mathematics

MDPI

Article

Balancing Efficiency, Patient Preference, and Operational Caregiver Burden in Dynamic Home Healthcare Scheduling: A Four-Objective NSGA-II Approach

Radhia Zaghdoud

Computer Science Department, College of Science, Northern Border University, Arar 91431, Saudi Arabia; radhia.zaghdouud@nbu.edu.sa

Abstract

In the home healthcare sector, patients often prefer familiar caregivers. Honoring these preferences can unbalance caregivers' workloads, which can overload some caregivers and disrupt the day's schedule. Patient cancellations, no-shows, emergencies, and caregiver delays and absences further disrupt schedules. This study proposes a four-objective mathematical model of the home healthcare scheduling and routing problem that jointly optimizes patient preference satisfaction, an operational caregiver burden index, travel distance, and workload balance; the index, built from duty time, patient acuity, and late visits, is a proxy, not a validated measure of burnout. The solution method combines NSGA-II with a local search procedure, and a dynamic extension re-optimizes the schedule when disruptions occur. The model is tested on 18 adapted Solomon instances spanning three scales and three distribution structures. A controlled ablation shows that optimizing the index lowers it in every instance (median change −25%, −15%, and −9% at 25, 50, and 100 patients, and −25% with seven caregivers on six 50-patient instances), but at the cost of 18%, 37%, and 39% more travel and, from 50 patients onward, worse patient preferences; it lowers the share of late visits but does not reduce overtime. NSGA-II is competitive with NSGA-III and MOEA/D. In 1080 simulated disruptions, re-optimization reduces travel, lateness, and overtime relative to keeping the plan, but it increases workload imbalance, does not lower the realized burden, and moves 9% to 11% of the patients to another caregiver (3% to 4% under keeping the plan).

Keywords: home healthcare; operational caregiver burden; patient preferences; dynamic scheduling; multi-objective optimization; NSGA-II

MSC: 90C59; 90B35; 90C29

Check for updates

Academic Editors: Ke Fu and Jiayan Xu

Received: 22 September 2026
Revised: 30 October 2026
Accepted: 2 November 2026
Published: 4 November 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Home healthcare occupies a challenging position within modern care systems. It is valued for its ability to provide elderly and mobility-limited patients with the care they require in their own homes, preserving their independence in ways that institutional care cannot replicate. However, the care that patients appreciate is often delivered by caregivers who can face demanding working conditions: traveling between multiple homes, adhering to tightly scheduled visits, and having little flexibility when unexpected issues arise. Honoring patients' preferences for familiar caregivers can add to this imbalance. When it continues, it can lead to caregiver burnout, which in turn can disrupt the daily schedule.

Scheduling and routing in home healthcare has been studied extensively. Reviews by [1,2] document a field that has matured considerably over the past two decades, moving from deterministic formulations toward dynamic and stochastic ones, and from single-objective cost minimization toward multi-objective models that weigh organizational, caregiver, and patient perspectives. Recent contributions have added uncertainty [3], learning-based solution methods [4], dynamic matching of caregiver–patient preferences [5], and explicit consideration of caregiver waiting time and workload fairness [6].

Among the studies we reviewed, none optimizes a caregiver burden measure directly. Burnout appears as motivation, as a constraint, as a measured outcome, or through proxies such as fatigue or workload variance. That omission has two consequences. First, it is not known how much burden-aware scheduling actually reduces the load of the most burdened caregiver. Second, it is not known what it costs: whether explicitly optimizing such a measure forces a meaningful sacrifice in travel efficiency or patient preference satisfaction, and whether any such trade-off holds across problem scales. A parallel gap runs through the dynamic literature. Dynamic models respond well to external uncertainty (new requests, cancellations by patients), but they rarely re-optimize in response to caregiver-side disruptions, such as delays or the unavailability of caregivers.

This paper addresses both gaps. We propose a four-objective home healthcare scheduling and routing model that jointly optimizes patient preference satisfaction, an operational caregiver burden index (a proxy for burnout), total travel distance, and workload balance. The solution approach combines Non-dominated Sorting Genetic Algorithm II (NSGA-II) [7] with a local search procedure and a dynamic extension that re-optimizes the schedule in response to disruptions on the patient side and on the caregiver side.

The contributions of this paper are as follows.

- We formulate a bounded operational burden index for the most burdened caregiver, built from duty time, patient acuity, and late visits, and optimize it as an explicit objective in a four-objective problem, in addition to travel distance, workload balance, and patient preferences. Skill requirements are hard constraints; time windows and shift ends are soft.
- We specify a complete NSGA-II with constrained domination and a local search, and compare it under equal evaluation budgets with a variant without local search, NSGA-III, and MOEA/D. Relative to the closest earlier studies [6,8], the burden objective, the dynamic extension, and this evaluation are new (Table 1).
- We extend the model with a dynamic re-optimization mechanism driven by the routes, with five event types, including the caregiver-side events delay and unavailability, and we compare it with a keep-the-plan policy.
- We quantify, on 18 adapted Solomon instances [9] with 20 runs each, what optimizing the index buys and costs, with effect sizes, confidence intervals, and corrections for multiple testing, and we test the sensitivity of the results to the index's weights.

Taken together, the results show that optimizing the burden index consistently lowers it at a clear price in travel distance and, from 50 patients on, in patient preference. They also show where the approach does not help: overtime on the larger instances, and the realized burden after re-optimization.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the problem, and Section 4 presents the mathematical model. Section 5 describes the solution approach, including the burden index computation, the dynamic re-optimization, NSGA-II and its local search, and the comparison algorithms. Section 6 describes the experimental study, and Section 7 reports the results. Section 8 discusses the findings and their limitations, and Section 9 concludes the paper.

Table 1. Comparison with the closest earlier studies, as described in Section 2.

Study	Objectives	Patient Preference	Caregiver Burden Measure	Disruptions
[8]	travel distance, workload imbalance, patient preference (health-aware)	yes	no	none (static)
[10]	travel time, workload fairness, waiting time (dynamic version of [6])	no	no	uncertain number of patients
[6]	travel time, workload fairness, waiting time	no	no	none (static)
[11]	arrival-time difference of two caregivers, workload balance, patient-satisfaction factors	yes	no	none (daily plan)
[12]	unified formulation of many real-world features	yes (preferred service times)	no	none (static)
[3]	service cost, patient satisfaction, workload balance	yes (satisfaction, interpersonal relationships)	no	uncertain departure time, demand, and travel time (robust)
[5]	caregiver–patient matching and satisfaction, patient priority	yes	no	dynamic matching
This paper	travel distance, workload balance, patient preference, burden index	yes	yes (operational proxy, not validated)	cancellation, no-show, emergency, caregiver delay, caregiver unavailability

2. Literature Review

2.1. Scope and Context

The home healthcare scheduling and routing problem (HHCSRP) is, at its core, a problem of assigning caregivers to patients and ordering visits under constraints that are rarely negotiable: time windows, skill matching, workload, and travel distance. The field has grown quickly over the past two decades, and three recent reviews [1,2,13] map it from different perspectives. These reviews trace a field that has moved away from static, deterministic models toward dynamic and stochastic formulations that better reflect the uncertainty home care agencies face. The literature is commonly organized by problem type, objective, constraint, benchmark, and solution method. Across these dimensions, a consistent picture emerges: cost efficiency, workload balance, and patient satisfaction are the objectives most often in conflict, while caregiver qualifications and time windows remain the most common constraints. Exact algorithms, heuristics, and metaheuristics all appear, but local search and hybrid algorithms dominate, with hybrids gaining ground since 2016. Recurring challenges include logistical complexity and workload imbalance, and the reviews point to adaptive scheduling, AI and IoT adoption, and uncertainty-aware, stakeholder-integrated models as the most promising directions. Running beneath all of these trends is a demographic shift: the population over 65 continues to grow, pushing care out of hospitals and into the home, and with it a renewed focus on caregiver–patient matchmaking and operational efficiency.

2.2. Multi-Objective Optimization

Multi-objective formulations have become the default rather than the exception. Ref. [14] examines the HHCSRP by combining online diagnosis with onsite home care, pursuing three objectives: operating cost, caregiver satisfaction, and patient satisfaction.

The proposed approach consists of a hybrid non-dominated sorted iterative greedy algorithm with adaptive destruction, a hybrid initial population strategy, six knowledge-based destruction operators, hybrid reconfiguration, and a problem-specific local search.

Not all multi-objective models treat uncertainty the same way. Ref. [8] studies a static HHCSRP minimizing travel distance and caregiver workload imbalance while maximizing patient preference satisfaction, which includes each patient's choice of caregiver. The authors propose a hybrid of NSGA-II and a heuristic that focuses on vulnerable patients, such as the elderly and those with chronic illnesses. Ref. [3], by contrast, builds a robust model that incorporates three uncertain parameters—caregiver departure time, patient demand, and travel time—while optimizing service cost, patient satisfaction, and caregiver workload balance under constraints related to skill matching, interpersonal relationships, and maximum working hours. Their solution is an improved multi-task constrained multi-objective algorithm that adds a multi-modal crossover operator, a local search operator, and an elite guidance strategy.

Learning-based approaches have entered this stream as well. The authors of Ref. [4] observe that most deep reinforcement learning (DRL) methods for HHCSRP are built for single-objective settings and handle conflicting objectives poorly. They respond with a weight-aware DRL framework for multi-depot routing that combines a Transformer-based pointer network with a weight-aware modulation mechanism and embeds Adaptive Large Neighborhood Search (ALNS) to improve both the quality and the diversity of the Pareto front.

Two further contributions extend the multi-objective view in different directions. Ref. [15] present a Decision Support System that combines fuzzy logic, gray relational analysis, and NSGA-II in three stages: facility location under uncertainty, cost minimization with attention to staff misallocation and service quality, and NSGA-II for scheduling and routing. The authors of Ref. [5] focus on dynamic matching of caregiver–patient preferences and patient priority. Satisfaction was quantified through prospect theory, and their two-stage MILP was solved with a hybrid that couples the Ivy Algorithm with an improved NSGA-III.

Across these studies, cost and patient satisfaction appear consistently, and caregiver well-being enters mainly as workload balance or satisfaction. No study in this group treats burnout as a distinct objective.

2.3. Caregiver Workload, Waiting Time, Fatigue, and Well-Being

Workload balance, waiting time, fatigue, and emotional well-being have shifted from the periphery toward the center, though they are still treated less systematically than cost and patient satisfaction. Ref. [6] proposes a mathematical model that optimizes three conflicting objectives—total travel time, fairness in workload allocation, and waiting time between patient visits—and solves it using a two-stage hybrid approach that combines NSGA-II with a tailored local search inserted at mid-generations to avoid premature convergence. Ref. [10] address the dynamic version of this problem with a two-stage approach: K-means++ clustering of patients and caregivers in the first stage, followed by a hybrid genetic algorithm in the second. Their model is solved under both deterministic and uncertain conditions, with the number of patients as the uncertain parameter.

Fatigue appears in a smaller set of studies. Ref. [11] addresses the HHCSRP with a daily scheduling and routing model for patients requiring two caregivers, minimizing the difference between the arrival times of the two caregivers. The model takes into account high-priority visits, workload balance, working hours, and factors related to patient satisfaction, such as preferred time windows, gender preferences, and caregiver–patient loyalty, which is influenced by caregiver fatigue and time-dependent travel times. The proposed

approach is a column-generation heuristic with a customized labeling algorithm for the routing subproblem.

The authors of Ref. [12] propose a unified formulation for the home healthcare routing and scheduling problem that consolidates real-world features from earlier models into a single framework. They provide an instance generator and a validation infrastructure and compare two solution methods, simulated annealing and a compact mixed-integer linear program, showing that general-purpose methods can outperform formulation-specific ones on many instances. Ref. [16] considers a multi-day staffing, routing, and scheduling problem with varying daily visit needs and caregiver qualification levels, optimizing additional caregiver shifts, emergency allocations, and paths under time windows and labor rules. Ref. [17] minimizes operational cost and maximize provider utilization under skill matching, work-time limits, and periodic scheduling, solving the model with an Adaptive Large Neighborhood Search and testing it on both random instances and a real case. Waiting time and workload fairness are now optimized explicitly, and fatigue has entered a few models. Healthcare rostering offers a related perspective. Reviews of nurse and healthcare-provider rostering [18,19] show that fatigue and workload are handled mainly through constraints and soft objectives, such as limits on consecutive working days and shifts, minimum rest times, and a fair distribution of unpopular shifts, while burnout itself is usually studied as an outcome to be measured with instruments [20] and not as a quantity to be optimized.

