider minimizes the operators' extra times to be paid and, consequently, the sustained variable costs. At the same time, such objective reduces the risk of burnout, which is related to the care volume exceeding the operator contract capacity. Burnout is a syndrome that can affect a broad range of professions (including physicians, nurses and educators) as a prolonged response to chronic job-related stressors (Cordes and Dougherty, 1993); this phenomenon causes decreased job performance and reduced job commitment, bringing workers to stress-related health problems and low career satisfaction. Another objective usually adopted by providers is to obtain a fair and balanced workload among the operators (Kovner et al., 2006; Lanzarone et al., 2012).

1.3. Contribution

The main difficulty in solving the assignment problem is to face the random events that affect the service delivery and give a high variability to the workloads charged to nurses and, consequently, to the overtime costs. This paper addresses the problem of assigning newly admitted patients to their reference nurse, while maintaining continuity of care. Specifically, this paper proposes an analytical policy for solving the nurse-to-patient assignment problem within the HC setting, taking into account the stochasticity of the new patient's demand and of the nurses' workloads.

The problem is formalized as a minimization of nurses' maximum overtimes in a lexicographic way: firstly, we minimize the highest value, then the highest value of the remaining ones, and so forth. Operatively, the goal is analytically pursued by minimizing the maximum increase of a stochastic cost function, which depends on the time spent by nurses when providing visits in surplus to their capacity. This objective also leads to balance the workloads of the nurses within districts when the HC structure is not underutilized. The simplicity of the proposed policy makes easy its implementation in practice.

Because of the assumptions introduced, the policy requires validation in real cases. When applied to a significant real case, the approach presented in this paper guarantees lower overtimes and better workload balancing when compared to a numerical approach based on mathematical programming (Lanzarone et al., 2012), to a policy which minimizes the expected value of the square overtimes (Lanzarone and Matta, 2012), and to the usual practice of HC providers. Specifically, lower overtimes and better balancing are obtained with the policy for the majority of patients, whose demand is characterized by a high variability. Only for patients whose demand is characterized by a low variability (e.g., palliative patients), results among the approaches are similar and show that considering uncertainty in the nurse assignment problem does not add significant benefits and simpler approaches can be successfully applied.

1.4. Structure of the paper

A literature analysis of the assignment problem is firstly presented in section 2. The formal statement of the nurse-to-patient assignment problem under continuity of care and the assumptions introduced are described in section 3. The proposed policy is then derived starting from the single-patient assignment (section 4) and extending to the multi-patient

assignment (section 5). Finally, section 6 reports the application of the proposed approach to a relevant real case, and the final conclusions are reported in section 7.

2. Literature review

The assignment problem is a very general problem with variety of applications, studied since 1952 (Pentico, 2007). In the most general formulation, namely the Generalized Assignment Problem (GAP), a set of *tasks* have to be optimally matched to a set of *agents*. In our problem, tasks refer to patients and agents to nurses.

Several variants of the GAP are present in literature and have been applied to different problems, as described in the recent survey of Pentico (2007). The nurse-to-patient assignment problem can be classified as a Bottleneck Assignment Problem (BAP) or as a Minimum Deviation Assignment Problem (MDAP). In the BAP, the objective function is the minimization of the maximum assigned workload, which in our case corresponds to reduce nurses' overtimes. In the MDAP, the objective is the minimization of the difference between the maximum and the average assigned workload, which in our case corresponds to balance the workload among the nurses.

The BAP is a well-known problem in manufacturing, where jobs have to be loaded on machines for processing. This is known in literature as the loading problem (Pinedo, 2012), in which one of the most utilized objective function is to minimize the completion time of the last processed job. The way in which overtimes are modeled is different from HC. In manufacturing, aim at minimizing the machine cost due to overtimes (e.g., operating in the third shift), it could be indifferent to assign two jobs to one single machine or two different machines, because the total cost may be the same. On the contrary, in nurse assignment problems, it is preferable to avoid high overtimes that could increase the risk of burnout.

The loading problem is faced as a subproblem in scheduling of manufacturing systems. Examples of recent papers dealing with loading problems in manufacturing are Abazari et al. (2012) and Kim et al. (2012). Randomness is often considered in manufacturing; in this case, the literature refers to stochastic load balancing problems and stochastic scheduling problems.

The solution methods proposed in the manufacturing literature are simple procedures based on Greedy strategies or complex algorithms based on mathematical programming or meta-heuristics (Grieco et al., 2001), which may also require the use of simulation for evaluating the system performance. The former approach is too simple for the problem addressed

in this paper, whereas the latter is of difficult implementation in the practice. In addition, this difficulty increases when the techniques are applied in health care organizations, because of the lack of operations management skills among the decision makers. Analytical rules are also defined for some specific simple cases (Cai and Zhou, 2005; Cai et al., 2007), but they cannot be easily extended to other problems, as the considered the nurse-to-patient assignment problem.

Finally, Pinedo (2012) includes a complete review of the assignment, loading and scheduling problems.

In the following, a review of the most relevant works dealing with the assignment problem in the health care domain is reported for both the widespread studied hospital case (section 2.1) and the HC context (section 2.2).

2.1. Assignment problem in hospitals

A large number of studies address the assignment and scheduling of nurses and medical staff in hospitals. In this section we only refer to review papers and to works that explicitly treat overtimes in hospitals.

Cheang et al. (2003) report a bibliographic survey of the methodologies that have been proposed to solve the nurse rostering problem. Also, Burke et al. (2004) propose a literature review of the problem, considering the nurse rostering problem within the global personnel scheduling in health care and drawing on the strengths and the weaknesses of the literature to outline the key issues for the nurse rostering research. Bard (2010) reports a recent literature review on nurse scheduling models.

Additionally, other works focus the analysis on the overtime to be paid. Brunner et al. (2009) develop a mixed-integer program model for a flexible shift scheduling of physicians in a German university hospital. Their objective is to find an assignment such that the total hours to be paid out as overtime are minimal under the restrictions given by the labor agreement. Brunner et al. (2011) also propose a methodology for solving the flexible shift scheduling problem of physicians in the presence of flexible start times, variable shift lengths, and overtime to cover demand.

Assignment problems in hospitals differ from those faced in HC in many aspects. First, caregivers in hospitals are assigned to time slots and not to patients, thus the continuity of care is not an issue in hospitals. Other differences are the territory, which is relevant in HC and not in hospitals, the relationships that caregivers activate at patient's home, and the

large variability of the home and social environment where patient lives.

2.2. Assignment problems in home care

In the HC literature, the scheduling and routing of human resources represent the most important volume of existing investigations within the planning context. Several studies address human resource planning in HC services without continuity of care or with partial continuity of care (Chahed et al., 2009; Eveborn et al., 2006, 2009; Bertels and Fahle, 2006; Thomsen, 2006; Akjiratikarl et al., 2007; Bennett and Erera, 2011).

Only a few papers take into account the continuity of care and the nurse-to-patient assignment problem, which is the focus of our paper. Among them, Borsani et al. (2006) study the scheduling of visits in two HC organizations operating under continuity of care. They propose an assignment model coupled to a scheduling one. The objective of the assignment is to ensure the workload balancing among the operators while respecting qualification requirements and geographical coherence constraints. Ben Bachouch et al. (2008) develop a mixed integer linear programming model to minimize the total distance traveled by nurses. This model is subject to several constraints, including visits' and nurses' time windows, nurses' meal breaks, continuity of care, each nurse's route beginning and ending at the HC facility, and the maximum distance between two consecutive visits by the same nurse. Hertz and Lahrichi (2009) propose two mixed programming models for allocating operators to patients in the Cotes-des-Neiges local community health clinic in Montreal, Canada. One model consists of linear constraints and a quadratic objective function, while the other includes nonlinear constraints and is solved by a Tabu search heuristic. Constraints related to maximum acceptable workloads and the assignment of each patient to exactly one nurse of each type are imposed, and the objective of the assignment is to balance the nurses' workloads by minimizing a weighted sum of the number of provided visits, of the assigned patients, and of the distances traveled. The possibility of assigning a patient to a nurse who does not belong to the patient's district is also considered. Lanzarone et al. (2012) propose a stochastic programming approach based on a mixed integer linear programming for solving the HC assignment problem while including the variability of patients' demands. Finally, Lanzarone and Matta (2012) propose an analytical policy to minimize the square value of the overtimes incurred by nurses. This policy will be considered in the graphical and numerical comparisons reported in section 4.4 and section 6, respectively.

