Visiting Nurses Assignment and Routing for Decentralized Telehealth Service Networks

Tan Yu[1], Yongpei Guan[1], Xiang Zhong[1*]

[1*]Department of Industrial and Systems Engineering, University of Florida, Stadium Road, Gainesville, 32611, Florida, USA.

*Corresponding author(s). E-mail(s): xiang.zhong@ufl.edu;
Contributing authors: yutan@ufl.edu; guan@ise.ufl.edu;

Abstract

As telehealth utilization for ambulatory and home-based care skyrockets, there has been a paradigm shift to a decentralized and hybrid care delivery modality integrating both in-person and telehealth services provided at different layers of the care delivery network, i.e., central hospitals, satellite clinics, and patient homes. The operations of such care delivery systems need to take into consideration patients' mobility and care needs, and rely on multiple types of nurses who can support and facilitate telehealth (with hospital physicians) in clinics and patient homes. We formulate an optimization problem, aiming at operationalizing the proposed care delivery network. Decisions regarding the type of care delivered, the location of care delivered, and the scheduling of all kinds of nurses are determined jointly to minimize operating costs while simultaneously satisfying patients' care needs. We propose a bi-level approximation that exploits the structure of the hybrid telehealth system, and develop column generation-based heuristic algorithms to identify the joint decision rules for clinic selection, patient assignment, and visiting nurse routing problems. Numerical experiment results demonstrate our algorithm's capability to achieve high-quality solutions in reasonable computation time, and is capable of solving instances with large patient sizes and time windows. Our work supports the efficient and effective operation of the proposed hybrid telehealth systems to improve patient access to care.

Keywords: Vehicle Routing Problem, Telehealth, Column Generation, Heuristic Algorithm

1 Introduction

An analysis of several national survey data estimated that 3.6 million people in the United States did not obtain sufficient medical care due to transportation barriers [1]. To address patient transportation issues, various strategies have been implemented by healthcare organizations, including 1) compensating their travel through public transportation, such as the provision of bus passes, taxi/transport vouchers, or transportation cost reimbursement, as well as 2) free agency services such as arranging or connecting patients to transportation, and 3) providing in-house transportation for patients (e.g., free shuttle service) [2]. Despite these efforts, the Health Research & Educational Trust identified several major problems that hospitals and patients still face, such as limited availability and routes, high cost of fares, long travel distances and lengthy wait times, and inconvenient time schedules. While these issues remained to threaten healthcare access during the pandemic, they also prompted a new pivot in the proliferation of telehealth.

Telehealth technology enables remote patient monitoring, communication, and delivery of health-related services, reducing the need for frequent in-person visits and associated travel burdens. According to the American Medical Association (AMA) 2021 Telehealth Survey Report, 85% of physician respondents now use some form of telehealth and 60% of clinicians agree or strongly agree that telehealth enabled them to provide high-quality care. The growth of telehealth has also facilitated the integration of telehealth into home healthcare (HHC), augmenting the traditional HHC with access to care providers residing in central hospitals and specialized services that visiting nurses cannot provide. The integration of telehealth into HHC implies a decentralized and hybrid care delivery network that has the potential to improve patient access to healthcare, especially for patients that suffer from transportation barriers such as the aging population. Yet, the realization of the proposed decentralized and hybrid care delivery requires careful planning and execution to ensure that it is implemented effectively, and the benefits are achieved without compromising the quality of care.

In this study, we examine a central hospital that manages a group of patients in a broad catchment area of the healthcare organization. These patients require regular healthcare services, including specialty care that is typically not provided at community clinics or patient homes as part of the HHC. In addition to the in-person appointments at the central hospital, we consider telehealth with hospital physicians provided at patient homes or nearby community clinics. Some patients could experience difficulties accessing healthcare facilities due to various factors, such as disability, lower income, older age, and rural residency (long-distance travel). For these patients, we consider employing travel nurses to visit their homes so they can receive basic nursing care as they would have when visiting the hospital in person. Patients also have the option to visit nearby clinics, and their nursing care (e.g., vital check, basic examinations) will be supported by nurses from participating community clinics prior to, during, or right after their telehealth with the hospital physician. The illustration of our design is shown in Fig. 1.

To operationalize the proposed hybrid telehealth system, the type of care (e.g., telehealth vs. in-person) delivered to individual patients, the location (e.g., patient

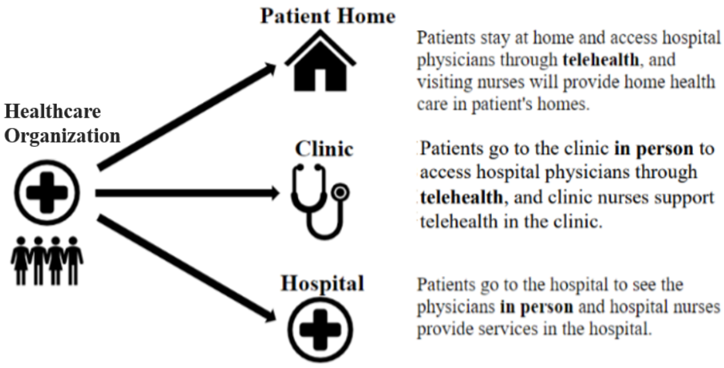


Fig. 1 The illustration of decentralized hybrid care delivery

home, community clinics) of care delivered, and the assignment of all kinds of nurses based on their work location, need to be determined. Lowering the operating cost while providing quality care is essential to ensure the sustainability of this service modality, envisioning the steady growth of telehealth service demands. In addition to minimizing operating costs, there is also a vested interest in maximizing social welfare by accounting for the disutility of patients who need to travel and thus bear the inconvenience or travel cost.

Our problem is related to the Vehicle Routing Problem with Time Window (VRPTW) and multi-depot VRP (MDVRP) [3] as the hospital assigns visiting nurses from various clinics to patients while observing capacity and time window constraints. The foremost challenge in our problem pertains to computational complexity, which is notably pronounced due to the inherent NP-hard nature of the VRPTW. This challenge is further compounded by the necessity to address the intricate aspects of clinic selection that affect patient assignment and visiting nurse routing, thereby engendering a complicated optimization problem. Consequently, novel approaches are required to incorporate patient features, cost factors, and time windows, to achieve the best assignments for all types of nurses, while maintaining computational efficiency. In this regard, we establish a set partitioning model, accompanied by the introduction of a highly effective column generation-based heuristic algorithm to tackle the problem at hand. This column generation-based heuristic algorithm encompasses a construction heuristic, a local search algorithm, and a heuristic labeling algorithm that identifies negative reduced-cost routes effectively. We further perform numerical experiments to demonstrate the superiority of the proposed approaches over commercial solvers using synthetic data mimicking the real-world problem setting.

The contributions of this paper are summarized as follows. First, we present a novel decentralized care delivery design for integrating telehealth services and HHC to improve patient access to care. A comprehensive mathematical formulation is developed to characterize the unique operational problem associated with this service network design. Second, to solve our problem with multiple decisions made jointly for a large size of patients, which entails a large mixed integer programming problem, we develop a bi-level column generation-based heuristic algorithm. The numerical analysis demonstrates that our algorithm is highly efficient, consistently yielding high-quality solutions within remarkably short computation times, even for instances involving up to hundreds of patients — a scale that significantly exceeds the solver's capability in providing practical solutions. Beyond these contributions, our solution offers decision-makers valuable insights into disparities arising from the inclusion or exclusion of patient disutility, as well as the impact of various operating factors. We believe this work holds significance in promoting patient access to care through better service system design and operation in the digital health era.

The remainder of this paper is organized as follows. Section 2 introduces the relevant literature. Section 3 provides an explicit description of the problem, along with the development of the mathematical formulation. Section 4 presents a bi-level model of the original problem and proposes a column generation-based heuristic algorithm for solving the problem. In addition to cost minimization, patient disutility is also examined as part of the objective function. Section 5 provides details about the numerical experiments and reports the computational results. Finally, discussions and concluding remarks are presented in Section 6.

2 Literature review

To the best of our knowledge, the integration of facility location problems, patient assignment problems, and vehicle routing problems with time windows (VRPTW) and multi-depot into a nested model has not been extensively explored in the literature. Therefore, we will briefly review the existing research on the classical VRPTW model, followed by a discussion of the relevant literature concerning column generation methods and heuristic strategies for solving VRPTW problems. This serves as the methodological foundation of our proposed model. Then, we explore studies that focus on HHC routing and scheduling issues as well as location selection problems.

The vehicle routing problem (VRP) was originally formulated by Dantzig and Ramser [4] in 1959. Over the past six decades, significant research efforts have been devoted to studying the optimization of routing problems and their extensions [5–7]. In recent years, new variants of the VRP with more complex objectives and constraints have emerged. One such variant is the VRPTW, where the service at each customer location starts within a given time window. It has been proven that the VRPTW is an NP-hard problem [8].

Many researchers have attempted to derive exact algorithms to solve the VRPTW. The column generation algorithm is a commonly used exact algorithm that has been successfully applied to solving large-scale combinatorial optimization problems, such as the production scheduling problem [9], the surgical scheduling problem [10], and the cutting stock problem [11], to name a few. In the context of VRPs, Desrochers et al. [12] proposed a column generation approach that accurately solved seven of Solomon's benchmark instances [3] exactly by decomposing the problem into sets of customers visited by the same vehicle and selecting the optimal routes between all possible ones.

Due to the exponential time complexity of exact approaches, it is unlikely to produce optimal solutions for practical-sized VRPTW with reasonable computational time. As a result, near-optimal solutions have been sought using heuristic approaches.

These approaches can be broadly categorized as construction heuristics, improvement heuristics, and composite heuristics. Construction heuristics are employed to construct feasible solutions by sequentially inserting unrouted customers into partially constructed routes until all customers are included. One such construction heuristic, the sequential insertion algorithm, was introduced by Solomon [3] and its parallel version was later implemented by Potvin and Rousseau [13]. Although construction heuristics often produce solutions rapidly, the quality of such solutions may not always be optimal. As a result, construction heuristics are typically used to generate initial solutions followed by improvement heuristics or other two-phase heuristics. Improvement heuristics, on the other hand, aim to improve upon an existing solution by performing local searches for better neighboring solutions, within the neighborhoods generated by node/edge-swapping operators. Edge-exchange heuristics are examples of improvement heuristics, and have been studied by various researchers including Potvin and Rousseau [14], Savelsbergh [8, 15], and Taillard et al. [16]. A composite heuristic, which combines route construction and improvement procedures, was proposed by Russell [17]. More recently, Yuan et al. [18] proposed a set partitioning model for the generalized vehicle routing problem and developed a heuristic algorithm based on column generation. This approach combines constructing heuristics, path optimization, local search operators, and a heuristic process to provide negative cost-reducing paths. Given the time window structure of the problem at hand, which is well-suited for a branch-and-price resolution scheme, and inspired by Yuan et al. [18], we develop a series of heuristics to design the route for visiting nurses in our study.

