A matheuristic algorithm for stochastic home health care planning

Nikzad, E., Bashiri, M. & Abbasi, B.

A Matheuristic Algorithm for Stochastic Home Health Care Planning

Erfaneh Nikzad[a], Mahdi Bashiri[b,a,*], Babak Abbasi[c]

[a]*Industrial Engineering Department, Faculty of Engineering, Shahed University, Tehran, Iran*
[b]*School of Strategy and Leadership, Faculty of Business and Law, Coventry University, Coventry, Uk*
[c]*Department of Information Systems and Business Analytics, RMIT University, Melbourne, VIC 3000, Australia*

Abstract

Efficient human resource planning is the cornerstone of designing an effective home health care system. Human resource planning consists of decisions on districting/zoning, staff dimensioning, resource assignment, scheduling, and routing. In this study, a two-stage stochastic mixed integer model is proposed that considers these decisions simultaneously. In the planning phase of a home health care system, the main uncertain parameters are travel and service times. Hence, the proposed model takes into account the uncertainty in travel and service times. Districting and staff dimensioning are defined as the first stage decisions, and assignment, scheduling, and routing are considered as the second stage decisions. A novel algorithm is developed for solving the proposed model. The algorithm consists of four phases and relies on a matheuristic-based method that calls on various mixed integer models. In addition, an algorithm based on the progressive hedging and Frank and Wolf algorithms is developed to reduce the computational time of the second phase of the proposed matheuristic algorithm. The efficiency and accuracy of the proposed algorithm are tested through several numerical experiments. The results prove the ability of the algorithm to solve large instances.

Keywords: Home health care, Matheuristic algorithm, Districting, Staff dimensioning, Progressive hedging algorithm, Fix and optimize method

1. Introduction

The United Nations has reported that 12.3 percent of the population were over 60 years of age in 2015, and forecast to be more than 21.3 percent in 2050. Therefore, many countries face an aging population. Providing the required health care services for an aging population is a challenge for health care systems in different countries. Many countries have invested in long-term care (LTC) services to improve the cost and quality of their health care services. Long-term care is also known as home health care (HHC) if the services are provided at home. Also, pandemic situations, such as what has been observed during the COVID-19 outbreak, significantly increase the demand of home healthcare as many people are asked to stay at home while they still need health and medical services. This has prompted many countries to further invest in their

---

*Corresponding author
Email addresses: `e.nikzad@shahed.ac.ir` (Erfaneh Nikzad), `mahdi.bashiri@coventry.ac.uk` (Mahdi Bashiri), `babak.abbasi@rmit.edu.au` (Babak Abbasi)

home healthcare services. Home Health Care News (HHCN) (Donlan, 2020) reported that the operations of 31% of home health care providers have been affected by the spread of COVID-19 over the past few weeks.

Home health care is defined as the provision of a wide range of medical and paramedical services to patients in their homes or in care facilities for elderly people. HHC services may include wound care for pressure sores or surgical wounds, injections, the monitoring of serious illnesses and unstable health statuses, etc. HHC is booming in many countries as it is less costly than hospital care and, in many cases, is more comfortable and effective than the care provided in a hospital (Akjiratikarl et al., 2007; Yuan et al., 2015).

Decisions related to human resource planning in home health care can be categorized into decisions on staff dimensioning, partitioning the territory into districts (districting), assigning caregivers to patients, and the scheduling and routing of caregivers (Yalçındağ et al., 2016; Sahin and Matta, 2015). A few articles have considered two levels of these decisions simultaneously, but, to the best of the authors' knowledge, there is no study focusing on all decision levels, while an integrated model considering all decision levels will lead to less cost and higher service quality. Therefore, in this study an integrated model is proposed to make integrated decisions on districting, dimensioning, assignment, and routing.

Districting is the initial phase of designing an HHC system, and has significant effects on improving the efficiency of care delivery. The aim of districting is to partition the basic units (sets of patients and caregivers) into a number of larger clusters so that caregivers are able to deliver care to the patients in their cluster more quickly. This ultimately increases the patients' and the caregivers' satisfaction. Moreover, staff dimensioning is an important tactical decision as it determines the number of caregivers who need to be hired in each district. This decision is critical because recruiting or laying off caregivers during the planning horizon is costly, and the decision also affects the caregivers' satisfaction by increasing their job security.

In the proposed model, the set of patients and caregivers is partitioned into districts based on the costs of the routing and staff dimensioning. The goal of the model is to determine the set of required caregivers and patients in each district and the required number of caregivers, while taking into account routing and hiring costs. The caregivers in a district can only be assigned to the patients in that district, and they start their visiting tour from their homes. Based on our review of real case situations, we found that HHC centers often try to allocate patients to the caregivers who live in the same residential zone and would like to design the schedule so that the starting and finishing locations for each caregiver are as close as possible. However, if this assumption is not imposed, the routing decisions can be replaced by the open vehicle routing problem in the proposed model. Caregivers have a predefined maximum number of working hours per day. Similarity between caregivers' skills and continuity of care, which are known to be important points in dimensioning, assignment, scheduling, and routing problems, are considered in the proposed model, to optimize the routing and assignment costs. Continuity of care guarantees that each patient is visited by one caregiver or a limited number of caregivers. Continuity of care decreases the data transmission between caregivers, increases patients' comfort, and, as a result, improves the relationships between the caregivers and the patients. In real situations, the values of some parameters are stochastic and can be characterized in

various scenarios. Hence, in this paper, a two-stage stochastic programming model is proposed to deal with uncertain travelling and servicing times. In several studies of stochastic vehicle routing problems, travel and service times are considered as uncertain parameters (Pillac et al., 2013). In the proposed two-stage model, the districting and staff dimensioning are the first stage decision variables, and the assignment and routing decisions are defined as the second stage variables.

To solve the proposed model, a new matheuristic algorithm is proposed. The proposed algorithm has of four phases. In the first phase, the problem size is decreased by determining the visiting sequence for the patients and obtaining a restricted set of the number of potential districts. An initial solution is determined in phase 2, based on the visiting sequence and the set of the numbers of potential districts obtained in phase 1. The progressive hedging based algorithm is employed to solve the proposed model in the second phase. In phase 3, an infeasible solution is repaired by using an iterative approach based on a fix and optimize method. The obtained solution is improved in the last phase of the algorithm by using a procedure based on the fix and optimize method.

This paper is organized as follows. Studies related to human resource planning in home health care are briefly reviewed in Section 2. The problem is defined and formulated in Section 3, and the proposed mathematical model based algorithm is explained in Section 4. Section 5 is devoted to numerical examples. Finally, the conclusions and some perspectives are provided in the last section. For a better illustration of the current research stream, a road map of the whole study is depicted in Figure 1.

2. Literature review

Home health care planning deals with different types of decision variables. Decisions related to home health care planning based on a planning horizon can be divided into three classes. These classes are the strategic, tactical, and operational decisions. Districting and staff dimensioning decisions are at the strategic and tactical levels, respectively, and assignment and routing decisions are at the operational level. The studies related to human resource planning in home health care are reviewed in this section on the basis of the type of decision.

2.1. Districting

A basic unit is defined as a postal area, a street, a geo-code address, etc. Using the definition of basic unit in HHC, patients are grouped into basic units to reduce the problem size in the districting phase (i.e., basic units are used for districting instead of each individual patient's address). Therefore, districting in HHC is defined as grouping basic units into larger clusters (districts) (Blais et al., 2003; Benzarti et al., 2013). In districting, a territory is partitioned into a number of districts and each caregiver is assigned to a district. Districting not only improves service quality but also increases patients' and caregivers' satisfaction. It decreases travel time, and consequently improves caregivers' efficiency in serving patients. Only a few studies in the context of home health care planning are devoted to districting. In these studies, several basic
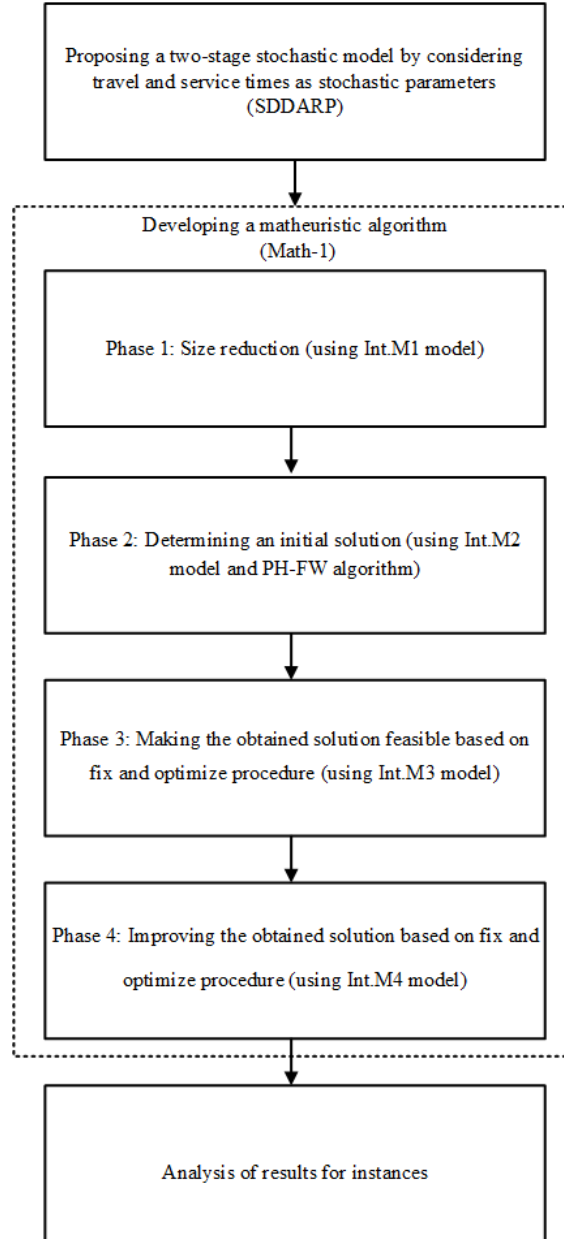
3

Figure 1: Road map of the this study.

units (where a basic unit is defined as a set of patients) are assigned to districts, by considering different criteria.

Blais et al. (2003) considered criteria such as the indivisibility of basic units, mobility, connectivity, workload equilibrium, and respect of borough boundaries in a home health care districting problem. The objective function of their model consisted of two terms, including the mobility measure and workload equilibrium, with the other criteria being included as hard constraints. The mobility measure determines the total distance travelled in one year. The authors used a Tabu Search algorithm to solve the problem.

Benzarti et al. (2013) proposed two models for an HHC districting problem by considering the indivisibility of basic units, compactness, workload balance between caregivers, and compatibility. Compatibility covers mobility and connectivity criteria. Balancing the workload of caregivers in different districts was the objective function of their first model, and minimizing the compactness measure (the maximum distance between basic units in the same district) was the objective function of their second model. In their models, caregivers' workload consists of the travelling time between patients and the time required to deliver the care service.

Blais et al. (2003) and Benzarti et al. (2013) defined a basic unit as a group of patients, but the starting nodes of the caregivers were not taken into account. It should be noted that the starting nodes can affect the efficiency of the home health care planning, and can help managers to decide on a cost effective system by decreasing the routing cost and the number of caregivers required. In addition, both papers assumed that caregivers are identical in terms of their level of skills, and the similarity between the skill level of the caregivers and the patients' requirements was not considered. Also, all parameters were deterministic in their models. A comparison between the current study and these studies is shown in Table 1.

Table 1: Comparison between the studies related to districting in HHC with the proposed model. The abbreviations are: MP: Multi-period, SP: Single period, TS: Two-stage stochastic programming, RP: Robust programming, PC: Probabilistic constraint (Chance constraint), IB: Indivisibility of basic units, WL: Workload equilibrium, CM: Compactness measure, AC: Accessibility, HC: Hiring Cost, RC: Routing Cost, E: Exact, MH: Matheuristic, TT: Travel time, ST: Service time.

Studies	Planning horizon		Considering uncertainty			Uncertain parameter	Criteria						Solution method
	MP	SP	TS	RP	PC		IB	WL	CM	AC	HC	RC	
Blais et al. (2003)		*				-	*	*	*	*	-	-	E
Benzarti et al. (2013)		*				-	*	*	*	*	-	-	E
This paper	*		*			TT, ST	*		*	*	*	*	MH

2.2. Staff dimensioning

Staff dimensioning is defined as determining the number of caregivers required in the home health care system to meet the demands of different types of patients. Caregivers' travelling time between patients, and factors such as the similarity between the patient's demands and the caregiver's level of skills, continuity of care and limitations on caregivers' working time, must be considered in staff dimensioning. In spite of the fact that staff dimensioning is critical in HHC planning, there is only one study in the literature dedicated to this problem. Rodriguez et al. (2015) proposed a two-stage stochastic model for staff dimensioning in

HHC when the demands of patients are uncertain. The number of caregivers at each skill level was the first stage decision variable, and the assignment of caregivers to patients was the second stage variable in their model. They developed an integrated model that simultaneously considers staff dimensioning, assignment, and routing decisions. They proposed a two-phase algorithm to solve their model. In the first phase, the minimum number of required caregivers at each skill level is determined to serve all demands under each scenario. Then, the number of caregivers required to satisfy a defined level of performance is obtained by solving a master problem. The authors assumed that all demands had to be met in a single period.

2.3. Scheduling and routing of caregivers

Most studies in home health care planning focus on scheduling and routing problems. Fikar and Hirsch (2017) and Cissé et al. (2017) reviewed the studies related to scheduling and routing problems in home health care. They classified the studies, based on the planning horizon, into two clusters. In the first cluster, a single period is considered, and most studies in the context of scheduling and routing problems are in this category. Fikar and Hirsch (2017) surveyed the articles according to different criteria, such as solution method, objective function, and constraints. They highlighted that mathematics-based (matheuristic) heuristic algorithms were proposed by Bertels and Fahle (2006), Bredström and Rönnqvist (2008), and Fikar and Hirsch (2015). Several articles on single period scheduling and routing problems in home health care considered uncertainty in different parameters, such as demand or service time. Lanzarone and Matta (2014) assumed the demands of the patients to be uncertain. Yuan et al. (2015) proposed a two-stage stochastic programming model for a scheduling and routing problem in home health care when the service times and skill requirements are stochastic. They solved their model using a branch-and-price algorithm.

In the second cluster, caregivers work on multiple days and the services required by patients are spread over different days. The assignment, scheduling, and routing decisions are considered to be integrated in this cluster of problems. Because the assignment of caregivers is considered and the services required by patients are spread over different days, some factors, such as continuity of care, are defined in this type of problem. Despite the importance of the continuity of care, only a few studies have paid attention to it. Continuity of care was considered as a term in the objective function in some studies, e.g. (Carello and Lanzarone, 2014; Wirnitzer et al., 2016) or as a soft or hard constraint in others e.g. (Duque et al., 2015; Cappanera and Scutellà, 2014).

Different solution methods have been proposed in the studies on assignment, scheduling, and routing problems. In most of these studies, heuristic and meta-heuristic algorithms have been proposed, e.g. (Nickel et al., 2012). Furthermore, in some studies, matheuristic algorithms were developed. Barrera et al. (2012) proposed a mathematics-based heuristic approach to solve a multi-objective assignment, scheduling, and routing problem. This approach is defined on the basis of lexicographic optimization, which was used as the method for solving multi-objective problems. Duque et al. (2015) proposed a two-phase mathematics-based heuristic solution method. In the first phase, the feasible visiting schemes were generated, and then an

optimal visiting scheme was determined for each patient using a mathematical model. A heuristic approach was then used to determine the routing decision, based on the scheme obtained in the previous phase.

There are several studies on assignment, scheduling, and routing problems focused on uncertainty, considering the demands, travel time, and servicing time as uncertain parameters. Yuan et al. (2018) proposed a two-stage stochastic programming model for when the servicing and travel times are stochastic. They also used a branch-and-price algorithm to solve the problem. Carello and Lanzarone (2014) proposed a robust model for assigning patients to caregivers when their demands are uncertain.

A comparison between the proposed model and the studies related to HHC with uncertain parameters is presented in Table 2. This table shows that in the context of resource planning in HHC there are few studies that consider uncertain parameters. It can be concluded that most studies are related to the scheduling and routing of the caregivers, and that there is no study considering the districting, staff dimensioning, scheduling, and routing decisions simultaneously.

Table 2: Comparison between the studies in HHC by considering uncertainty with the proposed model. The abbreviations are: Ro: Robust optimization, Sp: Stochastic programming, Dis: Districting, Dim: Staff dimensioning, AS: Assignment, RS: Routing and Scheduling, 2-s: Two-stage stochastic programming, M-s: Multi-stage stochastic programming, D: Demand, S: Skill requirement, TT: Travel time, ST: Servicing time, H: Heuristic, E: Exact, MH: Matheuristic, Meta: Metaheuristic.

