## Contents

1 **Residential Care** 3  
1.1 Overview of Home Care Delivery 3  
1.1.1 Home Care 4  
1.1.2 Home Health Care 4  
1.1.3 Operational Challenges 6  
1.2 An Overview of Optimization Technology 8  
1.2.1 Linear Programming 9  
1.2.2 Mixed Integer Programming 10  
1.2.3 Constraint Programming 10  
1.2.4 Heuristics and Dedicated Methods 11  
1.2.5 Technology Comparison 11  
1.3 Territory Districting 13  
1.4 Provider-to-Patient Assignment 15  
1.4.1 Workload Measures 15  
1.4.2 Workload Balance 16  
1.4.3 Assignment Models 17  
1.4.4 Assignment of New Patients 18  
1.5 Task Scheduling and Routing 18  
1.6 Perspectives 21  
1.6.1 Integrated Decision Making Under a New Business Model 21  
1.6.2 Home Telemetering Forecasting Adverse Events 22  
1.6.3 Forecasting the Wound Healing Process 23  
1.6.4 Adjustment of Capacity and Demand 24
1 Residential Care

Nadia Lahrichi, Louis-Martin Rousseau, and Willem-Jan van Hoeve

In this chapter we provide a perspective on the use of analytics in the context of home care delivery. In particular, we will concentrate on operational questions arising from nurse-to-patient assignments and employee scheduling and routing considerations. While these questions are highly relevant at the operational level, also tactical and strategic decision-makers can benefit from quantitative models to provide insight in the trade-offs that exist in health care organizations.

One of the most powerful analytical tools for formally representing (modeling) and solving the operational situations listed above is mathematical optimization. As the cornerstone of operations research, optimization-based decision support tools have been widely applied in industry, as well as in health care. One of our goals in this chapter is to provide an overview of the state of the art in optimization technology, and describe what models would be most suitable (and scalable) to home care decision making.

The chapter will conclude with outlining new perspectives for analytics in home care delivery, made possible by the emergence of mobile technology, based on massive and real-time data collection. The availability of such data, combined with efficient use of machine learning models and algorithms, opens the door to data-driven decision support system that will assist home care agencies in delivering the best possible care at the most efficient cost, preferably in real time.

1.1 Overview of Home Care Delivery

Home care programs differ from country to country (or even state or province) in the range of services offered and the coverage offered either by insurance companies or the public system. As of 2014, in the United States, there were an estimated 4,800 adult day services centers, 12,400 home health agencies (80.0% with for-profit ownership and 4.9 million patients who received and ended care any time), 4,000 hospices (60.2% with for-profit ownership and 1.3 million patients) [1].
Registered nurses are the most common employees in home health agencies (53.1%) and hospices (48.1%). In the US, home care is a $75 billion dollar a year industry, comprised of more than 1.5 million caregivers. Home care workers range from companion caregivers to skilled nurses and occupational therapists. In Canada, the prediction is that by 2020, two-thirds of nurses will be working in the community rather than hospitals, which is a drastic expected increase compared to the current value of 30% [2]. Private and public home care agencies will thus need to be equipped with efficient tools to deal with the increased demand in home care delivery.

In the following, we distinguish home care and home health care.

1.1.1

Home Care

Home care services (or in-home care) are commonly presented as a service to support activities of daily living (ADL) and instrumental activities of daily living (IADL). ADL refers to activities such as bathing, clothing, transfer, toilet use, feeding and walking, which reflect the patient’s ability to heal. IADL refers to everyday tasks such as light housework, meal preparation, taking medication, buying groceries or clothing, using the phone and managing money, which allow the patient to live independently in their community.

These services may be offered to help aging seniors or anyone suffering from an injury or chronic illness, an accident, or following surgery. It is increasingly an integral part of the post-hospital recovery process, particularly during the first few weeks after discharge.

Home care is usually provided by caregivers who are not certified medically.

