Home Hospitalization: A Bi-objective problem considering Travel Time and Continuity of Care, and accounting for Workload Balance

Margarida Arrais Martins

Instituto Superior Técnico
margarida.arrais.martins@tecnico.ulisboa.pt

The increase in life expectancy has resulted in an aging population with more need for hospital care. These factors have encouraged the emergence of new alternatives to conventional hospitalization, which is the case of Home Hospitalization. A bi-objective optimization problem is developed, aiming to minimize Travel Times and maximize Continuity of Care while ensuring Workload Balance. The purpose is to account for the three stakeholders' perspectives: managers, patients, and care workers. In the proposed model, nurses and physicians are assigned to teams that travel by car to locations representing the patients' homes, respecting some constraints. The epsilon-constraint method is used to find the approximate Pareto front of the two conflicting objectives. The trade-off between the objectives is analyzed through the computational results. Two different approaches are discussed to help decide on the Pareto front's best solution: Crowding distance and TOPSIS. Based on the real-world instances provided by the hospital under study, the proposed method can provide, on average, a reduction in Travel Time by 7.09%, an increase in Continuity of Care of 35.18% and an increase of 65.73% on Workload Balance. Thus, the three stakeholders' perspectives can be improved significantly with the use of the model. Numerical experiments are also conducted in literature instances.

Index Terms—**Home Health Care; Home Hospitalization; Staff Scheduling; Staff Routing; Pareto front; Multi-Objective Optimization.**

I. Introduction

The demand for hospital care has grown considerably due to the increase in life expectancy and the aging population. The average life expectancy at birth across Organization for Economic Cooperation and Development (OECD) countries was 80.7 years in 2017, ten years higher than in 1970 [21]. Additionally, the share of the population aged 65 years and older has nearly doubled from 9% in 1960 to more than 17% in 2017. As a consequence, more people around the world need ongoing health and social care. Indeed, the aging society led to an increase in the prevalence of chronic diseases in patients, resulting in overcrowding of emergency services.

In order to relieve the pressure in the acute hospital beds, together with an attempt to improve patient's satisfaction and expectations, some countries are adopting an alternative approach to conventional hospitalization. This alternative is called Home Hospitalization (HH). HH is defined as the clinical activity, which allows health care professionals to perform diagnostic and therapeutic procedures, usually associated with acute inpatient care, at the patient's home [5]. The evidence shows that this alternative enables to improve the life quality of patients and avoid the risk of hospital infections.

Hospital Garcia de Orta (HGO) implemented in November 2015 the first hospital unit in Portugal specifically dedicated to HH. In addition to patient and family satisfaction, this alternative has shown to be economically attractive as well. The cost savings for using HH were 45.83% in 2017 and 34.72% in 2018 [17]. The HGO's Home Hospitalization Unit (HHU) aims to increase the number of patients served each day, making more beds available for the hospital. However, to increase the number of patients served, the efficiency of operations and decisions must be improved. For instance, routing and scheduling decisions are made manually, taking approximately 30-45 minutes every morning, resulting in high organizational efforts (i.e., waste of time and resources). Moreover, all these decisions are based on the staff's experience with no systematic or decision support.

Planning decisions in health care organizations require a consideration of three key stakeholders: patients, care workers and managers. Nevertheless, these three key stakeholders have interests that most of the time come into conflict: quality of the service, workload balance and low costs, respectively. Considering that all the planning is currently done manually by HGO, it is difficult to account for all aspects simultaneously. As a result, the main focus of the HGO's HHU has been in containing costs.

The literature on HH is scarce, especially in logistical terms. However, another alternative with the same broad objectives of combating demographic shifts and with the same type of logistical operations is called Home Health Care (HHC). It is also about providing health care services in the patient's home. Nonetheless, HH concerns patients with acute or chronic pathologies who would require inpatient care in the absence of this service. By contrast, the literature on HHC regarding planning decisions is quite extensive, and it is possible to show that OECD countries are betting on this alternative.

The main goal of this dissertation is to improve the planning decisions of the HGO's HHU, considering the main challenges faced by this unit. For this purpose, the objective is to develop a decision support tool for scheduling and routing decisions, which can be generalized for other HHUs. The decision support tool should focus on the efficiency and on the quality of the service, showing the impact of the different features in the planning decisions.

II. LITERATURE REVIEW

The literature in HHC regarding scheduling and routing decisions is quite vast. Most of the works refer to this problem as the Home Health Care Routing and Scheduling Problem (HHCRSP), however, there is no single denomination for this type of problems, having received several different names in the literature. For detailed literature reviews on the field, highlighting the most relevant features, constraints, and objectives used by the different papers and other relevant sources, we refer the reader to three comprehensive surveys: Cissé et al. [6], Di Mascolo et al. [11] and Fikar, Hirsch [12]. However, most of the papers deal with home nursing services and not HHUs. Even though the general planning is similar, and some features are the same, there are aspects related to HHUs that are not considered in HHC problems. For instance, instead of having nurses visiting patients, visits are done by teams composed of physicians and nurses. In addition, patients have different care requirements than HHC services since they require the same type of care as they would have in a hospital. Thus, some patients may need to be seen by a physician depending on the patient's status, or patients may need to be seen in the early morning - first-visit; to do a blood test, for example.

As far as the author is aware, only one paper in the literature handles the case of an HHU which is Quintanilla et al. [23]. This paper was inspired by a hospital in Spain and considers teams of physicians and nurses visiting patients. However, the teams' characteristics (i.e., number of teams and respective composition) were already pre-established by the hospital, not requiring constraints related to this assignment. By contrast, in the HGO case, physicians are shared with other hospital units, and it is not possible to predict if there are physicians available on a specific day, meaning that teams' formation needs to happen every day. Additionally, Quintanilla et al. [23] does not consider Continuity of Care with teams and patients, some relevant features to explore in this case study.

Another lack in the literature regards the use of multi-objective functions to incorporate the interests of the various stakeholders. Nevertheless, some papers have already considered the three features: Continuity of Care, Workload Balance, and Travel Time; simultaneously in their models. Martinez et al. [18], addresses the problem in two different phases. In the first phase, for each patient, a set of know care workers is built to account for the Continuity of Care. In a second phase, Travel and waiting times are minimized, considering that the patients can only be assigned to the known care workers. Workload Balance is only considered when more than one care worker qualifies to be assigned to the patient. Thus, the chosen care worker is the one that has the shortest working time. Yalçındağ et al. [26], also refer to these three features. Nevertheless, Continuity of Care is ensured through a hard constraint, meaning that a unique care worker is assigned to each patient over the planning horizon, not allowing for flexibility. Once more, a two-stage approach is proposed: on the first stage, Workload Balance is optimized, and on the second stage, the Travel Time. Gomes, Ramos [13] also treats the problem as two independent sub-problems. One sub-problem that assigns teams to patients in a multi-objective way, considering Travel Time and Workload Balance. Here, the loyalty between caregivers and patients is ensured through soft-constraints, aiming to assign the same team to the same patient. The second sub-problem assigns caregivers to teams. Thus, as in this master thesis, Gomes, Ramos [13] addresses the aspect of teams' formation. An aspect which is not commonly addressed in the literature.