2.4. Dynamic, Stochastic, and Uncertain Environments

Static models assume that all requests are known before routing. Real operations rarely work that way, and a substantial literature now addresses uncertainty and dynamic requests. The authors of Ref. [21] formulate a robust routing–scheduling model with uncertain service and travel times, subject to temporal precedence, synchronization, sample-transfer timing, and multi-deployment constraints that allow a patient to be visited more than once per day. They compare simulated annealing, genetic algorithms, and memetic algorithms, finding the memetic algorithm to be the strongest on large instances, and they show that the robust model outperforms the deterministic one. They also propose a dynamic variant and test two dispatching policies on it.

Reinforcement learning has become a common tool for the online case. The authors of [22] address the Online Dynamic Home Health Care Scheduling Problem, in which the agency must decide whether to accept or reject each incoming request and, if it accepts, immediately adjust the schedule and routes. The objective is to maximize the number of patients served under limited resources, and because future requests are unknown and decision time is short, the authors model the problem as a Markov decision process and solve it with reinforcement learning.

Two further studies approach the dynamic problem from different angles. Ref. [23] studies the dynamic problem of prioritized routing and scheduling with cooperating service providers by using a lexicographic optimization framework to prioritize patient visits, travel distance, workload balance, and waiting time. The proposed approach is based on Greedy Randomized Adaptive Search Procedure followed by Variable Neighborhood Search. The authors of Ref. [24] study a related dynamic problem where referral counts and required visit numbers are uncertain, targeting the single-nurse case with an MDP that minimizes expected rejection, diversion, overtime, and travel costs. Because the standard approximate linear programming (ALP) approach only scales to very small instances, they derive a closed-form solution for the optimal ALP parameters in a special case and use it to motivate two heuristic reductions that handle large instances in acceptable time.

What these models share is a focus on external uncertainty—new requests, uncertain arrivals, and cancellations by patients. Caregiver-side disruptions, such as delays or the

unavailability of a caregiver, are not usually modeled as triggers for re-optimization. Decomposition and metaheuristic approaches have also been used for dynamic home healthcare planning in multi-service networks [25].

2.5. Solution Methods

The solution approaches vary from exact methods to metaheuristics and learning-based approaches, with a clear trend toward hybrids. The authors of Ref. [26] study a version of HHCSRP in which a shared vehicle moves multiple healthcare professionals between patients, with time windows, vehicle capacity, synchronized lunch breaks, and precedence and coupling constraints between visits and professional pickups. Their formulation is a pickup-and-delivery extension with each patient represented as a delivery–pickup node pair, which makes the coordination of visits, transport, and team deployment explicit. The solution is a matheuristic: a partial route is built over delivery nodes using a Traveling Salesman Problem with Time Windows (TSPTW) formulation, then an exact phase inserts pickup nodes while respecting the constraints, and a 2-opt-style local search explores alternative delivery sequences. Granular search and memory-based filtering keep the neighborhood exploration tractable.

Clustering-then-routing is another recurring pattern. Ref. [27] clusters patients based on their available time using K-means, generates initial routes within each cluster using the nearest neighbor method, and improves these routes through tabu search. The three-phase design trades some solution quality for computational simplicity, which the authors argue is useful when operational resources and processing time are limited. The authors of Ref. [28] propose a similar cluster-first, route-second scheme with lightweight temporal–spatial clustering and multi-route tabu search, organized as demand segmentation, greedy initial construction, and iterative improvement. They offer three variants: a baseline (M1), a sub-clustering extension (M2), and a priority-aware variant (M3) with representative-based initialization and differentiated penalties by patient criticality.

Across these approaches, NSGA-II paired with a problem-specific local search remains common, but we found no study that evaluates this pairing through both a controlled ablation and a dynamic simulation.

When four or more objectives are optimized, the selection pressure of Pareto dominance weakens, because a large share of the population becomes non-dominated. Two families of algorithms address this regime. Reference-point methods such as NSGA-III [29] replace the crowding distance of NSGA-II by the association of solutions with a set of well-spread reference points. Decomposition methods such as MOEA/D [30] turn the problem into a set of scalar subproblems that are solved cooperatively with the help of neighboring subproblems. Because the model of this paper has four objectives, we use NSGA-III and MOEA/D as comparison algorithms in Section 7.

2.6. Synthesis and Research Gap

Progress in HHCSRP has been made from multiple directions simultaneously, though not always in synchronization. Multi-objective models are now standard, and most weigh cost, patient satisfaction, and caregiver satisfaction or workload balance together [3,8,14]. Uncertainty has been addressed through dynamic, stochastic, and robust formulations covering demand, service time, travel time, and arrivals [10,21,22,24]. Solution methods have diversified as well: exact methods still appear, but metaheuristics, hybrid metaheuristics, and learning-based approaches now carry much of the field [4,26–28]. Caregiver-side objectives have followed more slowly. Waiting time and fatigue have entered the objective function in a handful of studies [6,10,11], and workload balance is now treated as a first-class objective rather than a by-product of cost [5,8].

The most conspicuous gap is burnout. Among the studies reviewed here, none optimizes the load of the most burdened caregiver directly. Burnout appears as motivation, as a constraint, as a measured outcome, or as a proxy, such as fatigue or workload variance, and the closest models balance such a proxy among the caregivers. The consequence is that we do not know how much burnout-aware scheduling helps, what it costs in travel efficiency or patient preference satisfaction, or whether the benefit holds as problem size grows. That last question matters more than it might seem. Dynamic and stochastic models handle uncertainty well, but they respond mostly to external events, such as new requests and cancellations by patients, and less often to caregiver-side events, such as delays or the unavailability of a caregiver, as triggers for re-optimization.

A related gap concerns patient preference. Time windows and skill matching are standard, and the choice of caregiver has been modeled explicitly [8], but it rarely appears alongside caregiver burnout in the same objective set. Methodologically, NSGA-II paired with a dedicated local search is common, yet the specific combination evaluated through both controlled ablation and dynamic simulation is uncommon. Taken together, these gaps point to a single question that the literature has raised but not answered: what does it cost, in travel efficiency and patient preference satisfaction, to treat caregiver burnout as a first-class objective rather than a constraint?

2.7. Present Study

This study addresses that question. We propose a mathematical model that jointly optimizes patient preference satisfaction, an operational caregiver burden index (a proxy for burnout), total travel distance, and workload balance, a four-objective formulation. The solution approach is NSGA-II combined with a local search procedure, and a dynamic extension re-optimizes the schedule in response to disruptions on the patient side (cancellations, no-shows, emergencies) and on the caregiver side (delays, unavailability). We evaluate the approach on 18 adapted instances of the Solomon benchmark, spanning three problem scales and three customer-distribution structures, with a controlled ablation, a comparison with NSGA-III and MOEA/D, a sensitivity analysis of the index, and a dynamic simulation compared with a keep-the-plan policy. Table 1 compares the present paper with the earlier studies.

3. Problem Description

The HHCSRP considered in this paper concerns the assignment of a set of patients, each requiring a home visit of a given duration and service type, to a set of caregivers, each with a skill level qualifying them for a subset of service types, and the sequencing of the visits of each caregiver. Every caregiver departs from and returns to a central depot and visits the assigned patients in sequence. The skill requirement is a hard constraint: a caregiver can serve a patient only if the caregiver's skill level is at least the level required by the service. Each patient has a time window. A caregiver who arrives early waits until the window opens, and a visit that starts after the window closes is called late. Each caregiver is available between a start time and an end time, and returning to the depot after the end time is called overtime. Windows and availability ends are treated as soft: the search does not forbid late visits or overtime, and every reported schedule is evaluated against them (Section 6). In the adapted benchmark instances, the number of caregivers is small (two, three, and four caregivers for 25, 50, and 100 patients), and on some of them, a complete day cannot be served within all shifts (Section 6); Section 7 reports the lateness and overtime of the schedules found.

Beyond this operational core, the problem incorporates three additional considerations motivated by the practical realities of home healthcare delivery. First, many patients express

a preference for a specific caregiver, reflecting familiarity and personal rapport. Honoring this preference is treated as an explicit scheduling objective. The health-aware mechanism of the earlier work [8] weights the strictness of enforcement by the health profile of the patient (age, blood pressure, and weight); in the present paper, this health-based strictness is used only in the construction of the initial population (Section 5.3), while the preference objective counts the share of patients served by their preferred caregiver, and the health profile enters the model through the acuity component of the burden index (Section 4). Second, the resulting distribution of workload across caregivers must remain balanced, and the burden it creates for the most loaded caregiver must be controlled. Concentrating visits, travel, and clinically demanding patients on a subset of caregivers creates a burden that this paper models and optimizes independently of raw workload balance, as two schedules with the same total working time per caregiver can differ considerably in the burden they place on the same caregiver. The burden index is an operational proxy for this load; it is not a validated measure of burnout. Third, the schedule is not assumed to be static. Over the course of an operational day, patients may cancel, fail to appear, or require emergency care, and caregivers may themselves become delayed or unavailable. Each such disruption invalidates part of the current schedule and requires timely re-optimization over the affected patients, without discarding the parts of the schedule that remain valid or changing the visits that have already been carried out.

The problem is therefore posed as a four-objective optimization problem, minimizing total travel distance, workload imbalance, preference dissatisfaction, and the caregiver burden index, solved initially for a full operational day and re-solved dynamically as disruptions occur.

4. Mathematical Model

We formulate the HHCSRP with a caregiver burden index as a multi-objective mathematical program with binary and continuous variables, extending the three-objective formulation of [8] with an explicit caregiver burden objective. The travel, workload, and preference objectives and all constraints are linear in the decision variables, whereas the burden objective is not: it contains ratios whose denominators depend on the assignment, the indicator of a late start, and a maximum over the caregivers. The model is therefore a nonlinear mixed-integer program and not a mixed-integer linear program; the burden objective is evaluated as a black box by the metaheuristic of Section 5, and no linearization is claimed.

4.1. Sets and Indices

- $P = \{1, \dots, n\}$: set of patients.
- $P' \subseteq P$: subset of patients who express a caregiver preference.
- $K = \{1, \dots, m\}$: set of all caregivers of the instance, including caregivers to whom no patient is assigned.
- $N = P \cup \{0\}$: set of nodes, where node 0 denotes the depot.
- S: set of service types; each service type requires a skill level.

4.2. Parameters

- $d_{ij} \geq 0$: travel distance between nodes $i, j \in N$.
- $t_{ij} \geq 0$: travel time between nodes $i, j \in N$; we set $t_{ij} = d_{ij}$ (unit speed).
- $TS_i > 0$: required service time for patient $i \in P$.
- $[e_i, l_i]$: time window for patient i; a caregiver who arrives before e_i waits, and the upper bound l_i is soft (a start after l_i is a late visit).
- $\sigma_i \in \{1, \dots, \Lambda\}$: skill level required to serve patient i (the level required by its service type), where Λ is the highest skill level.

- $\kappa_k \in \{1, \dots, \Lambda\}$: skill level of caregiver k; caregiver k is eligible to serve patient i iff $\kappa_k \geq \sigma_i$.
- $\pi_i \in K \cup \{\emptyset\}$: preferred caregiver of patient i ($\emptyset$ if the patient has no preference).
- A_i, W_i, BP_i: age, weight, and blood pressure of patient i, used to compute the acuity a_i of the burden index.
- s_k: start time of caregiver k (departure from the depot).
- E_k: end of the availability of caregiver k; returning to the depot after E_k is allowed and is reported as overtime.
- M: sufficiently large constant.

Time is measured in the time units of the Solomon instance, with $t = 0$ the start of the planning horizon at the depot.