Our work is closely related to the Home Healthcare Routing and Scheduling Problems (HHCRSP), which incorporates variants of VRPTW. HHCRSPs have to account for time windows (when patients are available) [19–22], and are subject to precedence constraints, such as multiple depots [20, 21, 23, 24] and the service skill preference [21, 25–31]. The incorporation of multiple depots enables nurses to commence their routes from various locations. Erdem and Koç [21] explored an HHCRSP that focuses on minimizing the total travel time across multiple depots with preference constraints. Bahadori-Chinibelagh et al. [32] proposed a multi-depot routing model to optimize healthcare logistics (nurse–patient–pharmacy–laboratory), aiming to minimize transportation costs and travel distances within constraints like vehicle capacity and patient time windows. Meanwhile, service preference includes both care worker and patient considerations. The first kind of such preference aims to ensure that the qualifications and expertise of nursing staff fulfill the stipulated requirements. Addressing the issue of variable nursing skill levels, Demirbilek et al. [33] developed a heuristic algorithm for a dynamic HHCRSP to accommodate different patient needs. The second one related to patient preference for caregiver (e.g., gender, language). In this context, Yadav and Tanksale [34] proposed a generalized model aimed at maximizing revenue, taking into account factors like patient preferences for caregiver gender and language.

Additionally, the locations of HHC facilities also have an impact on HHC operations. Within the context of HHC, this is commonly referred to as Location Routing Problems (LRP). In LRP, the location of facilities and the distribution routes are integral considerations. A commonly employed structure in LRP is a bi-level model. Fathollahi-Fard et al. [35] presented a location-allocation-routing model that integrates the location of pharmacies and laboratories, the assignment of patients, as well as the routing and scheduling of caregivers. Later, Fathollahi-Fard et al. [36] extended their research on the location-allocation-routing problem in the context of HHC. They formulated a bi-level programming model framed as a static Stackelberg game. Note that our problem is different from Fathollahi-Fard et al. [36], despite both studies concentrating on location-allocation-routing problems within the healthcare system. Unlike the single-nurse type model Fathollahi-Fard et al. [36], our framework incorporates multiple types of nurses, each restricted to providing services in specific places (hospital, community clinic, and patient's home). Additionally, the central hospital serves as the decision-maker in our model, in contrast to the bi-level decision-making involving both nurses and patients in Fathollahi-Fard et al. [36]. Dai et al. [37] proposed an extended model of traditional LRP, aiming to determine both the HHC center location and the caregiver route plan in a manner that minimizes construction costs, travel costs, and carbon emission costs. To summarize, Our model also integrates considerations for patient preferences and explores the implications for social welfare. These unique aspects render our problem worthy of further investigation, and set our work apart from existing studies in the field of HCC.

3 Problem Description

To improve patient access to care, we propose a decentralized care delivery design that enables patients with significant transportation barriers to receive telehealth services at home or nearby community clinics. To accommodate patient needs, we introduce two distinct patient classes: Type I patients, who are only eligible for receiving care within the familiarity of their homes, and Type II patients, who display no specific inclination towards either home-based care or alternate settings. We also assume that there are sufficient physicians at the central hospital to provide office visits or telehealth visits, as the physician's schedule will not be impacted if appointments are changed from in-person to virtual. The main operation process of the hybrid telehealth system can be outlined as follows:

In the planning phase, the healthcare organization will offer some time slots to patients based on hospital physicians' availability and then collect information from patients, including the patient's home address, anticipated service time, the level of mobility impairment, and other necessary information. This information can be employed to classify patients into two types. Type I patients demand exclusive home-based care, such as patients with mobility assessments falling under levels 1 and 2 [38]. Meanwhile, Type II comprises patients with the capability to access nearby healthcare facilities, making them eligible for a range of care options, including home-based care, clinic-based care, and hospital-based care.

In the next stage, the healthcare organization will determine which patients should receive telehealth services at home or at nearby clinics, or office visits at the central hospital, and determine the number and type of nurses needed. This is achieved by collaborating with community clinics and utilizing nurses at each participating clinic. If the central hospital decides to collaborate with a community clinic, a set-up cost is incurred. This cost can be interpreted as a contract fee for using the equipment and space of the clinic, separated from nurse salaries. In addition, nurse salaries vary based on their work locations. The healthcare organization aims to minimize its operating costs while ensuring high-quality care services tailored to each patient's type and scheduled within their preferred time windows. Cost reduction is a critical priority for ensuring long-term financial sustainability, even in not-for-profit hospitals.

On the day of operation, clinic and hospital nurses provide services to patients visiting clinics and the central hospital, while travel nurses visit patients' homes. Each visiting nurse is assigned a predetermined route, responsible for carrying out all service-related activities on that route. They start at a designated clinic, gather the required equipment, drive the vehicle to visit patients one by one according to the planned route, and return to the same clinic.

Fig. 2 illustrates the network structure of the hybrid telehealth system in our problem. The solid arrow lines depict patients who are assigned to the visiting nurse, while the dashed arrow lines represent patients assigned to the clinic nurse. The remaining patients are allocated to the central hospital. Note that this figure does not display the specific order in which the clinic nurse and hospital nurse attend to patients. It only illustrates the sequential order in which the visiting nurse cares for patients.

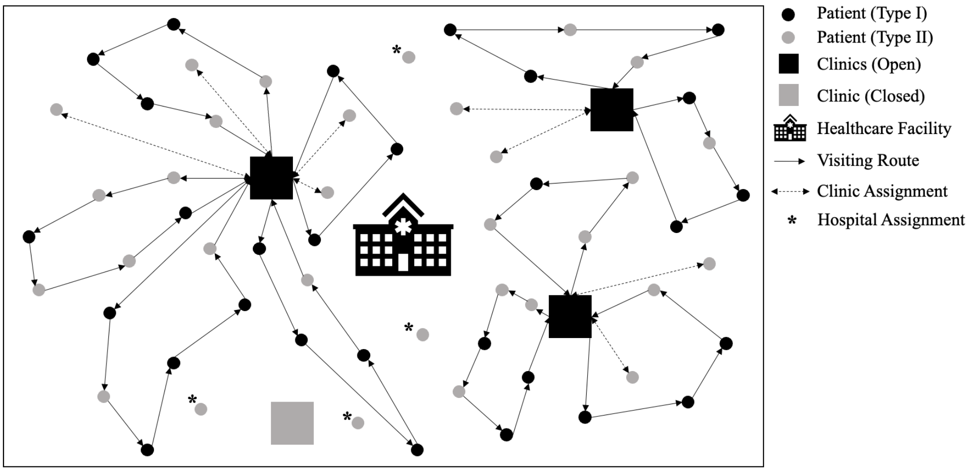


Fig. 2 The network structure of the assignment and routing problem

3.1 Model Setup and Notations

3.1.1 Model Parameters

We first introduce the parameters involved in the decision-making problem. Notably, each instance of assignment and routing is associated with a distinct set of parameters to represent daily operations, which are enumerated as follows:

- I: the set of all healthcare facilities located in the catchment area of the healthcare organization, comprising a central hospital indexed as 0, as well as several candidate community clinics to collaborate with, indexed as $i > 0$;
- J: the set of patients, who are further classified into two mutually exclusive groups, denoted as J_1 and J_2, for Type I and Type II patients, respectively;
- K: the set of all nurses at the disposition of the healthcare organization, which includes three distinct subsets: visiting nurses (K_v), clinic nurses (K_n), and hospital nurses (K_h);
- f_i: the collaboration fee charged by clinic i ($i \in I \setminus \{0\}$) in the event that the central hospital seeks to partner and share nurses with;
- $[a_j, b_j]$: time window of patient $j \in J$, and the time window is $[0, 12]$, i.e., a 12-hour window for all healthcare facilities;
- v_j: the service time of patient $j \in J$;
- l_k: the salary of nurse for nurse $k \in K$;
- e_{ij}: the traveling cost from clinic/patient i to clinic/patient j for visiting nurses;
- t_{ij}: the traveling time from clinic/patient i to clinic/patient j;
- g_{ij}: the disutility measure of the patient $j \in J$ when being assigned to travel to nearby clinics or the hospital $i \in I$, which is proportional to the distance traveled as a crude proxy of the monetary cost or the inconvenience incurred; the same disutility rate is applied to all patients.

For simplicity of exposition, we assume that $t_{ij} = t_{ji}, \forall i, j \in I \cup J$ and $e_{ij} = e_{ji}, \forall i, j \in I \cup J$. We also assume that once a nurse begins to serve patient j, s/he cannot quit before the end of the service. In addition, all the visiting and service times considered in this analysis are deterministic.

3.1.2 Decision Variables

Next, we define the set of decision variables:

- $z_i \in \{0, 1\}$: 1 if healthcare facility $i \in I$ is open, and 0 otherwise;
- $p_k^i \in \{0, 1\}$: 1 if nurse $k \in K$ needs to work in healthcare facility $i \in I$, and 0 otherwise;
- $w_j^{ik} \in \{0, 1\}$: 1 if nurse k in healthcare facility $i \in I$ is assigned to patient j, and 0 otherwise;
- $x_{j_1 j_2}^{ik} \in \{0, 1\}$: 1 if nurse $k \in K$ in healthcare facility $i \in I$ serves patient $j_1 \in J$ and then patient $j_2 \in J \setminus \{j_1\}$, and 0 otherwise;
- $s_j^k \in \mathbb{R}$: a variable that defines the instant in time at which nurse $k \in K$ will serve patient $j \in J$;

As a remark, f_i the collaboration fee here is independent of the number of nurses used. If the fee is proportional to the number of nurses used, we can set $f_i = 0$ and absorb the extra cost into the salary of nurses (l_k). The exact cost associated with community clinic partnership could vary significantly from being substantially greater than a nurse's daily salary to only slightly exceeding it. Specifically, if the community clinics are operated by the hospital, the partnership cost can be minimal. To account for this variability, we perform a sensitivity analysis of its impact on the solution structure in Section 5.2.3.

3.1.3 Model Constraints

For the constraints, it should be noted that for each type of nurse, there exists a candidate set from which we make assignments. Specifically, nurses $k \in K_v \cup K_n$ are associated with clinics $i \in I \setminus \{0\}$, whereas nurses $k \in K_h$ are associated with the hospital ($i = 0$). Additionally, visiting nurses $k \in K_v$ are assigned to patients $j \in J$, while clinic and hospital nurses $k \in K_n \cup K_h$ can only be assigned to patients $j \in J_2$. We first handle the redundant variables in the following manner:

$$z_0 = 1, \tag{1}$$

$$p_k^i = 0, \forall i \in I \setminus \{0\}, k \in K_h \tag{2}$$

$$w_j^{ik} = 0, \forall i \in I \setminus \{0\}, k \in K_h, j \in J \tag{3}$$

$$x_{j_1 j_2}^{ik} = 0, \forall i \in I \setminus \{0\}, j_1 \in J, j_2 \in J \setminus \{j_1\}, k \in K_h \tag{4}$$

Constraint (1) mandates that the hospital must continue its operation. Constraints (2), (3), and (4) are formulated to allocate hospital nurses exclusively to services rendered at the central hospital.