The only paper which considers the three features simultaneously as in our approach, without taking different stages, is Milburn, Spicer [19]. Milburn, Spicer [19] considers a multi-objective function with three different objectives, one for each feature. Continuity of care is measured through nurse consistency, i.e., the total number of different nurses seen by patients, and Workload Balance by the number of patients assigned to each care worker. However, in this paper, remote monitoring devices are used, which can substitute a nurse visit, enabling the improvement of the objectives considered.

There have been emerging some papers considering Continuity of Care. Most of them use hard constraints, forcing a patient to be visited by a specific care worker or limiting the number of different nurses assigned to each patient. Nickel et al. [20] uses a score that accounts for the number of different

	Features			Teams		WB		CC		TT	Objective Function		
Papers	**FV**	**Breaks**	**Skills**	**Assig.**	**Form.**	**Const.**	**Objec.**	**Const.**	**Objec.**	**Objec.**	**SO**	**WS**	**MO**
Gomes, Ramos (2019)		X		X	X		X	X		X			X
Grenouilleau et al. (2019)			X						X	X		X	
Liu et al. (2018)			X	X				X		X			X
Martinez et al. (2018)			X			X		X		X	X		
Milburn, Spicer (2013)							X		X	X			X
Nickel et al. (2012)			X						X	X		X	
Quintanilla et al. (2020)	X		X	X		X				X	X		
Yalçındağ et al. (2016a)			X				X	X		X	X		
This dissertation	**X**	**X**	**X**	**X**	**X**	**X**			**X**	**X**			**X**

FV – First-Visit; Assig. – Assignment, Form. – Formation; Const. – Constraint; Objec. – Objective; SO – Single-Objective; WS – Weighted Sum; MO – Multi-Objective

TABLE I: Summary and comparison of the most similar papers to the present dissertation.

nurses that have visited a patient, while Grenouilleau et al. [15] have a similar approach to our model, using a function based on the number of times a care worker has been assigned to a patient's past visits. However, the function has only three different values, for 0; 1 and 2; and more than 2 visits. Thus, for more than two visits, the value of the function is always the same. In our proposed model, it is the first time a total score is considered, accounting for all past visits of a care worker to a patient, aiming to assign the care worker with the highest number of visits with the respective patient.

Table I shows a summary and comparison, in terms of features and objective functions, of the most similar papers to the current dissertation. In conclusion, the model described in this master thesis makes three primary contributions to the literature in the HHCRSP:

1) The consideration of an HHU and its specific characteristics, namely teams' formation, first-visit requirements, and the patients' need to be seen by a physician.
2) The use of a multi-objective function accounting for Continuity of Care and Travel Time, enabling a trade-off study between these two features, and considering at the same time the feature Workload Balance.
3) The measure used for the Continuity of Care, accounting for all the times a care worker has visited a patient. Furthermore, aiming to maximize this value for both nurses and physicians.

III. Mathematical Model and Solution Approach

This section describes the proposed mathematical model and solution approach. First, the Home Hospitalization problem is formally described. Then, the mathematical model and the proposed solution approach are presented.

A. Problem Description

The HHCRSP consists of finding a schedule and route for each care worker to provide the planned care visits over a planning horizon. As most of the articles in this field, the problem is modeled as an extension of the Vehicle Routing Problem (VRP) augmented by new constraints specific to this problem. In the specific case of HH, patients need to be visited every day (since this is an alternative for traditional hospitalization; daily care must be ensured), and the available care workers - nurses and physicians - change daily, meaning that it is not possible to assign only one care worker to one patient for all the hospitalization period. For this reason, the three decisions, namely patient assignment, scheduling, and routing, are considered simultaneously.

All visits are performed during the morning period, by teams that travel by car to a set of patients' locations. All routes start and end at the hospital. Teams can be composed of a single nurse or by one nurse and one physician, depending on the number of physicians available on that day. The number of teams required to perform the home visits depends on the current capacity of the HHU, i.e., the number of vehicles available. All teams start the visits at the same time and cannot finish after the end of the shift. Some patients require to be the first ones to be visited in the morning due to specific conditions that cannot be predicted and are only known on the planning day. The remaining patients can be visited anytime within the time window for visits.

Additionally, when physicians are not available in all the teams, patients with a more severe case are allocated to teams with physicians. Physicians are all general practitioners, and there is no need to specify qualifications. Nonetheless, nurses may have different specialties that can be matched with the different patients' needs. In addition, nurses are allowed to take a break during the visits period but only after working a certain amount of time.

As was mentioned in Carello et al. [4], there are mainly three stakeholder perspectives in health care: patients (the quality of service), care workers (fair workloads), and managers (low costs). In our case, the quality of the service is measured through the concept of Continuity of Care, trying to assign the same team to the same patient to create a better relationship with the patient. Fair workloads are also an important measure pointed out by the HHU, which accounts for the balance in the total working time of the teams, i.e., since the teams leave the hospital until they come back again. Lastly, from a management perspective, the low costs, in this case, are only related to Travel Time, and by consequence, travel costs.

As a result, two objectives are considered: minimizing Travel Times (1) and maximizing Continuity of Care (2). Even though Workload Balance is not addressed as an objective in the objective function, it is ensured by hard constraints. Considering the three objectives in the objective function would bring a higher complexity to the model, not only from a computational point of view but also for interpreting the results from the HHU side.

B. Problem Formulation

The problem is modeled on a directed graph $\mathcal{G} = (\mathcal{N}', A)$, where $\mathcal{N}' = \mathcal{N} \cup 0$ is the node set, $\mathcal{N} = 1, ..., n$ denotes the set of patients to visit in different geographical locations and node 0 represents the hospital location i.e. the depot. A is the arc set between nodes and for an arc $(i, j) \in A$, t_{ij} is the Travel Time from node i to node j. Loops are not allowed. Let P and E denote the set of physicians and nurses available for HH visits on that day, respectively. Each care worker will have to be assigned to precisely one team $m \in M$ (corresponding to an individual vehicle), and each team can only have at most one nurse $e \in E$ and one physician $p \in P$. Additionally, each nurse $e \in E$ has specific skills $s \in S$.