4.3. Decision Variables

$$x_{ik} \in \{0,1\}, \quad \text{1 if patient } i \text{ is assigned to caregiver } k \tag{1}$$
$$y_{ijk} \in \{0,1\}, \quad \text{1 if caregiver } k \text{ travels directly from } i \text{ to } j,\ i \neq j \tag{2}$$
$$u_i \geq 0, \quad \text{service start time at patient } i \tag{3}$$
$$r_k \geq 0, \quad \text{return time of caregiver } k \text{ to the depot} \tag{4}$$
$$\ell_i \geq 0, \quad \text{lateness of the visit at patient } i \tag{5}$$
$$o_k \geq 0, \quad \text{overtime of caregiver } k \tag{6}$$
$$z_{\max} \geq 0, \quad \text{auxiliary for maximum workload} \tag{7}$$
$$z_{\min} \geq 0, \quad \text{auxiliary for minimum workload} \tag{8}$$

4.4. Objective Functions

f_1: Total Travel Distance (TD).

$$f_1 = TD = \sum_{k \in K} \sum_{i \in N} \sum_{j \in N, j \neq i} d_{ij}\, y_{ijk} \tag{9}$$

f_2: Workload Balance (WLB).

The workload of caregiver k is its total service time, $LW_k = \sum_{i \in P} x_{ik} TS_i$. Travel and waiting time are not part of this objective; they enter the burden index f_4 through the duty time. With the auxiliary variables $z_{\max}$ and $z_{\min}$,

$$f_2 = WLB = z_{\max} - z_{\min}. \tag{10}$$

The range is taken over all caregivers of K (constraints of the workload-balance block in Section 4.5). A caregiver without visits has workload zero, so the objective penalizes leaving a caregiver idle; the burden index gives such a caregiver the burden zero, and both definitions use the same set K. In the 18 instances of Section 6, no executed schedule leaves a caregiver without visits.

f_3: Patient Preference.

The objective is the share of the patients with a preference who are not served by their preferred caregiver, over the preference-holding patients $P' \subseteq P$:

$$f_3 = 1 - \frac{1}{|P'|} \sum_{i \in P'} x_{i,\pi_i}. \tag{11}$$

It is linear in the decision variables.

f_4: Caregiver Burden Index.

Let $TD_k = \sum_{i,j \in N, j \neq i} d_{ij}\, y_{ijk}$ be the travel distance, $NB_k = \sum_{i \in P} x_{ik}$ the number of visits, and $DT_k = LW_k + TD_k$ the duty time (service plus travel) of caregiver k. The acuity of patient i is the average of three min–max normalized health attributes,

$$a_i = \frac{1}{3}\left(\frac{A_i - A_{\min}}{A_{\max} - A_{\min}} + \frac{BP_i - BP_{\min}}{BP_{\max} - BP_{\min}} + \frac{W_i - W_{\min}}{W_{\max} - W_{\min}} \right) \in [0,1], \tag{12}$$

where the minima and maxima are taken over the patients of the instance (a term is zero if its range is zero). For a caregiver with $NB_k > 0$, we define three components in $[0,1]$:

$$\delta_k = \frac{DT_k}{\sum_{h \in K} DT_h}, \qquad \alpha_k = \frac{\sum_{i \in P} a_i\, x_{ik}}{\sum_{i \in P} a_i}, \qquad \lambda_k = \frac{1}{NB_k} \sum_{i \in P} x_{ik}\, 1[u_i > l_i], \tag{13}$$

namely the share of the caregiver in the total duty time, the share in the total patient acuity, and the share of the caregiver's visits that start after the window has closed. The burden of caregiver k is

$$B_k = w_1 \delta_k + w_2 \alpha_k + w_3 \lambda_k, \qquad w_1 = w_2 = w_3 = \tfrac{1}{3}, \tag{14}$$

with $B_k = 0$ if $NB_k = 0$, and the fourth objective is the burden of the most burdened caregiver:

$$f_4 = \max_{k \in K} B_k. \tag{15}$$

Rationale of the components. The duty share represents how much of the working time of the team, including driving, a caregiver carries. The acuity share represents the clinical demand of the assigned patients; we assume that older patients and patients with higher blood pressure or weight require more care effort, which is a simplifying assumption. The late share represents time pressure: a visit that cannot start before the window closes forces the caregiver to hurry or to postpone other work. Waiting time between visits is not a component, because its effect is ambiguous: a short pause can allow a caregiver to recover, whereas a schedule without pauses can increase stress. The index contains no thresholds, caps, or calibration constants: each component is a share or a proportion, bounded in $[0,1]$, and the only free choice is the vector of weights, whose influence is tested in Section 7. Because the shares δ_k and α_k each sum to one over the m caregivers, the index of any schedule is at least $2/(3m)$, that is, 0.33, 0.22, and 0.17 for $m = 2$, 3, and 4; this floor is a lower bound that is not attained in general, since the late share is non-negative and the skill requirements restrict the assignments. Their mean over the caregivers is therefore constant whatever the distribution of the load, so that an index based on the mean could be lowered without helping anyone, whereas the maximum is sensitive to how the load is distributed. The index is an operational proxy for the load placed on a caregiver; it has not been validated against caregiver-reported well-being, and we make no claim about burnout in the occupational-health sense.

Overall objective.

$$\min\,(f_1, f_2, f_3, f_4) = \min\,(TD,\ WLB,\ \text{Patient Preference},\ \text{Burden}) \tag{16}$$

4.5. Constraints

Assignment and eligibility.

$$\sum_{k \in K} x_{ik} = 1, \qquad \forall i \in P \tag{17}$$

$$x_{ik} = 0, \qquad \forall i \in P,\ k \in K : \kappa_k < \sigma_i \tag{18}$$

Routing flow conservation.

$$\sum_{j \in N, j \neq i} y_{ijk} = x_{ik}, \qquad \forall i \in P, k \in K \tag{19}$$

$$\sum_{i \in N, i \neq j} y_{ijk} = x_{jk}, \qquad \forall j \in P, k \in K \tag{20}$$

Depot.

$$\sum_{j \in P} y_{0jk} = \sum_{i \in P} y_{i0k} \leq 1, \qquad \forall k \in K \tag{21}$$

$$\sum_{j \in P} y_{0jk} \geq x_{ik}, \qquad \forall i \in P, k \in K \tag{22}$$

Time consistency.

$$u_i \geq s_k + t_{0i} - M(1 - y_{0ik}), \qquad \forall i \in P, k \in K \tag{23}$$

$$u_j \geq u_i + TS_i + t_{ij} - M(1 - y_{ijk}), \qquad \forall i, j \in P, i \neq j, k \in K \tag{24}$$

$$u_i \geq e_i, \qquad \forall i \in P \tag{25}$$

$$r_k \geq u_i + TS_i + t_{i0} - M(1 - y_{i0k}), \qquad \forall i \in P, k \in K \tag{26}$$

Soft criteria (evaluated and reported, not enforced).

$$\ell_i \geq u_i - l_i, \quad \ell_i \geq 0, \qquad \forall i \in P \tag{27}$$

$$o_k \geq r_k - E_k, \quad o_k \geq 0, \qquad \forall k \in K \tag{28}$$

Workload balance linearization.

$$z_{\max} \geq LW_k, \qquad \forall k \in K \tag{29}$$

$$z_{\min} \leq LW_k, \qquad \forall k \in K \tag{30}$$

5. Approach Description

This section describes how a candidate schedule is evaluated, including the computation of the caregiver burden index (Section 5.1), the dynamic re-optimization mechanism (Section 5.2), the NSGA-II procedure and its local search (Section 5.3), and the comparison algorithms (Section 5.4).

5.1. Burden Index Computation

Algorithm 1 computes the burden index of an evaluated individual, that is, of a schedule whose routes have been propagated in time: each caregiver starts at the depot at s_k, drives to the first patient, waits until e_i if the arrival is early, serves the patient, continues to the next patient, and returns to the depot after the last visit. The same evaluation yields the other three objectives, the number of skill violations, the lateness, and the overtime. The burden objective is therefore evaluated as a black box: the three shares of each caregiver and their maximum are computed directly from the resulting schedule (Section 4), and no linearization is needed.

Algorithm 1 Caregiver burden index computation.

procedure ComputeBurden(Individual I, Caregivers K, weights w_1, w_2, w_3)
  for $k \in K$ **do**
    $LW_k \leftarrow \sum_{i:\, x_{ik}=1} TS_i, \quad NB_k \leftarrow |\{i : x_{ik} = 1\}|$
    $TD_k \leftarrow$ travel distance of the route of k, including the legs from and to the depot
    $NL_k \leftarrow |\{i : x_{ik} = 1,\ u_i > l_i\}|$ ▷ number of late visits
  end for
  $DT_{\text{tot}} \leftarrow \sum_{k \in K}(LW_k + TD_k), \quad a_{\text{tot}} \leftarrow \sum_{i \in P} a_i$
  $f_4(I) \leftarrow 0$
  for $k \in K$ **do**
    if $NB_k = 0$ **then**
      $B_k \leftarrow 0$ ▷ caregiver without visits
    else
      $\delta_k \leftarrow (LW_k + TD_k)/DT_{\text{tot}}$ ▷ duty share
      $\alpha_k \leftarrow \sum_{i:\, x_{ik}=1} a_i \,/\, a_{\text{tot}}$ ▷ acuity share
      $\lambda_k \leftarrow NL_k/NB_k$ ▷ late share
      $B_k \leftarrow w_1\delta_k + w_2\alpha_k + w_3\lambda_k$
    end if
    $f_4(I) \leftarrow \max\big(f_4(I),\, B_k\big)$
  end for
  return $f_4(I)$
end procedure

5.2. Dynamic Re-Optimization

The initial plan of a simulated day is the executed schedule of a static optimization. Five types of events can occur: cancellation and no-show of a patient who has not yet been visited (a no-show may leave the caregiver with a wasted trip); an emergency, a new visit that is added to the plan; a delay of a caregiver, who is shifted by a given duration; and the unavailability of a caregiver, who leaves for the rest of the day. The last two are caregiver-side events. The parameters of the simulation are given in Section 6. A visit is committed at the time T of an event if its caregiver has left for it at or before T; this covers visits that are completed, in progress, or on the way. To determine this, the route of each caregiver is followed from the caregiver's start time and location, computing for each visit the departure, arrival, service start, and service end times; the first visit whose departure time is after T, and all later visits of the route, remain open. Committed visits keep their computed times and are never changed by a re-optimization. Each caregiver restarts from the location of the last committed visit and from the time at which that visit ends, or from the current time T if that is later (a caregiver without committed visits starts from the depot at time T). The reduced instance therefore contains the open visits, the available caregivers, and, for each caregiver, a starting location and a starting time; the times of the new routes are computed from these starting points, so that they are consistent with the committed visits. A delay shifts the time from which the affected caregiver is available, and an unavailable caregiver is removed from the set of available caregivers, with the open visits of that caregiver returned to the pool. Algorithm 2 describes one re-optimization.

Choice of the plan.The candidate is chosen from the skill-feasible first front by compromise programming: each objective is rescaled to $[0, 1]$ over the front, and the solution with the smallest Euclidean distance to the ideal point is selected, with ties broken by the smaller travel distance. The four objectives therefore count equally, and no single objective decides. The same rule chooses the initial plan of the day and the executed schedule reported in the static experiments. It is one possible rule among several: a provider that gives more weight to the burden of the caregivers, to patient preferences, or to travel cost could use a weighted distance or choose another point of the front, which remains available in full. The

feasibility of a candidate is checked before it is accepted: every remaining patient must be served and every assignment must satisfy the skill requirement. Patients that no remaining caregiver can serve, for example after a departure, are reported as unserved. The stability of the schedule is measured by the share of patients that change caregiver (Section 7); the re-optimization does not penalize such changes explicitly.

Algorithm 2 Re-optimization after an event.

Require: Plan in force, event e at time T
Ensure: Updated plan in force
1: ApplyEvent(e) ▷ CANCEL/NOSHOW remove a patient; EMERGENCY adds a visit; DELAY shifts a caregiver; UNAVAILABLE removes a caregiver
2: CommitVisits(T) ▷ visits whose caregiver has left for them at or before T are fixed
3: Reduced ← open visits, available caregivers, and their starting locations and times
4: Front ← NSGA-II(Reduced) ▷ population 100, 500 generations, initial population seeded with the plan in force and mutated variants
5: Candidate ← CompromiseChoice(Front) ▷ skill-feasible first front, closest to the ideal point
6: **if** Candidate serves all remaining patients **and** has no skill violation **then**
7: plan in force ← Candidate
8: **else**
9: keep the plan in force and log the rejection
10: **end if**
11: **return** plan in force

Keep-the-plan policy. As a baseline, visits that are no longer required are removed, a new or orphaned visit is appended to the route of the eligible available caregiver with the fewest visits, and nothing else changes. Both policies face the same events.