1.1.2

Home Health Care

Home health care covers a wide range of activities, which differ in the level of required expertise, frequency, and duration. They include wound care for pressure sores or a surgical wound, patient and caregiver education, intravenous or nutrition therapy, injections, and monitoring serious illness and unstable health status. Home care activities determine the patients' care plan which typically involves doctors, nurses and other care providers. They may be punctual, limited in duration or continuous, combining different types of care, sometimes extending several months. In what follows, we use the classification temporary care (or short term), chronic care (or long term), and specialized programs. Each of these may be accompanied by home care services to support the emotional needs and ADL of the patients.

The structures offering home health care differ by country. Home Care programs in Canada for example are managed by provincial governments through home and community office of the patients’ health authority [3][4]. They deliver a wide range of services, including short-term care, long-term care and other specialized programs. Services may vary from one community center to another.
In most European countries such as France, Italy, Portugal, Spain, and United Kingdom, home health care usually falls under the responsibility of a higher government authority, while these services are the responsibility of a local government or municipality in Denmark, Finland and Sweden. For example, in Belgium the federal public health insurance system finances a set of public and private organizations to provide home care assistance. In some cases home care and home health care may be overseen by different government levels [5].

In France, a distinction is made between home hospitalization (HAD) (often long-term for patients of all ages with acute or chronic, progressive and/or unstable pathologies who, if not treated by a HAD structure, would be hospitalized in a health facility with accommodation) and nursing home care (SSIAD) [6]. Specialized structures like palliative care, end of life [7] or rehabilitation services are also offered. In both France and the US, public structures and private institutions share the home health care market. Private institutions generally have more flexibility than public ones when accepting new patients, which impacts the decision-making process.

The needs assessment is usually based on medical expertise. Some countries may offer a single point of entry to the system: interdisciplinary teams or even agencies will provide guidance through the variety of services and providers [5].

Temporary care
Temporary care often refers to post-surgery or hospitalization support. The services address very specific needs: change dressings, help manage medications or ensure that the recommendations of the care team are being followed. Some structures (such as in France [6]) offer temporary care also in the case of non-stabilized disease (for example chemotherapy) or rehabilitation care at home.

Specialized programs
Programs including hospice care, palliative care, rehabilitation and chemotherapy are considered specialized programs. Hospice care is a bundle of comprehensive services for terminally ill patients with a medically determined life expectancy of 6 months or less. The provided care emphasizes the management of pain and symptoms. Hospice care [8] involves a team approach of expert medical care, pain management and emotional and spiritual support specifically tailored to the wishes of the patient. Emotional and spiritual support is also extended to the family.

Many hospice care programs nowadays also list palliative care to the range of care and services they provide, as hospice care and palliative care share the same core values and philosophies. Defined by the World Health Organization in 1990, palliative care seeks to address not only physical pain, but also emotional, social, and spiritual pain to achieve the best possible quality of life for patients and their families. Palliative care is acute care delivered in a holistic approach to the person with a severe, progressive or terminal illness. The goal of palliative care is to relieve physical pain and other symptoms, but also to take into account psychological, social and emotional needs.
spiritual suffering. Palliative care extends the principles of hospice care to a broader population that could benefit from receiving this type of care earlier in their illness or disease process. For example, in France more than 25% of the interventions offered are in palliative care [9].

1.1.3 Operational Challenges

In all the previously described forms, a doctor’s order is needed to start the care process [6,10]. After the request is received, an appointment is scheduled by the home health agency. The coordination of the care will then be handled by specific staff. In Canada for example, a case manager will be assigned to the patient, particularly in the context of chronic care. This will ensure the success of the coordination of care.

Continuity of care is a fundamental principal of the patient-nurse (or team) relationship. Studies show that continuity of care increases patient satisfaction and results in a better standard of care [11].

The major operational problems that a manager has to deal with are:

• who should be assigned as a case manager to each patient?
• how is the provider-to-patient assignment performed?
• how is the coordination between providers satisfied?
• who should determine the operationalization of the plan care?
• how is staff scheduled?
• how are patients visited, i.e. how are providers routed?
• is the case manager the principal provider?
• how is the continuity of care defined?
• how to deal with uncertainties?
• how to deal with dynamics (patients leave and others arrive)?
• what is the time horizon to consider?