There are six main types of decision variables. The binary routing variables $X_{ijm} = 1$ if team m goes from patient i to patient j (i.e., if edge $(i, j) \in A$ is included in the tour of team m), and 0 otherwise. The scheduling variables S_{im} define the starting time of the visit of patient i if team m is assigned to this patient; otherwise, S_{im} is zero. The Y_{ijm} binary variable is 1 if team m takes a break between patient i and j (otherwise is 0), while B_m defines the time when each team m takes the break. Finally, the binary assignment variables $Z_{em} \in (0, 1)$ and $Z_{pm} \in (0, 1)$ are equal to 1 when nurse e and physician p are assigned to team m and 0 otherwise, respectively. The different parameters and auxiliary variables used are going to be explained in more detail in the following subsections.

1) Routing

The following constraints (1) – (3) form the routing constraints based on the VRP [2]. Constraints (1) ensure that each patient is visited exactly once and by exactly one team. Constraints (2) and (3) define the routing network, ensuring that each team leaves and arrives at the hospital exactly once (constraint (2)), and that each patient is visited and left - flow conservation constraint (3).

$$\sum_{m \in M} \sum_{i \in N'} X_{ijm} = 1, \forall j \in N \tag{1}$$

$$\sum_{j \in N} X_{0jm} = \sum_{j \in N} X_{j0m} = 1, \forall m \in M \tag{2}$$

$$\sum_{j \in N'} X_{ijm} = \sum_{j \in N'} X_{jim}, \forall i \in N', m \in M \tag{3}$$

2) Scheduling

This subsection describes the scheduling constraints, particularly the traffic conditions and the constraints related to the break requirements. All the care workers in the HHU start to work at the same time - defined by the auxiliary variable sT, and need to finish before the end of the shift - defined by the auxiliary variable fT. Constraints (4) ensure that all visits happen within the period available for visits. Finally, constraints (5) ensure that all teams start the visits simultaneously. There are some regions characterized by heavy morning traffic, and thus, it is not suggested to go to these areas before a particular hour of the morning. This hour is given by the auxiliary variable $critT$. For this reason, if a patient i lives in a traffic area, $traf_i$ is equal to one, and the visits at this patient never start before the $critT$ (constraint (6)).

Regarding breaks, each team is allowed to have one break (constraint (7)), which is assumed to be taken near the patients' house, meaning that there is no need to drive to a break location. The duration of the break is defined by the auxiliary variable $breakT$ and is defined by the hospital. Constraints (8) impose that if a team m takes a break between patient i and j, this team must visit these patients (i.e., a break can only happen between two consecutive visits). The break time B_m of each team is defined by constraints (9) and (10). Moreover, the break should take place between a time window imposed by the working time: the teams should work a certain time before taking a break - $befBreak$; and the break should not take place as well at the end of the shift. Thus, the teams should work as well a certain time after the break -$aftBreak$ (constraints (11)).

For each patient, dur_i represents the time needed to perform the care, i.e., the time since the team parks the car until getting in the car again. As said before, the model does not have time windows, and as a consequence, there is no waiting time: a team goes straight from one patient to another and does not have to wait to start the service. Constraints (12) and (13) ensure the consistency of the patients visits with the break. Constraint (12) make sure that a team has enough time between two consecutive patients to perform the visit and travel to the next patient, and constraints (13) guarantee that the team goes from one patient immediately to another patient (or after the break, if one is taken at this point), avoiding any waiting times. The set of break constraints was adapted from Xiao et al. [25] and Trautsamwieser et al. [24].

$$S_{im} \geq \sum_{j \in N'} X_{ijm} \cdot sT + t_{0i} \cdot X_{0im}, \forall i \in N, m \in M \tag{4a}$$

$$S_{im} \leq \sum_{j \in N'} X_{ijm} \cdot fT - t_{i0} \cdot X_{i0m}, \forall i \in N, m \in M \tag{4b}$$

$$S_{im} \leq \sum_{j \in N'} X_{ijm} \cdot sT + t_{0i} \cdot X_{0im} + (1 - X_{0im}) \cdot K, \quad \forall i \in N, m \in M \tag{5}$$

$$(S_{im} - critT) \cdot traf_i + (1 - \sum_{j \in N} X_{ijm}) \cdot K \geq 0, \quad \forall i \in N, m \in M \tag{6}$$

$$\sum_{i \in N} \sum_{j \in N} Y_{ijm} = 1, \forall m \in M \tag{7}$$

$$Y_{ijm} \leq X_{ijm}, \forall i, j \in N, m \in M \tag{8}$$

$$S_{im} + dur_i \leq B_m + (1 - \sum_{j \in N} Y_{ijm}) \cdot K, \forall i \in N, m \in M \tag{9}$$

$$B_m + breakT \leq S_{jm} + (1 - \sum_{i \in N} Y_{ijm}) \cdot K, \quad \forall j \in N, m \in M \tag{10}$$

$$sT + befBreak \leq B_m \leq fT + aftBreak, \forall m \in M \tag{11}$$

$$S_{im} + breakT \cdot Y_{ijm} + (t_{ij} + dur_i) \cdot X_{ijm} \leq S_{jm} + (1 - X_{ijm}) \cdot K, \forall i, j \in N, m \in M \tag{12}$$

$$S_{jm} - (1 - X_{ijm}) \cdot K \leq S_{im} + breakT \cdot Y_{ijm} + (t_{ij} + dur_i) \cdot X_{ijm}, \forall i, j \in N, m \in M \tag{13}$$

3) Teams Formation

Apart from the scheduling and routing, the model must assign as well the different available care workers to teams. The team assignment constraints are modeled as follow: constraints (14) and (15) imply that each available care worker (physicians and nurses, respectively) is assigned to exactly one team. Constraints (16) and (17) guarantee that each team does not have more than one physician and exactly one nurse, respectively.

$$\sum_{m \in M} Z_{pm} = 1, \forall p \in P \tag{14}$$

$$\sum_{m \in M} Z_{em} = 1, \forall e \in E \tag{15}$$

$$\sum_{p \in P} Z_{pm} \geq 1, \forall m \in M \tag{16}$$

$$\sum_{e \in E} Z_{em} = 1, \forall m \in M \tag{17}$$

4) Patient Requirements

Patient requirements can be divided into three types: early morning visit, patients that need to be seen by a physician, and nurse qualifications. In HH, patients may need an early morning visit due to some unforeseen events. The binary parameter $first_i$ is equal to 1 if the patient i needs a first-visit in the morning, and 0 otherwise. Constraint (18) ensure that patients with the need for a first-visit are the first ones to be visited. Furthermore, if it is not possible to have a physician in each team, then, only those teams with one physician can be assigned to patients with parameter $q_i = 1$, which indicates whether patient i needs to be visited by a team with a physician (based on Cakrgil et al. [3] and Decerle et al. [10]). Thus, constraint (19) ensure that a patient who needs to be seen by a physician is assigned to a team with a physician. Lastly, the parameter r_{is} is 1 if patient i needs to be seen by a nurse with skill s (and 0 otherwise), and a_{es} is 1 if nurse e has skill s (and 0 otherwise). Constraint (20) impose that a patient who requires a treatment associated with a specific nurse skill must be visited by a nurse qualified with that skill (based on Cakrgil et al. [3]).