5.3. NSGA-II Procedure

The procedure follows NSGA-II [7], with the modifications described below. Representation. An individual is an ordered list of n patient–caregiver pairs, grouped by caregiver, in which the order inside the block of a caregiver is the visiting order. An individual is evaluated as in Section 5.1; every evaluation is counted, so that algorithms can be compared under equal evaluation budgets.

Constraint handling. An individual ω dominates an individual ω' if $v(\omega) < v(\omega')$, where v is the number of skill violations, or if $v(\omega) = v(\omega')$ and ω is no worse than ω' in the four objectives and better in at least one. A skill-feasible individual therefore always dominates an infeasible one, and no repair operator is needed. The time windows and the end of the shift are soft and are not repaired; their violations are measured.

Algorithm. Algorithm 3 gives the procedure used for the static day, with population size $N = 100$ and $G = 1000$ generations.

The mutation rate p_m is 20% in generations 0–499 and 40% from generation 500.

Initial population. Each individual is built by a randomized constructive procedure. The patients are processed in random order. A patient with a preferred caregiver who meets the skill requirement is assigned to that caregiver with probability η_i, the health-based strictness of the patient,

$$\eta_i = \min\{0.95,\ \max\{0.1,\ 0.5 + g_A(A_i) + g_{BP}(BP_i) + g_W(W_i)\}\},$$

where g_A equals 0.1, 0.2, 0.3, 0.4 for an age of at least 50, 60, 70, 80 years, g_{BP} equals 0.1, 0.2, 0.3 for a blood pressure of at least 140, 160, 180 mmHg, and $g_W = 0.1$ for a weight of at least 120 kg (and 0 below these limits). Otherwise, the patient is assigned to a caregiver who meets the skill requirement, chosen by a score that decreases with the number of

patients already assigned to the caregiver and with their clinical demand (age above 75, blood pressure above 160), plus a small random perturbation: the caregiver with the best score is taken if $\eta_i > 0.7$ or if only one caregiver is eligible, and otherwise a caregiver is sampled with probabilities proportional to the exponential of the scores. The visiting order of a route is the random order in which its patients were assigned. The same procedure creates the random individuals that replace duplicates in Algorithm 3. The strictness η_i and the thresholds belong to this heuristic only: they influence the starting points of the search and not the objectives.

Algorithm 3 NSGA-II with constrained domination and local search.

Require: Instance, population size $N = 100$, number of generations $G = 1000$
Ensure: Skill-feasible first front
create the initial population P_0 of N individuals and evaluate it
for $g = 0, \ldots, G-1$ **do**
    rank P_g by non-dominated sorting and compute the crowding distance within each front
    $Q \leftarrow \emptyset$
    while $|Q| < N$ **do**
        select two parents by binary tournament on (rank, crowding distance)
        create two children by crossover
        mutate a child if fewer than $p_m N$ offspring have been mutated in this generation; a mutant replaces its parent only if it dominates it
        add the two children to Q
    end while
    replace every offspring that duplicates another individual by a new random individual
    $R \leftarrow P_g \cup Q$ without duplicates, completed with random individuals up to $2N$; evaluate the new individuals
    rank R and keep the best N individuals by rank and crowding distance as P_{g+1}
    if $(g+1) \bmod 100 = 0$ **then**
        apply the local search to the first 10 individuals of P_{g+1} in front order
    end if
end for
return the skill-feasible first front of P_G

Crossover. The crossover is applied to every selected pair of parents. It is a two-point crossover on patient identity: the patients are sorted by identifier, two cut points are drawn, and each child takes the caregiver of a patient from its own parent outside the segment between the cut points and from the other parent inside it. Each child inherits the visiting order of its own parent.

Mutation. Two positions of the list of pairs are drawn. If they belong to the same caregiver, the two visits are swapped; otherwise, the two patients exchange their caregivers. A mutant replaces its parent only if it dominates it.

Local search. Every 100 generations, the first 10 individuals in front order (10% of the population) are improved: 50 neighbors are generated one after the other by the mutation operator, and a neighbor replaces the incumbent only if it dominates it.

Random numbers. The random-number generators are seeded from the system clock and from the operating system's entropy source, so that runs are independent but not bit-reproducible.

5.4. Comparison Algorithms

NSGA-II without local search is the proposed algorithm without the local-search step. NSGA-III [29] replaces the crowding distance by the association of the solutions with

reference points: two layers of reference points for four objectives (six and three divisions, 104 points), with random mating and the same operators. MOEA/D [30] decomposes the problem into 120 weight vectors (four objectives, seven divisions; with four objectives, the simplex-lattice design has 84 vectors with six divisions and 120 with seven, and no design gives exactly 100), keeps one individual per weight vector (a population of 120), scalarizes the normalized objectives by the Tchebycheff function, uses neighborhoods of 20 subproblems and at most two replacements per child, and the same operators and local search policy. All variants use the same evaluation, skill-constraint handling, decision rule, and number of generations, so that their numbers of objective evaluations per run are within 5% of each other (the number of evaluations of MOEA/D results from its 120 individuals and its 1000 generations; the matched quantity is the number of evaluations, not the population size, which is 100 for the other algorithms).

6. Experimental Study

This section describes the benchmark instances, algorithmic parameters, computing environment, evaluation metrics, and experimental protocol used to evaluate the proposed approach.

6.1. Benchmark Instances

Instances are adapted from the Solomon VRPTW benchmark suite, spanning three customer-distribution structures: clustered with tight time windows (C-series), randomly distributed (R-series), and mixed clustered/random (RC-series). For each series, two representative instances were selected (C101, C201; R109, R204; RC106, RC208), and each was instantiated at three problem scales: 25 patients served by two caregivers, 50 patients served by three caregivers, and 100 patients served by four caregivers, yielding 18 instances in total. Patient time windows and coordinates are taken directly from the corresponding Solomon instance, and distances and times are expressed in its coordinate units. The service time of every visit is 20 time units (the service times of the original instances are not used). The shift of every caregiver has the length of the planning horizon of the Solomon instance, and the skill levels are given in the caregiver file generated for each adapted instance. The remaining attributes were generated as follows, with a generator seeded by a fixed seed combined with the name of the instance, so that the data files can be regenerated.

- Health attributes. Age is drawn from a normal distribution with mean 72 and standard deviation 14 years, truncated to $[40, 95]$; systolic blood pressure from a normal distribution with mean 133 and standard deviation 19 mmHg, truncated to $[100, 200]$; weight from a normal distribution with mean 74 and standard deviation 14 kg, truncated to $[45, 120]$; the values are rounded to integers. These distributions are illustrative and are not calibrated to a specific population.
- Skill requirements. The service types are numbered in the order in which they first appear in the patient file, and the i-th type requires the skill level $(i \bmod m) + 1$, where m is the highest caregiver skill level of the instance (2, 3, or 4).
- Preferences. The preferred caregiver of each patient that holds a preference is drawn uniformly among the caregivers of the instance.

All data files were audited for consistency (identifiers, skill requirements, coordinates) before the experiments. The search does not enforce the time windows or the shift ends (Section 3); Section 7 reports the lateness and overtime of the schedules found. A necessary condition shows that the shift ends cannot all be respected on six of the 18 instances: the total service time plus the length of a minimum spanning tree over the depot and the patients (a lower bound on the total travel) must not exceed the total length of the shifts, and this fails on R109 and RC106 at all three sizes (ratios between 1.4 and 2.8). On these

instances, the shift limit could be met only with at least 4, 7, and 12 caregivers (R109) and 3, 6, and 11 caregivers (RC106) at 25, 50, and 100 patients. On the other twelve instances, the condition holds, and we do not know whether a schedule that respects all windows and shifts exists.

6.2. Parameter Settings

Table 2 summarizes the parameter settings used across all experiments, held constant across instances to ensure comparability. The parameter values follow the earlier study [6] and were not tuned separately for these instances. Every 100 generations, the first 10 individuals in front order (10% of the population) are improved by 50 successive mutations, each accepted only if it dominates the incumbent.

Table 2. Algorithmic parameters.

Parameter	Value
Population size	100
Generations (static optimization)	1000
Re-optimization after an event	population 100, 500 generations
Crossover	applied to every selected pair of parents
Mutation rate	20% of the offspring (generations 0–499), 40% (generations ≥ 500)
Local search	every 100 generations, first 10 individuals in front order, 50 neighbors each
Independent runs per instance and algorithm (static and dynamic)	20
Weights of the burden index (w_1, w_2, w_3)	(1/3, 1/3, 1/3)
NSGA-III reference points	104 (two layers, 6 and 3 divisions)
MOEA/D	population 120 (one individual per weight vector), neighborhood 20, at most 2 replacements
Hypervolume estimation	Monte Carlo, 100,000 samples

6.3. Computing Environment

All algorithms were implemented in C++and compiled with Microsoft Visual Studio 2022 in Release mode. The experiments ran on a laptop with an Intel Core i7-1355U processor (10 cores, 12 threads, 1.70 GHz base frequency), 16 GB of memory, and a 64-bit Windows operating system, with up to ten runs executed in parallel, one process per run. Because the experimental runs shared the machine, their computation times are not reported as algorithm speeds; algorithms are compared by their numbers of objective evaluations. The random-number generators are seeded from the system clock and from the operating system's entropy source, so that runs are independent but not bit-reproducible; the seed that generated the instance data is fixed. On an otherwise idle machine, with one process at a time, a run of the four-objective algorithm on the static day takes 17, 39, and 127 s at 25, 50, and 100 patients (instances C101-25-2, C101-50-3, and C101-100-4), and one re-optimization after an event (population 100, 500 generations) takes on average 2.4, 5.3, and 23.4 s; a simulated day with three events takes 24, 50, and 202 s. These are single runs, which give the order of magnitude and not the variability between instances. The metrics are defined in Section 6.5.

6.4. Experimental Protocol

Static evaluation.

For each of the 18 instances, 20 independent runs of each algorithm variant were executed. For every run, the skill-feasible first front of the last generation is recorded with its four objective values, the schedule to execute is chosen as defined in Section 6.5, and for

that schedule the share of late visits, the total lateness, the total overtime, and the number of caregivers without visits are computed. The number of objective evaluations of each run is also recorded.

Algorithm comparison.

The proposed NSGA-II is compared with NSGA-II without local search, NSGA-III, and MOEA/D (Section 5.4). All variants use the same evaluation, skill-constraint handling, reference point, decision rule, and number of generations, and the same population size of 100 except MOEA/D, whose population equals its 120 weight vectors, so that their numbers of objective evaluations per run are within 5% of each other. The comparison uses the normalized hypervolume of Section 6.5, tested for each instance and each comparison algorithm with the Mann–Whitney test, with the Holm correction applied over all instance-by-algorithm tests, and a sign test across the 18 instances.

Dynamic evaluation.

For each instance, 20 independent simulated days were executed, so that every pooled result rests on equal sample sizes ($18 \times 20 = 360$ days). All runs were generated by an automated driver that records every run; no run failed and none was discarded. Each day starts from the executed schedule of a static optimization; its length M_0 is the time at which the last caregiver finishes the last visit of that schedule. Three events occur at times drawn uniformly in $[0.15M_0, 0.75M_0]$, each of one of five types drawn with probabilities 0.30 (cancellation), 0.15 (no-show), 0.20 (emergency), 0.20 (caregiver delay), and 0.15 (caregiver unavailable). A cancellation or no-show removes a patient who has not yet been visited; an emergency adds a visit of 20 time units at the location of an existing patient, with the health attributes of that patient, a randomly drawn service type, no preferred caregiver, and a window of 60 time units from the time of the event; a delay shifts one available caregiver by 15 to 45 time units; an unavailability removes a caregiver for the rest of the day, and no event removes the last available caregiver. The patient concerned by a cancellation or a no-show is drawn uniformly among the patients that have not yet been visited, and the caregiver concerned by a delay or an unavailability is drawn uniformly among the available caregivers. The probabilities and durations are design choices that give a mix of patient-side and caregiver-side events; they are not estimated from operational data. The same events are replayed for both policies. A visit is committed once its caregiver has left for it and is never changed. After each event, the visits not yet started are re-optimized with the caregivers' current locations and times as starting points (Algorithm 2). Completion is not calibrated: whether a patient is served follows from the simulated routes, the travel and service times, the availability of the caregivers, and the event times.

Policies and measures.