Most providers currently address these problems using their experience. In practice, some managers would decide how the assignment is performed but the provider will be responsible of the scheduling and routing his or her patients.

In most cases presented in the literature, these problems are approached in three phases: First the assignment of patients is done (sometimes concurrently with a districting of the territory in case of large-scale problems), second the scheduling of providers will be performed, and finally the operational routing will be determined. The second and third phases are often addressed simultaneously since routing and scheduling are very similar in the home care context.

Mathematical models provide an efficient tool to solve these problems, especially when trying to solve multiple phases at the same time. These mathematical models can be adapted to each home care delivery context. For example, in temporary home care, coordination is usually not a challenge since a case manager is not mandatory. Therefore the major problem to address is the routing of staff. In the case of coordination of providers, if more than one person needs to visit the patient or to collect medicine/equipment, time windows and length of routes are introduced and
may become hard constraints. Each of these problems will be specifically addressed in the following sections. For the sake of generality, we will use the term provider for caregivers, nurses and any other professional/employee that may visit a patient in the course of the home care delivery.

Sahin and Matta [12] propose a categorization of the operations management decisions in home care structures.

- Long term: home care service offering specification (not covered in the literature) or global demand forecasting;
- Mid term: districting and allocation of capacity to districts;
- Short term: assignment of operators to visits or assignment of operators to patients;
- Very short term: routing.

While Sahin and Matta propose 1–3 years to characterize long-term decisions and 6–12 months for mid-term decisions, we nuance these durations and relate them to the home care context. In Quebec’s context for example, districting is a long-term decision as it involves the creation of teams (professionals, team leader, etc). Therefore the structure is too substantial to be modified regularly. We also point out that the assignment of operators to visits may be a short term decision; but in chronic and hospice care, this decision tends to be mid-term for follow-up requirements and continuity of care purposes. For example, in [13], the assignment problem is dealt with at the same time as the scheduling and routing of nurses to ensure continuity of care; they use a monthly horizon to assign, schedule and route the nurses. We next present a more detailed discussion on the choice of the planning horizon and the continuity of care.

Discussion on the planning horizon

Determining the length of the planning horizon is crucial for balancing the problem’s requirements and the efficiency of determining solutions. While choosing a planning horizon of one day allows for quick adjustments in the dynamic context of home care, it is very difficult to include continuity of care. Besides, a daily planning approach might lead to myopic decisions concerning mid-term working time restrictions. For example, in [14] a one-week horizon is used that takes into account routing and rostering decisions simultaneously, but disregards the continuity of care requirement for a longer horizon.

Works that focus on assignments only, usually prioritize continuity of care. Routing is performed as a second step, in which travel times are either ignored or estimated for overtime calculations. As a consequence, the assignment of nurses to patients might be infeasible due to working time restrictions or unavailabilities (e.g., because of holidays or training of a nurse). While some works impose the continuity of care as a strict requirement [15] [16], it is also possible to find robust assignments of nurses to patients over a long-term horizon (several months), in which the number of patient-to-nurse reassignments is minimized [17].

Another approach to consider continuity of care and routing is to first build weekly routes and then assign each of the routes to the nurses [13] [18]. For example, if a
patient is visited multiple times in a week, we can build routes such that all visits must be on the same route, and are therefore performed by the same nurse [18].

In [14], a mathematical programming approach is developed that aims to generate a mid-term nurse roster that seeks for the maximal continuity of care considering nurse availabilities, provider-to-patient compatibilities, and daily and monthly working time restrictions. On the basis of a weekly master schedule the solution provides the assignment of providers to master tours of patient visits throughout a planning horizon of one month.

Other approaches that recommend using a master schedule are presented in [14, 19], which again use a one-month planning horizon. The motivation for using a master schedule is that (in their applications) the home care provider may not want to create the schedule for the next week from scratch. They propose to find a repetitive tour plan considering all patient requests on a weekly basis.