3. Problem statement

In this section, we describe the specific nurse-to-patient assignment under continuity problem addressed in this paper (section 3.1), and we introduce the adopted formalism and the assumptions that are useful for deriving the assignment policy (section 3.2). All of the assumptions introduced are consistent with the real situation of several HC providers (Comondore et al., 2009; Chahed et al., 2009; Lanzarone et al., 2010).

3.1. Problem description

We analyze the case in which a set of new patients are admitted into service after an initial assessment of their clinical, social and psychological needs. Each newly admitted patient has to be assigned to only one reference nurse who is compatible with his/her needs, and this assignment is never changed to preserve the continuity of care. The problem is solved considering a single time period (e.g., a week or a month): this means that each patient's demand or nurse's workload refers to the time for visits in this future planning period. The problem of allocating visits to specific days of the week once the assignments are decided is not considered in this paper.

In general, the new patient is assigned to a nurse who belongs to his/her district. Furthermore, we consider that the district is not related to a vast geographical area. Thus, a fixed transportation time from a patient's home to another is assumed in the practice of planning. As a consequence, the demand that is related to the serviced patients can be expressed as the total amount of time requested for the visits in the planning horizon, including the transportation time of each visit.

A contract regulates each nurse's time capacity during the planning period. Moreover, nurses have been already given responsibility for the patients allocated in previous assignments and, consequently, have an initial assigned workload. The provision of visits represents a cost for the HC provider resulting in both fixed and variable costs. Fixed costs are not considered in the objective, because they are sustained independent of the decided assignments of patients to nurses once the workforce is enrolled. In contrast, the variable cost of each nurse is affected by the assigned patients, because it depends on the amount of care that the nurse provides in surplus to his/her capacity. Thus, patients are assigned aiming at minimizing the overtime incurred by nurses. The overtime is also kept low to avoid burn-out of operators.

The amount of service requested by all of the serviced patients (either newly admitted or already in care from previous periods) during the planning period is random. Randomness is due to the patient’s health and psycho-social conditions that may change over the time, increasing or decreasing the demand for service (Garg et al., 2010; Lanzarone et al., 2010). For instance, terminal patients usually increase their service request as their conditions worsen from day to day. Additionally, even if patient’s conditions do not change, the amount of service required within the planning period might be affected by short term fluctuations.

3.2. Modeling assumptions and notation

We study the single-district case in which a set of K newly admitted patients must be assigned to one reference nurse taken from the set Ω of compatible ones. Out-of-district assignments are not allowed, thus nurses can take care only of patients living in the same district.

The amount of service requested by newly admitted patient k (with $k = 1, \dots, K$) in the planning period is modeled as a random variable denoted with Y_k . Additionally, we assume that the duration of visits in the planning horizon (i.e., each variable Y_k) does not depend on the specific nurse-to-patient assignment.

The workload assigned to each nurse $i \in \Omega$ in the planning period is denoted with X_i . This is given by the sum of the initial workload X_i^0 and the demands of the newly assigned patients. X_i^0 is given by the amount of service time requested by the already assigned patients. These assignments must be kept for respecting the hard constraint of the continuity of care. X_i^0 is modeled as a positive continuous random variable with probability density function $\Phi_i(x_i^0)$. It is assumed that $\Phi_i(x_i^0)$ can be estimated using the most recent available information by means of appropriate stochastic models, as the one reported in Lanzarone et al. (2010).

In addition, each nurse i has a capacity v_i , which is expressed as the time for visits in the planning period.

The objective of the assignments is to minimize the maximum overtime among the nurses, then the next-to-maximum overtime, etc., in the lexicographic sense. Extending the classification of Pentico (2007), the problem can be formulated as a stochastic lexicographic bottleneck GAP. According to the introduced notation, the problem is stated as follows:

$$\text{lex min } \{ \max OVT_i \} \tag{1}$$

s.t.

$$\begin{aligned}
OVT_i &= \max \{X_i - v_i; 0\} & \forall i \\
X_i &= X_i^0 + \sum_{k=1}^K z_{ik} Y_k & \forall i \\
z_{ik} &\in \{0, 1\} & \forall i, k
\end{aligned}$$

where OVT_i is the overtime of nurse i , and z_{ik} is a binary decision variable equal to 1 if patient k is assigned to nurse i , and 0 otherwise.

The problem is first solved analytically in the simpler case of single-patient assignment (section 4). This analytical solution is then used in the multi-patient assignment after sorting the newly admitted patients (section 5).

4. Single-patient assignment problem

In this section we describe the proposed analytical approach for assigning one newly admitted patient. Since only one new patient is assigned ($K = 1$), the index k referring to the patient is omitted and his/her demand is simply denoted with Y in the remaining of this section.

Further assumptions are introduced for the analytical derivation, then the policy is enunciated and a robustness analysis is conducted. Finally, we also provide a graphical interpretation of the proposed policy, and we compare it with alternative ones.

4.1. Additional modeling assumptions

For the single-patient decision making purposes, we model $\Phi_i(x_i^0)$ as a triangular distribution with parameters a_i , b_i and c_i . Parameter a_i is the minimum value that the initial workload can assume, b_i is the mode of the distribution, and c_i is the maximum initial workload value (with $0 < a_i < b_i < c_i$). We denote this distribution with $\tilde{\Phi}_i(x_i^0)$. The use of the triangular distribution allows us to analytically derive the policy and to obtain simple rules, which can be given to planners and practitioners for their real implementation.

Additionally, the value of the nurse capacity v_i is assumed to be in the second part of the triangular distribution ($b_i \leq v_i \leq c_i$, $\forall i \in \Omega$) without loss of generality, because the other possibilities for v_i are not of interest for a practical application of an assignment policy. Indeed, the case $v_i < b_i$ refers to a highly overloaded nurse who should not be considered

for the assignment of another patient, while the case $v_i > c_i$ refers to a highly underloaded nurse, which is not frequent in real organizations in stationary conditions.

To be more general, we introduce a stochastic cost function that depends on the nurse's overtime. Given a workload X_i and a contract capacity v_i , we propose the following function:

$$C_i(X_i) = OVT_i^m = [\max\{X_i - v_i; 0\}]^m = \max\{(X_i - v_i)^m; 0\} \quad (2)$$

where the parameter m is a nonnegative constant used to penalize the overtime.

In providers where the cost of an extra visit is constant and does not depend on the overtime amount, $C_i(X_i)$ simply represents the nurse's overtime (i.e., $m = 1$). In providers that consider a rise of the cost while the overtime increases, $C_i(X_i)$ is a convex function of X_i (i.e., $m > 1$). In this case, a nurse who supplies two visits of one hour above v_i is considered more costly with respect to two nurses who supply one visit of one hour above v_i . The more the HC provider tends to avoid highly overloaded nurses (both to reduce costs and to maintain a wellbeing of nurses and, consequently, a high quality of visits), the more m assumes large values. The case $m < 1$ does not fit real HC providers and is not considered. Once the specific HC provider is defined, parameter m is assumed to be the same for all the nurses of the provider.

4.2. Proposed policy

The goal of the assignment is to minimize the overtimes in a lexicographic sense (1). This corresponds to assign the new patient to the nurse i with the lowest maximum overtime before the assignment. At the same time, this also corresponds to assign the new patient to the nurse i with the lowest value of the maximum cost $C_i(X_i^0)$ before the assignment, independent of the value of m .