The following mathematical formulation determines the schedule of the nurses while considering patients' features.

$$\sum_{i \in I} \sum_{k \in K_v} w_j^{ik} = 1, \forall j \in J_1 \tag{5}$$

$$\sum_{i \in I} \sum_{k \in K} w_j^{ik} = 1, \forall j \in J_2 \tag{6}$$

$$\sum_{i \in I} p_k^i \leq 1, \forall k \in K \tag{7}$$

$$p_k^i \leq z_i, \forall i \in I \setminus \{0\}, k \in K_v \cup K_n \tag{8}$$

$$w_j^{ik} \leq p_k^i, \forall i \in I, j \in J, k \in K \tag{9}$$

Constraints (5) ensure that patients classified as Type I are assigned to visiting nurses, and constraints (6) ensure that patients classified as Type II are assigned to one, and only one, type of nurse. We refer to these as mobility constraints, as they establish the connection between the patient features and the type of nurse assigned. To guarantee the optimal allocation of resources, constraints (7) impose the limitation that each nurse can only be assigned to a single job. Moreover, constraints (8) stipulate that only those visiting nurses and clinic nurses who are assigned to open clinics can be selected for a job. Additionally, the interrelation between nurse and patient assignments is governed by the constraints (9).

The routing must adhere to specific requirements, as indicated by the constraints below.

$$\sum_{j_2 \in J \setminus \{j_1\}} x_{j_1 j_2}^{ik} = w_{j_1}^{ik}, \forall i \in I, j_1 \in J, k \in K \tag{10}$$

$$\sum_{j_2 \in J \setminus \{j_1\}} x^{ik}_{j_2 j_1} = w^{ik}_{j_1}, \forall i \in I, j_1 \in J, k \in K \tag{11}$$

$$\sum_{j_1 \in J \cup \{i\} \setminus \{j\}} x^{ik}_{j_1 j} = \sum_{j_2 \in J \cup \{i\} \setminus \{j\}} x^{ik}_{j j_2}, \forall i \in I, j \in J, k \in K \tag{12}$$

$$\sum_{j \in J} x^{ik}_{ij} = p^i_k, \forall i \in I, k \in K \tag{13}$$

$$\sum_{j \in J} x^{ik}_{ji} = p^i_k, \forall i \in I, k \in K \tag{14}$$

$$s^k_{j_1} + t_{j_1 j_2} + v_{j_1} - M(1 - x^{ik}_{j_1 j_2}) \leq s^k_{j_2}, \forall i \in I, j_1 \in J, j_2 \in J \setminus \{j_1\}, k \in K_v \tag{15}$$

$$s^k_{j_1} + v_{j_1} - M(1 - x^{ik}_{j_1 j_2}) \leq s^k_{j_2}, \forall i \in I, j_1 \in J_2, j_2 \in J \setminus \{j_1\}, k \in K_n \cup K_h \tag{16}$$

Constraints (10) and (11) specify that nurses must visit each patient on their assigned route only once. Constraints (12) ensure flow conservation at each node for every nurse, while constraints (13) and (14) guarantee that each nurse departs from and returns to the same clinic after serving their designated patients. Due to the specific time window structure of our problem, the conventional subtour elimination constraints are unnecessary. Time constraints for nurse-patient interactions are defined in constraints (15), where the positive constant M is set to be sufficiently large (e.g. $M \geq |J|(\max_{j_1, j_2 \in J} t_{j_1 j_2} + \max_{j \in J} v_j)$). Similarly, constraints (16) define the time at which clinic and hospital nurses will attend to patients, without the need to take travel time into consideration.

3.1.4 Objective Functions

Since we assume the physician cost will be the same regardless of the service modality, we consider the total cost incurred by the nurses, Φ_{cost}, which encompasses their salaries as well as the travel expenses for the visiting nurses, and the clinic collaboration cost.

$$\Phi_{cost} = \sum_{i \in I} \sum_{k \in K} l_k p^i_k + \sum_{i \in I} \sum_{k \in K_v} \sum_{j_1 \in I \cup J} \sum_{j_2 \in I \cup J \setminus \{j_1\}} e_{j_1 j_2} x^{ik}_{j_1 j_2} + \sum_{i \in I} f_i z_i \tag{17}$$

Additionally, we consider the inclusion of patients' travel disutility to minimize the total cost from both the healthcare organization and the patient's side. This allows us to explore how the assignment and routing solutions would differ if the healthcare organization assumed the role of a social planner instead of a profit-driven entity. To facilitate clear reference, we denote the objective function in this context as Φ_{social}.

$$\Phi_{social} = \Phi_{cost} + \sum_{j \in J_2} \sum_{i \in I} \sum_{k \in K_n \cup K_h} w^{ik}_j g_{ij} \tag{18}$$

Notably, the two objective functions become the same when the patient disutility factor $g_{ij} = 0$. Furthermore, we exclude constraints (1) to (4) as they are predetermined and exempted from the optimization process. Therefore, the general mixed-integer programming (MIP) formulation for our cost-minimization problem is as follows:

$$\begin{aligned}
\min_{\mathcal{X}} \quad & \Phi_{social} \\
\text{s.t} \quad & (5) - (16) && (19) \\
& z_i \in \{0,1\}, \forall i \in I && (20) \\
& p_k^i \in \{0,1\}, \forall i \in I, k \in K && (21) \\
& w_j^{ik} \in \{0,1\}, \forall i \in I, j \in J, k \in K && (22) \\
& a_j \geq s_j^k \geq b_j, \forall j \in J, k \in K && (23) \\
& x_{j_1 j_2}^{ik} \in \{0,1\}, \forall i \in I, j_1 \in J \cup I, j_2 \in J \cup I \setminus \{j_1\}, k \in K && (24)
\end{aligned}$$

where $\mathcal{X}$ defines a joint vector of decision variables, i.e., $\mathcal{X} = (z, p, w, s, x)$. In real practice, the number of variables in the problem at hand is substantial, while the number of constraints is comparatively low. Given the significant number of variables involved, particularly those arising from multiple indices, the process of exhaustive enumeration can prove time-consuming even for readily available MIP solvers. In light of this, we opt to address the MIP problem through a column-generation approach.

3.2 Set Partitioning Formulation

Consider a directed graph $G = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ and $\mathcal{A}$ are its node and arc sets, respectively. Each node in $\mathcal{V}$ corresponds to either the location of a healthcare facility $i \in I$ or a patient $j \in J$. We define the neighborhoods for each patient j, denoted as $\mathcal{N}_j$ based on patient types. Specifically, for Type I patients, $\mathcal{N}_j = \{j_1 \in J | a_j + v_j + t_{jj_1} \leq b_{j_1}\}, \forall j \in J_1$, whereas for Type II patients, $\mathcal{N}_j = \{j_1 \in J_2 | a_j + v_j \leq b_{j_1}\} \cup \{j_1 \in J_1 | a_j + v_j + t_{jj_1} \leq b_{j_1}\}, \forall j \in J_2$. This definition of neighbourhoods allows us to narrow down to the pool of patients who can be visited afterward in the given time window. In addition, the arc set $\mathcal{A}$ defines possible routes between healthcare facilities and patients, as well as between individual patients, i.e., $\mathcal{A} = \{(i,j) | i, j \in I \cup J, i \neq j\}$, and we primarily consider the directed lines connecting patient j to any other patient i in their neighbourhood set $\mathcal{N}_j$.

To concurrently generate routes for visiting nurses and schedules for clinic and hospital nurses, we expand the notion of a "route" r to include not only the path traveled by visiting nurses but also the order in which they attend to patients in clinics and hospitals. Let Ω_v, Ω_n, and Ω_h denote the sets of feasible routes, respecting the time and patient mobility constraints, for visiting nurses, clinic nurses, and hospital nurses, respectively. The union of these sets, i.e., $\Omega_v \cup \Omega_n \cup \Omega_h$, represents the complete set of feasible routes, denoted by Ω. Note that for the hospital and clinic nurses, the route is nominal and is rather the sequence of patient appointments.

Let α_r^j, where $r \in \Omega$ and $j \in J$, denote the binary coefficients, being equal to 1 if patient j is visited on route r, and 0, otherwise. Let β_r^k, where $r \in \Omega$ and $k \in K$, denote the binary coefficients, being equal to 1 if nurse k is required to travel on route r. Each route $r \in \Omega$ is associated with a cost, denoted as ϕ_r. It is worth noting that the calculation of ϕ_r for $r \in \Omega_v$ and for $r \in \Omega_n \cup \Omega_h$ differ slightly. Specifically, for $r \in \Omega_v$, ϕ_r is the sum of the visiting nurse's salary and their traveling costs. On the other hand, for $r \in \Omega_n \cup \Omega_h$, the function ϕ_r represents the aggregate compensation received by the nurses employed at a healthcare facility. In the context of social welfare, ϕ_r accounts for both the aforementioned nurse compensation as well as the expenses borne by the patient for transportation. Lastly, we introduce the binary variable θ_r, with $r \in \Omega$, which is equal to 1 if and only if the route r is selected.

Note that the clinic selection problem must be resolved beforehand. This is because including the clinic collaboration fee at this stage would result in an inability to calculate the reduced cost of new routes using the dual variables corresponding to the relevant constraints. There are $2^{|I|-1}$ scenarios for clinic selection (does not include the hospital), each denoted as $\bar{s}$ and collectively represented by the set $\mathcal{S}$. Given a scenario $\bar{s} \in \mathcal{S}$, our column generation model, which is based on the Set Partitioning model, can be described as follows:

$$[MP(G)_{\bar{s}}] \quad \min \quad \sum_{r \in \Omega} \phi_r \theta_r \tag{25}$$

$$\text{s.t.} \quad \sum_{r \in \Omega_v} \alpha_r^j \theta_r = 1, \forall j \in J_1 \tag{26}$$

$$\sum_{r \in \Omega} \alpha_r^j \theta_r = 1, \forall j \in J_2 \tag{27}$$

$$\sum_{r \in \Omega} \beta_r^k \theta_r \leq 1, \forall k \in K \tag{28}$$

$$\theta_r \in \{0, 1\}, \forall r \in \Omega \tag{29}$$

The optimization problem described by formulations (25)-(29) aims to minimize the total cost subject to various constraints. Specifically, constraints (26) ensure that each Type I patient is visited exactly once, while constraints (27) guarantee the same for Type II patients. Constraints (28) ensure that each type of nurse is assigned to at most one route. It should be noted that by first addressing the clinic selection problem, the introduction of a constraint to define the interrelation between routes and open clinics becomes redundant, and the feasible route pool Ω only contains routes where the depots and destinations are limited to the open facilities.

Although the linear programming (LP) relaxation of the set partitioning formulation typically yields tight bounds, the number of potential routes in the set Ω can increase exponentially as the number of patients to be visited grows, rendering generating all feasible routes impractical. To address this challenge, we introduce a much smaller route pool, denoted as Ω', and solve a restricted version of the set partitioning model, considering only the routes in Ω'.

The route pool is dynamically populated following a branch-and-price procedure [39]. It consists of two main components: a pricing problem algorithm, used to generate new routes, denoted as columns, at each iteration, and branching rules, which determine how to partition the feasible region into subsets for recursive application of the algorithm until exhaustion. The restricted master problem (RMP), denoted by

RMP(Ω'), is defined as the LP relaxation of a subproblem consisting of a restricted set of columns generated so far. At each node of the search tree, the column generation method iteratively solves the RMP and a pricing problem. The goal of the pricing problem is either to generate columns with the most negative reduced costs based on the dual solution of the current RMP(Ω') or to prove that none exists. Newly generated columns are introduced to RMP(Ω') at each iteration, and the process terminates and a lower bound for the corresponding node is obtained whenever no additional column prices out favorably.