Studies	Decisions				Uncertain parameter				Ro	Sp		Solution method
	Dis	Dim	AS	RS	D	S	TT	ST		2-s	M-s	
Lanzarone and Matta (2014)			*	*	*				*			H
Carello and Lanzarone (2014)			*	*	*				*			E
Rodriguez et al. (2015)		*		*	*					*		MH
Yuan et al. (2015)				*		*		*		*		E
Yuan et al. (2018)				*			*	*		*		E
Shi et al. (2018)				*			*	*		*		E, Meta
Shi et al. (2019)				*			*	*	*			Meta
This paper	*	*	*	*			*	*		*		MH

2.4. Matheuristic algorithms

The development of mathematical programming based solution algorithms has focused on many studies related to the Vehicle Routing Problem. Li et al. (2019) designed a three phase matheuristic algorithm for the Production Routing Problem (PRP). An initial solution was generated by an iterative algorithm in the first phase of this matheuristic, then in the second phase the obtained solution was repaired to make it feasible, and in the last phase the feasible solution was improved. Yu et al. (2019) developed a matheuristic algorithm for the Orienteering Problem, in which an initial solution for routing was generated in the first phase and then a combination of local search and diversification procedures was used to improve the initial solution. Keskin and Çatay (2018) proposed a two-phase matheuristic algorithm for an Electric Vehicle Routing Problem. An initial solution was generated in the first phase using an adaptive large neighbourhood search (ALNS) algorithm, then the obtained solution was improved by optimizing a mathematical model in the second phase of the algorithm developed by Keskin and Çatay (2018). Archetti et al. (2017) proposed a matheuristic

algorithm for a Multivehicle Routing Problem. The initial solution was generated in the first phase, then a Tabu Search algorithm was used to generate a set of routes. A route-based mathematical model was proposed to improve the solutions. Campelo et al. (2019) designed a three-stage matheuristic algorithm for a Consistent Vehicle Routing Problem. The numbers of constraints and variables was decreased by node grouping in the first stage. In the second stage, a procedure was used for generating a suitable initial solution and, in the last stage, the solution was found using a fix and optimize algorithm. Crama et al. (2018) and Neves-Moreira et al. (2018) proposed matheuristic algorithms for an Inventory Routing Problem and a Production Routing Problem with perishable products, respectively. Grenouilleau et al. (2019) proposed a matheuristic algorithm based on a set-partitioning approach for routing and scheduling problem in the HHC. In this algorithm, a set of feasible routes is generated by a large neighbourhood search algorithm. Then a set partitioning formulation is used to determine the routing decisions. Moussavi et al. (2019) developed a decomposition based matheuristic algorithm for assignment and routing problems in the HHC. The algorithm includes three phases; the required number of caregivers is determined in the first phase. The set of services that must be done in each working day is obtained in the second phase and in the third phase the assignment and routing decisions are determined. Cinar et al. (2019) proposed a model for determining a priority for visiting patients and routing decisions by considering a single nurse in home healthcare. Furthermore, a matheuristic algorithm and an adaptive large neighbourhood search (ALNS) algorithm were developed. In their proposed matheuristic algorithm, the main model was decomposed based on each working day. Matheuristic algorithms were also developed for other problems such as a Shift Scheduling Problem (Hernández-Leandro et al., 2019), a Driver Scheduling Problem (Perumal et al., 2018), staff dimensioning in an emergency department (Andersen et al., 2019), pricing problem (Oliveira et al., 2019), a patient to bed assignment problem (Guido et al., 2018), and a vehicle reservation assignment (Oliveira et al., 2014).

Based on our literature review, the contribution of this study can be stated as follows: we propose a model to consider human resource planning decisions (districting, staff dimensioning, assignment, and routing) in HHC, simultaneously. Based on the reviewed papers in Tables 1 and 2, our study is the first one that considers these decisions simultaneously. We treat travel and service times as uncertain parameters and handle them by a two-stage stochastic programming, and consider the factors that affect the patients' satisfaction such as continuity of care. We propose a new mathematical model-based heuristic algorithm to solve the problem. Four novel intermediate models (Int.M1, Int.M2, Int.M3, and Int.M4) are proposed and used in this algorithm to decrease the computational time of the proposed matheuristic algorithm. Two clustering based models, Int.M1 and Int.M2, are used in the first and second phases of the matheuristic algorithm for dimension reduction and generating initial solution, respectively. An algorithm based on progressive hedging and Frank and Wolf algorithms (PH-FW) is used to decrease the computational time of the Int.M2 model. Also, a heuristic approach is developed to speed up the convergence of the PH-FW. A fix and optimize procedure is used in the third phase of the algorithm to repair any infeasibilty, and a novel intermediate model (Int.M3) is proposed. In the last phase of the proposed algorithm, the obtained solutions for the routing and assignment

decisions are improved by using a fix and optimize method.

3. Problem statement

3.1. Problem description

The stochastic districting−staff dimensioning−assignment−routing problem (SDDARP) is outlined as follows. A set of caregivers, a set of services to patients in their homes, and the basic units are given. Each basic unit is defined as a set of patients and caregivers. Basic units are partitioned into a number of districts based on routing, the costs to hire the caregivers, and the caregivers' and patients' satisfaction. The first goal of the SDDARP is to determine the set of caregivers and patients in each district by considering travel and service times as uncertain parameters. The caregivers in each district can only be assigned to the patients in that district, and the tour for each caregiver starts and ends at his/her home. Caregivers have a maximum working time, and continuity of care and similarity between the type of care required for the patients and the skills of the caregivers are taken into account in the SDDARP model. The problem is illustrated in Figure 2. This figure shows ten basic units, which are partitioned into four districts. Two caregivers are assigned to District 4. The patients in each district are visited only by the caregivers who live in the same district. The routes of each caregiver in a working day is presented in this figure. Each route can only start and finish at the caregiver's home, and sample tours for visiting patients are shown in this figure.
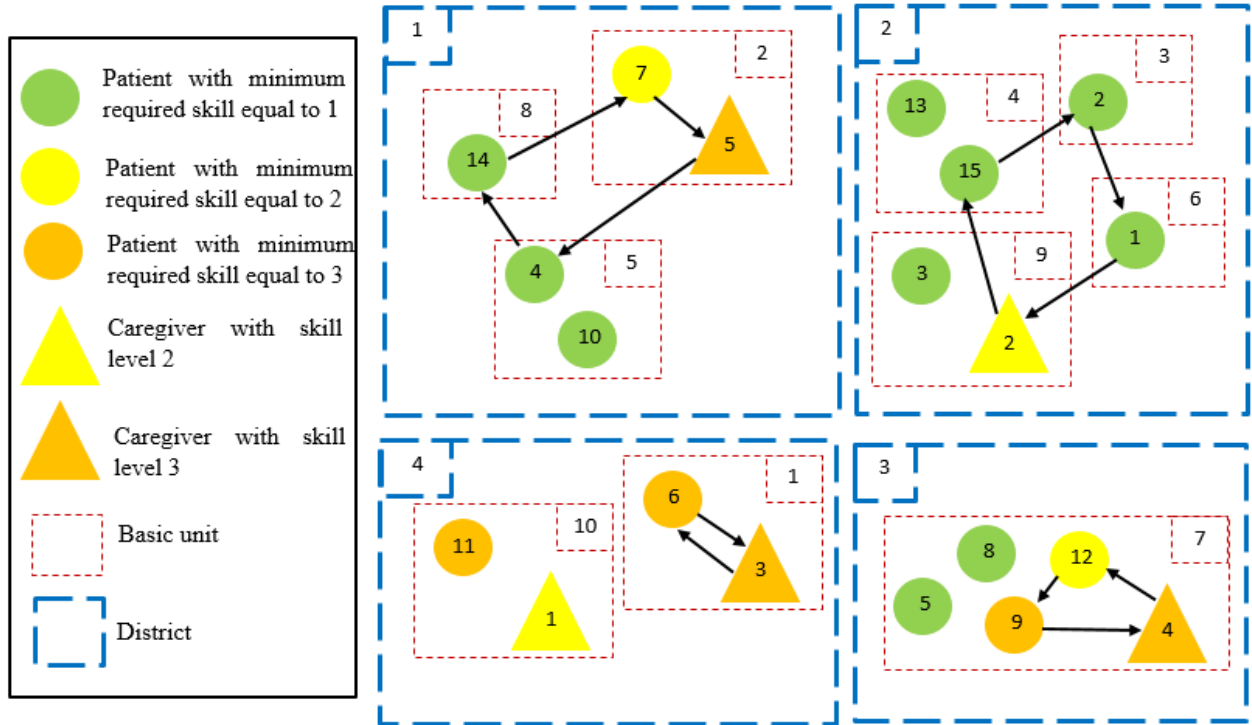

Figure 2: A schematic result of a districting-staff dimensioning-assignment-routing problem with four districts.

3.2. Problem formulation

In this section, the proposed mathematical model for districting, based on staff dimensioning and routing costs, is presented. In the proposed model, M is defined as the maximum number of districts, and K is the number of caregivers. This problem is defined by a graph $G(\mathcal{V}, \mathcal{A})$, where $\mathcal{V} = \{ 1, 2, ..., N, N+1, N+2, ..., N+K\}$ is the set of nodes and $\mathcal{A} = \{(i, j)|i, j \in \mathcal{V}, i \neq j\}$ is the set of arcs. $\mathcal{N} = \{ 1, 2, ..., N\}$ is the patient set, which determines the home location of each patient, and the set $\mathcal{N}_c = \{ N+1, N+2, ..., N+K\}$ is used to define the locations of the caregivers. Therefore, $\mathcal{N}_c$ defines the locations of the start and end point of the tours. The assumptions in the proposed model are as follows:

- The number of patients served by the HHC system is known and is not updated during the planning horizon.

- Each basic unit must be assigned to exactly one district.

- The caregivers in each district can only be assigned to patients in that district.

- The start and end points for each caregiver are his/her home.

- Patients in a given basic unit require different types of services. The lowest skill level that the caregiver should have to perform the required service for each patient is determined by assigning a number from the set $\{1, 2, \ldots, R\}$ to the caregiver.

- Patients need different number of visits in the planning horizon, based on their type of services. The schedule of visiting patients is determined by the mathematical model.

- Each patient can be visited at most once a day.

- Each caregiver can serve a limited number of patients. On the other hand, the working time of each caregiver is less than a predetermined value.

- Caregivers have different levels of skills, and they can serve patients if the required service matches their skills. The level of skill for each caregiver is determined by assigning a number from the set $\{1, 2, \ldots, Q\}$ $(Q \geq R)$ to the caregiver, where 1 and Q denote the lowest and highest level of skills, respectively. In addition, a caregiver can serve a patient when the minimum required level of skill for that patient is equal to or lower than her/his skill level. Therefore if the level of skill for a caregiver is equal to 2, patients with required services in levels 1 or 2 can be served by the caregiver.

- The optimal number of districts in the design is unknown in advance and is determined by the model. However, the maximum number of districts M is known and is defined by $M = \min (K,L)$. Where K is the number of available caregivers and L was defined above, as the number of basic units.

- Two basic units are incompatible if traveling between them is difficult or impossible. This may be because of geographical obstacles. This assumption is called "incompatibility" and considered as a hard constraint.

- The maximum allowable travel time between two basic units in a district is considered. This is called "neighbouring" and considered as a soft constraint. Two basic units are neighbours if the maximum travel time between them is less than a predefined value.

- Continuity of care is considered.

- Districting and staff dimensioning are strategic and tactical decisions and are fixed for a long time period.

- Travel and service times are considered as uncertain parameters and their values are determined for each scenario.

The sets, parameters, and decision variables of the proposed model are as follows:

Sets:

$\mathcal{V}$	Set of all nodes (patients and caregivers), $\mathcal{V} = \{ 1, 2, ..., N, N+1, N+2, ..., N+K \}$.
$\mathcal{D}$	Set of working days, $\mathcal{D} = \{ 1, 2, ..., D \}$.
$\mathcal{L}$	Set of basic units, $\mathcal{L} = \{ 1, 2, ..., L \}$.
$\mathcal{L}_l^c$	Set of basic units that are incompatible with basic unit l ($l \in \mathcal{L}$).
$\mathcal{L}_l^N$	Set of basic units that are not in the neighbourhood of lth ($l \in \mathcal{L}$) basic unit ($tr_{ll'} > td$).
$\mathcal{M}$	Set of potential districts, $\mathcal{M} = \{ 1, 2, ..., M \}$.
$\mathcal{K}$	Set of caregivers, $\mathcal{K} = \{ 1, 2, ..., K \}$.
$\mathcal{N}$	Set of patients, $\mathcal{N} = \{ 1, 2, ..., N \}$.
$\mathcal{N}_c$	Set of residential nodes of caregivers, $\mathcal{N}_c = \{ N+1, N+2, ..., N+K \}$.
$\mathcal{N}_c^k$	Set of residential node of the kth caregiver (The start and end points of the route corresponding to the kth caregiver).
Ξ	Set of scenarios, $\Xi = \{1, 2, ..., S\}$.

Parameters:

r_i	The lowest required skill level that a caregiver should have to serve the ith ($i \in \mathcal{N}$) patient.
q_k	The skill level of the kth ($k \in \mathcal{K}$) caregiver.
h_k	Hiring cost of the kth ($k \in \mathcal{K}$) caregiver.
v_{li}	1 if the ith ($i \in \mathcal{V}$) node is located at basic unit l ($l \in \mathcal{L}$), 0 otherwise.
λ_i	Number of required visits for the ith ($i \in \mathcal{N}$) patient.
$st_i(\xi)$	Service time for the ith ($i \in \mathcal{N}$) patient under scenario ($\xi \in \Xi$).
$tt_{ij}(\xi)$	Travel time between the ith and the jth ($i, j \in \mathcal{V}$) nodes under scenario ($\xi \in \Xi$).

11

$P(\xi)$ Probability of scenario ξ ($\xi \in \Xi$).

b_i Latest allowable servicing start time of ith ($i \in \mathcal{N}$) patient.

M_{max} A large number, $M_{max} = N + K$.

T_{max} A large number, $T_{max} = \max(b_i)$.

$tr_{ll'}$ The mean value of maximum travel time between the basic units l and l', $(l, l' \in \mathcal{L})$.

$$tr_{ll'} = \sum_{\xi \in \Xi} p(\xi) \times \max_{\substack{i \in \mathcal{V} \\ j \in \mathcal{V}}} \{ tt_{ij}(\xi) \times v_{li} \times v_{l'j} \}.$$

td Maximum allowable travel time between two basic units in the same district.

φ Maximum working time for each caregiver.

cc Maximum number of caregivers who can visit a patient in the planning horizon.

θ Transportation cost per unit of travel time.

α Cost of per unit violation in neighbouring criterion.

Decision variables:

$x^{dk}_{ijm}(\xi)$ 1 if caregiver k ($k \in \mathcal{K}$) travels along (i, j) ($i, j \in \mathcal{V}$) in district m ($m \in \mathcal{M}$) in day d ($d \in \mathcal{D}$) under scenario ξ ($\xi \in \Xi$), 0 otherwise.

y_{im} 1 if patient i ($i \in \mathcal{V}$) is assigned to district m ($m \in \mathcal{M}$), 0 otherwise.

u_{mk} 1 if caregiver k ($k \in \mathcal{K}$) is hired in district m ($m \in \mathcal{M}$), 0 otherwise.

$a_{ik}(\xi)$ 1 if patient i ($i \in \mathcal{N}$) is assigned to caregiver k ($k \in \mathcal{K}$) under scenario ξ ($\xi \in \Xi$), 0 otherwise.

z_{lm} 1 if basic unit l ($l \in \mathcal{L}$) is assigned to district m ($m \in \mathcal{M}$), 0 otherwise.

$t_{idk}(\xi)$ The time that servicing starts at the ith ($i \in \mathcal{N}$) patient by the kth ($k \in \mathcal{K}$) caregiver in day d ($d \in \mathcal{D}$) under scenario ξ ($\xi \in \Xi$).

$\pi_{ll'}$ 1 if basic units l and l' ($l, l' \in \mathcal{L}$) are assigned to the same district, 0 otherwise.

$w_{imkd}(\xi)$ 1 if node i (patient or caregiver) in district m ($m \in \mathcal{M}$) is assigned to the kth route (route corresponding to the kth caregiver) ($k \in \mathcal{K}$) in day d ($d \in \mathcal{D}$) under scenario ξ ($\xi \in \Xi$), 0 otherwise.