$$first_i \leq \sum_{m \in M} X_{0im}, \forall i \in N \tag{18}$$

$$\sum_{j \in N} X_{ijm} \cdot q_i \leq \sum_{p \in P} Z_{pm}, \forall i \in N, m \in M \tag{19}$$

$$\sum_{j \in N'} X_{ijm} \cdot r_{is} \leq \sum_{e \in E} Z_{em} \cdot a_{es}, \forall i \in N, m \in M, s \in S \tag{20}$$

5) Workload Balance

Another aspect to be considered in this model regards the working time definition and balance. The auxiliary variable working time w_m of a team m is defined as the interval of time between the time that a team m leaves (l_m) and arrives (u_m) at the hospital. Expressions (21) - (23) summarize these computations (based on [9]).

To balance the working times within teams, an auxiliary variable named average working time $\overline{W}$ is computed by adding the working time of all teams and dividing by the number of teams (which corresponds to the number of vehicles available, |M|) (expression (24)). Then, the working time w_m for all teams must be within the average plus or minus an acceptable deviation δ agreed with the hospital (constraints (25)) (based on [16]).

$$\begin{aligned} S_{im} - t_{0i} + (X_{0im} - 1) \cdot K \leq l_m \leq \\ S_{im} - t_{0i} - (X_{0im} - 1) \cdot K, \forall i \in N, m \in M \end{aligned} \tag{21}$$

$$\begin{aligned} S_{im} + dur_i + t_{i0} + (X_{i0m} - 1) \cdot K \leq u_m \leq \\ S_{im} + dur_i + t_{i0} - (X_{i0m} - 1) \cdot K, \forall i \in N, m \in M \end{aligned} \tag{22}$$

$$w_m = u_m - l_m, \forall m \in M \tag{23}$$

$$\overline{W} = \frac{\sum_{m \in M} \times w_m}{|M|} \tag{24}$$

$$\overline{W} - \delta \leq w_m \leq \overline{W} + \delta, \forall m \in M \tag{25}$$

6) Objective functions

The problem aims to optimize two different objectives. The first objective function (26) seeks to minimize the total Travel Time incurred by the medical teams, while the second (27) concerns the quality of the service, which is measured by the Continuity of Care. The Continuity of Care objective seeks to maximize the connection between care workers and patients. To measure the Continuity of Care, parameters CC_{ip} and CC_{ie} are used for physicians and nurses, respectively. The parameter is based on the number of times care work (physician p or nurse e) has been assigned to patient i in the past. Thus, the higher the value of the parameter, the higher the connection between care worker and patient. By maximizing these two parameters, the model is going to assign to a patient the care workers who visited him/her the most.

$$min \sum_{m \in M} \sum_{i \in N} \sum_{j \in N'} X_{ijm} \cdot t_{ij} \tag{26}$$

$$\begin{aligned} max \sum_{p \in P} \sum_{i \in N} \left(CC_{ip} \cdot \sum_{m \in M} \left(\sum_{j \in N'} X_{ijm} \cdot Z_{pm} \right) \right) \\ + \sum_{e \in E} \sum_{i \in N} \left(CC_{ie} \cdot \sum_{m \in M} \left(\sum_{j \in N'} X_{ijm} \cdot Z_{em} \right) \right) \end{aligned} \tag{27}$$

C. Solution approach

When an optimization problem involves only one objective function, the task of finding an optimal solution is called Single-Objective Optimization Problem (SOOP). Nevertheless, most of the real-world problems involve different and possibly conflicting objectives. In the proposed model, there are two objectives to optimize, thus, it may not exist a solution that optimizes both simultaneously. Instead of an optimal solution, one will search for trade-off solutions among objectives, where it is not possible to improve the value of one objective function without worsening the value of another. In this context, the concept of optimal solution is replaced by the concept of Pareto efficiency. The optimal trade-off solutions or non-dominated solutions constitute the Pareto front. The task of finding Pareto optimal solutions is known as Multi-Objective Optimization Problem (MOOP). For further details in MOOP refer to Deb et al. [7].

The solution approach proposed for this dissertation can be seen in Figure 1. It starts by finding the Pareto front of the two conflicting objectives using the ϵ-constraint method. This method generates single objective subproblems, called ϵ-constraint problems - $P_k(\epsilon)$, by selecting one objective to be minimized or maximized, while transforming the remaining objectives in constraints to be less/more or equal to a given target value. The algorithm can be found in Bérubé et al. [1].

After having the Pareto front, two different paths can be followed. The first one (Figure 1, top) uses the crowding distance metric to provide a smaller number of non-dominated solutions to the decision-maker. Crowding distance is a widely used metric to compare non-dominated solutions in the same Pareto-front, based on the extent of their proximity with other solutions. At this stage, the decision-maker may use his/her

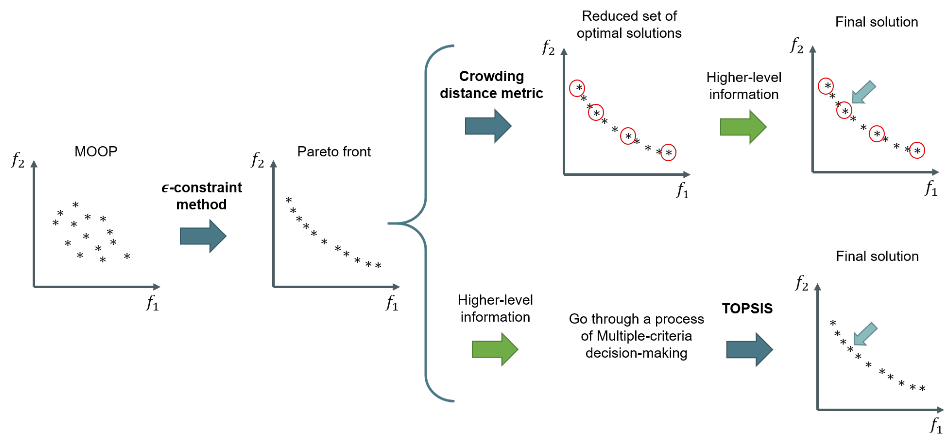

Fig. 1: Schematic solution approach for the MOOP

knowledge to compare the provided solutions and decide on the best one. For the full algorithm see Deb et al. [8].