Now we only consider the following dual variables of RMP(Ω'):

- π_j^1: dual variable corresponding to constraints (26) for Type I patient $j \in J_1$;
- π_j^2: dual variable corresponding to constraints (27) for Type II patient $j \in J_2$;
- π_k^3: dual variables corresponding to constraints (28) for nurses $k \in K$.

In each iteration of the branch-and-price process, the objective of the column generation subproblem is to identify a feasible nurse route originating from each open clinic, which has the minimum reduced cost with respect to the current dual solution of the RMP(Ω'). The reduced cost of the visiting, clinic, and hospital nurse routes, namely $\bar{c}_r^v$, $\bar{c}_r^n$, and $\bar{c}_r^h$, can be expressed as:

$$\bar{c}_r^v = \phi_r - \sum_{j \in J_1} \alpha_r^j \pi_j^1 - \sum_{j \in J_2} \alpha_r^j \pi_j^2 - \sum_{k \in K_v} \beta_r^k \pi_k^3, \ \forall r \in \Omega_v' \tag{30}$$

$$\bar{c}_r^n = \phi_r - \sum_{j \in J_2} \alpha_r^j \pi_j^2 - \sum_{k \in K_n} \beta_r^k \pi_k^3, \ \forall r \in \Omega_n' \tag{31}$$

$$\bar{c}_r^h = \phi_r - \sum_{j \in J_2} \alpha_r^j \pi_j^2 - \sum_{k \in K_h} \beta_r^k \pi_k^3, \ \forall r \in \Omega_h' \tag{32}$$

4 A Column Generation Based Heuristic Approach

In this section, we introduce our bi-level approach, which consists of an upper-level stage and a lower-level stage. The upper-level stage revolves around the determination of clinic openings, while the lower-level stage is aimed at deriving the best solution based on the determined clinic set. Within the lower-level stage, three key algorithms come into play: the construction heuristic algorithm, responsible for generating the initial route pool; the labeling algorithm, identifying routes with negative reduced costs; and the local search algorithm, aimed at improving the best solution and expanding the route pool. We will begin by introducing the algorithms in the lower-level stage, followed by the approach of the upper-level stage and the overall procedure.

4.1 The Construction Heuristic Algorithm

The heuristic algorithm presented in this section utilizes an iterative approach to generate a set of feasible initial routes to serve all patients. This approach builds upon the concept of *pivot* patients [18]. For our study, the *pivot* patients are identified as those who cannot be visited simultaneously by the same nurse, or those who are located far from both the healthcare facility and other patients. An assumption is posited wherein a sufficient number of nurses are available for allocation to each *pivot* patient, thereby ensuring the feasibility of the initial routes within the algorithmic framework. Specifically, we select at most $N_{\text{cardinality}}$ *pivot* patients using two criteria described below:

Criterion 1: Due to the time windows, some patients cannot be visited on the same route. Patients i and j are considered incompatible and denoted by the pair $\langle i, j\rangle$, if they cannot be served on the same route:

$$\langle i, j\rangle = \{i, j \in J_n | n \in \{1, 2\}, a_i + t_{ij} > b_j, a_j + t_{ij} > b_i\} \tag{33}$$

To identify the maximal set of patients that cannot be visited on the same routes, a graph $\bar{G}$ is constructed utilizing all incompatible pairs, and a maximum clique is sought. A recursive backtracking algorithm [40] is used to search for the maximal clique in graph $\bar{G}$.

Criterion 2: For each patient j, we determine a set of healthcare facilities as $\mathcal{H}_j$, which includes all the healthcare facilities connected with j, such that $i \in \mathcal{H}_j$ is the healthcare facility from which j can be reached within the time windows in the route design. At the same time, we also determine a set of patients $\mathcal{B}_j$, which includes patients from whom j can be reached, and patients who can be reached from j. Note that $\mathcal{B}_j$ is different from $\mathcal{N}_j$, which only includes the patients that can be reached from patient j. Formally, if j is a Type I patient, then $\mathcal{H}_j = \{i | t_{ij} \leq b_j, \forall i \in I\}$ and $\mathcal{B}_j = \{j_2 \in J | a_{j_2} + v_{j_2} + t_{j_2 j} \leq b_j\} \cup \{j_2 \in J | a_j + v_j + t_{j j_2} \leq b_{j_2}\}$. Otherwise, if j is a Type II patient, then $\mathcal{H}_j = \{i | i \in I\}$ and $\mathcal{B}_j = \{j_2 \in J_1 | a_{j_2} + v_{j_2} + t_{j_2 j} \leq b_j\} \cup \{j_2 \in J_1 | a_j + v_j + t_{j j_2} \leq b_{j_2}\} \cup \{j_2 \in J_2 | a_{j_2} + v_{j_2} \leq b_j\} \cup \{j_2 \in J_2 | a_j + v_j \leq b_{j_2}\}$. We then calculate the average traveling time from patient j to the healthcare facilities in $\mathcal{H}_j$, $w^j_{avg} = \frac{1}{|\mathcal{H}_j|} \sum_{i \in \mathcal{H}_j} t_{ij}$, and the minimal traveling time from patient j to the patients in $\mathcal{B}_j$, $w^j_{min} = \min_{j_2 \in \mathcal{B}_j} t_{j j_2}$. Then, we define a score w_j for each patient $j \in J$ as follows:

$$w_j = w^j_{avg} + w^j_{min}. \tag{34}$$

We rank patients from the highest value of w_j to the lowest; in other words, we prioritize patients who are either far away from the healthcare facility and/or far away from other patients. As a remark, $\mathcal{H}_j$ will be initially obtained with all clinics open and updated during the bi-level procedure.

Selection Procedure: Our goal is to curtail the size of the patient set to prioritize the assignment of the $N_{\text{cardinality}}$ most difficult patients for service. We introduce three sets, namely P_1 and P_2, obtained using Criterion 1 and Criterion 2, respectively, and the *pivot* patient set P_{new}, selected based on the following steps: If $|P_1| < N_{\text{cardinality}}$, the *pivot* patient set is defined as the union of P_1 and the top $N_{\text{cardinality}} - |P_1|$ patients from the set $P_2 \setminus (P_1 \cap P_2)$. If $|P_1| = N_{\text{cardinality}}$, then P_1 is designated as the *pivot* patient set P_{new}. Otherwise, P_1 is sorted based on the order in which the patients appear in P_2, and the top $N_{\text{cardinality}}$ patients in P_1 are selected as the *pivot* patient set P_{new}.

For the construction heuristic, for each patient p in P_{new}, a route with the lowest cost that only includes patient p is generated and added to the initial route pool Ω'.

Patients from the *non-pivot* patient set are then inserted into these routes following the pseudocode outlined in Appendix A.1.

4.2 Labelling Algorithm

In order to compute the low-cost path conforming to time windows, a labeling algorithm is developed. A patient j in route r is associated with a label L_j^r comprising four components: the cost, the reduced cost, the starting time of a feasible partial path starting at j, and the label of the previous patient i, who is in the same route r visited before the patient j, represented by $L_j^r = (c_j, \bar{c}_j, \bar{a}_j, L_i^r)$. The set of all labels associated with patient j is denoted as $\mathcal{L}(j) = \cup_{r\in\Omega_j} L_j^r$, where Ω_j is the set of all candidate routes including patient j. To improve the algorithm's efficiency, only non-dominated labels are retained. A label $L_j^1 \in \mathcal{L}(j)$ dominates a label $L_j^2 \in \mathcal{L}(j)$ if and only if $c_j^1 \leq c_j^2$ and $\bar{a}_j^1 \leq \bar{a}_j^2$. Due to the computational complexity and time constraints involved in generating an exhaustive set of routes in our problem, we limit the number of routes to ensure that the algorithm can complete its execution in a reasonable time. Specifically, we impose a restriction on the labeling algorithm, allowing it to generate a maximum of N_{labeling} new routes in each iteration.

4.3 Local Search

We outline the process to enhance the solution to our problem by tailoring three local search operators – insertion, deletion, and relocation. Each operator generates a type of move that applies to one of the current routes r to yield an alternative feasible route with a lower reduced cost, denoted as the "neighbour" of route r. These neighbours are then used to generate a set of up to N_{local} routes with the potential to improve the existing best solution.

In order to improve the efficiency of the local search for our problem, we introduce the concept of effective time windows. Specifically, we define a transportation route as an assignment of patients to a nurse, along with the order in which they are visited. The notation for a route with q patients being attended is $r = (i, h_1, h_2, ..., h_q, i)$, where i represents the starting clinic with the stipulation that $r[0] = r[q+1] = i$, and h_j, also denoted as $r[j]$, is the j-th patient on the route. For every feasible route r, the effective latest arrival times for each patient h_u included in r are computed. The computation of these effective latest arrival times on a given route r is carried out through the application of the following recursive principles.

For route r with q patients being attended in Ω_v', and for any patient h_u to be visited in that route, the upper bound for the effective latest arrival time for patient h_u is represented by the notation $\varsigma_{h_u}^{\text{Upper}}$:

$$\varsigma_{h_u}^{\text{Upper}} = \min\{\varsigma_{h_{u+1}}^{\text{Upper}} - v_{h_{u+1}} - t_{h_u,h_{u+1}}, b_{h_u}\},\ 0 \leq u \leq q-1, \tag{35}$$

$$\varsigma_{h_q}^{\text{Upper}} = b_{h_q}. \tag{36}$$

Similarly, for route r in $\Omega_n' \cup \Omega_h'$, the upper bound for the effective latest arrival time for patient h_u assigned to a clinic or hospital nurse is represented by the notation

$\tau_{h_u}^{\text{Upper}}$:

$$\tau_{h_u}^{\text{Upper}} = \min\{\tau_{h_{u+1}}^{\text{Upper}} - v_{h_{u+1}}, b_{h_u}\},\ 0 \leq u \leq q-1, \tag{37}$$

$$\tau_{h_q}^{\text{Upper}} = b_{h_q}. \tag{38}$$

Given a route $r \in \Omega'$ and a new patient h_{new}, the feasibility of inserting h_{new} into r or swapping it with an existing patient can be quickly determined based on the effective latest arrival time. The calculation of the effective latest arrival times can be done by traversing the patients in the route r only one time, regardless of the choice of h_{new}.

4.3.1 Insertion

A new route can be obtained by inserting a new patient h_{new} between two consecutive patients h_u and h_{u+1}, where $u + 1 \leq q$. The procedure of insertion in the construction heuristic algorithm is different from that in the pricing algorithm. In the next two paragraphs, we will expound upon the insertion procedures of both phases independently.