Here, a two-stage stochastic programming model is proposed for SDDARP when travel and service times are uncertain. Districting and staff dimensioning are strategic and tactical decision variables, respectively, and assignment and routing are operational decision variables. Therefore, districting and staff dimensioning are the first stage decision variables, and assignment and routing are defined as the second stage decision variables. The objective function consists of three terms, including the costs of hiring and of violating the neighbouring criterion, as well as the expected value of the transportation cost. The proposed two-stage stochastic programming model is defined as follows:

$$(\textbf{SDDARP}) \quad \min \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} h_k \times u_{mk} + \alpha \times \sum_{l \in \mathcal{L}} \sum_{l' \in \mathcal{L}_l^N} (tr_{ll'} - td) \times \pi_{ll'}$$

$$+ \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{V}} \sum_{\substack{j \in \mathcal{V} \\ j \neq i}} \sum_{d \in \mathcal{D}} \sum_{k \in \mathcal{K}} \sum_{\xi \in \Xi} \theta \times p(\xi) \times tt_{ij}(\xi) \times x_{ijm}^{dk}(\xi) \tag{1}$$

$$\text{s.t.:} \quad \sum_{m \in \mathcal{M}} z_{lm} = 1, \qquad\qquad \forall l \in \mathcal{L} \tag{2}$$

$$z_{lm} + z_{l'm} \leq 1, \qquad\qquad \forall l \in \mathcal{L}, l' \in \mathcal{L}_l^c, \ m \in \mathcal{M} \tag{3}$$

$$z_{lm} + z_{l'm} - 1 \leq \pi_{ll'}, \qquad\qquad \forall l \in \mathcal{L}, l' \in \mathcal{L}_l^N, \ m \in \mathcal{M} \tag{4}$$

$$\sum_{l \in \mathcal{L}} z_{lm} \times v_{li} = y_{im}, \qquad\qquad \forall i \in \mathcal{V}, m \in \mathcal{M} \tag{5}$$

$$M_{max} \sum_{k \in \mathcal{K}} u_{mk} \geq \sum_{i \in \mathcal{N}} y_{im}, \qquad\qquad \forall m \in \mathcal{M} \tag{6}$$

$$u_{mk} \leq \sum_{i \in \mathcal{N}} y_{im}, \qquad\qquad \forall k \in \mathcal{K}, m \in \mathcal{M} \tag{7}$$

$$\sum_{m \in \mathcal{M}} \sum_{d \in \mathcal{D}} \sum_{k \in \mathcal{K} | r_i \leq q_k} w_{imdk}(\xi) = \lambda_i, \qquad\qquad \forall i \in \mathcal{N}, \xi \in \Xi \tag{8}$$

$$\sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} w_{imkd}(\xi) \leq 1, \qquad\qquad \forall d \in \mathcal{D}, \ i \in \mathcal{V}, \ \xi \in \Xi \tag{9}$$

$$\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{V}} \sum_{m \in \mathcal{M}} st_i(\xi) x_{ijm}^{dk}(\xi) \leq \varphi, \qquad\qquad \forall d \in \mathcal{D}, \ k \in \mathcal{K}, \ \xi \in \Xi \tag{10}$$

$$2w_{imkd}(\xi) \leq y_{im} + y_{jm}, \qquad\qquad \forall d \in \mathcal{D}, i \in \mathcal{V}, j \in \mathcal{N}_c^k, k \in \mathcal{K}, m \in \mathcal{M}, \xi \in \Xi \tag{11}$$

$$w_{imkd}(\xi) \leq u_{mk}, \qquad\qquad \forall i \in \mathcal{N}, \ m \in \mathcal{M}, \ k \in \mathcal{K}, \ d \in \mathcal{D}, \ \xi \in \Xi \tag{12}$$

$$\sum_{i \in \mathcal{N}} w_{imkd}(\xi) \leq M_{max} w_{jmkd}(\xi), \qquad\qquad \forall j \in \mathcal{N}_c^{\ k}, \ k \in \mathcal{K}, \ d \in \mathcal{D}, \ \xi \in \Xi, \ m \in \mathcal{M} \tag{13}$$

$$2x_{ijm}^{dk}(\xi) \leq y_{im} + y_{jm}, \qquad\qquad \forall d \in \mathcal{D}, i \in \mathcal{V}, j \in \mathcal{V}, i = j, k \in \mathcal{K}, m \in \mathcal{M}, \xi \in \Xi \tag{14}$$

$$\sum_{j \in \mathcal{V}} x_{ijm}^{dk}(\xi) = w_{imkd}(\xi), \qquad\qquad \forall d \in \mathcal{D}, \ i \in \mathcal{N}, \ k \in \mathcal{K}, \ m \in \mathcal{M}, \ \xi \in \Xi \tag{15}$$

$$\sum_{j \in \mathcal{N}} x_{ijm}^{dk}(\xi) = w_{imkd}(\xi), \qquad\qquad \forall d \in \mathcal{D}, \ i \in \mathcal{N}_c^k, \ k \in \mathcal{K}, \ m \in \mathcal{M}, \ \xi \in \Xi \tag{16}$$

$$\sum_{j \in \mathcal{N}} x_{jim}^{dk}(\xi) = w_{imkd}(\xi) \qquad\qquad \forall d \in \mathcal{D}, \ i \in \mathcal{N}_c^k, \ k \in \mathcal{K}, \ m \in \mathcal{M}, \ \xi \in \Xi \tag{17}$$

$$\sum_{i \in \mathcal{V}} \sum_{m \in \mathcal{M}} x_{ijm}^{dk}(\xi) = \sum_{i \in \mathcal{V}} \sum_{m \in \mathcal{M}} x_{jim}^{dk}(\xi), \qquad \forall d \in \mathcal{D}, j \in \mathcal{N}, k \in \mathcal{K}, \xi \in \Xi \tag{18}$$

$$t_{jdk}(\xi) \geq t_{idk}(\xi) + st_i(\xi) + tt_{ij}(\xi) - T_{max} \left(1 - x_{ijm}^{dk}(\xi)\right),$$

$$\forall d \in \mathcal{D}, \ i \in \mathcal{V}, \ j \in \mathcal{N}, \ k \in \mathcal{K}, \ m \in \mathcal{M}, \ \xi \in \Xi \tag{19}$$

$$t_{idk}(\xi) \leq b_i, \qquad\qquad \forall d \in \mathcal{D}, \ i \in \mathcal{N}, \ k \in \mathcal{K}, \ \xi \in \Xi \tag{20}$$

$$a_{ik}(\xi) \geq \sum_{m \in \mathcal{M}} w_{imkd}(\xi), \qquad\qquad \forall i \in \mathcal{N}, \; k \in \mathcal{K}, \; d \in \mathcal{D}, \; \xi \in \Xi \tag{21}$$

$$\sum_{k \in \mathcal{K}} a_{ik}(\xi) \leq cc, \qquad\qquad \forall i \in \mathcal{N}, \; \xi \in \Xi \tag{22}$$

$$z_{lm}, \; y_{im}, \; u_{mk}, \; x^{dk}_{ijm}(\xi), \; w_{imkd}(\xi), \pi_{ll'} \in \{0,1\},$$

$$\forall l, l' \in \mathcal{L}, \; i, j \in \mathcal{V}, \; m \in \mathcal{M}, \; k \in \mathcal{K}, \; d \in \mathcal{D}, \; \xi \in \Xi \tag{23}$$

$$a_{ik}(\xi) \in \{0,1\}, \qquad\qquad \forall i \in \mathcal{N}, \; k \in \mathcal{K}, \; \xi \in \Xi \tag{24}$$

$$t_{idk}(\xi) \geq 0, \qquad\qquad \forall d \in \mathcal{D}, \; i \in \mathcal{V}, \; k \in \mathcal{K}, \; \xi \in \Xi \tag{25}$$

The objective function (1) minimizes the hiring, violation of the neighbouring criterion, and transportation costs, defined as a function of travel time. The constraints of the SDDARP model are classified into five groups including constraints related to partitioning the basic units into districts, hiring caregivers, assigning patients to caregivers, routing, and continuity of care. The constraints related to each class are defined here.

The constraints related to partitioning basic units to districts are as follows:

Constraints (2) guarantee that each basic unit is assigned to exactly one district. Constraints (3) guarantee the compatibility and maximum allowable distance between two basic units in a district. Constraints (4) determine that the basic units in terms of which violations of the neighboring criterion are defined. Constraints (5) state that each patient can be assigned to exactly one district, and this district must be the same as the district assigned to that patient's basic unit.

The constraints related to hiring caregivers are described as follows:

Constraints (6) ensure that if there are any patients in a district, at least one caregiver is assigned to that district. Constraints (7) state that caregivers are not hired for a district with no patients.

The constraints related to assigning patients to caregivers are explained in the following text:

Constraints (8) ensure that the required visits for each patient are scheduled in the planning horizon for different caregivers, considering the consistency between the services required for the patient and the skill levels of the caregivers. Constraints (9) guarantee that each patient is visited by at most one caregiver in each day. Constraints (10) guarantee that the working time of each caregiver is less than the maximum value for the working time of a caregiver. Constraints (11) ensure that a node can be assigned to a caregiver if there is a caregiver in the same district. Constraints (12) state that a patient can be assigned to hired caregivers. Constraints (13) ensure that a caregiver can be assigned to patients in a day if she/he leaves her/his home in that day.

The constraints related to routing decisions are as follows:

Constraints (14) guarantee that each node (patient and caregiver) j can be visited as a successor of node (patient and caregiver) i in a route if they are assigned to the same district. Constraints (15) state that if the ith patient is assigned to the kth caregiver in day d, she/he must be visited in the route of this caregiver on that day. Constraints (16) and (17) ensure that the start and end points of the kth route in the mth

district are from the kth caregiver's home if the kth caregiver is hired in this district. Constraints (18) are flow conservation constraints. Constraints (19) are sub-tour elimination constraints. Constraints (20) are time window constraints.

The constraints related to continuity of care are introduced here:

Constraints (21) determine the caregivers who are assigned to each patient. Constraints (22) guarantee continuity of care. Constraints (23) to (25) define the domain of the decision variables.

4. Multi-phase matheuristic algorithm

To solve the problem efficiently, a matheuristic algorithm is proposed, called Math-1. The proposed algorithm includes four main phases as depicted in Figure 3. In the first phase, the problem size is reduced by decreasing the potential number of districts ($\mathcal{M}$) and determining the proper sequence for the patient visits for each caregiver under each scenario. In the second phase, initial solutions are obtained for the different numbers of potential districts determined in the previous phase, taking the estimated routing and assignment costs into account, by using an intermediate model for the proposed algorithm. A progressive hedging based algorithm is employed to solve the intermediate model in the second phase. The feasibility of the solution obtained in the previous phase is examined in the third phase, then a procedure based on fix and optimize is used to recover the feasibility. In the last phase of the algorithm, the obtained solution is improved by using a fix and optimize method. The pseudo code of the proposed matheuristic algorithm is presented in algorithm 1.

Intermediate mathematical models (Int.M1, Int.M2, Int.M3, and Int.M4) are developed to be used among the phases of the algorithm, so additional sets, parameters, and decision variables in the models are defined here.
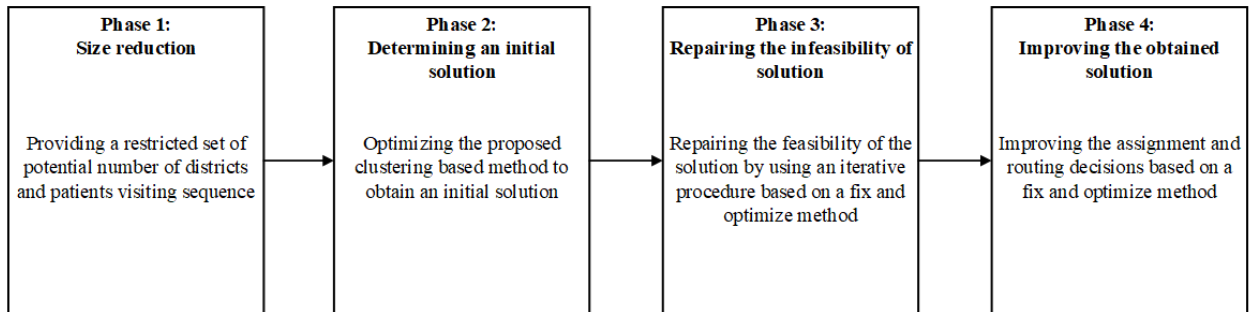

Figure 3: Overview of the proposed matheuristic algorithm (Math-1).

Sets:

$\mathcal{A}_{ik}(\xi)$ Set of nodes (caregivers and patients) that are visited after the ith $(i \in \mathcal{V})$ node in the obtained tour in Phase 1 of Math-1 algorithm for the kth $(k \in \mathcal{K})$ caregiver under scenario ξ $(\xi \in \Xi)$.

$\mathcal{B}_{ik}(\xi)$ Set of nodes (caregivers and patients) that are visited before the ith $(i \in \mathcal{V})$ node in the obtained tour in Phase 1 of of Math-1 algorithm for the kth $(k \in \mathcal{K})$ caregiver under scenario ξ $(\xi \in \Xi)$.

$\mathcal{G}_g$ Set of districts used in the first, second and third phases of matheuristic algorithm in the intermediate models Int.M1, Int.M2 and Int.M3.

L_{fe} Set of basic units assigned to the feasible district used in the third phase of matheuristic algorithm in the intermediate model Int.M3.

I_{fe} Set of nodes (patients and caregivers) assigned to the feasible district used in the third phase of matheuristic algorithm in the intermediate model Int.M3.

Parameters:

$\psi_{lm}^{1iter}(\xi)$ Weight matrix for the decision variables $z_{lm}(\xi)$ in the PH-FW algorithm at the $iter^{th}$ iteration.

$\psi_{mk}^{2iter}(\xi)$ Weight matrix for the decision variables $u_{mk}(\xi)$ in the PH-FW algorithm at the $iter^{th}$ iteration.

$\psi_{ll'm}^{3iter}(\xi)$ Weight matrix for the decision variables $_{ll'm}(\xi)$ in the PH-FW algorithm at the $iter^{th}$ iteration.

ρ_1 Penalty parameter for decision variables $z_{lm}(\xi)$ in the PH-FW algorithm.

ρ_2 Penalty parameter for decision variables $u_{mk}(\xi)$ in the PH-FW algorithm.

ρ_3 Penalty parameter for decision variables $_{ll'm}(\xi)$ in the PH-FW algorithm.

ε A threshold value for tolerance of solution obtained by the PH-FW algorithm that determines the stopping condition.

Decision variables:

$x_{ij}^{dk}(\xi)$ 1 if caregiver k $(k \in \mathcal{K})$ travels along (i, j) $(i, j \in \mathcal{V})$ in day d $(d \in \mathcal{D})$ under scenario ξ $(\xi \in \Xi)$, 0 otherwise.

$\mu_{ll'm}$ 1 if basic units l and l' $(l, l' \in \mathcal{L})$ are assigned to the mth district.

$w_{ikd}(\xi)$ 1 if node i $(i \in \mathcal{V})$ is assigned to the kth route (route corresponding to the kth caregiver) $(k \in \mathcal{K})$ in day d $(d \in \mathcal{D})$ under scenario ξ $(\xi \in \Xi)$, 0 otherwise.

F The objective value of obtained solution.

4.1. Phase 1: Dimension reduction

As mentioned previously, the maximum number of potential districts (M) is obtained on the basis of the numbers of basic units and caregivers $(M = min(K, L))$, so the number of potential districts must lie

---

Algorithm 1: The framework of Math-1 algorithm.

Input: Parameters of the model

Phase 1: Dimension reduction

for $k \in \mathcal{K}$ **do**

    for $\xi \in \Xi$ **do**

        Determine a preliminary reasonable patients visiting sequence by solving a Travelling
Salesman Problem (TSP) model considering all patients. Then define the sets $\mathcal{A}_{ik}(\xi)$ and
$\mathcal{B}_{ik}(\xi)$ for each combination of patients and caregivers under each scenario;

Construct g_{rest} empty vector for potential number of districts;

for $g \in \{2, 3, ..., M\}$ **do**

    $\mathcal{G}_g \leftarrow \{1, 2, ..., g\}$;

    Partition the basic units into g districts by optimizing Int.M1(g) model;

    if *The proposed model is feasible* **then**

        $g_{rest} \leftarrow g_{rest} \cup g$

$F_{best} \leftarrow \infty$;

$condition \leftarrow false$;

g $\leftarrow 2$

while *condition is false* **do**

    $\mathcal{G}_g \leftarrow \{1, 2, ..., g\}$;

    Phase 2: Determining an initial solution

    Optimize the Int.M2(g) according to sets $\mathcal{A}_{ik}(\xi)$ and $\mathcal{B}_{ik}(\xi)$ then determine an initial solution
$(\bar{z}_{lm}, \bar{u}_{mk}, \hat{w}_{ikd}(\xi)$ and $route_{kd}(\xi))$ when basic units are partitioned into g districts;

    if *Time window constraints are violated for at least one patient* **then**

        Phase 3: Repairing infeasible solution;

        condition2 $\leftarrow false$

        Algorithm 4 is used to repair infeasible solution;

    else

        Feasibility $\leftarrow true$

    if *Feasibility is true* **then**

        Phase 4: Improving the obtained solution;

        Algorithm 5 is used to improve the obtained solution;

        $F(g) \leftarrow$ The objective value of the obtained solution;

        if $F(g) \leq F_{best}$ **then**

            $F_{best} \leftarrow F(g)$;

            $g \leftarrow g + 1$;

        else

            $condition \leftarrow true$;

    else

        $g \leftarrow g + 1$;

    if $g > |g_{rest}|$ **then**

        $condition \leftarrow true$;

Result: F_{best}

$\mathcal{M} = \{1, 2, ..., M\}$. Because of the given compatibility assumption, the SDDARP model might be infeasible for some numbers of potential districts. Therefore, numbers of potential districts that make the SDDARP model infeasible should be removed from the set. To do this, an intermediate model (called Int.M1) is proposed to make it easier to assess the infeasibility of the SDDARP. This model is solved for different numbers of clusters $g = 2, 3, \ldots, M$. Therefore, the proposed model is solved $M - 1$ times, and the set of districts is defined as $\mathcal{G}_g = \{1, \ldots, g\}$. Finally, a restricted feasible set of the number of districts (g_{rest}) is extracted to be used in the next phases. The proposed Int.M1 model is as follows:

$$(\textbf{Int.M1 (g)}) \quad Min \quad \sum_{l \in \mathcal{L}} \sum_{l' \in \mathcal{L}} \theta \times tr_{ll'} \times \pi_{ll'} + \alpha \times \sum_{l \in \mathcal{L}} \sum_{\mathcal{L}_l^N} (tr_{ll'} - td) \times \pi_{ll'} \tag{26}$$

$$\text{s.t.: } (3),$$

$$\sum_{m \in \mathcal{G}_g} z_{lm} = 1, \qquad\qquad \forall l \in \mathcal{L} \tag{27}$$

$$\sum_{l \in L} z_{lm} \geq 1, \qquad\qquad \forall m \in \mathcal{G}_g \tag{28}$$

$$z_{lm} + z_{l'm} - \pi_{ll'} \leq 1, \qquad\qquad \forall l, l' \in \mathcal{L}, l = l', m \in \mathcal{G}_g \tag{29}$$

$$\sum_{j \in \mathcal{N}_c^k} y_{jm} \geq y_{im}, \qquad\qquad \forall i \in \mathcal{N} | r_i \leq q_k, m \in \mathcal{G}_g, k \in \mathcal{K} \tag{30}$$

$$z_{lm}, y_{im}, \pi_{ll'} \in \{0,1\} \qquad\qquad \forall i \in \mathcal{V}, l, l' \in \mathcal{L}, m \in \mathcal{G}_g \tag{31}$$

The objective function (26) minimizes the maximum travel cost between the basic units in the same district and cost of violation in neighbouring constraint. Constraints (27) guarantee that each basic unit is assigned to one district. Constraints (28) ensure that each district consists of at least one basic unit. Constraints (29) guarantee that two basic units can belong to a district if they are assigned to the same district. Constraints (30) ensure that a patient can be assigned to a district if a caregiver with the required skill has been assigned to the same district. Constraints (31) determine the domain of the intermediate model (Int.M1) decision variables.