Nevertheless, the decision-maker may not always be present to decide on the best solution, or it might still be difficult to choose only one solution. At this point, a method for Multiple Criteria Decision-Making (MCDM) should be used to obtain a ranking of the solutions. The method applied was Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), which allow to obtain a final solution from the initial Pareto front with only a few additional computations. The chosen solution is the one closer to the "ideal" solution and the farthest possible to the "negative-ideal" solution (see Opricovic et al. [22] for more details). For this method, it is necessary to determine weights for the different criteria so that the decision-maker could express his/her preferences in terms of the relative importance of criteria. In this stage, it would be important to tune these values with the decision-maker since the final solution depends on the vector of weights. Methods for weight assessment should be used. Nevertheless, it was not possible to apply those methods in our case study since it was not possible to discuss this further with the decision-maker. For this reason, in this dissertation, we use weights of 1/2 and 1/2 for both objectives simply to show, in this case, which solution would be chosen by the TOPSIS method.

IV. Numerical Experiments

This section presents an overview of the numerical experiments and discusses the main results obtained with the proposed solution approach. Experiments were conducted on real-world instances provided by HGO and on instances derived from the literature. All algorithms were coded in C++. For solving the Mixed-Integer Linear Programming, IBM ILOG CPLEX C++ Concert Technology version 12.9 was used.

A. *Real-world instances*

This work considers a real-world setting from the HHU of HGO, the case under study. A manual solution of a week is used to evaluate and compare the results provided by the model. These records indicate, for each day, all the nurses' visits, the corresponding starting time, and duration. Additionally, patients locations were provided in order to construct the matrix of Travel Time. Nevertheless, the data provided is not enough to test the whole model. There is no information regarding physicians' visits, traffic zones, and first-visit requests. Furthermore, information regarding breaks and skills was not available since these features are not considered in the current planning but is something the unit aims to incorporate. For this reason, a simplified model was used to run these instances and make a comparative analysis with the current planning. The simplified model comprises the two objective functions (26) and (27), without considering the physicians' parameters, and constraints (1) to (5); constraints (12) and (13) without the break parameter; constraints (15) and (17); and Workload Balance constraints (21) to (25).

There are 7 instances corresponding to a 7-day period from December 13 to December 19. On each day, there were 3 nurses and between 13 to 18 patients. The services' duration varies from 25 to 60 minutes. Locations correspond to patients real locations, and the Travel Time between patients was calculated using the postal-codes of the patients and Google Maps API, with the mode driving and traffic model equal to best guess [14]. Note that the Travel Time matrix is non-symmetric since it corresponds to real Travel Times. Additionally, the variables sT and fT were established as 9 am and 3 pm, respectively, corresponding to the current schedule of the HHU. The parameter δ, used to control the Workload Balance, was set by the HHU as 15 minutes, which means the maximum difference in working time between teams cannot be more than 30 minutes. The Continuity of Care score was computed taking into account the number of times a nurse visited a patient. For the first day of the test week, that is, for December 13, the Continuity of Care score was computed considering the assignments until that day. After December 13, the score was updated according to the solution given by the model, in order to see the improvement in Continuity of Care over a week. Thus, the Continuity of Care scores changes from the manual solution to the model solution since, in each instance, the assignment results from the manual and model solution may not necessarily be the same.

1) *Single-Objective analysis*

Before approaching the bi-objective model, a single objective analysis was performed to understand the model's behavior in each of the objectives. The model was solved for each objective individually using CPLEX with a time limit of 3600 seconds.

The graph in Figure 2 shows the percentage of improvement that the model allows in comparison with the current solution adopted by the hospital. In a single objective perspective, the Travel Time improvements range from 2.93 to 21.51% for the tested instances. As the hospital's operation area is relatively small, the improvements in Travel Time are limited to an extent. Nevertheless, more significant improvements can be achieved in the second objective, with the average increase in Continuity of Care of 37.11%.

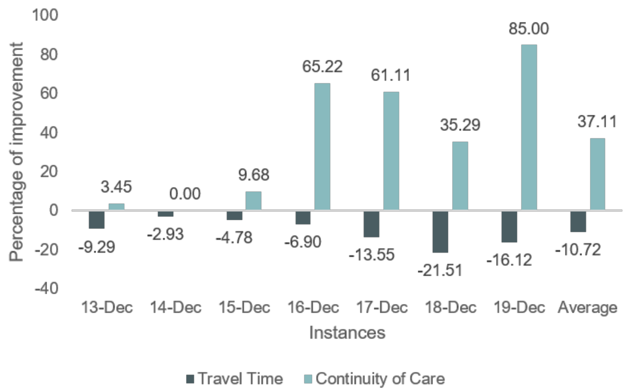

Fig. 2: Improvements in the single-objective values in relation to the current manual solution

The impact of the Workload Balance in the two objectives was also studied to better understand, alongside the HHU, which could be the best value for δ (i.e., the maximum deviation from the ideal average working time, which allows a difference in working times within teams). Figure 3 shows this analysis for the average of all instances. The graph on the top is related to Travel Time, while the one on the bottom represents the impact on Continuity of Care. Observing first the graph on the top, related to the Travel Time, the reference value is the maximum difference of 30 minutes. When this value decreases to 10 minutes, the Travel Time increases, on average, 4%. By contrast, when this value increases for 40 minutes or 60 minutes, the decrease in Travel Time is almost negligible. Thus, by increasing δ and giving more flexibility to the model to achieve lower Travel Times, it does not significantly improve after a maximum difference of 30 minutes (being this the value chosen by the HHU). Even so, one can conclude that Travel Time is slightly influenced by δ. On the other hand, looking for the graph related to the Continuity of Care on the bottom, it is possible to conclude that the Continuity of Care is not being influenced by δ.

2) HGO case study

The solution approach presented before was implemented in the real-world instances provided by HGO and compared with the hospital's results. Table II shows the results for the computed Pareto fronts. Column Time shows the total computation time in seconds for each instance, including finding the Ideal and Nadir values of each objective function and the

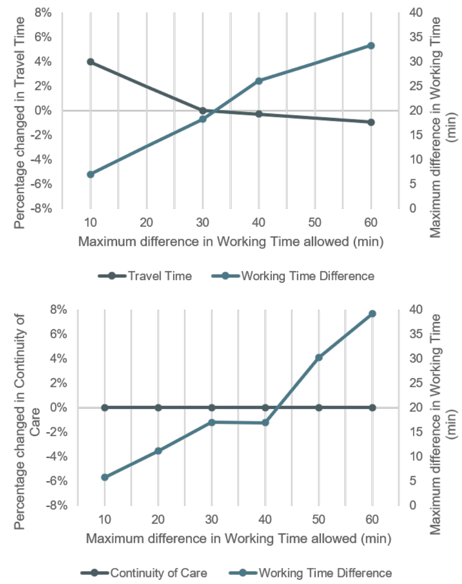

Fig. 3: Impact of the maximum working time difference allowed in Travel Time (graph on the top) and Continuity of Care (graph on the bottom)

computation of the ϵ-constraint subproblems. The next column represents the Pareto front size or, in other words, the number of non-dominated solutions found for each instance. Column $P_i(\epsilon_j)$ solved shows the number of ϵ-constraint subproblems solved to arrive at the final Pareto front. The two last columns show, respectively, the average computation time in seconds of the subproblems and the average gap, since a time limit of 3600 seconds was imposed. Thus, it is not possible to prove the optimality in all the subproblems run.