1. Insertion in initialization:
   We introduce an adapted insertion method for the construction heuristic algorithm to generate more feasible routes during the initialization phase. Specifically, this approach allows for the identification of patients feasible for insertion after a given patient h_u in route $r \in \Omega'$ through a prescribed procedure (Appendix A.2). It is noteworthy that this procedure takes any patient $h_u \in r$ as an input and provides either 1) an updated feasible route pool, indicating the construction of a new route by inserting the first feasible patient after h_u and the addition of it to the original feasible route pool, or 2) the original feasible route pool, indicating that there is no feasible insertion after $h_u \in r$.
2. Insertion in negative reduced cost generation:
   This method is designed to facilitate the generation of more feasible routes with a lower negative reduced cost during the pricing phase. The method involves a prescribed procedure (Appendix A.3) to identify patients feasible for insertion after a given patient h_u in route $r \in \Omega'_\xi$. Here the feasible route pools $\Omega'_\xi = \Omega'_{v,\xi} \cup \Omega'_{n,\xi} \cup \Omega'_{h,\xi}$ only include routes with a reduced cost that is less than a predetermined threshold ξ. Notably, only patients $j \in \mathcal{N}_{h_u}$ that can result in a negative reduced cost for the new route are considered, rather than all patients in $\mathcal{N}_{h_u}$. The procedure terminates once the reduced cost of the new route after inserting $j \in \mathcal{N}_{h_u}$ becomes positive. This is accomplished by sorting $\mathcal{N}_{h_u}$ in ascending order with respect to their dual values. If the reduced cost of the new route after inserting patient $j \in \mathcal{N}_{h_u}$ is positive, it indicates that the patient ranked after j in $\mathcal{N}_{h_u}$ could only make the reduced cost of the new route to be positive. The input and output of this procedure are the same as those of the previous insertion method.

4.3.2 Deletion

In the pricing phase, the patient associated with the maximal reduced cost is removed from route r. Subsequently, the reduced cost of the resulting new route is evaluated. If it remains negative and less than the original route r, the new route is added to the set Ω'.

4.3.3 Relocation

An alternative route can be generated by replacing a patient who has the maximum reduced cost in the given route with a new patient. Appendix A.4 presents the relocation procedure in detail.

4.4 Bi-level Structure

As previously mentioned, the collection of all possible combinations of open clinics is denoted by $\mathcal{S}$, where $|\mathcal{S}| = 2^{|I|-1}$. Due to the considerable computational cost of exploring every scenario $\bar{s} \in S$, we set all clinics to open initially, and close one clinic at a time iteratively until the stopping criterion is met.

A scoring scheme is introduced to identify clinics that are both expensive and have low utilization to be candidates for closure. The set of patients assigned to clinic i is denoted by $\mathcal{J}_i$, and the utilization ratio of clinic i is defined as the ratio of the number of patients assigned to it to the total number of patients, i.e., $u_i = \frac{|\mathcal{J}_i|}{|J|}$. Additionally, the level of expense associated with collaborating with clinic i is presented by $\bar{f}_i = \frac{f_i}{\sum_{i \in I} f_i}$. Finally, a score $score_i$ is defined as the ratio of the utilization ratio of clinic i to the expense level of collaborating with it, i.e., $score_i = \frac{u_i}{\bar{f}_i}$. We sort the clinics in ascending order according to their $score_i$ values and iteratively close the clinic with the lowest score until the termination criterion is met.

In the bi-level algorithm (Algorithm 1), we define the set of opened clinics as I_{open}. Then, we track two values, namely c_{pre_best} and c_{cur_best}, which respectively represent the best solutions for the previous stage and current stage. The algorithm terminates when further closures of clinics fail to produce an improved solution, and subsequently outputs the best solution attained in the final iteration. We proceed with the bi-level algorithm as follows:

1. We assume that all clinics are open, denoted by $I_{open} = I$. We introduce the iterators it and Δ to govern the stop criteria for solving the lower-level problem. Additionally, we initialize the best solutions for the current stage with c_{cur_best} set to a sufficiently large value, and the previous stage c_{pre_best} set to 0, respectively.
2. Given I_{open}, we initiate the solution process for the assignment and routing problem using the set of open clinics. Initially, we apply the construction heuristic algorithm to generate an initial route pool denoted as Ω' and generate a feasible initial solution. Subsequently, we apply a local search method to enhance the quality of the current starting solution and add up to $N_{\text{cardinality}}$ new routes into Ω'.
3. We solve the restricted master problem $RMP(\Omega')$. Following that, we employ local search algorithms to enhance the current best solution and incorporate up to N_{local} routes into the existing pool Ω'. We then utilize the heuristic labeling algorithm

to generate up to N_{labeling} routes with negative reduced costs and subsequently update the route pool Ω'. Lastly, we increment the value of it by 1. Additionally, we increment the value of Δ by 1 if the current solution is the same as the previous solution.

4. If the stopping criterion for the lower-level stage is not met, proceed to step 2; otherwise, proceed to the next step. The lower-level stage stops after N iterations (e.g., $\Delta > N$), or if the current best solution remains unchanged for the last ϵ iterations (e.g., $it > \epsilon$), or if no route with negative reduced cost has been found.
5. We update c_{cur_best} and c_{pre_best} with the total cost obtained from the aforementioned process, and update I_{open} by using the criterion for closing clinics discussed earlier in this section.
6. The algorithm terminates when further clinic closures do not yield an improved solution. In this case, we assign c_{pre_best} as the best solution; otherwise, we return to step 2.

Algorithm 1 Bi-level Algorithm

procedure BI-LEVEL ALGORITHM(N, ϵ)
    $I_{open} \leftarrow I$
    $it, \Delta \leftarrow 0$
    Initialize c_{cur_best}, c_{pre_best}
    while $c_{cur\_best} < c_{pre\_best}$ **do**
        $\Omega' \leftarrow$ Construction Heuristic Algorithm
        Expand Ω' by applying Local Search
        while $it \leq N$ or $\Delta \leq \epsilon$ **do**
            Solve RMP(Ω')
            Expand Ω' by applying Local Search
            Generate routes with negative reduced cost
            Check the current solution and update it and Δ
        end while
        Update c_{cur_best}, and c_{pre_best}
        Update I_{open} by using criterion for closing clinics
    end while
    return c_{pre_best}
end procedure

5 Computational Analysis

This section presents a detailed account of the test instances used in the study, followed by an analysis of the results of our computational experiments. The algorithm has been implemented in C++, and for solving linear programming problems, Gurobi 10.0 is employed. All experiments were conducted on a computer equipped with an AMD EPYC 7702 64-Core Processor, featuring a maximum clock speed of 3.35 GHz and a total RAM capacity of 1003GB. We assess the quality of our solutions while highlighting the efficacy and broad utility of our algorithm to investigate the operation of decentralized care delivery networks.

5.1 Data Generation and Experiment Design

The experimental data is motivated by real instances in a large healthcare organization serving north central Florida. The patient locations are generated at random with uniformly distributed coordinates within a circular area centered at the central hospital with a radius of 200 miles. The distances between nodes are calculated using the Euclidean distance matrix, and adjusted Euclidean distances are utilized to estimate the travel time between two points. Specifically, travel times are adjusted to fall within a range of 15 to 30 minutes for patients and 20 to 40 minutes for visiting nurses who require additional precautions while driving and carrying medical equipment [41]. Each patient's required service time is randomly selected from a range of 30 to 60 minutes, and the time windows are uniformly distributed throughout the day, from 08:00 AM to 08:00 PM, with a random duration ranging from 1.5 hours to 4 hours [42, 43]. The maximum daily workload for a nurse is restricted to 12 hours [44–46]. However, as time windows may impact the results, a sensitivity analysis of this parameter will be conducted later to evaluate its impact on the final results. In accordance with the labor statistics provided by the United States Department of Labor, the mean hourly remuneration for various categories of registered nurses stands at \$39.87 for visiting nurses, \$38.31 for clinic nurses, and \$43.56 for hospital nurses on average [47]. Giambruno et al. [48] and Branch and Nemeth [49] have found that transportation barriers affect healthcare access in sample populations ranging from 3% to 67%. In our analysis, we set the ratio of Type I patients to all patients in the range of $[0.4, 0.7]$, while the rest of the patients are Type II patients.

Furthermore, we need to determine the collaboration fee of individual clinics. A sensitivity analysis of the collaboration fee is performed to explore its impact on the solution structures. Specifically, we make the assumption that there are two possible scenarios with regard to the collaboration fee. In the first scenario, the collaboration fee slightly exceeds a nurse's daily salary, with a range of values between \$300 and \$450. In the second scenario, the collaboration fee significantly exceeds a nurse's daily salary, with a range of values between \$900 and \$1350. Intuitively, the model's complexity is expected to be greater under the first scenario, whereas under the second scenario, wherein the collaboration fee significantly surpasses a nurse's daily salary, the model is likely to prioritize opening the clinic that offers the lowest collaboration fee due to its dominant impact on the objective function. As a remark, when the collaboration fee is zero, there is no clinic selection (although it is still a multi-depot problem) and the complexity of the problem is also significantly reduced.

Our experiments encompass a triad of distinct data instances. These instances consist of the baseline dataset, a dataset marked by escalated clinic collaboration fees, and a dataset featuring extended time windows. All parameters remain constant across these instances, with the exception of the clinic collaboration fee or the time window allotted for patients. In the dataset with higher clinic collaboration fees, the adjusted clinic collaboration fees are set at three times the value of the initial collaboration fees. In the dataset with larger time windows, the new time windows extend from the end of the original baselines by an hour. If this one-hour extension exceeds the maximal limit of the planning horizon, which in our study is set as 12 hours, the new time windows are initialized to begin one hour before the beginning of the initial time windows in the baseline dataset. Each of these specified data instances further materializes into a tripartite arrangement: 1) cost minimization without explicit consideration of patient disutility, herein termed as "Non-Social Welfare"; 2) cost minimization considering patient travel cost with low perceived disutility of travel, labeled as "Social Welfare with Low Disutility Magnitude"; and 3), cost minimization considering patient travel cost with high perceived disutility of travel, denoted as "Social Welfare with High Disutility Magnitude". By scrutinizing the experiment outcomes across these three scenarios, an investigation into the interplay between model solutions and the magnitude of disutility becomes feasible.

5.2 Computational Results

In our computational analysis, we begin by examining the outcomes for the baseline dataset. We proceed by comparing the results of the baseline dataset with those obtained under elevated clinic collaboration fees. Subsequently, we contrast the baseline dataset with datasets featuring extended time windows. Finally, we delve into an exploration of the impact of patient disutility, focusing primarily on the baseline dataset.

5.2.1 Parameters

This section describes the values of the main parameters used in our algorithms to obtain the results reported in Section 5.2. From the algorithm 1, we have five parameters to set, including the number of pivot patients $N_{\text{cardinality}}$, the maximum number of routes generated by the labeling algorithm N_{labeling}, the maximal number of new routes that can be generated by local search N_{local}, the maximum iteration number in the lower-level stage N, and the maximal number of iterations that the current best solution remains unchanged ϵ.

During the initialization, we set $N_{\text{cardinality}}$ to be half of the total number of patients. This configuration aims to ensure that we generate enough initial routes to cover the patients that are difficult to serve, while also limiting the number of initial routes to a relatively small scale to avoid investing too much computational resource to the insertion of *non-pivot* patients into these routes. Regarding the efficiency of the labeling algorithm, we set the value of the maximal number of routes generated by the labeling algorithm N_{labeling} to be a hundred times the total number of patients. Our objective is to achieve a balance between preventing an excessively large pool while also ensuring that the algorithm generates an adequate collection of routes. Likewise, in relation to the efficiency of the local search, we set the maximal number of routes generated by local search N_{local} to be ten times the total number of patients. For the stop criterion of the lower-level stage, we fix the maximum iteration number at $N = 10$ as the algorithm already yields a sufficient number of solutions. Meanwhile, we set $\epsilon = 3$ as the threshold for the number of iterations in which the current best solution remains unchanged, preventing unnecessary attempts to enhance a solution that is unlikely to be further improved. This parameter setting is consistently applied to all types of datasets tested in the subsequent analysis.