Thus, the policy is stated as follows:

Given two nurses i and j ($i, j \in \Omega$) with parameters a_h, b_h, c_h (with $h = i, j$), the newly admitted patient has to be assigned to nurse i if $c_i - v_i < c_j - v_j$ or to nurse j if $c_j - v_j < c_i - v_i$.

As the initial workloads X_i^0 cannot be reallocated, this policy also corresponds to minimize the maximum increase of the cost function $C_i(X_i)$ in case $m > 1$, i.e., in providers with a rise of the cost while the overtime increases.

We remark that the maximum overtime refers to the triangular shape of $\tilde{\Phi}_i(x_i^0)$. When applied to a real case, starting from the available $\Phi_i(x_i^0)$, the right tail of the distribution

could be cut by the triangular fitting depending on the density shape. Hence, the policy is not too conservative because long right tails associated with low probability are not considered as they are eliminated by the fitting.

4.3. Robustness analysis

The policy proposed is conservative because it minimizes the maximum value of the cost function. However, it is possible to find the conditions under which the policy is optimal in the stochastic sense, i.e., it minimizes the entire $C_i(X_i^0)$ density function, and not only its maximum value. For this purpose, the minimization is provided by comparing two functions in terms of their entire probability density functions using stochastic order theory (Shaked and Shanthikumar, 1994, 2007). Then, the nurse to be chosen for the assignment is the one with the lowest density function of cost before the assignment, independent of Y .

As in the proposed policy, the lowest density function of cost before the assignment also corresponds to the lowest density function of cost increase in case of $m > 1$. This relationship holds from the independence between the demand Y of the newly admitted patient and the initial workloads $X_i^0 \forall i \in \Omega$.

The most important univariate stochastic orders are the usual stochastic order (ST), the hazard rate order (HR) and the likelihood ratio order (LR) (Shaked and Shanthikumar, 1994, 2007). Among them, we adopt the LR order because it implies the other two cases. The LR order affirms that, considering two continuous (or discrete) random variables A and B with densities (or discrete densities) $a(t)$ and $b(t)$, respectively, A is smaller than B (expressed as $A \leq_{lr} B$) if the ratio $b(t)$ to $a(t)$ increases in t over the union of the supports of A and B . This holds to:

Theorem 1. *Given two nurses i and j with initial workloads $X_i^0 \sim \tilde{\Phi}_i(x_i^0)$ and $X_j^0 \sim \tilde{\Phi}_j(x_j^0)$ before the assignment, respectively, then $C_i(X_i^0) \leq_{lr} C_j(X_j^0)$ and the robust decision is assigning the patient to nurse i if the following conditions are true:*

$$\begin{cases} c_i - v_i \leq c_j - v_j \\ \frac{c_i - v_i}{r_i(c_i - a_i)^2} \leq \frac{c_j - v_j}{r_j(c_j - a_j)^2} \end{cases} \quad (3)$$

and vice versa, where $r_i = 1 - r_{bi} - r_{vi}^2$ is defined as the shape index of the triangular density $\tilde{\Phi}_i(x_i^0)$, and

$$r_{bi} = \frac{b_i - a_i}{c_i - a_i}; \quad r_{vi} = \frac{c_i - v_i}{c_i - a_i}$$

are the asymmetry index and the surplus of workload index, respectively, with $0 < r_{bi} < 1$, $0 < r_{vi} < 1$ and $0 < r_i < 1$.

Proof. Let us consider the case $c_i - v_i < c_j - v_j$, so that the union of the supports of $\gamma_i(C_i(X_i^0))$ and $\gamma_j(C_j(X_j^0))$ is the interval $[0, (c_i - v_i)^m]$, with $\gamma_i(C_i(X_i^0))$ representing the probability density function of random variable $C_i(X_i^0)$. Functions $\gamma_i(C_i(X_i^0))$ and $\gamma_j(C_j(X_j^0))$, whose expression is derived in the Appendix, have a discontinuity, due to the Dirac δ function at the origin $\delta(0)$. Hence, the condition for having $C_i(X_i^0) \leq_{lr} C_j(X_j^0)$ assumes the following form (Shaked and Shanthikumar, 1994, 2007):

$$\frac{\gamma_i(q)}{\gamma_i(p)} \leq \frac{\gamma_j(q)}{\gamma_j(p)} \quad \forall p \leq q \quad (4)$$

Three alternatives are possible. (i) For $p = q = 0$, inequality (4) is always verified. (ii) For $p = 0$ and $q > p$, inequality (4) assumes the form $f_i(q) \geq f_j(q)$ with:

$$f_h(q) = \frac{q^{\frac{1}{m}} - (c_h - v_h)}{r_h (c_h - a_h)^2} \quad h = i, j \quad (5)$$

$f_h(q)$ varies from $-\frac{c_h - v_h}{r_h(c_h - a_h)}$ in $q = 0$ to 0 in $q = (c_h - v_h)^m$ with a linear trend if $m = 1$ or a root trend if $m > 1$. Hence, the solution is the following:

$$\begin{cases} c_i - v_i \leq c_j - v_j \\ \frac{c_i - v_i}{r_i(c_i - a_i)^2} \leq \frac{c_j - v_j}{r_j(c_j - a_j)^2} \end{cases} \quad (6)$$

(iii) For $q \geq p > 0$, inequality (4) gives the following result: $c_i - v_i \leq c_j - v_j$.

Summarizing all of the alternatives, the overall conditions are the same as in system (6). □

Also in this case the theorem is independent from parameter m . This means that there exist some conditions according to which it is always preferable to assign a patient to one operator instead of another whatever the importance of the overtime is.

4.4. Graphical representation

The aim of this section is to give a graphical representation of the policy, which allows a simple visualization of the assignment rule and an immediate comparison with alternative approaches.

A new patient to be assigned and two compatible nurses $i, j \in \Omega$, which are characterized by their initial workload density functions $\tilde{\Phi}_i(X_i^0)$ and $\tilde{\Phi}_j(X_j^0)$, are considered. For obtaining the representation, the density functions are first normalized with respect to the capacities v_i and v_j , considering the *assignable workload* (i.e., the time for visits that can be assigned to the nurse) or the *excess workload* (i.e., the time for visits that exceeds the nurse's capacity). Therefore, parameters a_i , b_i and c_i are translated and expressed in relative terms with respect to the nurse's capacity v_i :

$$\begin{cases} \tilde{a}_i = a_i - v_i \\ \tilde{b}_i = b_i - v_i \\ \tilde{c}_i = c_i - v_i \end{cases} \quad (7)$$

The domain of these parameters is determined according to the relationships imposed by the variables, i.e., $\tilde{a}_i \leq \tilde{b}_i \leq 0 \leq \tilde{c}_i$.

In this way, each nurse i is characterized by three parameters \tilde{a}_i , \tilde{b}_i and \tilde{c}_i and the graphical representation would involve a three-dimensional space. On examining data from several HC providers, it is frequently found that the nurses of a specific district (i.e., the set Ω of the nurses compatible with the newly admitted patient) have very similar asymmetries r_{bi} in their workload distributions $\tilde{\Phi}_i(X_i^0)$. Assuming the same value of r_{bi} for all of the nurses, they can be represented in the same plane $\{\tilde{c}, |\tilde{a}|\}$, where the abscissa \tilde{c} is the *maximum excess workload* and the ordinate is $|\tilde{a}|$ the *maximum assignable workload*. Each point in the feasible region that satisfies $\tilde{a}_i \leq \tilde{b}_i \leq 0 \leq \tilde{c}_i$ corresponds to a nurse with a specific initial workload (Figure 1).

The proposed policy is represented by two regions in the plane $\{\tilde{c}, |\tilde{a}|\}$. Points j to the left of point i determine the assignment of the new patient to nurse j , and vice versa (Figure 1A).