On average, each instance took 4.6 hours to solve through the ϵ-constraint method. More flexible models, depending on characteristics such as the Travel Time matrix, number of patients, and Continuity of Care score, allow for a higher number of possible solutions. Thus, more subproblems need to be solved, and the computation time increases as well.

Instance	Number of Patients	Total Comp. Time (h)	Pareto front size	Number of $P_i(\epsilon_j)$ solved	Avg. comp. Time of $P_i(\epsilon_j)$ (s)	Avg. Gap $P_i(\epsilon_j)$ (%)
13.12	18	6.72	3	4	3 600.00	15.38
14.12	17	1.00	1	0	-	-
15.12	18	5.28	3	3	2 733.49	9.12
16.12	18	11.00	7	8	3 600.00	12.22
17.12	14	2.28	3	1	733.81	0
18.12	14	5.63	4	6	1 620.14	0
19.12	13	0.26	1	0	-	-
Average		**4.60**	**3**	**3**	**2 457.49**	**7.34**

TABLE II: Computational Results for the real instances

Table III shows the results of the current solution of the hospital and the results obtained by the model focusing on the two objectives that are being optimized, namely Travel Time and Continuity of Care, and on Workload Balance. The final solutions chosen for the model were given by TOPSIS and considering the weight of 1/2 for each objective. For the hospital's current solution, the Travel Times were computed using the same Travel Time matrix used to compute the model solution and using the order of the visits provided by the hospital. Each team's workload was computed as for the model solution, adding the Travel Time of each team and the total time of visits duration. The highlighted columns show the variation from the current manual solution to the model solution. Note that Workload Balance is measured as the difference between the team which works more in a day and the team which works less. The lower the difference, the better is the Workload Balance.

	Travel Time (min)			Continuity of Care (score)			Workload Difference (min)		
Instance	Current Solution	Model Solution	Variation (%)	Current Solution	Model Solution	Variation (%)	Current Solution	Model Solution	Variation (%)
13.12	193.27	184.80	**-4.38**	29	30	**+3.45**	29.35	26.07	**-11.19**
14.12	177.38	172.18	**-2.93**	18	18	**0.00**	96.5	15.12	**-84.33**
15.12	186.60	181.65	**-2.65**	31	33	**+6.45**	105.47	27.52	**-73.91**
16.12	187.37	190.43	**1.64**	23	37	**+60.87**	35.85	12.87	**-64.11**
17.12	166.40	150.63	**-9.48**	18	29	**+61.11**	117.62	17.37	**-85.23**
18.12	197.37	166.40	**-15.69**	17	22	**+29.41**	89.85	14.93	**-83.38**
19.12	170.63	143.13	**-16.12**	20	37	**+85.00**	41.4	17.42	**-57.93**
Average			**-7.09**			**+35.18**			**-65.73**

TABLE III: Comparison between the current manual solution and the solution given by the model

According to Table III, the proposed approach improves the current hospital solutions in both objectives, Travel Time and Continuity of Care, and Workload Balance, except for instance 16.12, where the Travel Time increases 1.64% concerning the current solution. The results show, on average, a reduction in Travel Time of 7.09%. Nevertheless, this small improvement was accompanied by an average increase of 35.18% in Continuity of Care and an average decrease in working time difference within teams of 65.73%.

For the model solution, the patients are assigned to the nurse whom they have the highest Continuity of Care relation 94.64% of the times, while for the current solution, this happens 79.46%. Additionally, on the model solution, for the same patient, it does not happen more than once to be assigned to a nurse who is not the nurse with the highest Continuity of Care score. On the other hand, on the current manual solution, there are patients penalized four times in a period of one week. Concerning Workload Balance, for each nurse, the average working time at the end of the 7 days was computed. For the manual solution, the standard deviation of the averages is 42 minutes, while for the model solution, it is 27 minutes. By reducing the Travel Time, the teams' total working time also decreases since travel is being done more efficiently. For this reason, the total working time at the end of the 7 days was slightly higher in the manual solution (95.07h) than for the model solution (93.57h). Moreover, by improving Continuity of Care, it is expected that the duration of the service in each patient may decrease as well since nurses are more aware of the patients' status. Thus, the total working time may decrease even more.

To sum up, on average, the model was able to improve Travel Time by 7.09%. Additionally, it was possible to significantly improve the remaining objectives without compromising the Travel Time (except for instance 16.12). Nevertheless, the solutions given by the model are obtained using weights of 1/2 for each of the objectives. Thus, by changing the weights according to the decision-maker's preferences, it could be possible to obtain solutions that improve even more the decision-maker's preferred criteria.

B. Literature Instances

In the field of HHC, there are not many benchmark instances available, especially considering the Continuity of Care feature. The only instances available in the literature with Continuity of Care data were published by Grenouilleau et al. [15] and can be consulted in https://doi.org/10.17632/cbgt59hnhk.1. Nevertheless, these instances are provided to solve the HHCRSP for a weekly period, and not a single a day, which means that each instance can originate 7 instances for our daily problem. The number of caregivers used was based on the hospital information of 3 nurses per 18 patients. The information retrieved from the instances were location ids, travel time between locations, duration of the service, mandatory skills, Continuity of Care scores associated with each patient-caregiver pair, which corresponds to the number