Each simulated day is run under two policies that face the same events: the re-optimization policy and the keep-the-plan policy of Section 5.2. For each day, we record the numbers of required, served, and unserved patients, the number of late visits, the total lateness, the total overtime, the total travel, the realized objectives, the share of patients that change caregiver, and the numbers of accepted and rejected re-optimizations. For each event, we also record the change $\Delta F = F(\text{re-optimized plan}) - F(\text{plan in force})$ of the four objectives, where both plans are evaluated on the same reduced instance: the patients P_e still to be served after the event, the available caregivers, and their starting locations and times. The plan in force is the plan before the event adapted to P_e (visits that are no longer required are removed, and a new or orphaned visit is appended to the route of the eligible available caregiver with the fewest visits). All objectives are minimized, so a negative component of ΔF is an improvement, and a smaller patient set cannot by itself reduce the travel distance in this comparison. The policies are compared by paired differences over the

runs (Wilcoxon signed-rank test, Holm correction over the instances, bootstrap interval for the median difference).

Ablation study.

To isolate the contribution of the burden objective, all 18 instances were solved by an alternative configuration in which the burden index is excluded from non-dominated sorting, crowding-distance computation, duplicate detection, the comparison used by the local search, and the decision rule, while it is still computed for every individual in every generation. It is therefore computed identically in both configurations, but it influences selection only in the four-objective configuration. Twenty independent runs were executed per configuration and instance. The configurations are compared on the executed schedule (the four objectives, the share of late visits, the total lateness, and the total overtime) and on the mean and the minimum of the burden index over the first front. For each outcome, the 18 Mann–Whitney p-values (one per instance) are corrected by the Holm procedure, the difference of medians is given with a 95% bootstrap confidence interval and Cliff's delta, and an exact sign test over the 18 instances summarizes the direction of the effect. We do not correct across outcomes, because each outcome answers a separate question. Samples are aggregated at the level of the run rather than the individual solution, which avoids pseudo-replication arising from the correlated structure of the solutions within one front. Hypervolume was not used to compare the two configurations, because they optimize different numbers of objectives and their hypervolumes are therefore not directly comparable quantities.

Sensitivity analysis.

The weights (w_1, w_2, w_3) of the duty share, the acuity share, and the late share were varied over $(0.50, 0.25, 0.25)$, $(0.25, 0.50, 0.25)$, and $(0.25, 0.25, 0.50)$. First, the burden of the logged first fronts of the four-objective runs was recomputed under each set (no new optimization), and its rank correlation with the default score and the stability of the executed schedule were computed. Second, the optimization was rerun with each set on the 12 instances with 25 and 50 patients (20 runs each); the schedules found with the default weights and with the alternative weights were both scored with the alternative weights, and the difference (the regret of the default weights) was tested. Because the index has no age or health thresholds, no threshold sensitivity is needed.

6.5. Metrics and Statistics

Normalized hypervolume.

Let $r = (r_1, r_2, r_3, r_4)$ be the reference point of an instance, which covers all four objectives: r_1 is 1.5 times the largest travel distance among 10,000 random schedules of the instance (rounded up to a multiple of 10); r_2 is the total service time of all patients, the largest possible workload range; and $r_3 = r_4 = 1$, because the preference objective and the burden index are bounded by 1. For a finite set A of objective vectors in $[0, r]$,

$$HV(A) = \frac{\mathrm{vol}\big(\{z \in [0,r] : \exists\, q \in A,\ q \le z\}\big)}{\prod_{i=1}^{4} r_i} \in [0,1]. \tag{31}$$

It is estimated by drawing 100,000 points uniformly in $[0, r]$ and taking the fraction dominated by A. The estimator is unbiased, and its standard error is at most 0.0016, which is below the run-to-run standard deviation of every instance (0.004 to 0.013). The division by the volume of the reference box makes the values comparable across instances. All algorithms of an instance share the same reference point, and the hypervolume is used only to compare algorithms that optimize the same four objectives.

Executed schedule.

For the skill-feasible first front $\mathcal{F}$, each objective is rescaled to $[0,1]$ over $\mathcal{F}$,

$$\tilde{f}_i(\omega) = \frac{f_i(\omega) - \min_{\mathcal{F}} f_i}{\max_{\mathcal{F}} f_i - \min_{\mathcal{F}} f_i}$$

(zero if the range is zero), and the schedule with the smallest Euclidean distance to the ideal point is chosen,

$$\omega^\star \in \arg\min_{\omega \in \mathcal{F}} \Big(\sum_{i=1}^{4} \tilde{f}_i(\omega)^2 \Big)^{1/2}, \tag{32}$$

with ties broken by the smaller travel distance. This compromise-programming rule gives the four objectives equal weight, and it is used for every reported executed schedule and in the dynamic simulation.

Other measures.

The share of late visits, the total lateness $\sum_i \ell_i$, the total overtime $\sum_k o_k$, the number of caregivers without visits, the size of the first front, and the number of objective evaluations.

Effect sizes and tests.

For two samples $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$, Cliff's delta is

$$\delta_{\mathrm{C}} = \frac{\#\{(i,j) : x_i > y_j\} - \#\{(i,j) : x_i < y_j\}}{n_1 n_2} \in [-1,1]. \tag{33}$$

The p-values come from the two-sided Mann–Whitney test (independent samples) or the Wilcoxon signed-rank test (paired samples). The interval of a difference of medians is a percentile bootstrap interval with 2000 resamples. The Holm procedure corrects for multiple tests over the instances within each outcome (over all instance-by-algorithm tests for the algorithm comparison), and an exact sign test over the 18 instances summarizes the direction of an effect. Samples are aggregated at the level of the run.

7. Results

7.1. Static Scheduling Performance

Table 3 gives, for each instance, the normalized hypervolume, the size of the first front, the best travel distance, the lowest workload range and preference value reached, and the range of the burden index.

The normalized hypervolume lies between 0.17 and 0.42, and its run-to-run standard deviation is at most 0.013. The skill requirements make some objective values unreachable: in eight instances, the workload range cannot go below 40 to 140 time units, and the preference objective cannot go below 0.08 to 0.40 (the share of patients not served by their preferred caregiver), because patients who need a high skill level can only be served by some caregivers. Inside the first front, the burden index spans 0.13 to 0.35.

Correlations. Table 4 reports the rank correlation of the burden index with the other objectives inside the first front, for every instance. A lower burden goes with a longer travel distance in every instance, with a more even workload (positive correlation with the workload range), and with fewer satisfied preferences, more strongly at 50 and 100 patients (negative correlation with the preference objective, whose smaller values are better). In four instances, the correlation with the workload range reaches 0.9 or more (R109-50, RC106-50, R109-100, RC106-100).

Table 3. Four-objective NSGA-II, 20 independent runs per instance (mean ± standard deviation). HV: normalized hypervolume (Section 6.5). Front size: size of the first front in the combined population of $2N = 200$ at the last generation. Best TD: smallest travel distance of the front. WLB_{min} and $Pref_{min}$: lowest workload range and preference value found in the first front over all runs. Burden: mean and range (span) of f_4 in the first front.

Instance	HV	Front Size	Best TD	WLB_{min}	$Pref_{min}$	Burden Mean	Burden Span
C101-25-2	0.379 ± 0.013	101	229.0 ± 10.0	20	0.08	0.516 ± 0.017	0.286 ± 0.042
C201-25-2	0.396 ± 0.008	101	293.0 ± 13.9	20	0.12	0.459 ± 0.028	0.252 ± 0.070
R109-25-2	0.180 ± 0.005	104	484.0 ± 16.0	60	0.28	0.630 ± 0.010	0.166 ± 0.024
R204-25-2	0.248 ± 0.004	104	462.2 ± 18.9	60	0.28	0.452 ± 0.024	0.348 ± 0.051
RC106-25-2	0.280 ± 0.007	101	419.2 ± 26.2	20	0.16	0.607 ± 0.014	0.192 ± 0.045
RC208-25-2	0.419 ± 0.008	102	392.1 ± 28.6	20	0.12	0.438 ± 0.012	0.290 ± 0.029
C101-50-3	0.289 ± 0.009	113	681.5 ± 40.6	20	0.24	0.519 ± 0.011	0.240 ± 0.021
C201-50-3	0.333 ± 0.008	111	871.6 ± 41.3	20	0.16	0.471 ± 0.014	0.198 ± 0.027
R109-50-3	0.180 ± 0.004	112	983.0 ± 34.4	20	0.38	0.571 ± 0.006	0.183 ± 0.014
R204-50-3	0.255 ± 0.006	112	984.2 ± 49.6	20	0.38	0.392 ± 0.019	0.354 ± 0.050
RC106-50-3	0.177 ± 0.004	113	1121.7 ± 63.1	140	0.38	0.588 ± 0.007	0.207 ± 0.010
RC208-50-3	0.239 ± 0.010	116	1110.3 ± 87.3	140	0.36	0.460 ± 0.015	0.268 ± 0.038
C101-100-4	0.285 ± 0.008	119	2273.4 ± 111.2	0	0.18	0.501 ± 0.005	0.145 ± 0.011
C201-100-4	0.295 ± 0.006	121	2427.5 ± 90.8	0	0.20	0.474 ± 0.009	0.164 ± 0.014
R109-100-4	0.170 ± 0.004	120	2165.3 ± 69.0	60	0.40	0.531 ± 0.003	0.148 ± 0.009
R204-100-4	0.218 ± 0.008	120	2204.4 ± 79.0	60	0.37	0.437 ± 0.018	0.334 ± 0.016
RC106-100-4	0.192 ± 0.005	115	2702.0 ± 135.3	80	0.33	0.532 ± 0.007	0.129 ± 0.013
RC208-100-4	0.233 ± 0.006	119	2901.9 ± 80.8	40	0.32	0.464 ± 0.012	0.214 ± 0.012

Table 4. Spearman rank correlation of the burden index f_4 with the other objectives inside the first front (mean ± standard deviation over 20 runs per instance).

Instance	$\rho(f_4, f_1)$ Travel	$\rho(f_4, f_2)$ Workload Range	$\rho(f_4, f_3)$ Preference
C101-25-2	−0.85 ± 0.08	+0.30 ± 0.16	−0.19 ± 0.18
C201-25-2	−0.69 ± 0.13	+0.26 ± 0.18	−0.03 ± 0.23
R109-25-2	−0.78 ± 0.11	+0.83 ± 0.07	−0.77 ± 0.12
R204-25-2	−0.93 ± 0.02	+0.85 ± 0.07	−0.79 ± 0.15
RC106-25-2	−0.59 ± 0.18	+0.61 ± 0.12	−0.33 ± 0.30
RC208-25-2	−0.73 ± 0.13	+0.60 ± 0.13	−0.42 ± 0.13
C101-50-3	−0.76 ± 0.12	+0.77 ± 0.09	−0.72 ± 0.13
C201-50-3	−0.76 ± 0.12	+0.48 ± 0.17	−0.44 ± 0.15
R109-50-3	−0.70 ± 0.15	+0.91 ± 0.03	−0.90 ± 0.05
R204-50-3	−0.81 ± 0.07	+0.76 ± 0.09	−0.76 ± 0.09
RC106-50-3	−0.80 ± 0.11	+0.93 ± 0.02	−0.87 ± 0.05
RC208-50-3	−0.84 ± 0.10	+0.84 ± 0.04	−0.80 ± 0.07
C101-100-4	−0.65 ± 0.10	+0.88 ± 0.03	−0.89 ± 0.03
C201-100-4	−0.60 ± 0.17	+0.74 ± 0.09	−0.73 ± 0.06
R109-100-4	−0.67 ± 0.19	+0.96 ± 0.01	−0.94 ± 0.02
R204-100-4	−0.75 ± 0.13	+0.89 ± 0.05	−0.87 ± 0.05
RC106-100-4	−0.68 ± 0.20	+0.94 ± 0.01	−0.95 ± 0.02
RC208-100-4	−0.42 ± 0.18	+0.87 ± 0.05	−0.89 ± 0.03

Feasibility of the executed schedules. Table 5 reports the schedule to execute (Section 6.5) of the four-objective algorithm. A visit is late if its service starts after l_i, and overtime is the time a caregiver returns to the depot after E_k. No executed schedule of any of the 18 instances leaves a caregiver without visits. In 11 of the 18 instances, between 61% and 91% of the visits start after the window has closed, and overtime is absent in only three instances. These values describe the schedules found by a search that does not enforce

windows or shifts. They do not show that no better schedule exists, except on the six R109 and RC106 instances, where the shift limit cannot be met by any schedule (Section 6). All conclusions below on lateness and overtime refer to the schedules found by this search.