The stochastic order of costs for two nurses i and j , with $r_{bi} = r_{bj}$, is graphically represented by four regions in the plane, because of the presence of two conditions (6). A stochastic order is possible in two of these regions, while in the other two regions it is not possible to stochastically order the costs $C_i(X_i^0)$ and $C_j(X_j^0)$ (Figure 1B).

Figure 1 also shows other two assignment approaches, i.e., the expected cost increase policy and the expected available capacity approach.

The former minimizes the expected value of the cost increase when $m = 2$, as reported in Lanzarone and Matta (2012). Such policy also depends on the new patient's demand density

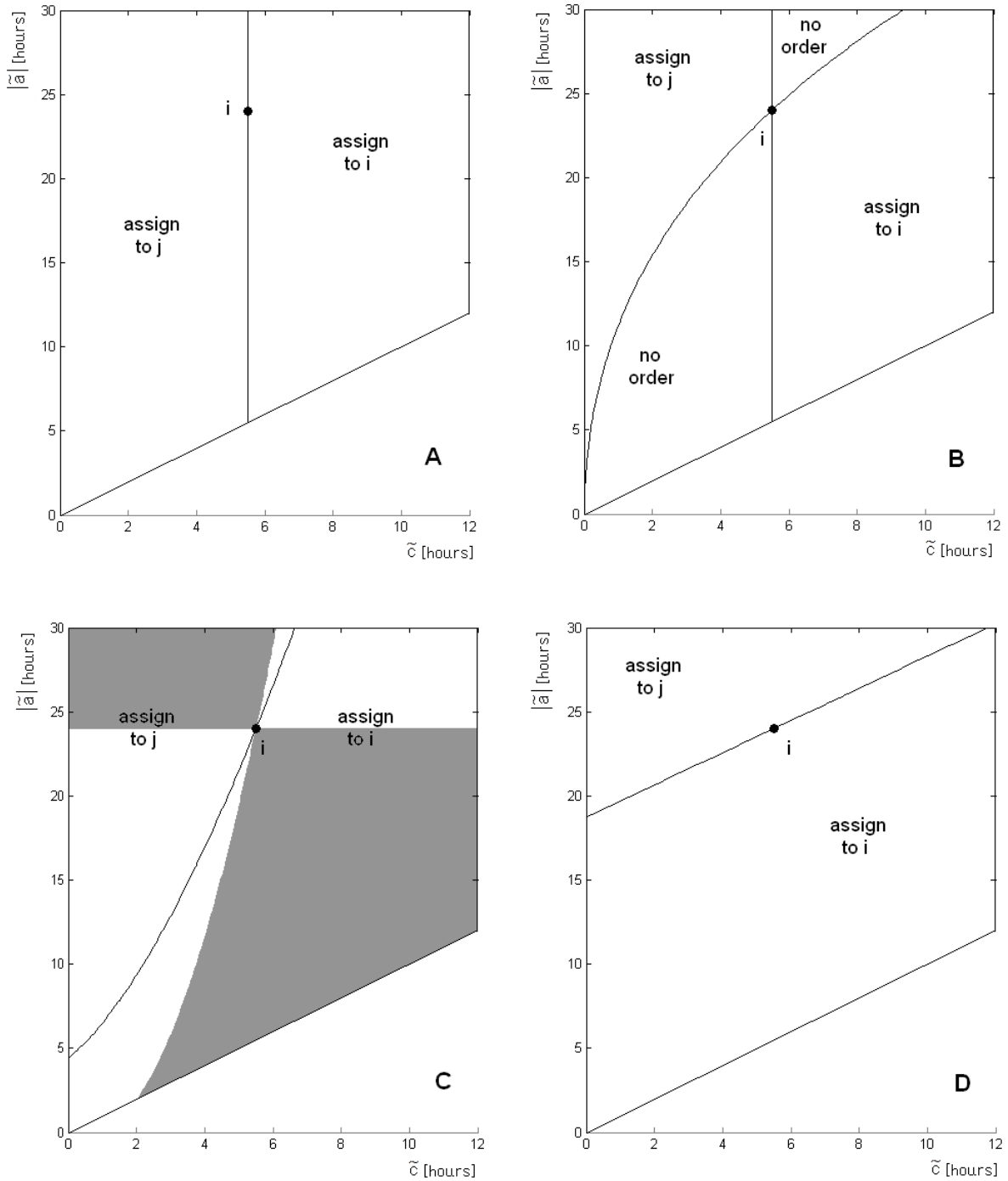


Figure 1: Graphical representation of the assignment for two nurses i and $j \in \Omega$ with $r_{bi} = r_{bj} = 0.5$: proposed maximum cost policy (A), stochastic order of $C_i(X_i^0)$ (B), expected cost increase policy with $m = 2$ (C), and assignment based on the expected available capacity W_i (D). Plot C is obtained with Y uniformly distributed between 2 and 4; the regions where the assignments are independent from new patient's demand are gray colored (Lanzarone and Motta, 2012).

function Y , which is assumed uniformly distributed. The policy separates the plane in two regions, whose division depends on the minimum and maximum values of Y . However, given two nurses i and j with $r_{bi} = r_{bj}$, there is a part within each region where the choice is independent from the new patient's demand characteristics (i.e., gray regions in Figure 1C).

The expected available capacity approach is frequently used in the practice of HC providers, where the variability of the future patient demand is usually neglected and the only information used for assigning each new patient to the reference nurse is the expected workload of each nurse (often estimated from reference standard values of demand for his/her patients in charge). The newly admitted patient is assigned to the nurse i with the highest expected available capacity W_i , given by the difference between the capacity v_i and the expected workload. Considering the triangular density function $\tilde{\Phi}_i(x_i^0)$, W_i is expressed as follows:

$$W_i = v_i - \frac{a_i + b_i + c_i}{3} = -\frac{\tilde{a}_i + \tilde{b}_i + \tilde{c}_i}{3} = -\frac{\tilde{a}_i(2 - r_{bi}) + \tilde{c}_i(1 + r_{bi})}{3} \quad (8)$$

and the corresponding graphical representation is in Figure 1D. This assignment is equivalent to apply the Graham's rule for load balancing, using expected values instead of deterministic service times, and modified to consider that the operators can have different contracts (Goel and Indyk, 1999).

The feasible region of the plane $\{\tilde{c}, |\tilde{a}|\}$ is divided into six regions by the three policies, neglecting the expected available capacity approach (Figure 2). The same assignment choice is provided by the three policies in the largest two regions, i.e., the regions of Figure 1B where a choice can be made are included in the regions of both Figure 1A and Figure 1C where the same choice is made.

In general, the stochastic order of costs $C_i(X_i^0)$ is the most robust approach, because it completely encompasses the variability of the problem (Shaked and Shanthikumar, 1994, 2007)). If this ordering can be applied, each other rule that extracts only a feature from the cost distributions (e.g., the maximum value as in the proposed policy) leads to the same assignments; thus, the other policies provide the same assignment, and this assignment is robust because the same choice is also confirmed by the stochastic order $\forall m$. In this case, the proposed maximum cost policy is the most suitable one, because it is the simplest one and does not require any assumption on Y nor a decision on m . In the four remaining regions, the stochastic order policy cannot be applied (Figure 2) and the selection of the

policy can be significant. Also in this case, the maximum cost policy is the most suitable choice, as it is more conservative and easier to apply than the other.

5. Multi-patient assignment problem

The policy proposed in the previous section and the other described approaches focus on the assignment of one single patient. When a certain number of newly admitted patients have to be assigned at the same time, i.e., when a multi-patient assignment is required, each approach is coupled with an ordering of new patients. In this way, the first patient of the ordered list is assigned with a single-patient assignment, his/her demand is included in the initial workload X_i^0 of the assigned nurse, and these two phases are repeated for all other remaining new patients until all of them are assigned.