5.2.2 Benchmark Comparison

The major outcomes of the comparison between our proposed approach and the standard solver solution from Gurobi are presented in Table 1. The column labeled *Instance Size* signifies the patient count. Problem instances encompassing patient counts from 10 to 50 are classified as small instances, from 60 to 100 patients, medium instances, and 150 and 200 patients, large instances. The two intermediate columns denoted as *Gurobi* present the results obtained from solving the original MIP formulation using the commercial solver Gurobi. The column labeled *benchmark* denotes the best achievable total cost value. This result may not always be optimal due to the computational time constraint imposed, which is one hour, or 3600 seconds, in this study. When instances become large, Gurobi is unable to generate results due to resource limitations, referred to as "OUT OF MEMORY". The subsequent column, labeled *time/s*, signifies the computation time in seconds required to obtain the best solution while not significantly exceeding the allowed computational time. These results do not involve the optimization of the problem structure for the specific purpose of accelerating Gurobi's computational speed. The outcomes generated utilizing the heuristic algorithm to solve for the bi-level approximation are under the *Heuristic CG* section. The *obj* column represents the objective value achieved, while the *time/s* column signifies the computation time in seconds for the respective solution. The final column indicates the relative gap between the outcomes derived from Gurobi and our algorithm. This relative gap is calculated as $GAP/\% = 100\% \times (benchmark - obj)/benchmark$.

Table 1 Computational Outcomes for the Baseline Dataset under the Non-Social Welfare Scenario

	Gurobi		Heuristic CG		
Instance Size	benchmark	time/s	obj	time/s	GAP/%
10	2523.13	3.13	2524.14	2.21	-0.04
20	3887.05	225.31	3887.05	10.91	0.00
30	4164.85	640.36	4165.27	12.49	-0.01
40	4893.43	3600.15	4895.88	36.31	-0.05
50	6332.17	3602.93	6341.04	38.94	-0.14
60	5849.77	3604.36	5854.45	43.43	-0.08
70	9018.52	3600.35	9037.46	45.76	-0.21
80	6918.17	3602.64	6937.11	51.76	-0.27
90	9111.49	3607.06	9134.36	63.01	-0.25
100	OUT OF MEMORY	OUT OF MEMORY	10340.51	74.2	-
150	OUT OF MEMORY	OUT OF MEMORY	13429.15	82.4	-
200	OUT OF MEMORY	OUT OF MEMORY	15763.81	130.64	-
Average	-	-	-	-	-0.12

Table 1 illustrates that the heuristic approach based on column generation achieves optimality in some small instances, and gets solutions close to the best achievable solutions in medium and large instances. The average gap is -0.12%, yet our algorithm exhibits significantly short computational times. For small and medium instance sizes, our algorithm consistently yields near-optimal cost values for each instance within a minute. Even as the instance sizes increase, computation time lengthens, yet our algorithm consistently converges to the near-optimal values within approximately 2 minutes. Our heuristic algorithm based on column generation in this work, showcases remarkable efficiency to yield exceptionally high-quality solutions within short computation times.

5.2.3 Results on Clinic Collaboration Fees and Time Windows

In this section, we delve into the impact of clinic collaboration fees and time windows on our experimental results. In Table 2, the last three columns present the optimal solution values derived from the Heuristic CG algorithm for the baseline dataset, the dataset with high clinic collaboration fees, and the dataset with extended time windows, respectively. When patient disutility is not explicitly considered, the influence of high clinic collaboration fees on the overall cost is evident, as it increases the total cost universally. Analyzing the best solution patterns between the baseline and high clinic collaboration fee scenarios, we observe that in most cases, the set of opened clinics remains consistent. However, there exist instances where an increase in clinic collaboration fees leads to a reduction in the number of opening clinics. This phenomenon can be attributed to a strategic shift where, with significant increases in clinic collaboration fees, a preference emerges for opening fewer clinics while accommodating more patients in each of these clinics.

Intuitively, when clinic collaboration fees are very high, the preference is to operate fewer clinics while concentrating more patients within each. Conversely, when clinic collaboration fees experience only modest increments, possibly beneath a specific threshold, maintaining the initial set of opening clinics is preferred. This threshold concept may be crucial for the decision-makers at the central hospital, aiding them in understanding the tipping point where such strategic shifts occur. While traditional solvers might prove time-intensive for this task, our algorithm, when combined with grid search, could offer an efficient means to identify these thresholds in reasonable times. This capacity underscores the broader utility of our algorithm.

Likewise, an extension of the time window results in a reduction of the total cost due to the expansion of the feasible region. Notably, there is an intuitive correlation between the larger time windows and the higher average computational time. This can be rationalized by the increased number of feasible routes to explore within the broader feasible region, thereby leading to lengthier computational processes. It is worth noting, however, that while the computational time does increase as the instance size grows, this escalation is not exponential in nature. This observation underscores the efficiency of our algorithm even in more intricate scenarios. In summary, these findings highlight the robustness of our algorithm to effectively handle complex situations.

5.2.4 Results on the Magnitude of Disutility

Within this section, we explore the influence of incorporating patient disutility into the analysis. By introducing the dimension of social welfare, we aim to unveil the alterations in the best solution and assess how it evolves in response to this pivotal

Table 2 Comparative Computational Results of Heuristic CG Algorithm for the Non-Social Welfare Scenario Across Various Datasets.

	Baseline		High Clinic Collaboration Fee		Large Time Window	
Instance Size	obj	time/s	obj	time/s	obj	time/s
10	2524.14	2.21	3261.52	2.71	2479.43	6.89
20	3887.05	10.91	4643.45	14.01	3733.45	37.36
30	4165.27	12.49	4850.40	13.29	4141.04	35.23
40	4895.88	36.31	5592.04	37.43	4681.89	109.95
50	6341.04	38.94	7083.56	38.51	6087.52	115.29
60	5854.45	43.43	6464.74	45.12	5842.89	113.84
70	9037.46	45.76	9913.01	54.31	8831.95	161.34
80	6937.11	51.76	8249.66	50.63	6763.37	142.10
90	9134.36	63.01	9963.85	73.12	7814.08	210.05
100	10340.51	74.23	12315.25	72.92	9921.41	200.25
150	13429.15	82.45	16120.51	87.41	11761.21	253.65
200	15763.81	130.64	17149.04	142.12	14845.12	409.23

consideration. Since the best solutions obtained by all three datasets (with variations in clinic collaboration fees and time windows) follow the same pattern, we only show the comparison among those three scenarios: the absence of social welfare consideration, the inclusion of social welfare with a low patient disutility rate, and the integration of social welfare with a high patient disutility rate, by using the baseline dataset.

Table 3 Comparative Computational Results of Heuristic CG Algorithm for Baseline Dataset with Varied Patient Disutility Rates.

	Non-Social Welfare	Low Patient Disutility Rate		High Patient Disutility Rate	
Instance Size	Φ_{cost} [1]	Φ_{cost}	Φ_{social} [2]	Φ_{cost}	Φ_{social}
10	2524.14	2684.94	2914.02	3156.29	3256.29
20	3887.05	4180.25	4302.62	4931.89	4931.89
30	4165.27	4623.65	4812.35	5779.12	5779.12
40	4895.88	5087.48	5371.54	6573.31	6582.50
50	6341.04	6547.97	6928.36	8793.55	8793.55
60	5854.45	6632.07	7031.53	7943.53	7943.53
70	9037.46	10301.23	10894.74	12415.31	12421.37
80	9937.11	10091.26	10424.31	10624.83	10624.83
90	9134.36	10046.68	10656.41	10743.90	10743.90
100	10340.51	11877.87	11980.23	11986.56	11986.56
150	13429.15	14121.67	14800.73	16216.57	16704.20
200	15763.81	16811.15	17135.70	17531.72	17712.42

[1] Φ_{cost} includes nurse salary, visiting nurse travel expenses, and clinic collaboration fees.

[2] Φ_{social} comprises the same that present in Φ_{cost}, and the aggregate patient disutility. This disutility is quantified through the incorporation of travel costs accrued by patients attending appointments at community clinics or hospitals.

In Table 3, the initial column presents the total cost Φ_{cost} under cost minimization. The subsequent two columns provide costs when the patient's disutility rate is at a lower level: one column signifies the total cost excluding patient disutility Φ_{cost}, while the adjacent column reflects the total cost incorporating patient disutility Φ_{social}. This comparison allows us to analyze how patient disutility impacts patient assignment. This analysis proves valuable not only for central hospital planners but also for decision-makers who aim to strike a harmonious balance between hospital expenditure and societal well-being. Similarly, the subsequent pair of columns portray an analogous comparison, differing only in the context of a higher patient disutility rate.

The observed trend indicates that the cost from the healthcare organization's side, Φ_{cost}, regardless of the magnitude of the patient disutility rate, is higher when patient travel cost is taken into account. This outcome is logically expected, as there is a tendency to allocate a greater number of patients to visiting nurses in these scenarios. As the patient disutility rate escalates to a certain threshold, compared with the low patient disutility rate scenario, the inclination shifts towards assigning more patients to visiting nurses, which significantly raises the cost from the healthcare organization's side. In certain instances, no patient is assigned to travel by themselves (see the column Φ_{cost} and Φ_{social} under the high patient disutility rate dataset in Table 3). While these findings may appear intuitive, they hold substantial significance in guiding the hospital's decision-making process. This allows for further scrutiny of strategies to coordinate between the decisions made by the social planner and the healthcare organization.

6 Discussions and Conclusions

6.1 Strategies for Addressing Temporary Changes in Patients' Schedules

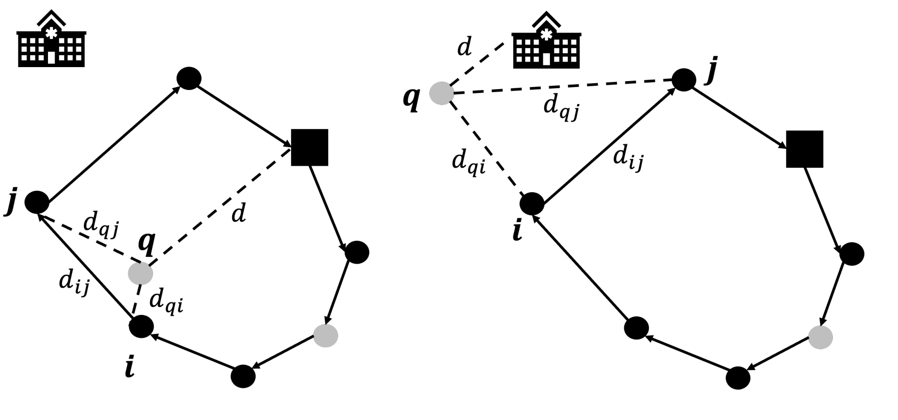
In the dynamic landscape of healthcare, the ability to effectively manage reschedules and same-day appointments has become a critical need. This entails providing swift access to medical care for individuals with urgent needs, without overhauling the original schedule, impacting all other patients. Therefore, we focus on the local adjustment of the schedule, assuming there is at least one open clinic in the region accessible to the patient with urgent needs.