In another step of the current phase, which is based on the algorithm proposed by Solyalı and Süral (2017), a Traveling Salesman Problem (TSP) is solved for each caregiver over all patients in all basic units under each scenario. Caregivers' locations are assumed to be the start and end points of their corresponding tours. On the other hand, the TSP model is solved $S \times K$ times and an initial sequence of patient visits for each caregiver is obtained. The succeeding and preceding nodes of each caregiver and patient (i) in an achieved tour (k) under each scenario (ξ) define the sets $\mathcal{A}_{ik}(\xi)$ and $\mathcal{B}_{ik}(\xi)$, respectively. An example is presented to clarify the definition of these sets. Two caregivers, five patients, and one scenario are considered, and it is assumed that the tours obtained for the first and second caregivers are 6-1-3-2-4-5-6 and 7-3-2-5-1-4-7, respectively. The sets obtained are shown in Table 3. In this example, nodes 6 and 7 are the first and second

caregivers' locations, respectively.

By considering the sets $\mathcal{A}_{ik}(\xi)$ and $\mathcal{B}_{ik}(\xi)$, the size of the problem in phase 2 is decreased, because sub-tour elimination constraints are removed. Sub-tour elimination constraints are known as constraints that make a problem harder to solve, and the number of these constraints increases as the number of patients and caregivers increases. Therefore, using these sets will decrease the computational time of the proposed clustering based model in phase 2.

Table 3: The sets $\mathcal{A}_{ik}(1)$ and $\mathcal{B}_{ik}(1)$ based on the tour obtained by solving the TSP in the illustrative example.

i	$\mathcal{A}_{i1}(1)$	$\mathcal{B}_{i1}(1)$	$\mathcal{A}_{i2}(1)$	$\mathcal{B}_{i2}(1)$
1	{3, 2, 4, 5, 6}	{6}	{4,7}	{7, 3, 2, 5}
2	{4, 5, 6}	{6, 1, 3}	{5, 1, 4, 7}	{7, 3}
3	{2, 4, 5, 6}	{6, 1}	{2, 5, 1, 4, 7}	{7}
4	{5, 6}	{6, 1, 3, 2}	{7}	{7, 3, 2, 5, 1}
5	{6}	{6, 1, 3, 2, 4}	{1, 4, 7}	{7, 3, 2}
6	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}
7	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}

4.2. Phase 2: Determining an initial solution

In this section, a clustering based intermediate model (Int.M2) is proposed, in which the potential visiting sequence for patients obtained in the previous phase is entered in the models. The sets $\mathcal{A}_{ik}(\xi)$ and $\mathcal{B}_{ik}(\xi)$ are used in the proposed model to determine the domain of the indices i and j in the routing constraints. This model is solved for different numbers of districts based on the restricted set for the potential number of districts ($g \in g_{rest}$). Therefore, the clustering based model is solved $|g_{rest}|$ times. In this model, clustering is performed based on minimizing the costs of hiring, violating the neighbouring constraint, and traveling. Then the basic units are partitioned into g districts. Determining an initial solution is the goal of this phase. The decision variables $x_{ijm}^{dk}(\xi)$ and $w_{imkd}(\xi)$ are replaced by $x_{ij}^{dk}(\xi)$ and $w_{ikd}(\xi)$, respectively in the model by adding new variables $\mu_{ll'm}$ to reduce the number of binary variables. The Int.M2 model is presented as follows:

$$(\textbf{Int.M2(g)}) \quad \min \quad \sum_{m \in \mathcal{G}_g} \sum_{k \in \mathcal{K}} h_k \times u_{mk} + \alpha \times \sum_{m \in \mathcal{G}_g} \sum_{l \in \mathcal{L}} \sum_{l' \in \mathcal{L}_l^N} (tr_{ll'} - td) \times \mu_{ll'm}$$

$$+ \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{A}_{ik}(\xi)} \sum_{d \in \mathcal{D}} \sum_{k \in \mathcal{K}} \sum_{\xi \in \Xi} \theta \times p(\xi) \times tt_{ij}(\xi) \times x_{ij}^{dk}(\xi) \tag{32}$$

$$\text{s.t.:} \quad \sum_{m \in \mathcal{G}_g} z_{lm} = 1, \qquad \forall l \in \mathcal{L} \tag{33}$$

19

$$\sum_{l\in\mathcal{L}} z_{lm} \geq 1, \qquad\qquad \forall m \in \mathcal{G}_g \tag{34}$$

$$z_{lm} + z_{l'm} \leq 1, \qquad\qquad \forall l \in \mathcal{L}, l' \in \mathcal{L}_l^c,\ m \in \mathcal{G}_g \tag{35}$$

$$\sum_{k\in\mathcal{K}} u_{mk} \geq 1, \qquad\qquad \forall m \in \mathcal{G}_g \tag{36}$$

$$\sum_{\substack{k\in\mathcal{K}|\ d\in\mathcal{D} \\ r_i\leq q_k}} w_{ikd}(\xi) = \lambda_i, \qquad\qquad \forall i \in \mathcal{N},\ \xi \in \Xi \tag{37}$$

$$\sum_{k\in\mathcal{K}} w_{ikd}(\xi) \leq 1, \qquad\qquad \forall d \in \mathcal{D},\ i \in \mathcal{V},\ \xi \in \Xi \tag{38}$$

$$\sum_{i\in N} st_i w_{ikd}(\xi) \leq \varphi, \qquad\qquad \forall d \in \mathcal{D},\ k \in \mathcal{K},\ \xi \in \Xi \tag{39}$$

$$2\mu_{ll'm} \leq y_{lm} + y_{l'm}, \qquad\qquad \forall l,l' \in \mathcal{L},\ l = l',\ m \in \mathcal{G}_g \tag{40}$$

$$\mu_{ll'm} \geq y_{lm} + y_{l'm} - 1, \qquad\qquad \forall l,l' \in \mathcal{L},\ l = l',\ m \in \mathcal{G}_g \tag{41}$$

$$w_{ikd}(\xi) \leq \sum_{m\in\mathcal{G}_g} \mu_{ll'm}\nu_{li}\nu_{l'(N+k)}, \qquad\qquad \forall i \in \mathcal{N},\ k \in \mathcal{K},\ l,l' \in \mathcal{L},\ d \in \mathcal{D},\ \xi \in \Xi \tag{42}$$

$$w_{ikd}(\xi) \leq \sum_{m\in\mathcal{G}_g} u_{mk}, \qquad\qquad \forall k \in \mathcal{K},\ d \in \mathcal{D},\ i \in \mathcal{N},\ \xi \in \Xi \tag{43}$$

$$\sum_{i\in\mathcal{N}} w_{ikd}(\xi) \leq M_{\max} w_{jkd}(\xi), \qquad\qquad \forall j \in \mathcal{N}_c^k,\ k \in \mathcal{K},\ d \in \mathcal{D},\ \xi \in \Xi \tag{44}$$

$$x_{ij}^{dk}(\xi) \leq \sum_{m\in\mathcal{G}_g} \mu_{ll'm}\nu_{li}\nu_{l'j}, \qquad\qquad \forall i \in \mathcal{V},\ j \in \mathcal{V},\ k \in \mathcal{K},\ d \in \mathcal{D},\ l,l' \in \mathcal{L},\ \xi \in \Xi \tag{45}$$

$$\sum_{j\in\mathcal{A}_{ik}(\xi)} x_{ij}^{dk}(\xi) = w_{ikd}(\xi), \qquad\qquad \forall i \in \mathcal{N},\ k \in \mathcal{K},\ d \in \mathcal{D}, \xi \in \Xi \tag{46}$$

$$\sum_{j\in\mathcal{B}_{ik}(\xi)} x_{ji}^{dk}(\xi) = w_{ikd}(\xi), \qquad\qquad \forall i \in \mathcal{N},\ k \in \mathcal{K},\ d \in \mathcal{D}, \xi \in \Xi \tag{47}$$

$$\sum_{j\in\mathcal{A}_{ik}(\xi)} x_{ij}^{dk}(\xi) = w_{ikd}(\xi), \qquad\qquad \forall d \in \mathcal{D},\ i \in \mathcal{N}_c^k,\ k \in \mathcal{K},\ \xi \in \Xi \tag{48}$$

$$\sum_{j\in\mathcal{B}_{ik}(\xi)} x_{ji}^{dk}(\xi) = w_{ikd}(\xi), \qquad\qquad \forall\, d \in \mathcal{D},\ i \in \mathcal{N}_c^k,\ k \in \mathcal{K},\ \xi \in \Xi \tag{49}$$

$$\sum_{i\in\mathcal{B}_{jk}(\xi)} x_{ij}^{dk}(\xi) = \sum_{i\in\mathcal{A}_{jk}(\xi)} x_{ji}^{dk}(\xi), \qquad\qquad \forall d \in \mathcal{D}, j \in \mathcal{N}, k \in \mathcal{K},\ \xi \in \Xi \tag{50}$$

$$a_{ik}(\xi) \geq w_{ikd}(\xi), \qquad\qquad \forall i \in \mathcal{N},\ k \in \mathcal{K},\ d \in \mathcal{D},\ \xi \in \Xi \tag{51}$$

$$\sum_{k\in K} a_{ik}(\xi) \leq cc, \qquad\qquad \forall i \in \mathcal{N},\ \xi \in \Xi \tag{52}$$

$$u_{mk} \leq \sum_{l\in\mathcal{L}} z_{lm}\nu_{l(N+k)}, \qquad\qquad \forall m \in \mathcal{G}_g,\ k \in \mathcal{K} \tag{53}$$

$$z_{lm},\ u_{mk},\ a_{ik}(\xi),\ w_{ikd}(\xi),\ \mu_{ll'm},\ x_{ij}^{dk}(\xi) \in \{0,1\} \quad \forall i \in \mathcal{N},\ k \in \mathcal{K},\ l \in \mathcal{L},\ m \in \mathcal{G}_g,\ \xi \in \Xi \tag{54}$$

Constraints (36) guarantee that at least one caregiver is hired in each district. Constraints (40) and (41) guarantee that $\mu_{ll'm}$ is equal to 1 if the basic units l and l' are assigned to the same district. Constraints (42)

state that the ith patient can be assigned to the kth caregiver if they are in the same district. Constraints (45) ensure that nodes i and j can be visited sequentially if they have been assigned to the same district. Constraints (53) ensure that the kth caregiver can be hired in the mth district if the caregiver is assigned to that district.

An advantage of the proposed model in this phase is that using $\mathcal{A}_{ik}(\xi)$ and $\mathcal{B}_{ik}(\xi)$ decreases the complexity of the problem by determining the sequence of patient visits and may lead to an improvement in the computational efficiency.

4.2.1. Progressive hedging based algorithm

The complexity of Int.M2 increases with increasing number of scenarios, therefore, a progressive hedging based algorithm is used to solve the proposed model. The progressive hedging algorithm was proposed by Rockafellar and Wets (1991) as a scenario based decomposition method and was based on an augmented Lagrangian theory by relaxing non-anticipativity constraints. It has been proved that this algorithm can obtain a global optimum solution when all decision variables are continuous, while in the case of mixed integer programming, finding the global optimum solution is not guaranteed. Nevertheless, the progressive hedging algorithm can be considered an efficient heuristic algorithm in the case of mixed integer programming (Guo et al., 2015). Boland et al. (2018) proposed an algorithm based on a combination of the progressive hedging and the Frank and Wolf (PH-FW) algorithms to strengthen the lower bounds obtained by the progressive hedging algorithm in the case of mixed integer programming. In the present paper, the PH-FW algorithm is used to solve Int.M2 and to obtain good quality solutions. The pseudo code of the PH-FW algorithm for Int.M2 is shown in Algorithm 2.

The PH-FW algorithm consists of two phases. In the first phase, the Int.M2 model is decomposed into a number of sub-problems by considering copies of the first stage decision variables under each scenario (Model-$1(\xi)$). Expression (55) illustrates the objective function of these sub-problems. Obtaining the values of the first stage decision variables for calculating weight matrices is the aim of the first phase. The second phase of the PH-FW algorithm includes two mathematical models, Model-$2(\xi)$ and Model-$3(\xi)$. In this phase, at first the weight matrices $\psi_{lm}^{1iter}(\xi)$, $\psi_{mk}^{2iter}(\xi)$, and $\psi_{ll'm}^{3iter}(\xi)$ are updated based on the obtained values for the first stage decision variables and corresponding penalty parameters ρ_1, ρ_2, and ρ_3. Then the sub-problems Model-$2(\xi)$ are optimized by considering a penalty term for relaxing non-anticipative constraints in the objective function. The objective function of the sub-problems in the second phase is presented in Expression (56). Obtaining a set of solutions for Int-M2 when non-anticipative constraints are relaxed is the aim of optimizing these sub-problems. The sub-problems extracted from a scenario-based decomposition of the augmented Lagrangian by relaxing non-anticipative constraints (Model-$3(\xi)$) are optimized at the end of the second phase of the PH-FW algorithm. The Simplicial Decomposition Method (SDM), as an extension of the Frank and Wolf algorithm, is used to solve these sub-problems. The objective functions of these sub-problems are illustrated in Expression (57) and the integrality of the decision variables is relaxed. The optimal values of the decision variables in each subproblem are determined based on a convex combination of optimal solutions

obtained by Model-1(ξ) and Model-2(ξ). The second phase of the PH-FW algorithm is solved iteratively and the tolerance is defined based on the variance of the obtained solutions for the first stage decision variables under different scenarios and calculated in each iteration. If the tolerance in an iteration is less than a predefined threshold (ε), then the termination condition is satisfied.

As mentioned previously, the districting decisions are considered as binary first stage decision variables. In this paper it is assumed that those basic units assigned to the same district are important in spite of the labels used for defining the districts, and permutations of labels used for defining districts are considered as alternative solutions. Two solutions for districting ten basic units are illustrated in Figure 4. In these solutions, the basic units assigned to the same district are similar, while the labels of districts are different in these solutions. For example, the basic units 2, 5 and 8 are in the same district in both solutions but the label of the district is 1 in the first solution whereas it is 3 in the second solution. Therefore, these solutions for districting decisions are alternative solutions, because the basic units in a district are constant and the labels of the districts are alternative. The labels of districts replaced with each other in these solutions are shown in the same colors in this figure. These alternative solutions can affect the efficiency of the PH-FW algorithm. If the solutions illustrated in Figure 4 are obtained by optimizing Int.M2 for each $\xi \in \{1, 2\}$, it is considered that the achieved solutions are different and the PH-FW algorithm tries to make them similar and decrease the violations between the obtained solutions under each scenario and their mean value. Therefore, alternative solutions affect the efficiency of the Frank and Wolf algorithm and as a result affect the computational time of the PH-FW algorithm. An example is presented to clarify the effect of alternative solutions on the efficiency of the PH-FW algorithm. The obtained solution for the decision variable $\hat{z}_{lm}(\xi)$ under two different scenarios and their mean values are shown in Figure 5. In this example, ten basic units are partitioned into four districts. The basic units in each district are represented in this figure. As mentioned previously, the progressive hedging algorithm tries to find equal values for the first stage decision variables. Therefore, the mean values of the first stage decision variables are calculated and the weight matrices are updated based on them. The weight matrices $\psi_{lm}^{12}(\xi)$ based on the solutions presented in Figure 5 have been calculated when ρ_1 is equal to 0.5 and are illustrated in Figure 6. As shown in Figure 6, $\psi_{lm}^{12}(1)$=- $\psi_{lm}^{12}(2)$. Therefore, the obtained solutions in different iterations of the PH-FW algorithm do not converge to a unique solution. It can be concluded that the alternative solutions can affect the efficiency of the PH-FW algorithm in obtaining a near optimal solution.

The basic units are assigned to each district are shown in Table 4 based on the obtained solution for decision variable $\hat{z}_{lm}(\xi)$ under two different scenarios. The solutions in Table 4 are considered as alternative solutions of Int.M2 for the districting decision. As long as they are treated as different solutions in the PH-FW algorithm, the algorithm cannot converge to a unique solution, or at least in some cases the computational time increases significantly.

Algorithm 2: The framework of PH-FW algorithm for solving Int.M2 model in the second phase of Math-1 algorithm.