The idea is to maintain an analytical approach for each single-patient assignment, fitting the triangular density $\tilde{\Phi}_i(x_i^0)$ each time an assignment has to be decided. Then, before deciding the next assignment, the density function of the latest chosen nurse's workload is numerically updated: the real density $\Phi_i(x_i^0)$ is taken, and a convolution is computed with the real shape of the density function of the latest assigned patient's demand. In this way, the approximation induced by the triangular fitting of the densities is not propagated. The approach is clearly suboptimal. However, in this way, the most time consuming phase (i.e., the assignment decision) is analytically treated so as to exploit the benefits of the proposed single-patient policy.

More precisely, the overall approach consists of the following steps:

1. Sort new patients starting from the highest demanding one.
2. Compute the densities of the initial workloads $\Phi_i(x_i^0)$, and fit the corresponding triangular densities $\tilde{\Phi}_i(x_i^0)$.
3. Assign the first patient of the ordered list according to a single-patient assignment policy.
4. Remove the assigned patient from the list, and numerically compute (convolution) the density $\Phi_i(x_i^0)$ of initial workload for the assigned nurse.
5. Fit the triangular density $\tilde{\Phi}_i(x_i^0)$ for the assigned nurse.
6. If the list is empty, exit; otherwise, go to step 3.

We remark that, for each subsequent single-patient assignment, the actual initial workloads X_i^0 shall respect the assumptions of the policies. In particular, the condition $b_i \leq v_i \leq$

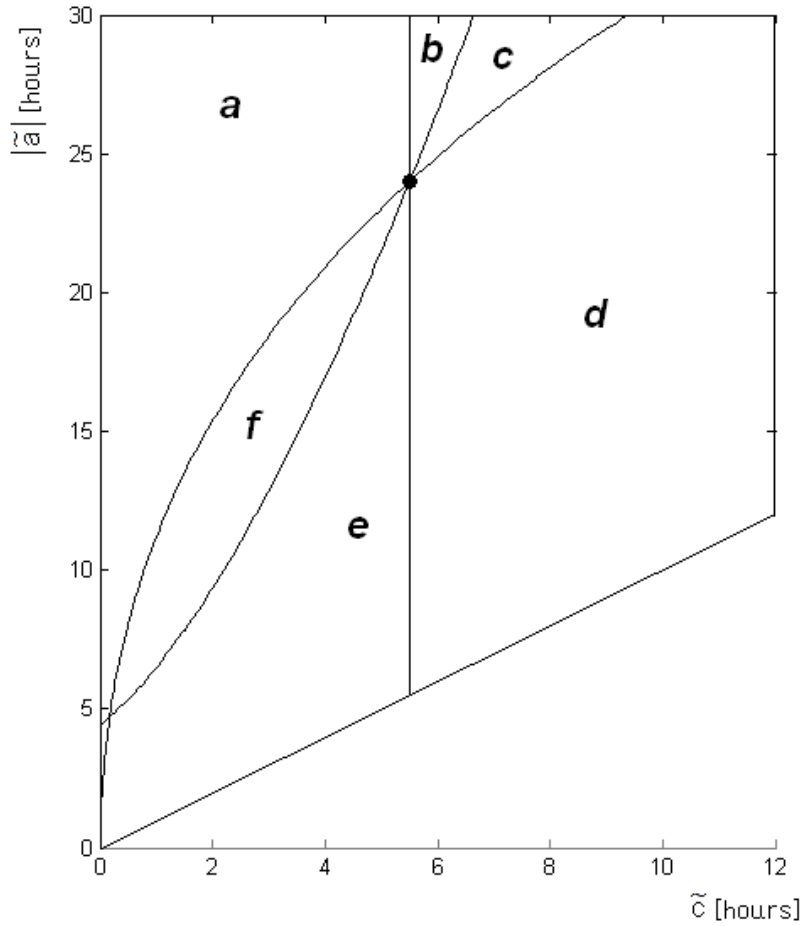


Figure 2: Graphical representation of the three policies. In region *a*, all policies assign the new patient to nurse *j*. In region *b*, the proposed maximum cost policy assigns nurse *i* and the expected cost increase policy assigns nurse *j*, while no stochastic order of costs is possible. In region *c*, no stochastic order of costs is possible and the other two policies assign the new patient to nurse *i*. In region *d*, all policies assign the new patient to nurse *i*. In region *e*, the proposed maximum cost policy assigns to nurse *j* and the expected cost increase policy assigns to nurse *i*, while no stochastic order of costs is possible. In region *f*, no stochastic order of costs is possible and the other two policies assign nurse *j*.

District	Skills	Territory	Number of nurses	Capacities [weekly hours]
NPA	Non-Palliative	A	8	35,40,45,50,50,50,50,50
PA	Palliative	A	3	20,30,30
NPB	Non-Palliative	B	4	30,35,50,50
PB	Palliative	B	1	35
NPC	Non-Palliative	C	5	30,35,40,50,50
PC	Palliative	C	1	35

Table 1: Districts of the analyzed division; only four districts have more than one nurse (i.e., NPA, PA, NPB, NPC).

c_i has to be respected. However, it can happen that one or more nurses do not respect this condition. The case $b_i > v_i$ corresponds to a highly overloaded nurse who should not be considered for other assignments. Such nurses are excluded from the set of available ones and, in case many nurses are excluded, this is an indicator for planners about an insufficient workload capacity. The case $c_i < v_i$ corresponds to a highly underloaded nurses. Such nurses are the first to be considered for the assignment independently of the policies.

6. Real case analysis

The behavior of the proposed policy is evaluated in a real HC provider, considering the data collected from one of the largest Italian public HC providers. This provider operates in the north of Italy, covering a region of about 800 km². About 1,000 patients are assisted at the same time by about 50 nurses. The provider includes three separate divisions and the analysis is carried out for the nurses of the largest division. The provider pursues the continuity of care; therefore, each newly admitted patient is assigned to only one reference nurse. Patients in charge are divided in two classes: palliative and non-palliative care patients. The skills of nurses (for palliative care and non-palliative care) and their territorial distribution are taken into account in the assignment. 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 (Table 1). The analysis is conducted in four of these districts, where more than one nurse is present.

6.1. Experimental setup

The details of the experiments are provided in the following sections.

6.1.1. Patients, planning horizon and frequency of the assignment

The activity of the provider is analyzed over a period of 26 weeks. An initial assignment of the reference nurse is performed at the initial week (named week 0) for all of the patients in the charge, while the other assignments are performed on a rolling basis for every successive week: at the beginning of each week, the new patients admitted in the service during the previous week are assigned.

The weekly arrivals of new patients, the number of patients in charge and patients' classes are taken at each week from the real historical data of the provider. A total of 1,046 patients are present in the division in the simulated weeks: 581 patients are in the charge at week 0, while 465 are assigned from week 1 through week 25. Patients' demand distributions, including an average value of transportation time for each visit, are estimated with the patient stochastic model of Lanzarone et al. (2010). At each rolling week, the current data of the patients are used as inputs to the stochastic model, and the estimates of patients' demands are obtained in terms of their probability density functions (considering the evolution of patients' demands after one week).

For fitting the triangular densities $\tilde{\Phi}_i(x_i^0)$, data collected from the analyzed provider show that r_{bi} has a mean value of 0.467 and a standard deviation of 0.070 (estimations made on the basis of a sample of 22 nurses). Hence, triangular distributions $\tilde{\Phi}_i(x_i^0)$ are fitted assuming $r_{bi} = 0.467 \forall i \in \Omega$, and based on the expected value and the variance of the corresponding distribution $\Phi_i(x_i^0)$. Parameter a_i is required to be higher or equal to 0 $\forall i \in \Omega$; in case $a_i < 0$, $a_i = 0$ is set, and b_i and c_i are obtained with the imposed r_{bi} maintaining the expected value and underestimating the variance.