Instances	Total Time (h)	Ideal* TT Time (s)	Ideal* TT Gap (%)	Ideal* CC Time (s)	Ideal* CC Gap (%)	Nadir* TT Time (s)	Nadir* TT Gap (%)	Nadir* CC Time (s)	Nadir* CC Gap (%)	Pareto Size	N of Sub problems	Avg. Comp. time (s)	Average Gap (%)
1_F_14	0.20	76.55	0	141.27	0	205.54	0	101.44	0	4	2	105.81	0
2P_F_14	0.15	98.29	0	161.96	0	48.17	0	55.34	0	4	2	95.76	0
1_G_16	6.41	128.91	0	3600	16.67	3600	No sol.	1520.15	0	6	6	2373.11	4.89
2N_G_16	5.94	134.95	0	3600	0	3600	0	135.22	0	4	5	2201.51	4.63
1_A_18	3.98	346.54	0	3600	No sol.	-	-	322.26	0	4	5	2009.52	0.78
2S_A_18	8.39	267.54	0	3600	3.13	155.23	0	254.29	0	19	19	1364.63	0.31
3N_A_18	1.60	238.15	0	3600	3.33	92.61	0	449.51	0	2	2	683.95	0
1_B_18	1.51	3600	2.06	25.26	0	42.89	0	1679.38	0	3	2	40.48	0
2F_B_18	3.55	3600	7.86	593.42	0	45.26	0	3600	No sol.	5	5	988.81	1.35
3T_B_18	1.66	3600	3.76	81.12	0	11.53	0	1854.57	0	5	4	106.04	0
1_C_18	4.31	99.83	0	3600	No sol.	-	-	388.69	0	4	5	2282.12	0.33
2P_C_18	3.15	359.14	0	3600	No sol.	-	-	888.92	0	2	3	2169.73	0
1_D_22	6.00	3600	12.02	3600	35.38	3600	13.61	3600	No sol.	2	2	3600	9.16
2FT_D_22	20.00	3600	12.24	3600	18.42	3600	14.48	3600	No sol.	10	16	3600	14.54
1_E_24	19.00	3600	18.04	3600	12.24	3600	7.57	3600	No sol.	9	15	3600	13.75

TABLE IV: Computational results for approximate Pareto fronts (the symbol * is used to show that some obtained values can be approximate values) of times the caregiver has been assigned to the patient's past visits. The remaining parameters (i.e., first-visit, need for physician, traffic and continuity of care score of physicians) were generated randomly based on data and information provided by HGO.

The instances were named following the logic CHARACTERISTIC_DATA_NUMBER OF PATIENTS. Each letter on column Data represents a different set of patients, nurses, and physicians. To study the impact of the different parameters on the model, instances with a different number of patients and different data were generated. Moreover, for the same set of data and number of patients, characteristics such as number of first-visits (F), number of patients needing a physician (N), number of skills available (S), number of patients living in areas with heavy traffic (T) and number of physicians available (P) were changed. For example, instance 1_F_14 is composed of 14 patients, using the data set F and the characteristics of what is considered the base instance using data set F. Then instance 2P_F_14 is the second instance using data set F, but the characteristic P was modified from the base instance.

1) Experiments

Tests were performed in instances with 14, 16, 18, 22, and 24 patients. There was an attempt to conduct tests in larger instances, but the model could not provide a solution within the time limit of one hour for instances with 26 and more patients. The time limit of one hour was imposed for each intermediate run, i.e., to find the Ideal and Nadir values, and for each of the ϵ-constraint subproblems. Note that it was not always possible to obtain a gap of 0%, meaning that the final result is not a Pareto front but an approximate Pareto front. The results of the instances can be found in Table IV.

The first difference one can note compared to the real instances' results is that until 18 patients, the optimization of Travel Time can happen within 400 seconds, except for instances with data set B. This may be due to the fact that now all the features of the model are being considered, which may facilitate the assignment and routing process by limiting the number of possible solutions (e.g., first visits and physician requirements). By contrast, when optimizing the Continuity of Care objective, in general, it is not possible to arrive at a gap of 0% within 3600 seconds. In real instances, the model could find the optimal value of Continuity of Care. However, there are more restrictions to respect in the literature instances, and, thus, it is harder to optimize the Continuity of Care.

The instances were defined in order to help understand the impact of the different parameters used in the model. For example, from instance 1_B_18 to instance 2F_B_18, it changes the number of patients needing a first-visit. The first-visit parameter facilitates the routing process since it restricts the possible solutions as some patients need to be the first ones on the routing solution. For this reason, having fewer patients needing a first-visit increased the computation time from 1.51 hours to 3.55 hours. Similarly, instance 3T_B_18 does not have patients living in a traffic area, which also slightly increased the computation time to 1.66 hours. The impact of these two features together can be seen in instances with data D. From instance 1_D_22 to 2FT_D_22, the number of patients needing a first-visit and living in an area with heavy traffic changed to zero. As a consequence, the total computation time increased from 6 hours to 20 hours. The impact can also be seen in the size of the Pareto front, which increased from 2 to 10 due to the increase in the instances' flexibility.

Looking now at instance 1_A_18, it was solved in a total computation time of 3.98 hours. The final approximated Pareto front found has 4 non-dominated solutions- Figure 4 in red. Instance 2S_A_18 has the same patients, physicians, and nurses' characteristics identified by letter A. Nevertheless, the nurses' skills and the patients' requirements for those skills were not considered. As a result, the total computational time increased from 3.98 to 8.39 hours, and the approximate Pareto front from 4 to 19 non-dominated solutions - Figure 4 in green. This happens because there are more possibilities for assigning patients to nurses, increasing the number of possible non-dominated solutions. In the third instance, 3N_A_18, the number of patients needing a physician rose from 4 to 8. As there are only 2 physicians to assign to the 3 teams, increasing the number of patients who need to be seen by a physician facilitates the assignment process. Thus, the computational time decreased to 1.6 hours, and the exact Pareto front - since the gap was 0% - has only 2 non-dominated solutions - Figure 4 in blue. From this analysis, it is possible to conclude that the Pareto front's size and structure are highly dependent on the instances' characteristics.

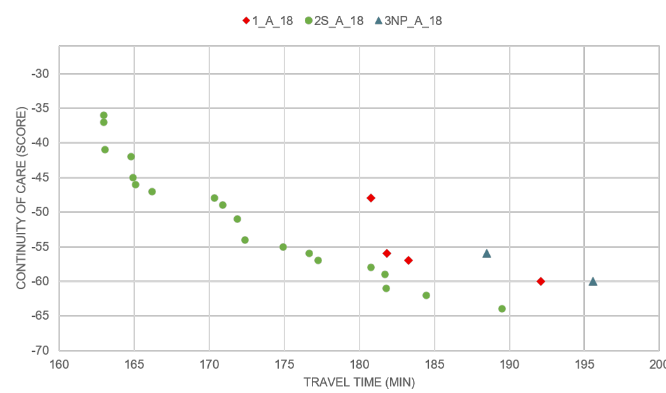

Fig. 4: Approximate Pareto fronts for three different instances with the same patients, nurses and physicians

V. Conclusions and Future Research

Portugal has seen in HH an opportunity to decrease hospital congestion. Encouraged by the government to adapt the health care system to the growing needs of an aging population, HH has been expanding rapidly over the last few years. HGO was the first hospital to implement an HHU in Portugal and has already confirmed the unit's benefits. Nevertheless, the scheduling and routing decisions are currently done manually by the HGO's HHU, requiring a long time to obtain a valid schedule.