Input: ρ_1, ρ_2, ρ_3, g, ε, parameters of Int.M2 model

$iter \leftarrow 1$, $V_0(\xi) \leftarrow []$, $\psi_{lm}^{1iter}(\xi) \leftarrow 0$, $\psi_{mk}^{2iter}(\xi) \leftarrow 0$, $\psi_{ll'm}^{3iter}(\xi) \leftarrow 0$;

Phase 1:

for $\xi \in \Xi$ **do**

Optimize the following model then determine the decision variables $(\hat{z}_{lm}(\xi)\ \hat{u}_{mk}(\xi)\ \hat{\mu}_{ll'm}(\xi)$
$\hat{a}_{ik}(\xi), \hat{w}_{ikd}(\xi)\ \hat{x}_{ij}^{dk}(\xi))$ when basic units partitioned into g districts;

$$\min \sum_{m\in\mathcal{G}_g\ k\in\mathcal{K}} h_k \times u_{mk}(\xi) + \alpha \times \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}\ l'\in\mathcal{L}_l^N} (tr_{ll'} - td) \times \mu_{ll'm}(\xi)$$
$$+ \sum_{i\in\mathcal{V}\ j\in\mathcal{A}_{ik}(\xi)\ d\in\mathcal{D}\ k\in\mathcal{K}} \theta \times tt_{ij}(\xi) \times x_{ij}^{dk}(\xi) \qquad (55)$$

s.t. (33)-(54)

Using Algorithm 3 to obtain the adjusted solution (The value of first stage decision variables are input of this algorithm and adjusted versions of first stage decision variables are output of this algorithm);

$V_{iter}(\xi) \leftarrow V_{iter-1}(\xi) \cup (\hat{z}_{lm}(\xi), \hat{u}_{mk}(\xi), \hat{\mu}_{ll'm}(\xi), \hat{a}_{ik}(\xi), \hat{w}_{ikd}(\xi), \hat{x}_{ij}^{dk}(\xi))$;

converge3 $\leftarrow$ *false*

Phase 2:

while *converge3 is false* **do**

$\bar{z}_{lm} = \sum_{\xi\in\Xi} p(\xi)\hat{z}_{lm}(\xi)$, $\bar{u}_{mk} = \sum_{\xi\in\Xi} p(\xi)\hat{u}_{mk}(\xi)$, $\bar{\mu}_{ll'm} = \sum_{\xi\in\Xi} p(\xi)\hat{\mu}_{ll'm}(\xi)$, $iter \leftarrow iter + 1$,

$\psi_{lm}^{1iter}(\xi) = \psi_{lm}^{1iter-1}(\xi) + 2\rho_1(\hat{z}_{lm}(\xi) - \bar{z}_{lm})$, $\psi_{mk}^{2iter}(\xi) = \psi_{mk}^{2iter-1}(\xi) + 2\rho_2(\hat{u}_{mk}(\xi) - \bar{u}_{mk})$,

$\psi_{ll'm}^{3iter}(\xi) = \psi_{ll'm}^{3iter-1}(\xi) + 2\rho_3(\hat{\mu}_{ll'm}(\xi) - \bar{\mu}_{ll,m})$;

for $\xi \in \Xi$ **do**

Optimize the following model then determine the decision variables

$$\min \sum_{m\in\mathcal{G}_g\ k\in\mathcal{K}} h_k \times u_{mk}(\xi) + \alpha \times \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}\ l'\in\mathcal{L}_l^N} (tr_{ll'} - td) \times \mu_{ll'm}(\xi)$$
$$+ \sum_{i\in\mathcal{V}\ j\in\mathcal{A}_{ik}(\xi)\ d\in\mathcal{D}\ k\in\mathcal{K}} \theta \times tt_{ij}(\xi) \times x_{ij}^{dk}(\xi) + \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}} \psi_{lm}^{1iter}(\xi) \times z_{lm}(\xi)$$
$$+ \sum_{m\in\mathcal{G}_g\ k\in\mathcal{K}} \psi_{mk}^{2iter}(\xi) \times u_{mk}(\xi) + \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}\ l'\in\mathcal{L}} \psi_{ll'm}^{3iter}(\xi) \times \mu_{ll'm}(\xi) \qquad (56)$$

s.t. (33)-(54)

Using Algorithm 3 to obtain the adjusted solution;

$V_{iter}(\xi) \leftarrow V_{iter-1}(\xi) \cup (\hat{z}_{lm}(\xi), \hat{u}_{mk}(\xi), \hat{\mu}_{ll'm}(\xi), \hat{a}_{ik}(\xi), \hat{w}_{ikd}(\xi), \hat{x}_{ij}^{dk}(\xi))$;

for $\xi \in \Xi$ **do**

Optimize the following model then determine the decision variables

$$\min \sum_{m\in\mathcal{G}_g\ k\in\mathcal{K}} h_k \times u_{mk}(\xi) + \alpha \times \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}\ l'\in\mathcal{L}_l^N} (tr_{ll'} - td) \times \mu_{ll'm}(\xi) +$$
$$\sum_{i\in\mathcal{V}\ j\in\mathcal{A}_{ik}(\xi)\ d\in\mathcal{D}\ k\in\mathcal{K}} \theta \times tt_{ij}(\xi) \times x_{ij}^{dk}(\xi) + \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}} \psi_{lm}^{1iter}(\xi) \times z_{lm}(\xi) + \sum_{m\in\mathcal{G}_g\ k\in\mathcal{K}} \psi_{mk}^2(\xi) u_{mk}(\xi) +$$
$$\sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}\ l'\in\mathcal{L}} \psi_{ll'm}^3(\xi) \mu_{ll'm}(\xi) + \sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}} \rho_1(z_{lm}(\xi) - \bar{z}_{lm})^2 + \sum_{m\in\mathcal{G}_g\ k\in\mathcal{K}} \rho_2(u_{mk}(\xi) - \bar{u}_{mk})^2 +$$
$$\sum_{m\in\mathcal{G}_g\ l\in\mathcal{L}\ l'\in\mathcal{L}} \rho_3(\mu_{ll'm}(\xi) - \bar{\mu}_{ll'm})^2 \qquad (57)$$

s.t. $(\hat{z}_{lm}(\xi)\ \hat{u}_{mk}(\xi)\ \hat{\mu}_{ll'm}(\xi)\ \hat{a}_{ik}(\xi), \hat{w}_{ikd}(\xi)\ \hat{x}_{ij}^{dk}(\xi)) \in conv(V_{iter}(\xi))$

$tolerance = \sum_{\xi\in\Xi} p(\xi)\left[(\hat{z}_{lm}(\xi) - \bar{z}_{lm})^2 + (\hat{u}_{mk}(\xi) - \bar{u}_{mk})^2\right]$;

if *tolerance* $\leq \varepsilon$ **then**

converge3 $\leftarrow$ *true*

Result: $\bar{z}_{lm}$, $\bar{u}_{mk}$, $\bar{\mu}_{ll'm}$, $\hat{w}_{ikd}(\xi)$, $route_{kd}(\xi)$

(a) Districting decision for solution 1

(b) Districting decision for solution 2
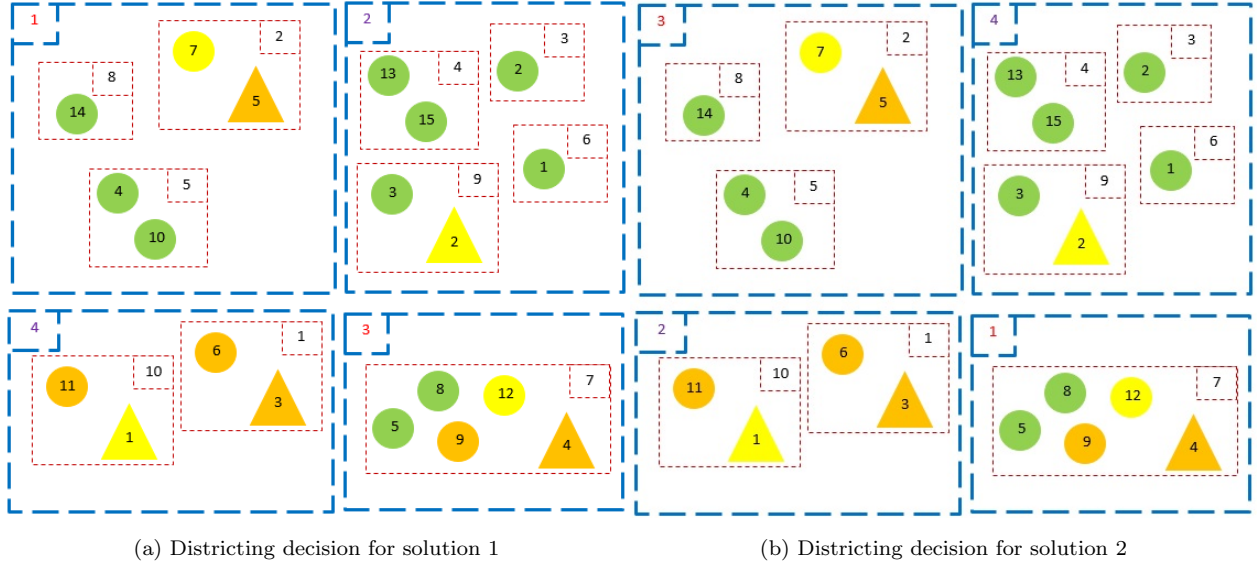
Figure 4: Alternative solutions for districting decision variable.

(a) Districting decision for solution under scenario 1

(b) Districting decision for solution under scenario 2

(c) Mean value of obtained solutions

Figure 5: Alternative solutions for districting decision variable and their mean value.

Figure 6 contains two weight matrices.

(a) Weight matrix for solution related to scenario 1

Basic units	Districts 1	2	3	4
1	0	-0.5	0	0.5
2	0.5	0	-0.5	0
3	0	0.5	0	-0.5
4	0	0.5	0	-0.5
5	0.5	0	-0.5	0
6	0	0.5	0	-0.5
7	-0.5	0	0.5	0
8	0.5	0	-0.5	0
9	0	0.5	0	-0.5
10	0	-0.5	0	0.5

(b) Weight matrix for solution related to scenario 2

Basic units	Districts 1	2	3	4
1	0	0.5	0	-0.5
2	-0.5	0	0.5	0
3	0	-0.5	0	0.5
4	0	-0.5	0	0.5
5	-0.5	0	0.5	0
6	0	-0.5	0	0.5
7	0.5	0	-0.5	0
8	-0.5	0	0.5	0
9	0	-0.5	0	0.5
10	0	0.5	0	-0.5

(a) Weight matrix for solution related to scenario 1   (b) Weight matrix for solution related to scenario 2

Figure 6: Weight matrices for the solutions in Figure 5.

Table 4: Values of districting decision variables for two different scenarios.

District	Scenario 1	Scenario 2
1	{2 5 8}	{7}
2	{3 4 6 9}	{1 10}
3	{7}	{2 5 8}
4	{1 10}	{3 4 6 9}

An algorithm is proposed to make the obtained solutions similar. In this algorithm, a solution is considered as a reference solution. In this paper, the obtained solution for the districting decision variables for the first scenario in the first iteration of PH-FW algorithm is considered as a reference solution. A criterion is defined to determine the degree of similarity between basic units in different pairs of districts. The value of this criterion $(ind_{mm'})$ for districts m and m' is obtained based on the difference between the number of similar basic units in these districts and number of different basic units in them. For example, for the first district of scenario 1 and second district of scenario 2 (m and m' are equal to 1 and 2, respectively), there are no common basic units in these districts and they have five different basic units. Therefore, the value of ind_{12} is equal to $0 - 5 = -5$. The value of ind for different pair of districts for the solutions presented in Table 4 are shown in Table 5.

Table 5: Values for the "*ind*" for the solutions in Table 4 when solution for the first scenario is reference solution.

			scenario 2			
	District	1	2	3	4
scenario 1	1	-4	-5	3	-7
	2	-5	-6	-7	4
	3	1	-3	-4	-5
	4	-3	2	-5	-6

The obtained values for $ind_{mm'}$ are used to determine the degree of similarity between the basic units in each pair of districts in a reference solution and a given solution. At first, a district from the given solution is selected (e.g., the m'^{th} district is selected), then the district from the reference solution that has the maximum degree of similarity with the selected district is identified by $m_{m'}^{maxdegree} = \arg\max_{m \in \mathcal{G}_g}\{ind_{mm'}\}$. For example, if the first district of the solution related to the second scenario is selected, then a district from the reference solution with the maximum degree of similarity with the selected district is determined based on the values for *ind* reported in Table 5. Therefore, the third district of the reference solution is selected as $m_1^{maxdegree}$ because it has the maximum degree of similarity (the greatest value of *ind*). Then an adjusted version of the solution is constructed by placing the basic units in the selected district of the solution in the district with maximum degree of similarity. In the mentioned example, basic unit 7 is in the first district of the solution related to the second scenario and this district has the maximum degree of similarity with the third district in the reference solution. Therefore, basic unit 7 is located in the third district in the adjusted version of the solution. The solutions of the first and second scenarios and adjusted version of this solution are represented in Table 6. As shown in Table 6, the adjusted version of the solution for the second scenario is completely similar to the reference solution (the solution corresponding to the first scenario). Considering the adjusted version of the solution for the second scenario instead of the obtained solution can affect the efficiency of the Frank and Wolf algorithm. Therefore, it can be concluded that employing this algorithm increases the efficiency of the PH-FW algorithm for the proposed model. The pseudo code of the proposed algorithm is presented in Algorithm 3.

Table 6: Values of districting decision variables for the second scenario and the adjusted version of this solution.

District	Scenario 1 (Reference solution)	Scenario 2	Adjusted version of scenario 2
1	{2 5 8}	{7}	{2 5 8}
2	{3 4 6 9}	{1 10}	{3 4 6 9}
3	{7}	{2 5 8}	{7}
4	{1 10}	{3 4 6 9}	{1 10}

Algorithm 3: The framework of proposed procedure to speed up thePH-FW convergence in the second phase of Math-1 algorithm.

Input: $\hat{z}_{lm}(\xi)$, $\hat{u}_{mk}(\xi)$, $\hat{\mu}_{ll'm}(\xi)$, g

$sol_{ref} \leftarrow \hat{z}_{lm}(1)$;

for $\xi \in \Xi$ **do**

    for $m' \in \mathcal{G}_g$ **do**

        for $m \in \mathcal{G}_g$ **do**

            $ind_{mm'}$=number of common basic units in m^{th} district of sol_{ref} and m'^{th} district of $\hat{z}_{lm}(\xi)$-number of different basic units in m^{th} district of sol_{ref} and m'^{th} district of $\hat{z}_{lm}(\xi)$;

        $m_{m'}^{maxdegree} = \underset{m \in \mathcal{G}_g}{\arg\max}\{ind_{mm'}\}$;

        if $|m_{m'}^{maxdegree}| = 1$ **then**

            $m^{adj} \leftarrow m_{m'}^{maxdegree}$;

            $\hat{z}_{lm^{adj}}^{adj}(\xi) \leftarrow \hat{z}_{lm'}(\xi)$;

            $\hat{u}_{m^{adj}k}^{adj}(\xi) \leftarrow \hat{u}_{m'k}(\xi)$;

            $\hat{\mu}_{ll'm^{adj}}^{adj}(\xi) \leftarrow \hat{\mu}_{ll'm'}(\xi)$;

        else

            An appropriate value for m^{adj} is selected from the set m'' randomly;

            $\hat{z}_{lm^{adj}}^{adj}(\xi) \leftarrow \hat{z}_{lm'}(\xi)$;

            $\hat{u}_{m^{adj}k}^{adj}(\xi) \leftarrow \hat{u}_{m'k}(\xi)$;

            $\hat{\mu}_{ll'm^{adj}}^{adj}(\xi) \leftarrow \hat{\mu}_{ll'm'}(\xi)$;

$\hat{z}_{lm}(\xi) \leftarrow \hat{z}_{lm}^{adj}(\xi)$;

$\hat{u}_{mk}(\xi) \leftarrow \hat{u}_{mk}^{adj}(\xi)$;

$\hat{\mu}_{ll'm}(\xi) \leftarrow \hat{\mu}_{ll'm}^{adj}(\xi)$;

Result: $\hat{z}_{lm}(\xi)$, $\hat{u}_{mk}(\xi)$, $\hat{\mu}_{ll'm}(\xi)$

4.3. Phase 3: Repairing infeasibility

As mentioned previously, sub-tour elimination constraints are omitted in Int.M2 and as a result the time window constraints are removed from the proposed model, too. Therefore, the obtained solution from the second phase may be infeasible and should be repaired. An iterative procedure based on the fix and optimize method is used in this section to repair any infeasibility of the obtained solutions. The pseudo code of the proposed procedure is represented in Algorithm 4. In this procedure, the districting and assignment decisions are fixed in the districts that do not face a problem with infeasibility. But the districting, staff dimensioning and assignment decisions are re-optimized for those basic units and patients in districts facing this problem of infeasibility. The model Int.M3 is proposed to obtain feasible solutions for the districting, staff dimensioning and assignment decisions, based on an estimation of the routing cost. The intermediate model in this phase is presented as follows:

Algorithm 4: The framework of proposed procedure to repair infeasibility in the third phase of Math-1 algorithm.

Input: condition2, parameter of Int.M3 model

while *condition2 is false* **do**

    for $m \in \mathcal{G}_g$ **do**

        if *the mth district faces by infeasibility* **then**

            for $l \in \mathcal{L}$ **do**

                if $\bar{z}_{lm} = 1$ **then**

                    for $i \in \mathcal{N}$ **do**

                        if $\bar{z}_{lm} \times v_{li} = 1$ **then**

                            for $m' \in \mathcal{G}_g$ **do**

                                for $k \in \mathcal{K}$ **do**

                                    for $d \in \mathcal{D}$ **do**

                                        for $\xi \in \Xi$ **do**

                                            if $\bar{u}_{m'k} = 1$ **then**

                                                if $\sum\limits_{j \in \mathcal{N}} \hat{w}_{jkd}(\xi) > 1$ **then**

                                                if $\hat{w}_{ikd}(\xi) = 1$ **then**

                                                if *i^{th} patient faces by infeasibility* **then**

                                                    $\sigma_{ikd}(\xi) = M_{max}$

                                                else

                                                $\sigma_{ikd}(\xi) = \theta \times (tt_{i^-i}(\xi) + tt_{ii^+}(\xi) - tt_{i^-i^+}(\xi)),$

                                            else

                                            The insertion cost corresponding to the ith patient in the jth (cc_{ij}) position of $route_{kd}(\xi)$ is obtained by $cc_{ij} = \theta \times (tt_{i_j^- i}(\xi) + tt_{ii_j^+}(\xi) - tt_{i_j^- i_j^+}(\xi)),$ where i_j^- and i_j^+ denote predecessor and successor nodes of patient i in position j.