6.1.2. Compared assignment methods

Experiments are conducted comparing the proposed policy with other three approaches. Indeed, four types of assignments are considered: *Maximum Cost (MC)* according to the maximum cost policy proposed in this paper, *Expected Cost (EC)* according to the expected cost increase policy proposed in Lanzarone and Matta (2012), *Expected Available Capacity (EAC)* referring to the expected available capacity approach described in section 4.4, and *Mathematical Programming (MP)*.

MP uses a Mixed Integer Linear Programming model, which is reported in Lanzarone et al. (2012), to find the optimal assignments. The comparison considers the solution obtained with *Model II* and the expected value approach presented in that paper. This configuration refers to independent districts, and to the expected value for each patient’s demand and nurse’s workload. The goal of such mathematical programming model is different from that pursued by the policies proposed in this paper, as it consists of minimizing the expected range of utilizations within each district. The main difference with the other experiments (*MC*, *EC* and *EAC*) is that *MP* assigns all new patients of a rolling week together, whereas the approaches are associated with a ranking process and assign only one patient at a time.

As for *EC*, it depends on Y_k (Lanzarone and Matta, 2012), which is assumed uniformly distributed between α_k and β_k . In our experiment, these two values are estimated based on the expected value and the variance of the corresponding distribution, obtained from the patient’s stochastic model (Lanzarone et al., 2010). Parameter α_k is required to be higher or equal to 0 $\forall k$; in case $\alpha_k < 0$, $\alpha_k = 0$ is set, and β_k is obtained maintaining the expected value and underestimating the variance.

6.1.3. Initialization

For the initialization at week 0, all of the patients in charge are considered as new patients to be assigned, and all of the nurses $i \in \Omega$ have a null workload X_i^0 before the assignments. Therefore, this situation does not fit with the realistic HC assignment problem analyzed in this paper, and it is not included in the assumptions. Hence, only at week 0, assignments are provided considering the minimization of the expected cost increase, taking all of the densities directly from the estimated distributions of the patient stochastic model without any fitting.

6.1.4. Execution of the assignments

The results of each experiment are the newly provided assignments all over the rolling weeks and the planned workloads of each nurse for the each week. These planned assignments are then executed on a number of sample paths, where a sample path is a set of the time for visits required by all of the patients in each week. These paths are generated with a mix between a Monte Carlo drawing and the real number of required visits. The Monte Carlo approach is used for the majority of patients to extract values from their demand distributions. On the contrary, the demands of long-term non-palliative patients are taken from the real historical data of the provider. These patients show a very low variability over

the time, they do not represent an uncertainty source, and the variability of the entire mix of patients is not reduced by taking their real historical demands.

The execution of the planned assignments is performed considering a number of sample paths equal to 30. Finally, the executed assignments are evaluated by computing the performance indicators described in the following section.

6.2. Performance indicators

The first performance indicator is the Total Overtime (*TOVT*) incurred by the nurses of a district; this is obtained as the sum of the overtimes of each nurse who belongs to the district, from week 1 through week 25, divided by the number of these nurses.

Another measure that is relevant in the practice is the mean utilization \bar{u}_i of nurse i . This measure is obtained as the ratio between the workload provided by nurse i and his/her capacity v_i for each week, averaged from week 1 through week 25. Even if the goal of the proposed policies is the minimization of the overtimes, the utilizations give information about the workload balancing performance of the assignments. In fact, the range of \bar{u}_i among the nurses of a district (named Z) is an indicator of the workload balancing performance in the district: the more a strict Z is obtained, the more a higher workload balancing is performed.

Finally, a third indicator is the total cost (*TC*) of a district, obtained as the sum of the values assumed by the cost function for each nurse who belongs to the district, from week 1 through week 25, divided by the number of these nurses. The function is computed with $m = 2$ for comparing with the *EC* policy.

All of the performance indicators are calculated from the executions of the assignments with the data of each specific path.

6.3. Results

Mean utilizations \bar{u}_i obtained from the executions of the experiments are reported in Table 2, where the minimum, the mean and the maximum values of \bar{u}_i among the 30 sample paths are reported for each nurse i . In district NPA, no nurse is over allocated (i.e., $\bar{u}_i > 1$) with the *MC* policy, while one nurse is over allocated with the *EC* approach, three nurses with the *EAC* approach, and two nurses with the *MP* approach (considering the maximum values). In districts NPB and NCP, *MC* and *EC* show a limited maximum \bar{u}_i with respect to *EAC*, where a lower number of nurses have \bar{u}_i higher than 1 but with higher values. Finally, no noticeable differences are obtained in district PA.

The average values of $TOVT$, Z and TC among the paths (together with the confidence intervals) are reported in Table 3. Very similar average values are observed between MC and EC in all districts. Moreover, the confidence intervals are highly overlapped, resulting in the absence of significant differences. This means that the majority of the assignments are not in the regions of the plane $\{\tilde{c}, |\tilde{a}|\}$ where a different choice is provided between MC and EC (Figure 2).

Results also show that smaller overtimes and better workload balancing are obtained with MC and EC with respect to EAC and MP for non-palliative patients, whereas a similar behavior is shown for palliative patients. This is highly satisfactory, considering that the policies suffer from the assumptions introduced, i.e., the triangular distribution $\tilde{\Phi}_i(x_i^0)$ and the repeated single-patient assignment coupled with an ordering process, while MP assigns all the new patients of a rolling week together. Moreover, all policies do not require any optimization software, which may be expensive for a real HC provider and require adequate hardware to avoid long computational times.

The lower $TOVT$ with MC and EC in the presence of non-palliative patients show the added value of considering the tail of the workload distributions above the capacity in the presence of patients with a relevant variability. On the contrary, the similar results in the case of palliative patients are given by the low variability of their demands with respect to those of non-palliative patients; in this case, the inclusion of the variability (i.e., the probability density function) does not add value to the solution.

Nurse	<i>MC</i>			<i>EC</i>			<i>EAC</i>			<i>MP</i>			
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	
NPA	1	0.7120	0.8333	0.9735	0.7790	0.8800	1.0073	0.7940	0.9227	1.0290	0.7765	0.9574	1.0873
	2	0.6886	0.7772	0.9010	0.6704	0.7882	0.9002	0.5696	0.6740	0.7580	0.7070	0.7983	0.9164
	3	0.6860	0.8314	0.9671	0.7044	0.8217	0.9447	0.7464	0.8698	1.0011	0.6573	0.7645	0.9244
	4	0.5408	0.6422	0.7362	0.5476	0.6334	0.7580	0.4904	0.5784	0.6974	0.4662	0.6244	0.7284
	5	0.7043	0.8133	0.9509	0.6583	0.7554	0.9183	0.9291	1.0299	1.1949	0.7414	0.8528	0.9894
	6	0.7270	0.8254	0.9364	0.7368	0.8423	0.9338	0.7450	0.8094	0.9166	0.8184	0.9027	1.0148
	7	0.8266	0.9014	0.9922	0.8156	0.8930	0.9828	0.7940	0.8624	0.9580	0.6436	0.7592	0.8850
	8	0.6944	0.8096	0.8814	0.7174	0.8107	0.9320	0.6488	0.7737	0.8604	0.6652	0.8042	0.9152
NPB	1	0.9667	1.1121	1.3543	0.9823	1.1343	1.3493	1.0777	1.3781	1.5817	0.5340	0.6406	0.7973
	2	0.6860	0.8093	0.9443	0.6617	0.8102	0.9129	0.7220	0.8712	0.9903	0.9240	1.0084	1.0903
	3	0.8316	0.9556	1.0698	0.8030	0.9318	1.0592	0.7784	0.8818	1.0142	0.9512	1.1612	1.2750
	4	0.8568	0.9453	1.0696	0.8694	0.9552	1.0420	0.7470	0.8162	0.8862	0.7784	0.8833	1.0836
NPC	1	0.7312	0.8952	1.0178	0.7860	0.9081	1.0308	0.6610	0.7972	0.9094	0.8552	0.9951	1.0890
	2	0.8214	0.9115	1.0702	0.7666	0.8772	1.0278	0.6990	0.7760	0.9146	0.7538	0.8744	0.9864
	3	0.8997	0.9595	1.0314	0.8831	0.9557	1.0186	1.0040	1.0903	1.1926	0.8277	0.9349	1.0943
	4	0.8363	0.9501	1.1043	0.9230	1.0229	1.1765	0.8285	1.0064	1.1863	0.7178	0.8104	0.8938
	5	0.7133	0.9142	1.0933	0.6657	0.8573	1.0493	0.8453	1.0758	1.2920	0.9137	1.0246	1.1757
PA	1	0.3920	0.6235	0.8015	0.4880	0.7241	0.9250	0.5710	0.7636	0.9790	0.3020	0.4972	0.6660
	2	0.5597	0.7684	0.9823	0.4577	0.7594	0.9933	0.4467	0.7031	0.8843	0.5203	0.7246	0.9297
	3	0.3420	0.4973	0.7047	0.3100	0.4392	0.5903	0.3053	0.4692	0.6743	0.4137	0.6254	0.9147