Our analysis is based on several assumptions to simplify the problem while still being able to generate operational insights. First, we assume only one patient requiring a same-day appointment is accommodated at any given time. With multiple requests, the potential strategy involves adopting a sequential rescheduling approach, accommodating one patient at a time. We also consider the more general case that the patient belongs to Type II. As a result, three scenarios can unfold: incorporation of the patient into the nearest feasible route, to the clinic that is the nearest and accessible, and to the hospital. Secondly, there exist nurses capable of providing care to the additional patient without the need to significantly deviate from their existing schedule. This assumption alleviates the requirement to factor in nurse salaries and costs incurred by changes in other patients' schedules when obtaining the best solution for accommodating the additional patient's request. Thirdly, we assume the unit travel costs for visiting nurses and patients are equivalent, and employ the Euclidean metric as the only criterion for distance-based cost measurement. Consequently, the application of the triangle inequality principle becomes feasible under these assumptions.

Ignoring patient disutility, the optimal solution would consistently entail assigning the patient to the nearest healthcare facility (either an open clinic or the hospital). Therefore, we mainly focus on the solution with patient disutility in our deliberations, and we obtain a proposition that offers straightforward solutions with minimal computational effort.

Preceding this analysis, we introduce a set of notations that are applied in the subsequent discourse. The Type II patient who seeks a same-day appointment is denoted as q, and the nearest feasible route segment into which patient q can be inserted is represented as (i, j). Here, i and j correspond to the patients/clinics positioned at the beginning and the end of this segment respectively. We represent the distance from patient q to the clinic as d_{clinic} and to the hospital as d_{hos}, respectively. Further, let the minimal distance from patient q to either the clinic or hospital be represented as $d = \min\{d_{\text{clinic}}, d_{\text{hos}}\}$. Finally, we denote the distances from patient q to patients i and j as d_{qi} and d_{qj}, and the distance between patient i and j as d_{ij}.

The depiction of these notations is presented in Fig. 3. We show two examples of the problem. In the first example, shown in Fig. 3(a), the patient q is near the clinic, and the distance from patient q to the clinic is greater than the distance from both patients i and j to patient q, i.e., $d \geq d_{qi}$ and $d \geq d_{qj}$. Based on this example, we will discuss when d is greater than at least one of d_{qi} and d_{qj}, what is the best solution for patient q, and we will provide a straightforward solution under this situation in Proposition 1. In the other example, shown in Fig. 3(b), the patient q is near the hospital, and the distance from patient q to the hospital is less than the distance from both patients i and j to patient q, i.e., $d < d_{qi}$ and $d < d_{qj}$. In this situation, a straightforward solution is lacking, necessitating computational evaluation and comparison.



(a) $d \geq d_{qj} \geq d_{qi}$

(b) $d_{qj} \geq d_{qi} \geq d$

Fig. 3 The examples of patient q under different situations

Proposition 1. *If at least one of the distances from patient i and j to the patient q is not larger than d, represented as $d \geq \min\{d_{qi}, d_{qj}\}$, the best decision is to allocate the Type II patient q to the visiting nurse.*

Proof. With the aforementioned assumptions, the problem reduces to comparing the costs of inserting the Type II patient q into the closest feasible segment route and assigning patient q to the clinic or the hospital. These costs are denoted as $c_{\text{Visit}} = d_{qi} + d_{qj} - d_{ij}$ and $c_{\text{NoVisit}} = 2d$. Applying the triangle inequality theorem, we have the following inequalities:

$$d_{ij} \geq |d_{qi} - d_{qj}|.$$

Without loss of generality, we assume that $d_{qi} \geq d_{qj}$, then we have $d_{ij} \geq d_{qi} - d_{qj}$. Then, we compare the values of c_{Visit} and c_{NoVisit}, aiming to identify the conditions that induces the inequality $c_{\text{Visit}} \leq c_{\text{NoVisit}}$.

$$\begin{aligned} c_{\text{Visit}} &\leq c_{\text{NoVisit}}, \\ d_{qi} + d_{qj} - d_{ij} &\leq 2d, \\ \text{Since } d_{qi} + d_{qj} - d_{ij} &\leq d_{qi} + d_{qj} - d_{qi} + d_{qj}, \\ \text{we need } d_{qi} + d_{qj} - d_{qi} + d_{qj} &\leq 2d, \\ d_{qj} &\leq d. \end{aligned}$$

The same deduction applies for $d_{qj} \geq d_{qi}$. We can infer that when $d \geq \min\{d_{qi}, d_{qj}\}$, $c_{\text{Visit}} \leq c_{\text{NoVisit}}$ is always satisfied.

Therefore, when $d \geq \min\{d_{qi}, d_{qj}\}$, we will always insert the Type II patient q into the existing route. □

Based on the aforementioned property, we are able to efficiently reach a decision for the Type II patient seeking a same-day appointment under certain conditions with simple computations.

Initiating the procedure, we commence by identifying the nearest patient, denoted as q_{nei} in relation to patient q. We evaluate the feasibility of incorporating patient q into the route r containing the nearby patient q_{nei}, while ensuring that time window constraints for other patients on route r are not violated. If infeasible, we continue to evaluate the next closest patient and their route. Once a feasible route is found, a comparison is made between d and $\min\{d_{qi}, d_{qj}\}$, and if they conform to the condition set in Proposition 1, patient q is assigned to the visiting nurse. If not met, a further calculation and comparison between $d_{qi} + d_{qj} - d_{ij}$ and $2d$ are carried out, and the best solution is determined. If a feasible route cannot be found, patient q is allocated to the nearest healthcare facility.

6.2 Implications for Hybrid Telehealth Service Network Implementation

The emerging trend of hybrid telehealth services is expected to gain prominence due to their enhanced efficiency and convenience. In this design, specific patients are allowed to remain at home while others whose home-based care is not necessary might be directed to either the hospital or nearby clinics, and can also be assigned to a visiting nurse, taking advantage of being close to a route where a nurse has to visit regardless.

To realize the potential of this hybrid network service system, maintaining a low operating cost is essential.

Our proposed approach can be used to support the operation of the proposed hybrid service network, helping the operator to solve a combination of facility location, assignment, and vehicle routing problems with the goal of minimizing total operational costs. Furthermore, the knowledge of patient disutility (the precise assessment of patient travel costs) would empower us to perform a careful analysis of the implementability of the proposed decentralized care delivery network. This analysis could encompass an assessment of the overall benefit of the service design from the cost-saving perspective. For instance, we will be able to consider the cost of shuttle service or travel compensations provided prior to the reform to decide if it is better off to send nurses. Further, we could compare the social welfare encompassing patients' travel costs prior to and post the reform. In terms of the current sensitivity analysis, we have demonstrated the impact of several operating factors like the clinic collaboration fee. The future direction could be exploring other operating factors like nurse salary, travel compensation, among others. Our model has the capacity to perform a range of sensitivity analyses regarding the parameters and could incorporate other parameters.

6.3 Limitations and Future Research Directions

Because the proposed hybrid telehealth service system has not yet achieved wide adoption, our numerical experiments are primarily based on synthetic data. However, the existing systems, such as the home-based and clinic-based telehealth programs established by Veterans Affairs (VA) healthcare, might offer valuable information and data that can enable a more precise estimation of our model parameters, such as travel-related expenses and time for visiting nurses, and patient care preferences. Understanding the operations of these existing systems and incorporating their operating features will enhance the model's capacity to closely mirror real-world conditions. As leveraging telehealth technologies to ameliorate barriers to healthcare access is gaining uptake, we posit that our proposed system holds the potential for future implementation.

A direct extension of our current model involves the consideration of a more heterogeneous patient population. While our current work includes only two types of patients, real-world scenarios could feature multiple types of patients. For instance, certain patients may require specialized care that necessitates a physical visit to a specific healthcare facility, owing to the non-portability of specialized medical equipment. The intricate variability in patient needs underscores the importance of expanding the model to consider multiple types of patients, and each type of patient is characterized by distinct healthcare requirements and constraints.

Another potential avenue for future exploration is delving into the dynamic variant of the routing problem within visit services. For example, the traveling costs associated with visiting nurses could be non-deterministic, and the service time for each patient could fluctuate within a predetermined interval. Those variations entail decision-making under uncertainty, which may be formulated as either stochastic or robust optimization problems, and demand computationally efficient solutions.

Additionally, incorporating fairness into our model also presents an intriguing direction for further work. The notion of fairness in our context could be constructed as the equitable and balanced distribution of visits and distances among the available nurses and patients. This conceptualization of fairness serves to elevate employee contentment and potentially improve the quality of service rendered to patients. For example, the fairness objective or constraint might entail minimizing the longest traveling distance among patients and nurses, or restricting each nurse services approximately the same number of patients.

Appendix A Algorithms

A.1 The Construction Heuristic Algorithm

Algorithm 2 The Construction Heuristic Algorithm

procedure INITIALIZATION
    $\Omega' \leftarrow$ A set of feasible routes
    $P_1 \leftarrow$ Obtain a set of patients by criterion 1
    $P_2 \leftarrow$ Obtain a set of patients according to criterion 2 ranked based on w_j in descending order
    $P_{\text{new}} \leftarrow$ Obtain the pivot patient set following the selection procedure
    for all $p \in P_{\text{new}}$ **do**
        The labeling set algorithm is called to determine the best route r that visits only patient p
        The route r is added to the set of Ω'
    end for
    for all $p \in J \setminus P_{\text{new}}$ **do**
        Insert p to current feasible routes in Ω'
        Update Ω' with new routes
    end for
end procedure

A.2 Insertion Algorithm in Initialization Phase

For each patient h_u in r, we obtain its starting time $\bar{a}_{h_u}$ in route r by applying the labeling algorithm, and obtain its neighbourhood set $\mathcal{N}_{h_u}$ as defined in Section 3.2.

Algorithm 3 Insertion Procedure (Initialization)

procedure INSERTION($h_u \in r$)
    Rank all the patients in $\mathcal{N}_{h_u}$ in ascending order with respect to their distance from h_u
    Compute $\varsigma_{h_{u+1}}^{\text{Upper}}$ or $\tau_{h_{u+1}}^{\text{Upper}}$
    for all $j \in \mathcal{N}_{h_u}$ **do**
        if $r \in \Omega'_v$ **then**
            if $\bar{a}_{h_u} + v_{h_u} + t_{h_u,j} \leq b_j$ **then**
                if $\bar{a}_{h_u} + v_{h_u} + t_{h_u,j} + v_j + t_{j,h_{u+1}} \leq \varsigma_{h_{u+1}}^{\text{Upper}}$ **then**
                    return j
                end if
            end if
        else
            if $\bar{a}_{h_u} + v_{h_u} \leq b_j$ **then**
                if $\bar{a}_{h_u} + v_{h_u} + v_j \leq \tau_{h_{u+1}}^{\text{Upper}}$ **then**
                    return j
                end if
            end if
        end if
    end for
    return $\emptyset$
end procedure

A.3 Insertion Algorithm in Pricing Phase

In order to simplify the notation, the dual variables for both Type I and Type II patients are denoted by π_j, and the reduced cost for any route r is denoted by $\bar{c}_r$.