                                            if *Inserting the ith patient make $route_{kd}(\xi)$ infeasible* **then**

                                                $cc_{ij} = cc_{ij} \times M_{max}$

                                            $\sigma_{ikd}(\xi) = \min\limits_{j}\{cc_{ij}\}$

                                    else

                                      $\sigma_{ikd}(\xi) = 2 \times \theta \times tt_{i(N+k)}(\xi)$

                                else

                                    if $\bar{z}_{lm'} \times v_{l(N+k)} = 1$ **then**

                                      $\sigma_{ikd}(\xi) = 2 \times \theta \times tt_{i(N+k)}(\xi)$

  Optimize the Int.M3(g) according the estimated routing cost and determine the value of decision variables ($\bar{z}_{lm}$, $\bar{u}_{mk}$ and $\hat{w}_{ikd}(\xi)$);

  Solve series of TSP models in each district for each caregiver at each day under each scenario (TSP($mkd\xi$)) and determine $route_{kd}(\xi)$;

  if *Time window constraint is not violated* **then**

    Feasibility $\leftarrow true$

    condition2 $\leftarrow true$

  else

    if *Maximum number of iteration achieved* **then**

      condition2 $\leftarrow true$

Result: $\bar{z}_{lm}$, $\bar{u}_{mk}$ and $\hat{w}_{ikd}(\xi)$, Feasibility and $route_{kd}(\xi)$

$(\textbf{Int.M3(g)}) \quad \min \quad \sum_{m \in \mathcal{G}_g} \sum_{k \in \mathcal{K}} h_k \times u_{mk} + \alpha \times \sum_{m \in \mathcal{G}_g} \sum_{l \in \mathcal{L}} \sum_{l' \in \mathcal{L}_c^N} (tr_{ll'} - td) \times \mu_{ll'm}$

$$+ \sum_{i \in \mathcal{N}} \sum_{d \in \mathcal{D}} \sum_{k \in \mathcal{K}} \sum_{\xi \in \Xi} p(\xi) \times \sigma_{ikd}(\xi) \times w_{ikd}(\xi) \tag{58}$$

s.t.: $(33) - (39), (40) - (44), (51) - (53),$

$$z_{lm} = \bar{z}_{lm}, \qquad\qquad\qquad\qquad \forall l \in L_{fe}, \ m \in \mathcal{G}_g \tag{59}$$

$$w_{ikd}(\xi) = \hat{w}_{ikd}(\xi), \qquad\qquad\qquad \forall i \in I_{fe}, \ k \in \mathcal{K}, \ d \in \mathcal{D}, \ \xi \in \Xi \tag{60}$$

$$u_{km} \geq \bar{u}_{km}, \qquad\qquad\qquad\qquad \forall k \in K, \ m \in \mathcal{G}_g \tag{61}$$

$$z_{lm}, \ u_{mk}, \ a_{ik}(\xi), \ w_{ikd}(\xi), \ \mu_{ll'm} \in \{0,1\} \qquad \forall i \in \mathcal{N}, \ k \in \mathcal{K}, \ l \in \mathcal{L}, \ m \in \mathcal{G}_g, \ \xi \in \Xi \tag{62}$$

The objective function (58) minimizes the hiring, violation of the neighbouing constraint, and estimated routing costs. As mentioned previously, L_{fe} is defined as the set of basic units assigned to the feasible districts. On the other hand, the basic units in this set are assigned to the districts in which the time window constraints are not violated for any patients in them. Constraints (59) ensure that the districting decision is fixed for $l \in L_{fe}$. Constraints (60) guarantee that the assignment decision is fixed for the nodes (patients and caregivers) assigned to the feasible districts ($i \in I_{fe}$). The set I_{fe} consists of the nodes in the districts which the time window constraints are not violated for any patients in that district. Constraints (61) state that the caregivers used in the obtained solution from Int.M2 must be used in the repaired solution. Constraints (62) determine the domains of the decision variables.

As mentioned previously, the districting, staff dimensioning and assignment decisions are determined by the estimated routing cost in Int.M3. In this model, the parameter $\sigma_{ikd}(\xi)$ determines the cost corresponding to visiting the ith patient by the kth caregiver in day d under scenario ξ, and is obtained in each iteration based on the achieved routes in the previous iteration. In the first iteration, $\sigma_{ikd}(\xi)$ is calculated based on the routes obtained from the second phase. While in other iterations a series of TSPs are solved in each district for each caregiver at each day under each scenario TSP($mkd\xi$) based on the obtained assignment decision from solving Int.M3 and the corresponding routes determined. The proposed procedure to calculate this parameter is presented in Algorithm 4. Based on the proposed procedure, if the ith node exists in the route corresponding to the kth caregiver in day d under scenario ξ, the parameter $\sigma_{ikd}(\xi)$ is calculated by $\sigma_{ikd}(\xi) = \theta \times (tt_{i^-i}(\xi) + tt_{ii^+}(\xi) - tt_{i^-i^+}(\xi))$, where i^- and i^+ denote the predecessor and successor nodes of patient i in this route, respectively. But under the conditions that the ith node is not visited by the kth caregiver in day d under scenario ξ, the ith node is inserted in different positions of that route and the insertion cost related to each position is calculated. Then the cheapest insertion cost is taken as $\sigma_{ikd}(\xi)$. Also to guarantee the feasibility of the obtained solutions from Int.M3, a large value of $\sigma_{ikd}(\xi)$ is assigned to any position that makes the obtained route infeasible.

4.4. Phase 4: Improving the obtained solutions

The routing and assignment decisions are determined based on estimated routing cost in the second and third phases of the algorithm. Therefore, these decisions are improved in this section by using an iterative procedure and the fix and optimize method. In this procedure some patients are selected and their corresponding assignment and routing decisions are improved when the first stage decision variables and routing and assignment decisions are fixed for the other patients. These patients are selected based on σ. First, the set of patients for improvement in their corresponding decisions $(N^*(\xi))$ must be determined. For this reason, the value of $\sigma_{ikd}(\xi)$ is computed for each patient. In the first iteration, $\sigma_{ikd}(\xi)$ is computed based on the routing decision obtained from the second or third phases and the method mentioned in Algorithm 4 is used for calculating $\sigma_{ikd}(\xi)$. In other iterations, $\sigma_{ikd}(\xi)$ is computed from the solutions achieved from the previous iteration according to the procedure proposed in Algorithm 4. Then the caregiver and working day corresponding to the minimum value of $\sigma_{ikd}(\xi)$ are determined by $(k_i^*(\xi), d_i^*(\xi)) = \underset{\substack{k \in \mathcal{K} \\ d \in \mathcal{D}}}{argmin}\{\sigma_{ikd}(\xi)\}$ for the ith patient under scenario ξ. The ith patient is added to the set $N^*(\xi)$ if $\hat{w}_{ik_i^*(\xi)d_i^*(\xi)}(\xi)$ is equal to zero.

The intermediate model (Int.M4) is solved for each patient in $N^*(\xi)$ based on the value of $\sigma_{i^*kd}(\xi)$. In this model the assignment decision is optimized based on an estimation of the routing cost. Then the routing decision variable is obtained by solving a series of TSPs. The pseudo code of the proposed procedure is shown in Algorithm 5. The Int.M4 model is as follows:

Algorithm 5: The framework of proposed procedure to improve routing and assignment decisions in the fourth phase of Math-1 algorithm.

Input: parameter of Int.M4 model, $\hat{w}_{ikd}(\xi)$

for $\xi \in \Xi$ do

    while $N^*(\xi)$ *is not empty* do

        i^* is selected randomly from the set $N^*(\xi)$;

        i^* is removed from the set $N^*(\xi)$;

        The parameter σ_{i^*kd} for patient i^* is calculated based on the procedure in Algorithm 4;

        Solve the Int.M4 model for patient i^* and determine $\hat{w}_{ikd}(\xi)$;

        Solve series of TSPs in each district for each caregiver at each day under each scenarios $(TSP(mkd\xi))$ and determine $route_{kd}(\xi)$. ;

Result: $\hat{w}_{ikd}(\xi)$ and $route_{kd}(\xi)$

$$(\textbf{Int.M4(g, m, }\xi)) \quad \min_{i \in \mathcal{N} \, d \in \mathcal{D} \, k \in \mathcal{K}} \sigma_{ikd}(\xi) w_{ikd}(\xi) \tag{63}$$

$$\text{s.t.: } (37) - (39), (42) - (44), (51), (52),$$

$$w_{ikd}(\xi) = \hat{w}_{ikd}(\xi), \qquad\qquad \forall i \in \mathcal{V} \setminus \{i^*\}, \ k \in \mathcal{K}, \ d \in \mathcal{D}, \tag{64}$$

$$u_{km} = \bar{u}_{km}, \qquad\qquad \forall k \in \mathcal{K} \tag{65}$$

$$\mu_{ll'm} = \bar{\mu}_{ll'm}, \qquad\qquad \forall l, l' \in \mathcal{L} \tag{66}$$

$$u_{mk}, \ a_{ik}(\xi), \ w_{ikd}(\xi), \mu_{ll'm} \in \{0,1\}, \qquad\qquad \forall i \in \mathcal{N}, \ k \in \mathcal{K}, \ l \in \mathcal{L}, \ \xi \in \Xi \tag{67}$$

The objective function (63) minimizes the estimated routing cost. Constraints (64) ensure that the assignment decision is fixed for patients other than the selected one. Constraints (65) and (66) guarantee that the first stage decision variables are fixed. Constraints (67) determine the domains of the decision variables.

5. Computational results

In this section, the validity of the proposed main model (SDDARP) is examined through numerical examples. The efficiency of the proposed mathematics-based algorithm is also illustrated. Our instances were generated based on real-world data provided by the Austrian Red Cross (Fikar and Hirsch, 2015). The data was generated based on five instances in an urban region by considering 75 nodes. As mentioned previously, the patients that are located in close proximity of each other form a basic unit to decrease the size of the problem. Therefore, to generate instances, the k-means algorithm (MacQueen, 1967) is used to determine the nodes located close to each other and construct the basic units. The nodes related to patients and caregivers are partitioned into L groups (L basic units) by using the k-means algorithm. Three sizes of problem are examined in this section. Their specifications are shown in Table 7.

Table 7: Specifications of different sizes of problem.

Size of problem	N	K	L	D	S	Number of instances
Small	10, 15	5	10	5	2, 5	20
Medium	20, 25, 30	5	10	5	2, 5	30
Large	35, 40, 45	5	10	5	2, 5	30

The parameters that are used in the proposed model are not defined in the dataset, so the modified values for the parameters are shown in Table 8. The travel times between nodes are generated based on driving distances reported in the dataset (d_{ij}) by $tt_{ij}(\xi) = d_{ij} \times \alpha(\xi) \, \forall i, j \in \, , \xi \in \Xi$. Where, $\alpha(\xi)$ are random

numbers drawn in the range [5, 15]. In this section, five instances (Ins 1, Ins 2,..., Ins 5) are considered for each combination of number of patients and scenarios, as a result, 80 instances are examined in this section. The generated instances can be found at https://github.com/bashirimahdi/SDDARP-for-HHC. The computations were performed on a 3.5 GHz Workstation with 32 GB RAM and 6 cores operating with Windows 10 (64-bit) and using the Julia software. Cplex 12.7.1 has been used as a solver.

Table 8: Parameters' values in numerical instances.

Parameters	Corresponding values or distribution
q_k	Discrete uniform [1, 3]
λ_i	r_i
h_k	1000
cc	1
φ	600
td	57
$st_i(\xi)$	Discrete uniform [30, 120]
θ	1
α	100

5.1. Examining the necessity of the simultaneous consideration of decisions

In this section, the necessity of the simultaneous consideration of decisions related to human resource planning is investigated. In most of the previous studies related to districting in HHC, the objective function minimizes the distances between the basic units, and overlooks routing and assignment decisions. We show that the aggregation of these decisions decreases the routing cost. The results for different numbers of patients and considering one scenario are presented in Table 9. The Int.M1 model is optimized for different numbers of districts (g) to obtain districting decisions without considering the staff dimensioning and routing costs. Then the achieved districting decision for each value of g is fixed in the SDDARP model and it is solved to obtain the objective value as well as the values of the other decision variables. The routing cost corresponding to the best objective value amongst the obtained objective values for different values of g is shown in Table 9. In addition, the relative cost reduction (RCR) is indicated. The results indicate that by considering the decisions simultaneously, the routing cost decreases 18.45% on average.

Table 9: Effect of simultaneous consideration of human resource related decisions in HHC.

N	Instance	Routing cost		RCR (%)	N	Instance	Routing cost		RCR(%)
		Classical model	SDDARP model				Classical model	SDDARP model	
	Ins 1	615	505	21.78		Ins 1	596	539	10.58
	Ins 2	507	470	7.87		Ins 2	462	369	25.20
10	Ins 3	623	537	16.01	20	Ins 3	707	511	38.36
	Ins 4	575	352	63.35		Ins 4	529	502	5.38
	Ins 5	353	353	0.00		Ins 5	704	577	22.01
	Ins 1	587	518	13.32		Ins 1	826	703	17.50
	Ins 2	476	367	29.70		Ins 2	499	488	2.25
15	Ins 3	607	474	28.06	25	Ins 3	705	613	15.01
	Ins 4	572	469	21.96		Ins 4	692	645	7.29
	Ins 5	668	555	20.36		Ins 5	778	756	2.91

## 5.2. Tuning the parameters of the progressive hedging algorithm

The parameters ρ_1, ρ_2, ρ_3 and ϵ in the progressive hedging algorithm should be tuned. It is assumed that the values of the maximum number of iterations, ϵ, and ρ_1 are equal to 50, 0.001, and 0.5, respectively. Then four combinations are considered for ρ_2 and ρ_3. These combinations are examined for different numbers of patients and scenarios when the number of districts (g) is equal to 3. The gap between the obtained objective values of Int.M2 by the commercial solver and the progressive hedging algorithm for different combinations of parameters is reported in Table 10. The computation times of the progressive hedging algorithm for different combinations of parameter are also reported in this table. A paired t-test is used to test the hypothesis that the objective values for Int.M2 obtained by the commercial solver are equal to those obtained by the progressive hedging algorithm. A paired t-test considering a two-tailed alternative hypothesis has been designed and the p-value corresponding to each combination of parameters is shown in Table 10. The obtained p-value is 0.233 when ρ_2 and ρ_3 are equal to 50 and 0.5, respectively, and it can be concluded that the solutions obtained by the PH-FW algorithm are almost equal to the solutions obtained by the commercial solver. Also, the individual values of the differences for each combination of parameters are illustrated in Figure 7. The differences between the objective values obtained by the commercial solver and those using the PH-FW algorithm for different combinations of parameters (ρ_2 and ρ_3) are indicated in this figure. Based on the obtained results, it is shown that the progressive hedging algorithm finds a better solution by considering the combinations (50, 0.5) and (25, 2), but the average computational time of combination (50, 0.5) is 1.5 times longer than the average computational time of combination (25, 2). Therefore, $\rho_2 = 25$ and $\rho_3 = 2$ are selected to use in the other instances.

5.3. Evaluating the effects of the proposed procedure for speeding up the PH-FW

As mentioned previously in the description of the progressive hedging algorithm, the efficiency of the PH-FW algorithm is improved by proposing an acceleration procedure. The tolerance of the PH-FW algorithm in each iteration of this algorithm is compared with the tolerance of the PH-FW algorithm in the case of using an acceleration procedure and depicted in Figure 8. As illustrated in this figure, the proposed procedure can help to speed up the convergence of the PH-FW algorithm. In the case of using the proposed procedure, the PH-FW algorithm converges within 16 iterations. However, late convergence is observed in 41 iterations when the acceleration procedure is not used in the PH-FW algorithm.