For this reason, this dissertation proposes a mathematical model and solution approach to help in the complex operations decisions. The bi-objective model accounts for the three main stakeholders' perspectives existing in HH: management (Travel Time), patients (Continuity of Care), and care workers (Workload Balance). The Continuity of Care was considered for both nurses and physicians. After looking at the results, there is significant room for improvement from the manual solution to the solution provided by the model in the three stakeholders perspectives. Furthermore, it is possible to improve the Travel Time and, consequently, the costs while still improving the Continuity of Care and the Workload Balance. The results show, on average, a reduction in Travel Time of 7.09%. Nevertheless, this small improvement was accompanied by an average increase of 35.18% in Continuity of Care and an average decrease in working time difference within teams of 65.73%.

The main limitation of the proposed model and solution approach is the computation time needed to find all the Pareto front and the fact that the model can only afford small instances. Additionally, the application of the ϵ-constraint method does not always allow to find the exact Pareto front, but an approximate Pareto front. This dissertation contributed to the understanding of the trade-off and parameters used in this problem. Nevertheless, from a practical point o view, a lower computation time may be preferable. Opportunities for future research should incorporate the study of possible heuristics algorithms to couple with larger instances and find non-dominated solutions within a lower computation time. Furthermore, other variants of the Continuity of Care score may also be studied. For example, the HHU may prefer to privilege the Continuity of Care for patients who have been hospitalized for less time because the patient's condition may not yet be well defined, and ensuring follow-up is crucial. The Continuity of Care score can then also be set according to the length of stay. Additionally, the Continuity of Care score is not easy to interpret, and it should also be thought of as a way to communicate this measure better to the hospital. Additionally, online decisions should be studied and proposed. In HH, care workers are faced with real-time information such as a patient who calls for immediate visits. The capability of reacting to these emergencies by determining which care worker is closer to that patient location or immediately available to visit and redefining routes or readjusting the initial plan needs to be incorporated in more robust models.

REFERENCES

[1] J.-f. Bérubé and M. Gendreau. An Exact ϵ -constraint Method for Bi-objective Combinatorial Optimization Problems - Application to the Traveling Salesman Problem with Profits. (February), 2007.

[2] D. Bredström and M. Rönnqvist. Combined vehicle routing and scheduling with temporal precedence and synchronization constraints. *European Journal of Operational Research*, 191(1):19–31, 2008.

[3] S. Çakırgil, E. Yücel, and G. Kuyzu. An integrated solution approach for multi-objective, multi-skill workforce scheduling and routing problems. *Computers and Operations Research*, 118, 2020.

[4] G. Carello, E. Lanzarone, and S. Mattia. Trade-off between stakeholders' goals in the home care nurse-to-patient assignment problem. *Operations Research for Health Care*, 16:29–40, 2018.

[5] J. Cheng, M. Montalto, and B. Leff. Hospital at Home. *Clinics in Geriatric Medicine*, 25(1):79–91, 2009.

[6] M. Cissé, S. Yalçındağ, Y. Kergosien, E. Şahin, C. Lenté, and A. Matta. OR problems related to Home Health Care: A review of relevant routing and scheduling problems. *Operations Research for Health Care*, 13-14:1–22, 2017.

[7] K. Deb. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001.

[8] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[9] J. Decerle, O. Grunder, A. Hajjam El Hassani, and O. Barakat. A general model for the home health care routing and scheduling problem with route balancing. *IFAC-PapersOnLine*, 50(1):14662–14667, 2017.

[10] J. Decerle, O. Grunder, A. Hajjam El Hassani, and O. Barakat. A memetic algorithm for a home health care routing and scheduling problem. *Operations Research for Health Care*, 16:59–71, 2018.

[11] M. Di Mascolo, M. L. Espinouse, and Z. E. Hajri. Planning in Home Health Care Structures: A literature review. *IFAC-PapersOnLine*, 50(1):4654–4659, 2017.

[12] C. Fikar and P. Hirsch. Home health care routing and scheduling: A review. *Computers and Operations Research*, 77:86–95, 2017.

[13] M. I. Gomes and T. R. P. Ramos. Modelling and (re-)planning periodic home social care services with loyalty and non-loyalty features. *European Journal of Operational Research*, 277(1):284–299, 2019.

[14] Google. Google Maps APIs, 2020.

[15] F. Grenouilleau, A. Legrain, N. Lahrichi, and L. M. Rousseau. A set partitioning heuristic for the home health care routing and scheduling problem. *European Journal of Operational Research*, 275(1):295–303, 2019.

[16] A. Hertz and N. Lahrichi. A patient assignment algorithm for home care services. *Journal of the Operational Research Society*, 60(4):481–495, 2009.

[17] HGO. Unidade de Hospitalização Domiciliária Indicadores de Atividade. Technical report, 2019.

[18] C. Martinez, M. L. Espinouse, and M. Di Mascolo. Continuity of care in home services: A client-centered heuristic for the home health care routing and scheduling problem. *2018 5th International Conference on Control, Decision and Information Technologies, CoDIT 2018*, pages 1045–1050, 2018.

[19] A. B. Milburn and J. Spicer. Multi-objective home health nurse routing with remote monitoring devices. *International Journal of Planning and Scheduling*, 1(4):242, 2013.

[20] S. Nickel, M. Schröder, and J. Steeg. Mid-term and short-term planning support for home health care services. *European Journal of Operational Research*, 219(3):574–587, 2012.

[21] OECD. *Health at a Glance 2019: OECD Indicators*. OECD Publishing, Paris, 2019.

[22] S. Opricovic and G. H. Tzeng. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research*, 156(2):445–455, 2004.

[23] S. Quintanilla, F. Ballestín, and Pérez. *Mathematical models to improve the current practice in a Home Healthcare Unit*, volume 42. Springer Berlin Heidelberg, 2020.

[24] A. Trautsamwieser and P. Hirsch. A Branch-Price-and-Cut approach for solving the medium-term home health care planning problem. *Networks*, 64(3):143–159, 10 2014.

[25] L. Xiao, M. Dridi, and A. H. El Hassani. Mathematical Model for the Home Health Care Scheduling and Routing Problem with Flexible Lunch Break Requirements. *IFAC-PapersOnLine*, 51(11):334–339, 2018.

[26] S. Yalçındağ, P. Cappanera, M. Grazia Scutellà, E. Şahin, and A. Matta. Pattern-based decompositions for human resource planning in home health care services. *Computers and Operations Research*, 73:12–26, 2016.