Table 5. Executed schedule of the four-objective algorithm (mean ± standard deviation over 20 runs). Total lateness and total overtime are in time units.

Instance	Late Visits (%)	Total Lateness	Total Overtime
C101-25-2	41.6 ± 6.1	4429 ± 1224	0 ± 0
C201-25-2	23.0 ± 7.1	8870 ± 2280	0 ± 0
R109-25-2	74.0 ± 5.4	4369 ± 441	681 ± 50
R204-25-2	13.2 ± 5.4	943 ± 438	23 ± 44
RC106-25-2	66.8 ± 5.2	3225 ± 358	569 ± 26
RC208-25-2	6.2 ± 5.1	141 ± 143	0 ± 0
C101-50-3	74.3 ± 6.4	22648 ± 3641	384 ± 159
C201-50-3	62.7 ± 7.6	42474 ± 6882	12 ± 35
R109-50-3	83.9 ± 3.1	14062 ± 1001	1634 ± 98
R204-50-3	23.3 ± 7.8	3084 ± 1092	363 ± 174
RC106-50-3	85.7 ± 2.6	15859 ± 866	1864 ± 100
RC208-50-3	41.8 ± 9.3	4893 ± 1577	385 ± 139
C101-100-4	86.2 ± 2.0	67698 ± 7218	2173 ± 301
C201-100-4	79.5 ± 6.4	128316 ± 19482	1547 ± 418
R109-100-4	91.0 ± 2.1	50404 ± 2973	3835 ± 156
R204-100-4	47.6 ± 6.1	18235 ± 3998	1855 ± 270
RC106-100-4	90.5 ± 1.6	58967 ± 2908	4418 ± 151
RC208-100-4	60.6 ± 4.4	29458 ± 3847	2360 ± 244

7.2. Ablation Study: Contribution of the Burden Objective

The ablation compares the four-objective configuration with the configuration in which the burden index is computed but ignored by the search (Section 6), on the executed schedule, for all 18 instances. Table 6 reports the burden index of each configuration, the difference with its 95% confidence interval, and the effect size.

Optimizing the burden index lowers it in all 18 instances. Cliff's delta is between −0.82 and −1.00, no confidence interval contains zero, every difference is significant after Holm correction, and an exact sign test over the 18 instances gives $p < 0.0001$. The absolute differences lie between 0.036 and 0.190, which corresponds to a median relative reduction of 25% at 25 patients, 15% at 50 patients, and 9% at 100 patients, so that the effect shrinks with the size of the instance. Measured against the gap between the three-objective index and the floor $2/(3m)$ of Section 4, the four-objective configuration closes a median of 59%, 25%, and 12% of the gap at 25, 50, and 100 patients (ranges 25% to 69%, 15% to 48%, and 11% to 25%, computed from the medians of Table 6), so that the decrease with size is not an artifact of the floor, which is a bound and not an attainable value. The mean and the minimum of the index over the first front are also lower in all 18 instances (the mean significantly in 17). These are changes in an operational measure of the load placed on the most burdened caregiver, and not demonstrated improvements in caregiver well-being. They come at a cost in travel distance and in patient preference, which is reported below. Table 7 reports the other objectives for both configurations, so that the price of the lower burden can be read directly, and Table 8 summarizes all outcomes by problem size.

Table 6. Burden index f_4 of the executed schedule (median over 20 independent runs) in the four-objective and in the three-objective configuration, difference of medians with its 95% bootstrap confidence interval, and Cliff's delta δ_C (negative values mean a lower burden with the four-objective configuration). All differences are significant after Holm correction of the Mann–Whitney tests over the 18 instances.

Instance	4 Objectives	3 Objectives	Difference	95% CI	δ_C
25-patient instances					
C101-25-2	0.487	0.669	−0.182	[−0.196, −0.151]	−1.00
C201-25-2	0.420	0.592	−0.172	[−0.193, −0.141]	−1.00
R109-25-2	0.622	0.718	−0.096	[−0.118, −0.080]	−1.00
R204-25-2	0.444	0.634	−0.190	[−0.213, −0.091]	−0.95
RC106-25-2	0.566	0.657	−0.091	[−0.108, −0.060]	−0.97
RC208-25-2	0.385	0.498	−0.113	[−0.141, −0.097]	−0.99
50-patient instances					
C101-50-3	0.498	0.585	−0.086	[−0.097, −0.071]	−0.98
C201-50-3	0.447	0.530	−0.083	[−0.105, −0.067]	−0.99
R109-50-3	0.552	0.612	−0.060	[−0.082, −0.049]	−0.93
R204-50-3	0.345	0.459	−0.114	[−0.147, −0.101]	−0.99
RC106-50-3	0.566	0.629	−0.062	[−0.076, −0.050]	−0.98
RC208-50-3	0.438	0.536	−0.098	[−0.139, −0.086]	−1.00
100-patient instances					
C101-100-4	0.483	0.523	−0.040	[−0.048, −0.030]	−0.99
C201-100-4	0.453	0.489	−0.036	[−0.052, −0.023]	−0.82
R109-100-4	0.509	0.561	−0.051	[−0.057, −0.045]	−1.00
R204-100-4	0.393	0.468	−0.075	[−0.106, −0.050]	−1.00
RC106-100-4	0.514	0.558	−0.045	[−0.056, −0.038]	−1.00
RC208-100-4	0.427	0.497	−0.070	[−0.086, −0.062]	−0.97

Table 7. Executed schedule of the four-objective and of the three-objective configuration (median over 20 independent runs, shown as 4 objectives/3 objectives) for travel distance f_1, workload range f_2, and preference objective f_3, with the difference of medians (4 objectives minus 3 objectives). A dagger marks a difference that is significant after Holm correction of the Mann–Whitney tests over the 18 instances.

Instance	f_1 (4/3)	Δf_1	f_2 (4/3)	Δf_2	f_3 (4/3)	Δf_3
25-patient instances						
C101-25-2	284.3/223.5	+60.8 †	60/60	0	0.160/0.160	0.000
C201-25-2	345.4/303.4	+41.9 †	20/20	0	0.160/0.200	−0.040 †
R109-25-2	548.5/468.9	+79.6 †	100/140	−40 †	0.360/0.340	+0.020 †
R204-25-2	540.8/455.2	+85.6 †	180/220	−40 †	0.400/0.380	+0.020
RC106-25-2	501.7/430.9	+70.8 †	60/100	−40 †	0.280/0.240	+0.040
RC208-25-2	497.4/393.9	+103.5 †	100/100	0	0.240/0.260	−0.020
50-patient instances						
C101-50-3	929.2/629.9	+299.3 †	80/110	−30 †	0.360/0.340	+0.020
C201-50-3	1216.9/787.2	+429.7 †	60/70	−10	0.240/0.240	0.000
R109-50-3	1189.1/924.0	+265.0 †	180/220	−40 †	0.540/0.500	+0.040 †
R204-50-3	1199.6/910.0	+289.6 †	240/320	−80†	0.520/0.500	+0.020 †
RC106-50-3	1481.3/1088.1	+393.1 †	200/280	−80 †	0.540/0.480	+0.060 †
RC208-50-3	1471.5/1060.6	+411.0 †	230/330	−100 †	0.520/0.460	+0.060 †
100-patient instances						
C101-100-4	2761.7/1889.2	+872.5 †	150/260	−110 †	0.350/0.290	+0.060 †
C201-100-4	2831.8/2017.9	+813.9 †	120/170	−50 †	0.335/0.280	+0.055 †
R109-100-4	2557.2/1897.3	+659.9 †	330/460	−130 †	0.550/0.490	+0.060 †
R204-100-4	2549.1/1937.9	+611.2 †	410/530	−120 †	0.525/0.470	+0.055 †
RC106-100-4	3220.6/2346.0	+874.7 †	330/440	−110 †	0.500/0.430	+0.070 †
RC208-100-4	3440.4/2373.4	+1067.0 †	300/400	−100 †	0.460/0.400	+0.060 †

Table 8. Four-objective minus three-objective configuration, executed schedule. The entries are medians over the six instances of each size of the difference of medians (20 runs per configuration). The last two columns give the number of instances (of 18) in which the four-objective configuration is lower or higher, and the exact sign-test p-value over the 18 instances. The row "share of the gap to the floor closed" gives, for each size, the median over its six instances of the difference between the three-objective and the four-objective index divided by the difference between the three-objective index and the floor $2/(3m)$.

Outcome	25 Patients	50 Patients	100 Patients	Lower/Higher	Sign Test p
Burden index f_4 (relative change)	−25%	−15%	−9%	18/0	<0.0001
Burden index f_4 (share of the gap to the floor closed)	59%	25%	12%		
Travel distance f_1 (relative change)	+18%	+37%	+39%	0/18	<0.0001
Workload range f_2 (time units)	−20	−60	−110	15/0 (3 equal)	0.0001
Preference objective f_3 (points)	+1.0	+3.0	+6.0	2/14 (2 equal)	0.004
Share of late visits (points)	−25	−8.5	−2.5	17/0 (1 equal)	<0.0001
Total lateness (relative change)	−55%	−13%	−4%	14/4	0.03
Total overtime (time units)	0	−17	+554	5/11	0.21

The price of the lower burden is travel distance, which is higher in all 18 instances (significantly in all), and, from 50 patients on, the preference objective (significantly higher in 11 instances and significantly lower in one). The workload range is lower in 15 instances (significantly in 14). The share of late visits falls in 17 instances (significantly in 14), but the effect on total lateness is mixed (significantly lower in eight instances and significantly higher in two, RC106-50 and R109-100), and overtime is higher in 11 instances (significantly in seven, all with 50 or more patients) and lower in five (significantly in three). The index counts late visits and not the minutes by which they are late; it therefore reduces how many visits are late, while tolerating longer delays and overtime when the day is overloaded.

7.3. Sensitivity to the Weights of the Burden Index

Recomputing the burden of the logged first fronts with the weights $(0.50, 0.25, 0.25)$, $(0.25, 0.50, 0.25)$, and $(0.25, 0.25, 0.50)$ gives a rank correlation with the default score between 0.89 and 0.996 over the solutions of a front; the executed schedule is unchanged in 30% to 90% of the runs, depending on the instance and the weights. The recomputation reproduces the logged index to within 8×10^{-7}. Rerunning the optimization with the alternative weights on the 12 instances with 25 and 50 patients (20 runs each), and scoring both groups of schedules with the alternative weights, the schedules found with the default weights are on average worse by 0.0034 (median regret; largest 0.056, R204-50 with weights 0.25/0.25/0.50). The difference is significant after Holm correction in one of 36 comparisons for the index and in one for travel distance (R204-25, weights 0.25/0.50/0.25, travel 6% lower). The weighting that stresses late visits matters most: the regret is positive in nine of the 12 instances, with median 0.011. The conclusions are therefore insensitive to the weights within the tested range, with the exception of a heavier weight on late visits.

7.4. Comparison with Other Algorithms

All variants optimize the same four objectives with the same evaluation, constraint handling, reference point, and decision rule, and use between 155,694 and 162,966 objective evaluations per run (a spread of 4.7%). Table 9 summarizes the comparison of the normalized hypervolume.

The proposed algorithm has the best average rank, and it is never significantly worse than NSGA-III or than the variant without local search. It is significantly better than NSGA-III in four instances (R109-50, R204-50, R109-100, RC106-100). Against MOEA/D, it wins in one instance (C201-25) and loses in one (RC208-100). None of the sign tests across the instances is significant, so the data do not show that any variant is better than the others:

they show that the proposed algorithm is competitive and that the local search brings little, at the cost of about 3% more evaluations.

Table 9. Comparison of the algorithms on the normalized hypervolume over the 18 instances (20 runs per instance and algorithm). Average rank (1 = best, by median); mean number of objective evaluations per run; mean size of the first front (taken from the combined population of 200 individuals for the NSGA variants, and from the population of 120 individuals of MOEA/D); number of instances in which the proposed algorithm is significantly better, not different, or significantly worse (Holm correction over all 54 tests); number of instances in which its median is higher, and exact sign-test p-value over the 18 medians.