Table 10: Tuning the parameters ρ_2 and ρ_3 in the PH-FW algorithm.

| N	S	Instance	$(\rho_2, \rho_3)=(50, 0.5)$		$(\rho_2, \rho_3)=(50, 2)$		$(\rho_2, \rho_3)=(25, 0.5)$		$(\rho_2, \rho_3)=(25, 2)$	
			gap(%)	time	gap(%)	time	gap(%)	time	gap(%)	time
10	2	Ins 1	0.20	181.97	0.00	74.51	0.00	140.33	0.00	70.53
		Ins 2	0.00	68.06	0.00	99.17	0.00	48.39	0.00	65.99
		Ins 3	0.00	42.41	0.00	43.36	0.00	32.25	0.00	36.42
		Ins 4	0.04	42.2	0.00	83.99	0.00	50.84	0.00	89.41
		Ins 5	0.00	96.21	0.00	61.76	0.44	238.04	0.00	68.99
	5	Ins 1	0.00	260.08	0.00	92.94	0.00	128.18	0.00	99.52
		Ins 2	0.04	311.28	0.04	114.18	0.00	210.58	0.00	322.77
		Ins 3	0.00	393.42	0.00	95.93	0.00	467.40	0.00	122.74
		Ins 4	0.00	91.61	0.00	65.94	0.00	109.07	0.00	82.49
		Ins 5	0.00	383.78	0.05	577.66	0.10	270.42	0.10	429.22
	Average		0.03	187.10	0.01	130.94	0.05	169.55	0.01	138.81
15	2	Ins 1	0.00	67.81	0.06	56.11	0.06	392.87	0.06	64.18
		Ins 2	0.00	297.35	0.38	55.07	0.00	193.68	0.00	178.58
		Ins 3	0.00	62.97	0.00	51.86	0.00	58.82	0.00	44.79
		Ins 4	0.00	118.59	0.00	54.75	0.00	94.32	0.00	70.17
		Ins 5	0.11	77.06	0.11	52.67	0.11	96.06	0.11	75.16
	5	Ins 1	0.00	680.93	0.08	512.98	0.00	932.86	0.00	366.47
		Ins 2	0.00	583.26	0.75	785.02	0.00	609.46	0.00	762.8
		Ins 3	0.00	393.84	0.17	266.61	0.00	475.01	0.17	200.94
		Ins 4	0.00	225.32	0.00	132.13	0.00	202.11	0.00	181.8
		Ins 5	1.73	1258.4	1.73	1198.04	2.00	527.42	2.00	58.33
	Average		0.18	376.55	0.33	221.03	0.22	425.32	0.23	247.23
	p-value		0.233		0.08		0.196		0.235	
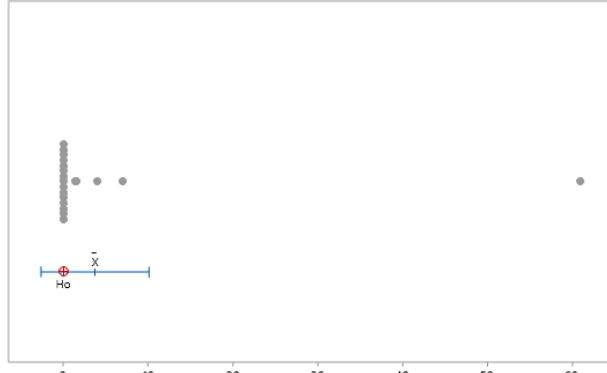
(a) The differences between the obtained objective values by the commercial solver and the PH-FW algorithm when $(\rho_2, \rho_3)=(50, 0.5)$
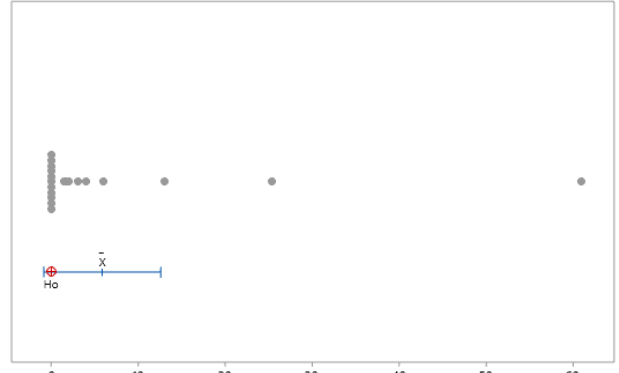
(b) The differences between the obtained objective values by the commercial solver and the PH-FW algorithm when $(\rho_2, \rho_3)=(50, 2)$
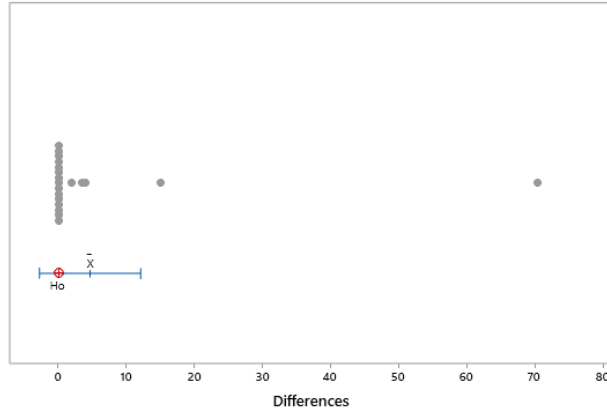
(c) The differences between the obtained objective values by the commercial solver and the PH-FW algorithm when $(\rho_2, \rho_3)=(25, 0.5)$
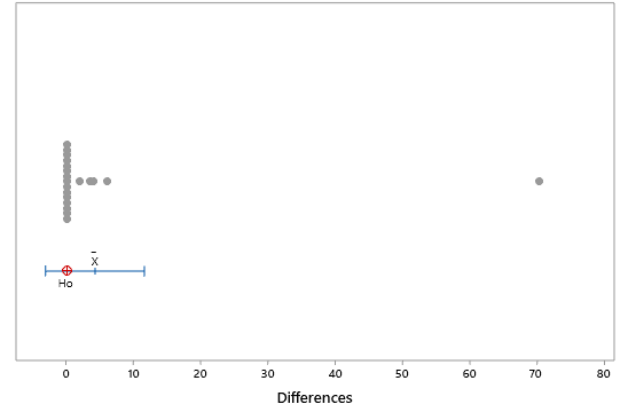
(d) The differences between the obtained objective values by the commercial solver and the PH-FW algorithm when $(\rho_2, \rho_3)=(25,2)$

Figure 7: Individual value of differences for different combinations of (ρ_2, ρ_3).
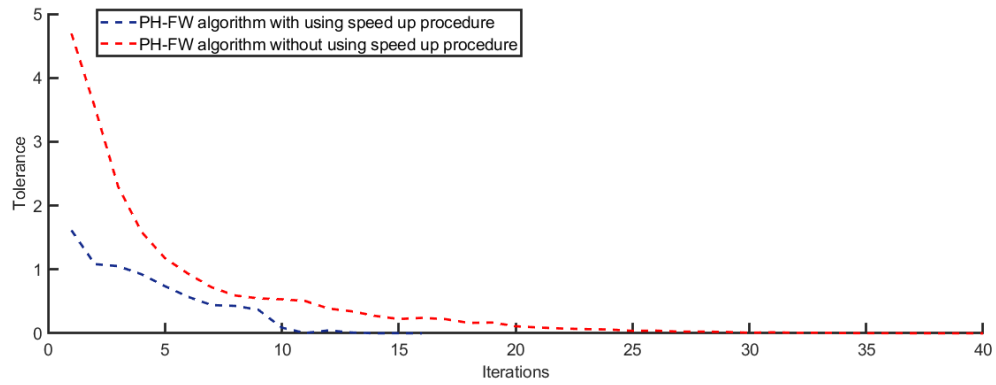

Figure 8: The effect of the proposed acceleration algorithm to speed up the convergence of the PH-FW algorithm.

35

5.4. An evaluation of the efficiency of the proposed mathematics-based algorithm

The efficiency of the proposed mathematics-based algorithm (Math-1) for solving the proposed model is presented in this section. The results obtained from the proposed matheuristic algorithm are compared with the result using a commercial solver for small size instances (up to 15 patients), as reported in Table 11. The best objective function value of the SDDARP model obtained by the commercial optimization software in 10,000 seconds and the obtained objective function values achieved by the matheuristic algorithm with and without using the progressive hedging algorithm are reported in columns 4, 6, and 9, respectively. If the optimal solution was not obtained by solving the SDDARP model in 10,000 seconds, the best solution provided by the commercial solver is denoted in column 4. Gap_1 and Gap_2 represent the deviations of the values of the objective function from the values obtained by the solver: $Gap_1(Gap_2) = (Obj_{Math-1}(Obj_{Math-1}^{PH-FW}) - Obj_{CS})/Obj_{CS} \times 100\%$. Also, the computational times of the commercial solver and the matheuristic algorithm with and without using the progressive hedging algorithm are reported in Table 11.

The results obtained show the efficiency of the proposed matheuristic algorithm. The proposed algorithm can find near-optimal solutions, and the average gap is almost equal for matheuristic algorithm with and without using the progressive hedging algorithm, but the average computational times of the matheuristic with and without using the PH algorithm are 420.42 seconds and 289.69 seconds, respectively. Therefore, it can be concluded that the matheuristic algorithm without using the PH-FW algorithm has a better performance than the matheuristic algorithm using the PH-FW algorithm in instances with a small size.

More analyses were performed by using the paired t-test to examine the equality of the obtained objective values by the matheuristic algorithm and matheuristic algorithm with the PH-FW algorithm when the alternative hypothesis assumes that the mean value of the differences between the objective values obtained from matheuristic algorithm with and without using PH-FW algorithm is greater than zero. The box plot of differences between the obtained objective values and p-values are illustrated in Figure 9 for different numbers of patients and scenarios. This figure indicates that the PH-FW algorithm can obtain good quality solutions in instances of a small size. Also, the effects of using the PH-FW algorithm in matheuristic for increasing numbers of scenarios with ten patients are shown in Figure 10. The mean values of the objective and of the computational times of the matheuristic algorithm with and without the PH-FW algorithm are compared in this figure when the number of scenarios are equal to 10 and 30. The results show that the efficiency of the matheuristic algorithm with PH-FW increases with an increasing number of scenarios. The efficiency of the proposed matheuristic algorithm for medium and large numbers of instances are reported in Tables 12 and 13, respectively. In these tables, Obj_{CS} is the best objective value obtained by the commercial solver in 10,000 seconds.

Table 11: Computation results for small size instances.

| N	S	Instance	Commercial solver (CS)		Math-1 without PH-FW			Math-1 with PH-FW		
			Obj_{CS}	$Time_{CS}(s)$	Obj_{Math-1}	$Time_{Math-1}$	$Gap_1(\%)$	Obj_{Math-1}^{PH-FW}	$Time_{Math-1}^{PH-FW}$	$Gap_2(\%)$
10	2	Ins 1	3519	1122.91	3524	22.98	0.14	3526.5	148.76	0.21
		Ins 2	3322.5	1318.47	3323.5	19.03	0.03	3323.5	143.83	0.03
		Ins 3	3490.5	1264.82	3497	16.61	0.19	3497	104.31	0.19
		Ins 4	3345.5	8994.6	3353	15.91	0.22	3354.5	154.2	0.27
		Ins 5	3364	1832.3	3375.5	19.14	0.34	3375.5	411.31	0.34
	5	Ins 1	3505.2	10000	3506.2	66.45	0.03	3507.4	445.56	0.06
		Ins 2	3323.2	10000	3323.6	44.6	0.01	3323.6	584.19	0.01
		Ins 3	3411.6	10000	3413.4	42.53	0.05	3413.4	272.93	0.05
		Ins 4	3354.8	10000	3356	29.67	0.04	3354.8	243.57	0.00
		Ins 5	3362.8	10000	3366.6	53.33	0.11	3371.6	718.46	0.26
15	2	Ins 1	3525	10000	3532.5	142.44	0.21	3526.5	276.18	0.04
		Ins 2	3364	10000	3372	30.67	0.24	3370.5	258.95	0.19
		Ins 3	3488	10000	3493	25.6	0.14	3493	277.4	0.14
		Ins 4	3454	10000	3455.5	37.34	0.04	3456.5	277.6	0.07
		Ins 5	3559	10000	3573.5	99.93	0.41	3564	352.83	0.14
	5	Ins 1	3538	10000	3549.4	1977.62	0.32	3550.4	617.01	0.35
		Ins 2	3381.2	10000	3381.4	1419.18	0.01	3381.2	1211.75	0.00
		Ins 3	3440	10000	3444.2	1143.45	0.12	3447	443.73	0.20
		Ins 4	3477.8	10000	3485.6	85.68	0.22	3481.8	515.22	0.11
		Ins 5	3534.2	10000	3560.4	501.7	0.74	3558	954.71	0.67
		Average				289.69	0.18		420.62	0.17
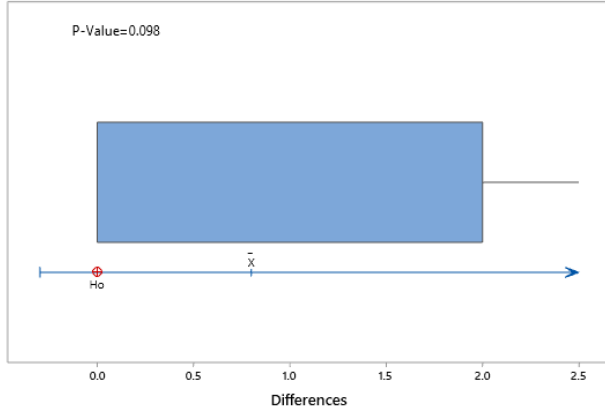
(a) 10 patients and 2 scenarios
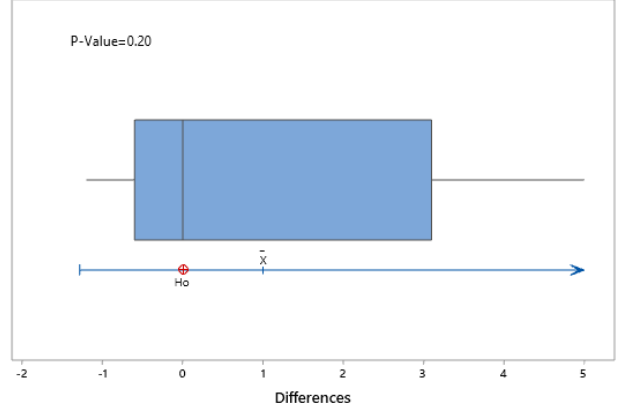
(b) 10 patients and 5 scenarios
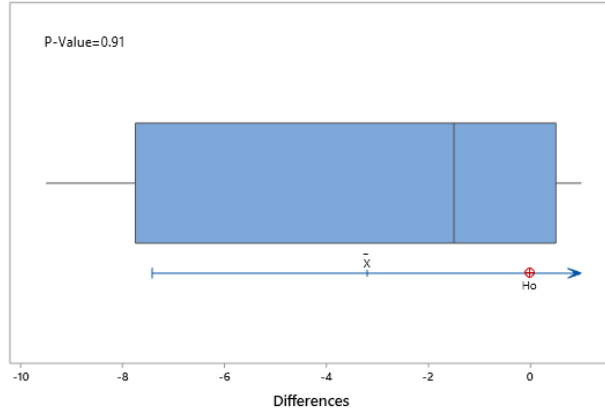
(c) 15 patients and 2 scenarios
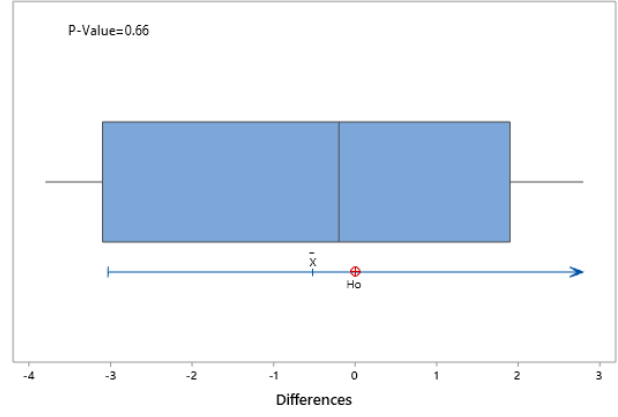
(d) 15 patients and 5 scenarios

Figure 9: Box plots and p-values of differences between objective values under different number of patients and scenarios when
H_0: mean value of $(Obj^{PH-FW}_{Math-1}-Obj_{Math-1})=0$
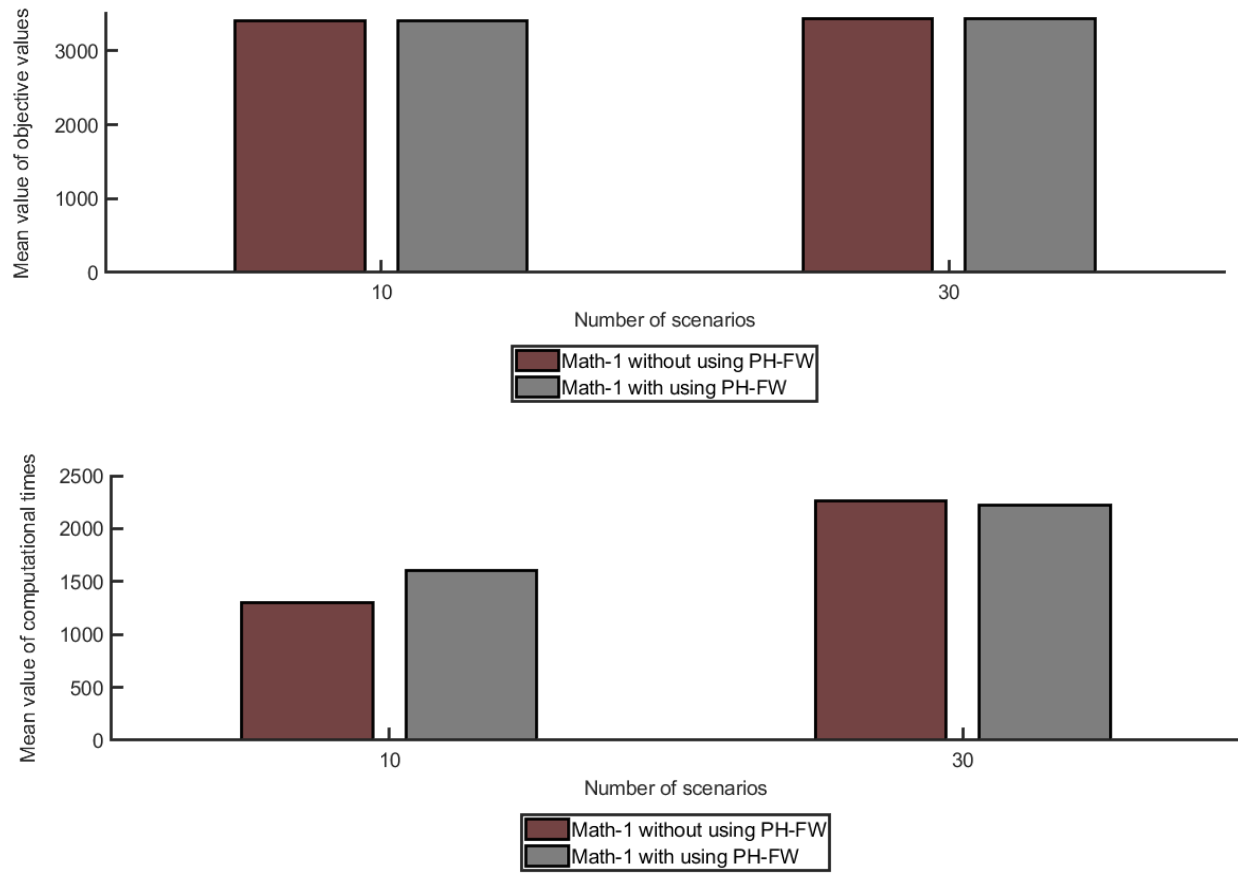H_1: mean value of $(Obj^{PH-FW}_{Math-1}-Obj_{Math-1})>0$.