Table 2: Minimum, mean and maximum value of the utilizations \bar{u}_i among the 30 sample paths for each nurse i .

District	Variable	<i>MC</i>	<i>EC</i>	<i>EAC</i>	<i>MP</i>
NPA	<i>TOVT</i>	22.82 ± 2.16	22.42 ± 2.03	31.95 ± 2.86	29.51 ± 2.75
	<i>Z</i>	0.2790 ± 0.0217	0.2901 ± 0.0234	0.4552 ± 0.0235	0.3569 ± 0.0285
	<i>TC</i>	211.45 ± 28.28	198.19 ± 24.58	353.47 ± 40.85	309.02 ± 42.65
NPB	<i>TOVT</i>	65.81 ± 5.84	66.65 ± 5.80	93.56 ± 6.86	85.84 ± 5.81
	<i>Z</i>	0.3060 ± 0.0444	0.3245 ± 0.0361	0.5761 ± 0.0412	0.5206 ± 0.0347
	<i>TC</i>	762.58 ± 100.86	772.85 ± 104.88	1577.99 ± 180.82	1247.45 ± 121.95
NPC	<i>TOVT</i>	50.43 ± 4.59	53.64 ± 5.43	68.18 ± 5.45	58.25 ± 4.64
	<i>Z</i>	0.1704 ± 0.0221	0.2151 ± 0.0270	0.3777 ± 0.0271	0.2444 ± 0.0277
	<i>TC</i>	510.34 ± 67.49	564.13 ± 86.45	776.79 ± 88.23	674.16 ± 77.70
PA	<i>TOVT</i>	13.26 ± 3.19	14.71 ± 3.37	16.62 ± 3.48	12.19 ± 2.63
	<i>Z</i>	0.2904 ± 0.0410	0.3629 ± 0.0397	0.3285 ± 0.0438	0.2753 ± 0.0426
	<i>TC</i>	102.25 ± 33.73	115.42 ± 36.25	138.37 ± 41.23	104.24 ± 34.05

Table 3: *TOVT* (in hours), *Z* and *TC* for the four experiments, executed for the 30 sample paths (mean value ± half-width 95% confidence interval). The normality of the data was assessed with the Anderson-Darling test.

7. Conclusions

The topic of assigning the reference nurse in HC under continuity of care is fundamental to maintaining a high quality of the provided services. At the same time, this topic is also challenging, because in the literature the high variability of patients' demands is still neglected.

In this paper, we propose a solution of the assignment problem under continuity of care that minimizes the overtimes incurred by nurses. This approach is innovative with respect to the usual practice of HC providers, where new patients are assigned to the nurse with the highest expected available capacity, defined as the difference between his/her capacity and the actual expected workload. Specifically, the proposed policy considers the probability that the workload of a nurse is above the capacity, while the current practice of HC providers only takes into account the expected workload. Moreover, with respect to the previous approach proposed in Lanzarone and Matta (2012), the policy proposed in this paper is more general and require less assumptions.

Different advantages derive from the adoption of the policy. First, an analytical based policy can solve the problem with a limited computational effort and without requiring ex-

pensive software applications. Moreover, this limited computational effort can easily include the high variability of the demand. Finally, the structure of the optimal policy could help to search the optimum value in the heuristic-based algorithms that have to be adopted for large-scale problems.

Results from the implementation in a real case study show that the choices proposed by proposed policy are often discordant with those that consider the expected available capacity. This results into significantly lower overtimes, associated with a higher workload balancing, in case of the application of the policy for patients with high demand variability. Considering the typical classification of HC patients (palliative care and non-palliative care patients), this result refers to non-palliative patients. In the case of palliative patients, lower demand uncertainty is present and, therefore, to consider uncertainty in the assignment of the reference nurse does not add significant benefit.

The policy proposed in this paper has some limitations that can be subject of further investigation in future works. Firstly, the performance of the policy could decrease for larger numbers of patients to assign, because the approach presented in this paper requires an ordering of the new patients and to repeat the single-patient assignment for each one of them. Moreover, some assumptions can be removed. We will take into account a more general distribution for $\Phi_i(x_i^0)$ instead of the triangular one, and we will include several time slots in the planning horizon so as to manage durations longer than the demand evolution of the patients in the charge. We will also consider multi-district assignment; this will make possible to manage all the districts together and to allow out-of-district assignments. In the current configuration of the approach, with separated districts, the nurse-to-patient compatibility is a hard constraint and out-of-district assignments are not permitted. In an integrated management, they would be treated as soft constraints which can be violated with a certain penalty cost.

Finally, we emphasize that a software application containing the proposed policy and the other compared approaches is currently used by the analyzed provider to assign workloads to the operators taking into account the variability of patient demands.

Appendix: Probability density function of $C_i(\mathbf{X}_i^0)$

The probability density function $\gamma_i(C_i(X_i^0))$ is derived in this Appendix, according to (2), and based on v_i and the triangular workload density function $\tilde{\Phi}_i(x_i^0)$. The probability that the variable cost $C_i(X_i^0)$ of nurse i before the considered assignment assumes a certain

value is the following:

$$\begin{cases} P[C_i(X_i^0) < 0] = 0 \\ P[C_i(X_i^0) = 0] = P[0 \leq X_i^0 \leq v_i] = \int_0^{v_i} \tilde{\Phi}_i(x_i^0) dx \\ P[0 < C_i(X_i^0) \leq \bar{C}_i] = P[v_i < X_i^0 \leq v_i + \sqrt[m]{\bar{C}_i}] = \int_{v_i}^{v_i + \sqrt[m]{\bar{C}_i}} \tilde{\Phi}_i(x_i^0) dx \end{cases} \quad (.1)$$

where $v_i + \sqrt[m]{\bar{C}_i}$ is the value of X_i^0 that corresponds to the cost \bar{C}_i . Considering a workload X_i^0 that varies from a_i to c_i , the maximum cost associated with $\tilde{\Phi}_i(x_i^0)$ is equal to $(c_i - v_i)^m$. Hence, the cumulative density function $\Gamma_i(C_i)$ of $C_i(X_i^0)$ is as follows:

$$\Gamma_i(C_i) = \begin{cases} 0 & C_i < 0 \\ \int_0^{v_i} \tilde{\Phi}_i(x_i^0) dx & C_i = 0 \\ \int_0^{v_i} \tilde{\Phi}_i(x_i^0) dx + \int_{v_i}^{v_i + \sqrt[m]{C_i}} \tilde{\Phi}_i(x_i^0) dx & 0 < C_i < (c_i - v_i)^m \\ 1 & C_i \geq (c_i - v_i)^m \end{cases} \quad (.2)$$