Algorithm	Avg. Rank	Evaluations	Front Size	Better/Same/Worse	Higher in	Sign Test p
NSGA-II (proposed)	2.06	160,178	111.4			
MOEA/D	2.44	161,129	90.5	1/16/1	12 of 18	0.24
NSGA-III	2.67	162,966	109.9	4/14/0	11 of 18	0.48
NSGA-II without local search	2.83	155,694	111.1	1/17/0	12 of 18	0.24

7.5. Dynamic Disruption-Handling Performance

Each of the 18 instances was simulated for 20 days with three events per day (1080 events: 308 cancellations, 168 no-shows, 229 emergencies, 224 delays, and 151 caregiver departures); the re-optimization policy and the keep-the-plan policy faced the same events (Section 6). Every re-optimized plan passed the acceptance check of Algorithm 2. Table 10 compares the two policies over the simulated day. The plans of the initial day and of every re-optimization are selected by the compromise rule of Section 5.2, which gives the four objectives equal weight; the comparison of Table 10 depends on this choice, and another rule would select other plans.

Table 10. Re-optimization minus keep-the-plan over the simulated day. For each instance, the median of the 20 paired differences is computed (same events for both policies); the entries are medians of these values over the six instances of each size. Negative values mean that re-optimization is lower. The column "Lower/higher" counts the instances (of 18) in which re-optimization is lower or higher. Distances and times are in the units of the Solomon instances.

Outcome	25 Patients	50 Patients	100 Patients	Lower/Higher
Total travel	−34	−137	−306	18/0
Total lateness	−271	−769	−867	18/0
Total overtime	−13	−112	−262	14/0
Late visits (number)	−0.5	−2.0	−1.8	12/0
Preference objective	0.00	−0.035	−0.040	14/1
Workload range	0	+40	+70	0/14
Realized burden index	−0.005	+0.005	+0.005	8/10
Unserved patients	0	0	0	no difference
Patients that change caregiver (%): re-optimization vs. keep the plan	9.9 vs. 3.6	9.0 vs. 3.0	11.3 vs. 3.1	

Re-optimization lowers travel in all 18 instances (significant after Holm correction in all) and total lateness in all 18 (significant in seven). The number of late visits is lower in 12 instances (significantly in eight), equal in six, and never higher. Overtime is never higher: it is lower by the median difference in 14 instances and equal in four, and it is significantly lower after Holm correction in 15 instances (in the two significant instances with equal medians, re-optimization lowers overtime on 40% and 45% of the days and never raises it). The preference objective is lower in 14 instances (significantly in nine), higher in one, and equal in three. Re-optimization worsens the realized workload range in 14 instances (significantly in 11), by 10 to 110 time units (per-instance medians). The increase is concentrated at 50 and 100 patients: the per-size medians in Table 10 are 0, 40, and 70 time units at 25, 50, and 100 patients, and 10 of the 11 significant instances have 50 or

100 patients. The re-optimization has no systematic effect on the realized burden index (higher in 10 instances and lower in eight; significant after Holm correction in three, two lower and one higher; sign test $p = 0.81$). The number of unserved patients is the same for both policies: these are patients whom no remaining caregiver can serve after a departure. Re-optimization moves 9% to 11% of the patients to another caregiver, against 3% to 4% for the keep-the-plan policy, where the moves are forced by departures. Table 11 gives the effect of one re-optimization for each event type.

Table 11. Effect of one re-optimization per event type: re-optimized plan minus the plan in force, on the same set of required visits and from the same caregiver positions and times, pooled over all 18 instances and 20 runs (mean). Negative values are improvements.

Event	Events	Δ Travel	Δ Workload Range	Δ Preference	Δ Burden	Patients Moved (%)
Cancellation	308	−34.2	+7.7	−0.016	−0.012	8
Delay	224	−32.5	+2.0	−0.012	−0.008	7
Emergency	229	−50.1	+9.7	−0.016	−0.009	7
No-show	168	−40.6	+4.4	−0.024	−0.010	8
Caregiver unavailable	151	−143.1	+7.0	−0.015	−0.021	30

At the level of a single event, the re-optimized plan improves travel, preference, and the burden index and slightly worsens the workload range; all 1080 re-optimized plans were accepted. Over the whole day, the improvement in the burden index vanishes while the workload range worsens; a possible explanation, which we did not test, is that locally good re-plans accumulate an imbalance that nothing corrects later in the day.

7.6. Realistic Staffing

To see what the index does when the day is closer to feasible, the six 50-patient instances were solved with seven caregivers instead of three (R109-50-7, C101-50-7, C201-50-7, R204-50-7, RC106-50-7, and RC208-50-7), with the protocol of Section 6: 20 runs per configuration, the four-objective and the three-objective configurations, and the same parameters. The patients are those of the three-caregiver instances; the four added caregivers (skill levels 1, 2, 3, and 3) start at the same depot with the same shift, and the preferred caregiver of each patient was redrawn uniformly among the seven caregivers. The necessary condition of Section 6 does not rule out a schedule that respects the shifts on any of these instances (the ratio is at most 0.88, for R109-50). Table 12 reports the executed schedules.

The index is lowered in all six instances (exact sign test $p = 0.031$, the smallest value possible with six instances), by a median of 25% (from 11% to 38%), against 15% on the same instances with three caregivers. Measured against the gap between the three-objective index and the floor $2/(3m) = 0.095$ of Section 4, the four-objective configuration closes a median of 36% of the gap (from 14% to 52%), against 25% with three caregivers. The costs are of the same order as with three caregivers: the travel distance is higher in all six instances (from 17% to 44%, median 31%), and the preference objective is worse in all six (from 3 to 10 points, median 5, significantly in five). The gain in workload balance observed with three caregivers largely disappears: the workload range is lower in four instances, but significantly in only two (C201-50-7 and RC106-50-7).

Staffing reduces late visits and overtime but does not remove them. With the four-objective configuration, the mean shares of late visits over the 20 runs are 48%, 38%, 64%, 3%, 67%, and 12% in C101-50-7, C201-50-7, R109-50-7, R204-50-7, RC106-50-7, and RC208-50-7, against 74%, 63%, 84%, 23%, 86%, and 42% with three caregivers (Table 5); the mean overtime is 0, 0, 1112, 13, 1426, and 1 time units, against 384, 12, 1634, 363, 1864, and 385. Late visits thus remain below 15% only in R204-50-7 and RC208-50-7, and overtime is absent

or negligible in four instances and remains above 1100 time units in R109-50-7 and RC106-50-7, the two instances for which the shift condition is tightest. As with three caregivers, optimizing the index lowers the share of late visits (in all six instances, by 3 to 22 points, significantly) and the total lateness (in all six, significantly in three), but not the overtime, which is significantly higher in R109-50-7 and RC106-50-7 (by 141 and 347 time units) and not significantly different in the other four instances (zero in three of them). The logs of these runs keep the burden components of the first four caregivers only; over the solutions of the first fronts, the other three caregivers carry between 44% and 52% of the duty time, so that the added caregivers do work, but we do not report the share of schedules with an idle caregiver for these instances.

Table 12. Six 50-patient instances with seven caregivers. Executed schedule of the four-objective and of the three-objective configuration (median over 20 runs), shown as 4 objectives/3 objectives for the burden index f_4, the share of late visits (%), and the total overtime; differences of medians (4 objectives minus 3 objectives) for the burden index (with Cliff's delta in parentheses), the travel distance (relative change), the preference objective (points), and the workload range. A dagger marks a difference that is significant after Holm correction of the Mann–Whitney tests over the six instances.

Instance	f_4 (4/3)	Δf_4 (δ_C)	Δf_1	Δf_3	Δf_2	Late Visits (4/3)	Overtime (4/3)
C101-50-7	0.297/0.398	−0.101 † (−0.99)	+37% †	+3.0	+20	45/64 †	0/0
C201-50-7	0.259/0.421	−0.162 † (−0.99)	+44% †	+10.0 †	−30 †	38/60 †	0/0
R109-50-7	0.352/0.395	−0.042 † (−0.90)	+24% †	+4.0 †	−20	65/71 †	1125 †/984
R204-50-7	0.140/0.188	−0.048 † (−0.76)	+17% †	+6.0 †	+10	3/6 †	0/12
RC106-50-7	0.359/0.425	−0.065 † (−0.95)	+34% †	+6.0 †	−30 †	66/70 †	1397 †/1050
RC208-50-7	0.192/0.255	−0.064 † (−0.79)	+28% †	+4.0 †	−10	12/16 †	0/0

8. Discussion

8.1. The Burden–Efficiency–Preference Trade-Off

Optimizing the burden index lowers it in every instance, but the gain is not free. The executed schedule needs between 14% and 55% more travel (median 33%), and from 50 patients on, it satisfies fewer patient preferences (three to six percentage points). The relative gain shrinks with the size of the instance (−25%, −15%, and −9% at 25, 50, and 100 patients), while the cost in travel does not (+18%, +37%, and +39%). The dependence of the effect on instance size is a property of this index and of these instances, not a general statement about problem scale: an index with a different aggregation, for example a mean over the caregivers, may show a different dependence, and the index used here has a lower bound, $2/(3m)$, that falls as the number of caregivers m grows (Section 4). On these instances, the trade-off is most favorable for small and medium teams. The Pareto front gives the decision maker the whole range; Table 3 shows how wide it is (burden spans of 0.13 to 0.35).

The index combines three things: duty time, the clinical load of the assigned patients, and punctuality. Because duty time is part of it, a lower burden goes with a more even workload (rank correlation up to 0.96 inside a front, Table 4; the workload range is lower in 15 of 18 instances after optimization), which is a benefit but also an overlap. The two objectives are not redundant, since the four-objective configuration lowers the index by a further 7% to 30% relative to the three-objective configuration, which already balances the workload. The index counts late visits and not the minutes by which they are late; this is why the share of late visits falls in 17 instances, while total lateness falls only slightly at the larger sizes and overtime is not reduced and is even higher in several large instances. A variant that penalizes the minutes of lateness is a natural extension. Waiting time is not a component of the index, because a short pause between visits may help a caregiver recover while a day without pauses may increase stress; overtime, rest periods, and the number of consecutive visits are further factors that a richer index could include.

8.2. Late Visits, Overtime, and Staffing

Table 5 shows that most schedules of the 50- and 100-patient instances are late: 61% to 91% of the visits in 11 instances, and overtime of 1500 to 4400 time units at 100 patients. Two causes must be distinguished. On R109 and RC106, the shift limit cannot be met by any schedule with these numbers of caregivers (Section 6); at 100 patients, at least 12 and 11 caregivers would be needed, so that adding a few caregivers would not remove the overtime. On the other twelve instances, the necessary condition does not rule out a feasible schedule, and the late visits reflect that the search does not enforce the windows; whether the windows can be met there with two, three, or four caregivers is not known. Section 7.6 shows that seven caregivers on the 50-patient instances remove or nearly remove the overtime on four of six instances and bring the late visits of two instances to 3% and 12%, but not those of the others. The conclusions on lateness and overtime must be read in that light. The instances without overtime (C101-25, C201-25, and RC208-25) are those where the effect of the index is among the largest (-23% to -29%). The skill requirements also bind: they put floors under the workload range and the preference objective (Table 3), so that a perfectly even workload or fully satisfied preferences are not attainable in several instances.

8.3. The Optimizer and the Dynamic Policy

NSGA-II with local search is competitive with NSGA-III and MOEA/D at an equal budget and has the best average rank, but the differences between the variants are small and not significant across instances (Table 9), and the local search adds little. The contribution of this study therefore lies in the model, in the analysis of what the burden objective does, and in the dynamic framework, which do not depend on the optimizer. The hypervolume was used only to compare algorithms that optimize the same four objectives.

In the dynamic simulation, re-optimizing after each event is better than keeping the plan for travel, lateness, overtime, and preferences, but it makes the workload less even and does not reduce the realized burden over the day (Table 10). Each re-optimization uses the same compromise rule as the first plan and does not look ahead to the remaining events. It also moves 9% to 11% of the patients to another caregiver, against 3% to 4% for the keep-the-plan policy; the re-optimization does not penalize such changes, so that the stability of the schedule is not guaranteed, and a stability-aware variant, with a penalty on the number of changed visits or an explicit target for the workload range, is left for future work. The simulation is a controlled stress test: it applies three events per day, of five types, with probabilities and durations fixed by design and not estimated from operational data, so that it shows how the two policies behave under a defined disruption load and not how often such events occur in a real agency.