Figure 10: Evaluation the effects of increasing in the number of scenarios on the efficiency of matheuristic algorithm with using the PH-FW.

Table 12: Computation results for medium size instances

N	S	Instance	Obj_{CS}	Math-1 without PH-FW			Math-1 with PH-FW		
				Obj_{Math-1}	$Time_{Math-1}$	$Gap_1(\%)$	Obj_{Math-1}^{PH-FW}	$Time_{Math-1}^{PH-FW}$	$Gap_2(\%)$
20	2	Ins 1	3570.5	3580	151.9	0.27	3579	252.8	0.24
		In 2	3398	3402	3947.11	0.12	3402	583.93	0.12
		Ins 3	3541	3532	80.52	-0.25	3532.5	475.26	-0.24
		Ins 4	3478.5	3496	1944.57	0.50	3489	404.43	0.30
		Ins 5	3550	3564	128.66	0.39	3544	168.86	-0.17
	5	Ins 1	5937	3567.2	2524.54	-39.92	3567	3156.24	-39.92
		Ins 2	3449.6	3425.6	5632.8	-0.70	3425.2	1116.34	-0.71
		Ins 3	3556.4	3564.6	1207.47	0.23	3564.2	658.96	0.22
		Ins 4	3536.4	3533.4	505.49	-0.08	3547	950.32	0.30
		Ins 5	3661.2	3661.8	1210.94	0.02	3667.6	2209.59	0.17
25	2	Ins 1	3701.5	3724.5	1474.28	0.62	3730.5	360.52	0.78
		Ins 2	3792.5	3492.5	1648.2	-7.91	3489.5	2486.22	-7.99
		Ins 3	3597	3600	856.76	0.08	3599.5	812.76	0.07
		Ins 4	3663.5	3645	293.01	-0.50	3641.5	494.42	-0.60
		Ins 5	3781.5	3721.5	517.68	-1.59	3770	615.86	-0.30
	5	Ins 1	17417	6348.2	688.56	-63.55	6342.8	1158.88	-63.58
		Ins 2	7022.6	3511	7315.31	-50.00	3513	7343.92	-49.98
		Ins 3	20960.4	7101.8	297.81	-66.12	7099.4	374.56	-66.13
		Ins 4	4347.85	3533.2	7444.17	-18.74	3532.4	2449.14	-18.76
		Ins 5	6153.6	4001.8	156.27	-34.97	4003.4	337.05	-34.94
30	2	Ins 1	11460	6506.5	313.53	-43.22	6505	345.23	-43.24
		Ins 2	10441	10453	220.24	0.11	10454.5	760.54	0.13
		Ins 3	7140.5	7146	104.12	0.07	7155.5	214.04	0.21
		Ins 4	10295	10267.5	793.15	-0.27	10237	863.89	-0.56
		Ins 5	17062.5	16118.5	190.83	-5.53	16128	318.29	-5.48
	5	Ins 1	42007.2	13308.2	1134.91	-68.32	13327	760.51	-68.27
		Ins 2	-	8608	846.93	-	8713.4	1329.6	-
		Ins 3	-	7363.2	886.14	-	7324	2152.42	-
		Ins 4	-	3569	8315.59	-	3567	5005.95	-
		Ins 5	-	4007.8	132.44	-	4008.4	202.07	-
Average					1698.80	-24.95		1278.75	-24.89

For the medium and large sizes, a limited computational time (3600 seconds) is considered for Int-M2. Table 12 shows that the computational time of the matheuristic algorithm with the PH-FW algorithm is better than the computational time of the matheuristic algorithm without the PH-FW algorithm. Therefore, the matheuristic algorithm with the PH-FW algorithm has a better performance than the matheuristic algorithm without the PH-FW algorithm for solving the proposed (SDDARP) problem. The results in Table 12 show that the average value of the solutions obtained by the matheuristic algorithm without using the PH-FW algorithm is 0.05% better than the average value of the solutions obtained by the matheuristic algorithm with the use of the PH-FW algorithm, while the computational time of the matheuristic algorithm without using the PH-FW algorithm is almost 1.3 times longer than the computational time of the matheuristic algorithm with the PH-FW algorithm. Also, the commercial solver cannot find any feasible solution in 4 instances when 30 patients and 5 scenarios are considered. But the proposed matheuristic algorithm finds solutions in a reasonable time for these instances. The solutions achieved by the matheuristic algorithm with and without the PH-FW algorithm are on average 25% better than the solutions obtained with the commercial solver in 10,000 seconds.

In Table 13, Gap represents the deviation of the values of the objective function from the value obtained by the matheuristic algorithm without the FW-PH algorithm: $Gap = (Obj_{Math-1}^{PH-FW} - Obj_{Math-1})/Obj_{Math-1} \times 100\%$. The results reported in Table 13 indicate that the commercial solver can not find any feasible solutions in 10,000 seconds, while the proposed matheuristic algorithm with and without using the PH-FW algorithm obtain good quality solution. The average computational time of the matheuristic algorithm without using the PH-FW algorithm is 1.25 times greater than the computational time of the matheuristic algorithm with the PH-FW algorithm. Also, the matheuristic algorithm with the PH-FW algorithm can find better solutions in most cases, the average gap between the obtained solutions is -0.68. Therefore, it can be concluded that using the PH-FW algorithm for solving Int.M2 increases the efficiency of the proposed matheuristic algorithm for medium and large instances.

Table 13: Computation results for large size instances

N	S	Instance	Obj_{CS}	Math-1 without PH-FW		Math-1 with PH-FW		$Gap(\%)$
				Obj_{Math-1}	$Time_{Math-1}$	Obj_{Math-1}^{PH-FW}	$Time_{Math-1}^{PH-FW}$	
35	2	Ins1	21000	8744	272.99	8747.5	419.65	0.04
		Ins 2	3884	3783.5	7166.35	3779.5	2707.14	-0.11
		Ins 3	11120	4516.5	132.41	4474.5	328.32	-0.93
		Ins 4	29748	5169	294.84	5247	467.53	1.51
		Ins 5	22627	7533	110.44	7526.5	164	-0.09
	5	Ins 1	-	5873.6	1406.54	5864	674.44	-0.16
		Ins 2	-	12453.8	2604.75	12438.4	4693.98	-0.12
		Ins 3	-	5030	1289.04	5018.6	2198.83	-0.23
		Ins 4	-	3683.8	5442.05	3690.8	5510.63	0.19
		Ins 5	-	4732.2	263.31	4726.2	481.62	-0.13
40	2	Ins 1	20409.5	5504	358.23	5519	1091.81	0.27
		Ins 2	9681.5	3805	4463.48	3805	3176.46	0.00
		Ins 3	6877.5	4366.5	601.8	4370.5	1325.58	0.09
		Ins 4	-	4064	541.43	4063	607.29	-0.02
		Ins 5	10775.5	4366	487.64	4382.5	1281.63	0.38
	5	Ins 1	-	5123.6	6132.89	4039	2158.4	-21.17
		Ins 2	-	7723.2	5001.04	7898	2971.25	2.26
		Ins 3	-	6839.2	7422.72	6642.6	3316.56	-2.87
		Ins 4	-	3859.4	5099.4	3877	4327.71	0.46
		Ins 5	-	6014.4	1671.46	6019	2286.51	0.08
45	2	ins 1	-	6675.5	2187.19	6781	3764.85	1.58
		ins 2	24313.5	7675.5	1494.17	7670.5	1861.45	-0.07
		ins 3	21805.5	7210.5	1339.92	7201.5	2537.87	-0.12
		ins 4	-	6439	208.96	6448.5	594.97	0.15
		ins 5	-	7828	753.61	7822	683.49	-0.08
	2	ins 1	-	7952.4	6429.24	7923	6924.81	-0.37
		ins 2	-	5920.6	4247.87	5905.8	6173.06	-0.25
		ins 3	-	10688.2	6136.04	10663.6	4882.14	-0.21
		ins 4	-	5232.6	8663.96	5235.2	3314.38	0.05
		ins 5	-	5386.6	7890.05	5363.8	1287.73	-0.42
	Average				3003.794		2407.136333	-0.68

6. Conclusion

In this paper, an integrated stochastic model is proposed to consider the districting, staff dimensioning, assignment, and routing decisions required in the delivery of home health care (HHC), simultaneously. Taking all the human resource planning decisions into consideration in HHC has significant effects on being able to plan in a cost effective manner. However, no study has been done in this context. In this study, districting and staff dimensioning are accomplished based on routing and hiring costs, and some constraints, such as continuity of care and similarity between the type of care required by patients and the caregivers' skills, are taken into account to make more realistic decisions.

A two-stage stochastic programming model has been developed to accommodate uncertain travel and service times, which are important and uncertain parameters in HHC. In the proposed model, districting and staff dimensioning are the first stage variables, and routing and assignment are the second stage variables.

Moreover, considering these decisions in an integrated way increases the complexity of the proposed model. Hence, a mathematical programming based algorithm is proposed as the solution strategy. This algorithm consists of four phases. Also, the PH-FW algorithm is used for solving the intermediate model in the second phase (Int.M2) and an acceleration procedure is proposed to speed up the convergence of this algorithm. The efficiency and validity of the algorithm are explored using numerous instances. The proposed algorithm can find optimal or near-optimal solutions in a reasonable computational time, even for large instances that could not be solved by the commercial solver. The results show that the matheuristic without using the PH-FW algorithm performs better in small size problems, while the matheuristic with the PH-FW algorithm is more efficient when the problem size increases. Also, the commercial solver cannot find any feasible solution in 10,000 seconds when the number of scenarios increases even beyond five, but the matheuristic algorithm can obtain near optimal solutions in less than 10,000 seconds in these cases. In this research, the proposed matheuristic is constructed by decomposing the main problem, which paves the way for further extensions by including other decomposition approaches in our proposed algorithm for solving the HHC and other resource planning problems.

References

Akjiratikarl, C., Yenradee, P., Drake, P.R., 2007. Pso-based algorithm for home care worker scheduling in the uk. Computers & Industrial Engineering 53, 559–583.

Andersen, A.R., Nielsen, B.F., Reinhardt, L.B., Stidsen, T.R., 2019. Staff optimization for time-dependent acute patient flow. European Journal of Operational Research 272, 94–105.

Archetti, C., Boland, N., Grazia Speranza, M., 2017. A matheuristic for the multivehicle inventory routing problem. INFORMS Journal on Computing 29, 377–387.

Barrera, D., Velasco, N., Amaya, C.A., 2012. A network-based approach to the multi-activity combined timetabling and crew scheduling problem: Workforce scheduling for public health policy implementation. Computers & Industrial Engineering 63, 802–812.

Benzarti, E., Sahin, E., Dallery, Y., 2013. Operations management applied to home care services: Analysis of the districting problem. Decision Support Systems 55, 587–598.

Bertels, S., Fahle, T., 2006. A hybrid setup for a hybrid scenario: combining heuristics for the home health care problem. Computers & Operations Research 33, 2866–2890.

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

Boland, N., Christiansen, J., Dandurand, B., Eberhard, A., Linderoth, J., Luedtke, J., Oliveira, F., 2018. Combining progressive hedging with a frank–wolfe method to compute lagrangian dual bounds in stochastic mixed-integer programming. SIAM Journal on Optimization 28, 1312–1336.

Bredström, D., Rönnqvist, M., 2008. Combined vehicle routing and scheduling with temporal precedence and synchronization constraints. European journal of operational research 191, 19–31.

Campelo, P., Neves-Moreira, F., Amorim, P., Almada-Lobo, B., 2019. Consistent vehicle routing problem with service level agreements: A case study in the pharmaceutical distribution sector. European Journal of Operational Research 273, 131–145.

Cappanera, P., Scutellà, M.G., 2014. Joint assignment, scheduling, and routing models to home care optimization: A pattern-based approach. Transportation Science 49, 830–852.

Carello, G., Lanzarone, E., 2014. A cardinality-constrained robust model for the assignment problem in home care services. European Journal of Operational Research 236, 748–762.

Cinar, A., Salman, F.S., Bozkaya, B., 2019. Prioritized single nurse routing and scheduling for home healthcare services. European Journal of Operational Research . doi: 10.1016/j.ejor.2019.07.009.

Cissé, M., Yalçındağ, S., Kergosien, Y., Şahin, E., Lenté, C., Matta, A., 2017. Or problems related to home health care: A review of relevant routing and scheduling problems. Operations Research for Health Care 13, 1–22.

Crama, Y., Rezaei, M., Savelsbergh, M., Woensel, T.V., 2018. Stochastic inventory routing for perishable products. Transportation Science 52, 526–546.

Donlan, A., 2020. Hhcn poll: 31% of home-based care providers have been affected by covid-19. URL: https://homehealthcarenews.com.

Duque, P.M., Castro, M., Sörensen, K., Goos, P., 2015. Home care service planning. the case of landelijke thuiszorg. European Journal of Operational Research 243, 292–301.

Fikar, C., Hirsch, P., 2015. A matheuristic for routing real-world home service transport systems facilitating walking. Journal of Cleaner Production 105, 300–310.

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

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

Guido, R., Groccia, M.C., Conforti, D., 2018. An efficient matheuristic for offline patient-to-bed assignment problems. European Journal of Operational Research 268, 486–503.

Guo, G., Hackebeil, G., Ryan, S.M., Watson, J.P., Woodruff, D.L., 2015. Integration of progressive hedging and dual decomposition in stochastic integer programs. Operations Research Letters 43, 311–316.

Hernández-Leandro, N.A., Boyer, V., Salazar-Aguilar, M.A., Rousseau, L.M., 2019. A matheuristic based on lagrangian relaxation for the multi-activity shift scheduling problem. European Journal of Operational Research 272, 859–867.

Keskin, M., Çatay, B., 2018. A matheuristic method for the electric vehicle routing problem with time windows and fast chargers. Computers & Operations Research 100, 172–188.

Lanzarone, E., Matta, A., 2014. Robust nurse-to-patient assignment in home care services to minimize overtimes under continuity of care. Operations Research for Health Care 3, 48–58.

Li, Y., Chu, F., Chu, C., Zhu, Z., 2019. An efficient three-level heuristic for the large-scaled multi-product production routing problem with outsourcing. European Journal of Operational Research 272, 914–927.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA. pp. 281–297.

Moussavi, S.E., Mahdjoub, M., Grunder, O., 2019. A matheuristic approach to the integration of worker assignment and vehicle routing problems: Application to home healthcare scheduling. Expert Systems with Applications 125, 317–332.

Neves-Moreira, F., Almada-Lobo, B., Cordeau, J.F., Guimarães, L., Jans, R., 2018. Solving a large multi-product production-routing problem with delivery time windows. Omega . doi: 10.1016/j.omega.2018.07.006.

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

Oliveira, B.B., Carravilla, M.A., Oliveira, J.F., Costa, A.M., 2019. A co-evolutionary matheuristic for the car rental capacity-pricing stochastic problem. European Journal of Operational Research 276, 637–655.

Oliveira, B.B., Carravilla, M.A., Oliveira, J.F., Toledo, F.M., 2014. A relax-and-fix-based algorithm for the vehicle-reservation assignment problem in a car rental company. European Journal of Operational Research 237, 729–737.

Perumal, S.S., Larsen, J., Lusby, R.M., Riis, M., Sørensen, K.S., 2018. A matheuristic for the driver scheduling problem with staff cars. European Journal of Operational Research . doi: 10.1016/j.ejor.2018.11.011.

Pillac, V., Gendreau, M., Guéret, C., Medaglia, A.L., 2013. A review of dynamic vehicle routing problems. European Journal of Operational Research 225, 1–11.

Rockafellar, R.T., Wets, R.J.B., 1991. Scenarios and policy aggregation in optimization under uncertainty. Mathematics of operations research 16, 119–147.

Rodriguez, C., Garaix, T., Xie, X., Augusto, V., 2015. Staff dimensioning in homecare services with uncertain demands. International Journal of Production Research 53, 7396–7410.

Sahin, E., Matta, A., 2015. A contribution to operations management-related issues and models for home care structures. International Journal of Logistics Research and Applications 18, 355–385.

Shi, Y., Boudouh, T., Grunder, O., 2019. A robust optimization for a home health care routing and scheduling problem with consideration of uncertain travel and service times. Transportation Research Part E: Logistics and Transportation Review 128, 52–95.

Shi, Y., Boudouh, T., Grunder, O., Wang, D., 2018. Modeling and solving simultaneous delivery and pick-up problem with stochastic travel and service times in home health care. Expert Systems with Applications 102, 218–233.

Solyalı, O., Süral, H., 2017. A multi-phase heuristic for the production routing problem. Computers & Operations Research 87, 114–124.

Wirnitzer, J., Heckmann, I., Meyer, A., Nickel, S., 2016. Patient-based nurse rostering in home care. Operations Research for Health Care 8, 91–102.

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

Yu, Q., Fang, K., Zhu, N., Ma, S., 2019. A matheuristic approach to the orienteering problem with service time dependent profits. European Journal of Operational Research 273, 488–503.

Yuan, B., Liu, R., Jiang, Z., 2015. A branch-and-price algorithm for the home health care scheduling and routing problem with stochastic service times and skill requirements. International Journal of Production Research 53, 7450–7464.

Yuan, B., Liu, R., Jiang, Z., 2018. Daily scheduling of caregivers with stochastic times. International Journal of Production Research 56, 3245–3261.