Algorithm 4 Insertion Procedure (Pricing)

procedure INSERTION($h_u \in r$)
  Rank all the patients in $\mathcal{N}_{h_u}$ in ascending order with respect to their dual values
  Compute $\varsigma_{h_{u+1}}^{\text{Upper}}$ or $\tau_{h_{u+1}}^{\text{Upper}}$
  for all $j \in \mathcal{N}_{h_u}$ **do**
    if $r \in \Omega'_{v,\xi}$ **then**
      if $\bar{a}_{h_u} + v_{h_u} + t_{h_u,j} \leq b_j$ **then**
        if $\bar{a}_{h_u} + v_{h_u} + t_{h_u,j} + v_j + t_{j,h_{u+1}} \leq \varsigma_{h_{u+1}}^{\text{Upper}}$ **then**
          if $\bar{c}_r + \pi_j + e_{h_u,j} + e_{j,h_{u+1}} < 0$ **then**
            return j
          end if
        end if
      end if
    else
      if $\bar{a}_{h_u} + v_{h_u} \leq b_j$ **then**
        if $\bar{a}_{h_u} + v_{h_u} + v_j \leq \tau_{h_{u+1}}^{\text{Upper}}$ **then**
          if $\bar{c}_r + \pi_j + g_{h_u,j} + g_{j,h_{u+1}} \leq 0$ **then**
            return j
          end if
        end if
      end if
    end if
  end for
  return $\emptyset$
end procedure

A.4 Relocation Algorithm in Pricing Phase

Algorithm 5 Relocation Procedure (Pricing)

procedure RELOCATION(r)
  $h_u \leftarrow$ the patient with the maximal reduced cost in route r
  Rank all the patients in $\mathcal{N}_{h_{u-1}}$ in ascending order with respect to their dual values
  Compute $\varsigma_{h_{u+1}}^{\text{Upper}}$ or $\tau_{h_{u+1}}^{\text{Upper}}$
  for all $j \in \mathcal{N}_{h_{u-1}}$ **do**
    if $r \in \Omega'_v$ **then**
      if $\bar{a}_{h_{u-1}} + v_{h_{u-1}} + t_{h_{u-1},j} < b_j$ **then**
        if $\bar{a}_{h_{u-1}} + v_{h_{u-1}} + t_{h_{u-1},j} + v_j + t_{j,h_{u+1}} \leq \varsigma_{u+1}^{\text{Upper}}$ **then**
          return j
        end if
      end if
    else
      if $\bar{a}_{h_{u-1}} + v_{h_{u-1}} \leq b_j$ **then**
        if $\bar{a}_{h_{u-1}} + v_{h_{u-1}} + v_j \leq \tau_{u+1}^{\text{Upper}}$ **then**
          return j
        end if
      end if
    end if
  end for
  return $\emptyset$
end procedure

References

[1] Wallace, R., Hughes-Cromwick, P., Mull, H., Khasnabis, S.: Access to health care and nonemergency medical transportation: two missing links. Transportation research record **1924**(1), 76–84 (2005)

[2] Starbird, L.E., DiMaina, C., Sun, C.-A., Han, H.-R.: A systematic review of interventions to minimize transportation barriers among people with chronic diseases. Journal of community health **44**(2), 400–411 (2019)

[3] Solomon, M.M.: Algorithms for the vehicle routing and scheduling problems with time window constraints. Operations Research **35**(2), 254–265 (1987)

[4] Dantzig, G.B., Ramser, J.H.: The truck dispatching problem. Management Science **6**(1), 80–91 (1959)

[5] Bodin, L., Golden, B.: Classification in vehicle routing and scheduling. Networks **11**(2), 97–108 (1981)

[6] Eksioglu, B., Vural, A.V., Reisman, A.: The vehicle routing problem: A taxonomic review. Computers Industrial Engineering **57**(4), 1472–1483 (2009)

[7] Laporte, G.: Fifty years of vehicle routing. Transportation Science **43**(4), 408–416 (2009)

[8] Savelsbergh, M.W.: Local search in routing problems with time windows. Annals of Operations research **4**(1), 285–305 (1985)

[9] Chen, Z.-L., Powell, W.B.: A column generation based decomposition algorithm for a parallel machine just-in-time scheduling problem. European Journal of Operational Research **116**(1), 220–232 (1999)

[10] Fei, H., Meskens, N., Chu, C.: A planning and scheduling problem for an operating theatre using an open scheduling strategy. Computers & Industrial Engineering **58**(2), 221–230 (2010)

[11] Vance, P.H., Barnhart, C., Johnson, E.L., Nemhauser, G.L.: Solving binary cutting stock problems by column generation and branch-and-bound. Computational optimization and applications **3**, 111–130 (1994)

[12] Desrochers, M., Desrosiers, J., Solomon, M.: A new optimization algorithm for the vehicle routing problem with time windows. Operations Research **40**(2), 342–354 (1992)

[13] Potvin, J.-Y., Rousseau, J.-M.: A parallel route building algorithm for the vehicle routing and scheduling problem with time windows. European Journal of Operational Research **66**(3), 331–340 (1993)

[14] Potvin, J.-Y., Rousseau, J.-M.: An exchange heuristic for routeing problems with time windows. Journal of the Operational Research Society **46**(12), 1433–1446 (1995)

[15] Savelsbergh, M.W.: The vehicle routing problem with time windows: Minimizing route duration. ORSA journal on computing **4**(2), 146–154 (1992)

[16] Taillard, É., Badeau, P., Gendreau, M., Guertin, F., Potvin, J.-Y.: A tabu search heuristic for the vehicle routing problem with soft time windows. Transportation science **31**(2), 170–186 (1997)

[17] Russell, R.A.: Hybrid heuristics for the vehicle routing problem with time windows. Transportation science **29**(2), 156–166 (1995)

[18] Yuan, Y., Cattaruzza, D., Ogier, M., Semet, F., Vigo, D.: A column generation based heuristic for the generalized vehicle routing problem with time windows. Transportation Research Part E: Logistics and Transportation Review **152**, 102391 (2021)

[19] Afifi, S., Dang, D.-C., Moukrim, A.: Heuristic solutions for the vehicle routing problem with time windows and synchronized visits. Optimization Letters **10**, 511–525 (2016)

[20] Lin, C.-C., Hung, L.-P., Liu, W.-Y., Tsai, M.-C.: Jointly rostering, routing, and rerostering for home health care services: A harmony search approach with genetic, saturation, inheritance, and immigrant schemes. Computers & Industrial Engineering **115**, 151–166 (2018)

[21] Erdem, M., Koç, Ç.: Analysis of electric vehicles in home health care routing problem. Journal of Cleaner Production **234**, 1471–1483 (2019)

[22] Hassani, M.R., Behnamian, J.: A scenario-based robust optimization with a pessimistic approach for nurse rostering problem. Journal of Combinatorial Optimization **41**, 143–169 (2021)

[23] Liu, M., Yang, D., Su, Q., Xu, L.: Bi-objective approaches for home health-care medical team planning and scheduling problem. Computational and Applied Mathematics **37**(4), 4443–4474 (2018)

[24] Shi, Y., Boudouh, T., Grunder, O.: 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 (2019)

[25] Lanzarone, E., Matta, A.: A cost assignment policy for home care patients. Flexible Services and Manufacturing Journal **24**, 465–495 (2012)

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

[27] Riazi, S., Wigström, O., Bengtsson, K., Lennartson, B.: A column generation-based gossip algorithm for home healthcare routing and scheduling problems. IEEE Transactions on Automation Science and Engineering **16**(1), 127–137 (2018)

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

[29] Liu, W., Dridi, M., Fei, H., El Hassani, A.H.: Hybrid metaheuristics for solving a home health care routing and scheduling problem with time windows, synchronized visits and lunch breaks. Expert Systems with Applications **183**, 115307 (2021)

[30] Liu, W., Dridi, M., Fei, H., El Hassani, A.H.: Solving a multi-period home health care routing and scheduling problem using an efficient matheuristic. Computers & Industrial Engineering **162**, 107721 (2021)

[31] Li, Y., Xiang, T., Szeto, W.Y.: Home health care routing and scheduling problem with the consideration of outpatient services. Transportation research part E: logistics and transportation review **152**, 102420 (2021)

[32] Bahadori-Chinibelagh, S., Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M.: Two constructive algorithms to address a multi-depot home healthcare routing problem. IETE Journal of Research **68**(2), 1108–1114 (2022)

[33] Demirbilek, M., Branke, J., Strauss, A.K.: Home healthcare routing and scheduling of multiple nurses in a dynamic environment. Flexible Services and Manufacturing Journal **33**(1), 253–280 (2021)

[34] Yadav, N., Tanksale, A.: An integrated routing and scheduling problem for home healthcare delivery with limited person-to-person contact. European Journal of Operational Research **303**(3), 1100–1125 (2022)

[35] Fathollahi-Fard, A.M., Govindan, K., Hajiaghaei-Keshteli, M., Ahmadi, A.: A green home health care supply chain: New modified simulated annealing algorithms. Journal of Cleaner Production **240**, 118200 (2019)

[36] Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M., Tavakkoli-Moghaddam, R., Smith, N.R.: Bi-level programming for home health care supply chain considering outsourcing. Journal of Industrial Information Integration **25**, 100246 (2022)

[37] Dai, Z., Zhang, Z., Chen, M.: The home health care location-routing problem

with a mixed fleet and battery swapping stations using a competitive simulated annealing algorithm. Expert Systems with Applications **228**, 120374 (2023)

[38] A Bedside Mobility Assessment Tool for Nurses. https://cdn-links.lww.com/permalink/jncq/a/jncq_2019_05_29_arnold_1800211_sdc5.pdf

[39] Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W., Vance, P.H.: Branch-and-price: Column generation for solving huge integer programs. Operations research **46**(3), 316–329 (1998)

[40] Östergård, P.R.J.: A fast algorithm for the maximum clique problem. Discrete Applied Mathematics **120**(1-3), 197–207 (2002)

[41] Ray, K.N., Chari, A.V., Engberg, J., Bertolet, M., Mehrotra, A.: Opportunity costs of ambulatory medical care in the united states. The American journal of managed care **21**(8), 567–574 (2015)

[42] Trautsamwieser, A., Hirsch, P.: Optimization of daily scheduling for home health care services. JAOR **3**, 124–136 (2011)

[43] Kergosien, Y., Lenté, C., Billaut, J.-C.: Home health care problem an extended multiple traveling salesman problem (2009)

[44] Cohen, B., Hyman, S., Rosenberg, L., Larson, E.: Frequency of patient contact with health care personnel and visitors: Implications for infection prevention. Joint Commission journal on quality and patient safety / Joint Commission Resources **38**, 560–565 (2012)

[45] Tai-Seale, M., McGuire, T.G., Zhang, W.: Time allocation in primary care office visits **42**, 1871–1894 (2007)

[46] Bae, S.-H., Fabry, D.: Assessing the relationships between nurse work hours/overtime and nurse and patient outcomes: Systematic literature review. Nursing Outlook **62**(2), 138–156 (2014)

[47] Bureau of Labor Statistics, Occupational Employment and Wage Statistics. https://www.bls.gov/oes/current/oes291141.htm

[48] Giambruno, C., Cowell, C., Barber-Madden, R., Mauro-Bracken, L.: The extent of barriers and linkages to health care for head start children. Journal of Community Health **22**, 101–114 (1997)

[49] Branch, L.G., Nemeth, K.T.: When elders fail to visit physicians. Medical Care **23**(11), 1265–1275 (1985)