The density function $\gamma_i(C_i)$ is the derivative of $\Gamma_i(C_i)$ with respect to C_i , as follows:

$$\gamma_i(C_i) = \begin{cases} \delta(0) \int_0^{v_i} \tilde{\Phi}_i(x_i^0) dx & C_i = 0 \\ \frac{\partial}{\partial C_i} \int_{v_i}^{v_i + \sqrt[m]{C_i}} \tilde{\Phi}_i(x_i^0) dx & 0 < C_i < (c_i - v_i)^m \\ 0 & elsewhere \end{cases} \quad (.3)$$

where $\delta(0)$ is the Dirac delta function on $C_i = 0$. Considering the triangular function for $\tilde{\Phi}_i(x_i^0)$, $\gamma_i(C_i)$ assumes the following form:

$$\gamma_i(C_i) = \begin{cases} \delta(0) \left[1 - \frac{(c_i - v_i)^2}{(c_i - a_i)(c_i - b_i)}\right] & C_i = 0 \\ \frac{2C_i^{\frac{1-m}{m}} (c_i - v_i - C_i^{\frac{1}{m}})}{m(c_i - a_i)(c_i - b_i)} & 0 < C_i < (c_i - v_i)^m \\ 0 & elsewhere \end{cases} \quad (.4)$$

References

- Abazari, A.M., Solimanpur, M., Sattari, H., 2012. Optimum loading of machines in a flexible manufacturing system using a mixed-integer linear mathematical programming model and genetic algorithm. *Computers and Industrial Engineering* 62, 469 – 478.
- Akjiratikarl, C., Yenradee, P., Drake, P., 2007. PSO-based algorithm for home care worker scheduling in the UK. *Computers and Industrial Engineering* 53, 559–583.

- American Geriatrics Society AGS, 2006. Geriatric healthcare statistics: demography of aging, in: Reported resources, Statistics.
- Bard, J., 2010. Nurse scheduling models. John Wiley & Sons, Inc.
- Ben Bachouch, R., Fakhfakh, M., Guinet, A., Hajri-Gabouj, S., 2008. Planification de la tournée des infirmiers dans une structure de soins à domicile, in: Conference Francophone en Gestion et Ingénierie des Systèmes Hospitaliers (GISEH).
- Bennett, A., Erera, A., 2011. Dynamic periodic fixed appointment scheduling for home health. *IIE Transactions on Healthcare Systems Engineering* 1, 6–19.
- Bertels, S., Fahle, T., 2006. A hybrid setup for a hybrid scenario: combining heuristics for the home health care problem. *Computers and Operations Research* 33, 2866–2890.
- Borsani, V., Matta, A., Beschi, G., Sommaruga, F., 2006. A home care scheduling model for human resources, in: Proceedings of ICSSSM06 (International Conference on Service Systems and Service Management), Troyes, France, pp. 449–454.
- Brunner, J., Bard, J., Kolisch, R., 2009. Flexible shift scheduling of physicians. *Health Care Management Science* 12, 285–305.
- Brunner, J., Bard, J., Kolisch, R., 2011. Midterm scheduling of physicians with flexible shifts using branch and price. *IIE Transactions* 43, 84–109.
- Burke, E., Causmaecker, P., Berghe, G., Landeghem, H., 2004. The state of the art of nurse rostering. *Journal of Scheduling* 7, 441–499.
- Cai, X., Wang, L., Zhou, X., 2007. Single-machine scheduling to stochastically minimize maximum lateness. *Journal of Scheduling* 10, 293–301.
- Cai, X., Zhou, X., 2005. Single-machine scheduling with exponential processing times and general stochastic cost functions. *Journal of Global Optimization* 31, 317–332.
- 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, 179–191.
- Cheang, B., Li, H., Lim, A., Rodrigues, B., 2003. Nurse fostering problems—a bibliographic survey. *European Journal of Operational Research* 151, 447–460.
- Chevreur, K., Com-Ruelle, L., Midy, F., Paris, V., 2005. The development of hospital care at home: an investigation of Australian, British and Canadian experiences. Technical Report 1610. Institut de Recherche et Documentation en Économie de la Santé (IRDES).
- Comondore, V., Devereaux, P., Zhou, Q., Stone, S., Busse, J., Ravindran, N., Burns, K., Haines, T., Stringer, B., Cook, D., Walter, S., Sullivan, T., Berwanger, O., Bhandari, M., Banglawala, S., Lavis, J., Petrisor, B., Schunemann, H., Walsh, K., Bhatnagar, N., Guyatt, G., 2009. Quality of care in for-profit and not-for-profit nursing homes: systematic review and meta-analysis. *British Medical Journal* 339, b2732.
- Cordes, C., Dougherty, T., 1993. A review and an integration of research on job burnout. *Academy of Management Review* 18, 621–656.
- Davies, C., Dale, J., 2003. Paediatric home care for acute illness: I. GPs' and hospital-at-home staff views.

- International Journal of Health Care Quality Assurance 16, 361–366.
- Eveborn, P., Flisberg, P., Ronnqvist, M., 2006. LAPS CARE-an operational system for staff planning of home care. *European Journal of Operational Research* 171, 962–976.
- Eveborn, P., Ronnqvist, M., Einarsdottir, H., Eklund, M., Liden, K., Almroth, M., 2009. Operations research improves quality and efficiency in home care. *Interfaces* 39, 18–34.
- Garg, L., McClean, S., Meenan, B., Millard, P., 2010. A non-homogeneous discrete time markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system. *Health Care Management Science* 13, 155–169.
- Goel, A., Indyk, P., 1999. Stochastic load balancing and related problems, in: *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pp. 579–586.
- Grieco, A., Semeraro, Q., Tolio, T., 2001. A review of different approaches to the fins loading problem. *International Journal of Flexible Manufacturing Systems* 13, 361–384.
- 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.
- Hertz, A., Lahrichi, N., 2009. A patient assignment algorithm for home care services. *Journal of the Operational Research Society* 60, 481–495.
- Kim, H.W., Yu, J.M., Kim, J.S., Doh, H.H., Lee, D.H., Nam, S.H., 2012. Loading algorithms for flexible manufacturing systems with partially grouped unrelated machines and additional tooling constraints. *The International Journal of Advanced Manufacturing Technology* 58, 683–691.
- Kovner, C., Brewer, C., Wu, Y.W., Cheng, Y., Suzuki, M., 2006. Factors associated with work satisfaction of registered nurses. *Journal of Nursing Scholarship* 38, 7179.
- Lanzarone, E., Matta, A., 2012. A cost assignment policy for home care patients. *Flexible Services and Manufacturing Journal* 24, 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 System Man and Cybernetics - Part A* 42, 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, 3–25.
- Matta, A., Chahed, S., Sahin, E., Dallery, Y., 2012. Modelling home care organisations from an operations management perspective. *Flexible Services and Manufacturing Journal*, epub ahead of print.
- Medicare Payment Advisory Commission MEDPAC, 2011. Home health services (Chapter 8) in Report to the Congress: medicare payment policy. Technical Report.
- Pentico, D.W., 2007. Assignment problems: A golden anniversary survey. *European Journal of Operational Research* 176, 774–793.
- Pinedo, M., 2012. *Scheduling: theory, algorithms, and systems - 4th edition*. Springer-Verlag.
- Shaked, M., Shanthikumar, J., 1994. *Stochastic orders and their applications*. Academic Press, UK.
- Shaked, M., Shanthikumar, J., 2007. *Stochastic orders*. Springer series in statistics, NY, USA.
- Thomsen, K., 2006. Optimization on home care. Technical Report. Technical University of Denmark. Thesis in Informatics and Mathematical Modeling.

Yalcindag, S., Matta, A., Sahin, E., 2011. Human resource scheduling and routing problem in home health care context: a literature review, in: Proceedings of ORAHS 2011 (37th Conference on Operational Research Applied to Health Services), Cardiff, UK, pp. 8-22.