8.4. Practical Implications

The results do not support recommending the index for deployment. They show a method to quantify a trade-off: for a given team and day, the front shows how much travel and how many preferences must be given up for a given reduction of the burden of the most loaded caregiver. All experiments use synthetic data derived from a benchmark, so that the figures describe the behavior of the model and not that of a real agency. A provider should inspect several points of the front, check the lateness and overtime of the chosen schedule, and consider adding capacity before optimizing the burden on an overloaded day.

8.5. Limitations

Several limitations should be considered when interpreting the results. First, the burden index (Equation (14)) is an operational proxy built from duty time, patient acuity, and late visits, with equal weights fixed by assumption; it is not a clinically validated psy-

chometric instrument and has not been validated against caregiver-reported data, so that it should not be read as a measure of caregiver burnout in the occupational-health sense. The results are insensitive to the weights within the tested range, with the exception of a heavier weight on late visits. Second, all instances are synthetic adaptations of the Solomon benchmark, with a fixed service time of 20 time units and generated skill requirements, health attributes, and preferences, and the schedules of the larger instances are mostly late and in overtime, so that the findings describe that regime. Third, the algorithms were run for a fixed number of generations, and we did not test whether longer runs would improve the fronts. Fourth, the executed schedule is chosen by one rule; other rules would give other points of the front. Fifth, the dynamic study uses three events per day with fixed probabilities and does not model uncertain travel or service times; computation times were measured only in one dedicated run per instance size on an idle machine (Section 6), and not in the experiments, which shared the machine, so that the study gives the order of magnitude of the computation time and does not support a claim about real-time responsiveness in a given deployment. Sixth, the statistics rest on 20 runs per setting, and the corrections for multiple testing are conservative.

A further limitation concerns possible bias if the method were used in practice. The acuity component of the burden index uses the age, blood pressure, and weight of the patients as a proxy for the complexity of care. A scheduling rule of this kind could disadvantage older or sicker patients, for example by influencing which caregiver serves them, and it could encode assumptions about which patients are burdensome without evidence that these attributes predict the effort that a visit requires. The index uses no age or health cutoffs, which would have no empirical basis, but the choice of the three attributes and of their equal weighting remains an assumption. The construction heuristic that builds the initial population also uses age and blood-pressure thresholds to decide how strictly a preferred caregiver is respected (Section 5.3); this influences only the starting points of the search and not the objectives, but the same remark applies. Because all patient attributes in our experiments are synthetic, the study cannot show whether such a rule would disadvantage older or sicker patients in practice. In a deployment, the attributes should be replaced by validated clinical acuity measures defined with clinicians, the assignments should be audited for systematic differences in how older or sicker patients are treated, and the use of such a measure in scheduling should be reviewed for fairness before it affects care.

9. Conclusions and Future Work

We studied the four-objective HHCSRP in which the load of the most burdened caregiver, measured by an operational index of duty time, patient acuity, and late visits, is optimized together with travel distance, workload balance, and patient preference, with skill requirements as hard constraints and soft time windows. The problem was solved with NSGA-II and a local search, and the algorithm was compared with NSGA-III and MOEA/D under equal budgets.

On 18 adapted Solomon instances with 20 runs each, optimizing the index lowered it in every instance (by 25%, 15%, and 9% at 25, 50, and 100 patients), improved the workload balance and the share of late visits, and cost 18% to 39% more travel and, from 50 patients on, fewer satisfied preferences; it did not reduce overtime. The results were insensitive to the weights of the index, and the optimizer was competitive with the alternatives rather than better. In a route-driven simulation with 1080 disruptions, re-optimization reduced travel, lateness, and overtime compared with keeping the plan, at the cost of a less even workload, without lowering the realized burden, and with 9% to 11% of the patients moved to another caregiver (3% to 4% when keeping the plan). The search does not enforce the time windows or the shift ends: 61% to 91% of the visits were late in 11 instances, and the shift limit cannot

be met in six of them. With seven instead of three caregivers on six 50-patient instances, the index was still lowered in all of them (median 25%), at a travel cost of 17% to 44% and a preference cost of 3 to 10 points; the added staff reduced late visits and overtime but did not remove them, and optimizing the index still did not reduce overtime.

Future work should enforce the time windows and the shift ends as hard constraints with staffing levels that make them feasible, validate the index against caregiver-reported data, add the minutes of lateness, overtime, and rest periods to the burden, constrain the re-optimization to protect the workload balance and to limit the number of changed visits, extend the dynamic study to uncertain travel and service times and to more events, and test the approach on real operational data.

Funding: The author extends her appreciation to the Deanship of Scientific Research at Northern Border University, Arar, Saudi Arabia, for funding this research work through the project number "NBU-FPEJ-2026-1714-01".

Data Availability Statement: The benchmark instances (patient, location, caregiver, and service files for all 24 instances) supporting the conclusions of this article will be made available by the authors on request. The instances are adapted from the Solomon VRPTW benchmark [9]. The health attributes, skill requirements, and preferences were generated synthetically as described in Section 6.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Um-in, N.; Tharmmaphornphilas, W. A Comprehensive Review of Home Health Care Routing and Scheduling Optimization. *Eng. J.* **2025**, *29*, 39–64. [CrossRef]
2. Atta, S.; Basto-Fernandes, V.; Emmerich, M. A Concise Review of the Home Health Care Routing and Scheduling Problem. *Oper. Res. Perspect.* **2025**, *15*, 100347. [CrossRef]
3. Li, W.; Du, G. A Study of Multi-Objective Home Healthcare Routing Problem Considering Interpersonal Relationships and Complex Uncertainty Scenarios. In *Annals of Operations Research*; Springer: Berlin/Heidelberg, Germany, 2026; pp. 1–38.
4. Liu, Y.; Li, Y.; Tian, J.; Liu, H.; Zhang, X.; Liu, H. Weight-Aware Deep Reinforcement Learning for Multi-Depot Home Healthcare Routing Optimization. In *Proceedings of the 2025 International Conference on New Trends in Computational Intelligence (NTCI)*; IEEE: New York, NY, USA, 2025; pp. 172–176.
5. Ge, X.; Ye, H.; Yin, Q.; Li, M. Bi-Objective Home Health Care Routing and Scheduling Problem Based on Caregiver-Patient Matching and Satisfaction. *Expert Syst. Appl.* **2026**, *308*, 131138. [CrossRef]
6. Zaghdoud, R.; Zidi, I.; Rhaiem, O.B.; El Khediri, S.; Mesghouni, K. A Hybrid NSGA-II and Delayed Local Search Approach for Home Healthcare Scheduling and Routing Optimization. *Appl. Soft Comput.* **2026**, *191*, 114617. [CrossRef]
7. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
8. Zaghdoud, R. A Health-Aware Multi-Objective Approach to Home Healthcare Routing and Scheduling. In Proceedings of the ICAART (3), Marbella, Spain, 5–7 March 2026; pp. 2875–2882.
9. Solomon, M.M. Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints. *Oper. Res.* **1987**, *35*, 254–265. [CrossRef]
10. Zaghdoud, R.; Rhaiem, O.B.; Amara, M.; Mesghouni, K.; Galet, S. A Hybrid Genetic Algorithm Approach Based on Patient Classification to Optimize Home Health Care Scheduling and Routing. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 15099–15105. [CrossRef]
11. Amini-Rarani, M.; Nikbakhsh, E.; Mahnam, M.; Zegordi, S.H. A Column Generation-Based Heuristic Algorithm for Home Healthcare Scheduling and Routing with Synchronization and Workload Balance. *Eur. J. Oper. Res.* 2026, *in press*. [CrossRef]
12. Ceschia, S.; Da Ros, F.; Gaspero, L.D.; Mancini, S.; Maniezzo, V.; Montemanni, R.; Rosati, R.M.; Schaerf, A. A Unified Formulation for Home Healthcare Routing and Scheduling Problems. *Int. Trans. Oper. Res.* **2025**, *early view*. [CrossRef]
13. Amri, E.B.; Marzouki, B.; Driss, O.B. A Survey of Home Healthcare Optimization Problem. In *Proceedings of the 2025 IEEE Afro-Mediterranean Conference on Artificial Intelligence (AMCAI)*; IEEE: New York, NY, USA, 2025; pp. 1–6.
14. Li, W.; Du, G.; Yue, X. Multi-Objective Home Healthcare Scheduling with Online–Onsite Collaboration. *Comput. Ind. Eng.* **2026**, *222*, 112369. [CrossRef]
15. Chaieb, M.; Bouhachem, M.A.; Rouin, A.; Jaghdam, I.H. Multi-Objective Optimization for Home Health Care Scheduling under Uncertainty Using Fuzzy Logic and Non-Dominated Sorting Genetic Algorithm II. *Neural Comput. Appl.* **2026**, *38*, 164. [CrossRef]

16. Algendi, A.; Urrutia, S.; Hvattum, L.M.; Helgheim, B.I. Home Healthcare Staffing, Routing, and Scheduling Problem with Multiple Shifts and Emergency Considerations. *Networks* **2025**, *85*, 223–242. [CrossRef]
17. Li, B.; Pan, J.; Wei, Q. An Adaptive Large Neighborhood Search Algorithm with Customized Search Operators for Solving Home Healthcare Scheduling and Routing Problem with Skill Matching and Periodic Visits. *Appl. Soft Comput.* **2026**, *197*, 115152. [CrossRef]
18. Abdelwanis, M.; Ouda, E.; Absi, D.A.; Rahmadani, F.; Simsekler, M.C.E.; Sleptchenko, A. Variable Neighborhood Search for Healthcare Providers Rostering: A Comprehensive Review and Future Research Agenda. *J. Heuristics* **2026**, *32*, 2. [CrossRef]
19. Burke, E.K.; De Causmaecker, P.; Vanden Berghe, G.; Van Landeghem, H. The State of the Art of Nurse Rostering. *J. Sched.* **2004**, *7*, 441–499. [CrossRef]
20. Dall'Ora, C.; Ball, J.; Reinius, M.; Griffiths, P. Burnout in Nursing: A Theoretical Review. *Hum. Resour. Health* **2020**, *18*, 41. [CrossRef] [PubMed]
21. Shahnejat-Bushehri, S.; Tavakkoli-Moghaddam, R.; Boronoos, M.; Ghasemkhani, A. A Robust Home Health Care Routing-Scheduling Problem with Temporal Dependencies under Uncertainty. *Expert Syst. Appl.* **2021**, *182*, 115209. [CrossRef]
22. Ta-Dinh, Q.; Pham, T.-S.; Hà, M.H.; Rousseau, L.-M. A Reinforcement Learning Approach for the Online Dynamic Home Health Care Scheduling Problem. *Health Care Manag. Sci.* **2024**, *27*, 650–664. [CrossRef] [PubMed]
23. Parçaoğlu, M.; Salman, F.S.; Araz, O.M. Multi-Objective Dynamic Prioritized Routing and Scheduling for Home Healthcare Services with Cooperating Service Providers. *Health Care Manag. Sci.* **2025**, *28*, 759–786. [CrossRef] [PubMed]
24. Khorasanian, D.; Patrick, J.; Sauré, A. Dynamic Home Care Routing and Scheduling with Uncertain Number of Visits per Referral. *Transp. Sci.* **2024**, *58*, 841–859. [CrossRef]
25. Mohammadi, S.B.; Radman, M. Dynamic Home Healthcare Planning Using Decomposition and Metaheuristic Approaches in Multi-Service Networks. *Ann. Oper. Res.* **2026**, *358*, 1015–1051. [CrossRef]
26. Pemberthy-R, J.I.; Rivera, J.C. A Matheuristic Framework for Shared-Vehicle Routing in Home Healthcare Delivery. *Healthc. Anal.* **2026**, *10*, 100479. [CrossRef]
27. Fkiri, A.; Rjaibi, N.; Sakhri, M.S.A. A Hybrid Metaheuristic Approach to the VRPTW in Home Healthcare Logistics. In *Proceedings of the 2026 IEEE International Conference on Complex Systems and Intelligent Computing (ICCSIC)*; IEEE: New York, NY, USA, 2026; pp. 188–195.
28. Amri Sakhri, M.S. Efficient Home Health Care Vehicle Routing with Time Windows via Hybrid Clustering and Tabu Search. *Oper. Res. Forum* **2026**, *7*, 85. [CrossRef]
29. Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601. [CrossRef]
30. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.