Lastly, Bennett and Erera [20] consider a one-year rolling horizon for planning daily visit schedules. They maintain continuity of care by decomposing the problem into single-nurse problems. In their context, each nurse serves one district.

**Home care planning problem**

To summarize, the home (health) care planning problem (H(H)CP) consists of:

- the scheduling of the visits to each patient in a time slot (i.e., a day and a time period, morning or afternoon, in the planning horizon),
- the assignment of a caregiver to each of the visits, and,
- the sequence in which the caregivers visit each of their assigned patients in a time slot considering work regulations.

While each of these problems can be addressed individually, many applications address also routing and provider-to-patient allocation simultaneously.

### 1.2

**An Overview of Optimization Technology**

As the focus of this chapter is the application of mathematical optimization to model and solve home care delivery problems, we first provide an overview of the available optimization technology, and the current capabilities.

Optimization methods can be classified in several ways. First, there are *exact methods* that can return provably optimal solutions, and *heuristic methods* that may be able to find good solutions, but have no guarantee of optimality. Second, there are *generic optimization methods* that need the problem description to be in a specific format, and *dedicated methods* which are developed for a specific application. All generic optimization methods require that the problem is represented as a mathematical optimization model, using decision variables, constraints, and one single objective function to be optimized (minimize or maximize). The generic methods differ in
Table 1.1 An overview and characterization of optimization technologies. The third column ‘Scalability’ represents the typical problem size that is expected to be optimally solved within a reasonable amount of time, for example several minutes to one hour, as of 2017.

the restrictions they impose on the model; for example, linear programming requires that all constraints, and the objective, are linear expressions. This choice impacts the performance of the associated solving methodology; for example, linear programming be solved efficiently for almost any practical problem that fits in the computer’s memory.\[1\]

In Table 1.1 we list and characterize the optimization technologies that will be used in this chapter. We next describe each technology in more detail and provide guidelines for choosing the right technology for the application at hand.

1.2.1

Linear Programming

Linear programming is among the most efficient and scalable optimization technologies available. It requires that all variables are continuous, and the constraints and objective are linear. This structure allows to develop highly efficient solution methods that provide provably optimal solutions; see \[21\] for an overview of the development of linear programming solvers. Nowadays, practical linear programming problems that contain millions of variables and constraints may be solved in seconds or sometimes minutes.

For some problems a linear program may provide an integer optimal solution, but in general the solution will be fractional. Since many of the problems discussed in this chapter demand integer decisions, linear programming alone may not be sufficient and we need a different methodology.

1) There exist linear programming models of modest size for which solvers may need more time, although these are rare in practice \[21\].
1.2.2 Mixed Integer Programming

In mixed integer programming (MIP), we can extend linear programming models with integer variables (the ‘mixed’ stands for the mix of continuous and integer variables). While general integer variables are sometimes useful, the most commonly used integer variables are binary variables. A binary variable \( x \) can represent, for example, whether a specific provider will be allocated to a patient \((x = 1)\), or not \((x = 0)\). MIP solvers are designed to be exact and thus provide a provably optimal solution. They can also be executed to terminate within a given time limit, in which case they provide a heuristic solution (together with a guaranteed maximum distance from the optimum, or optimality gap).

The addition of integer constraints to linear programming makes these problem fundamentally more difficult to solve than LPs. Yet, there has been tremendous progress in solving MIP models over the last decades, and nowadays MIP models with tens of thousands of variables and constraints can often be solved in a matter of minutes. By the nature of this type of problem more difficult MIP models do exist, but the ongoing improvements in optimization technology continuously push the frontier of scalability and applicability; see [22] and [23] for an overview of progress in MIP solving technology.

To date, the most powerful MIP solvers are IBM ILOG CPLEX, Gurobi, and Fico Xpress. Other MIP solvers of good quality include SCIP (which can also handle CP models), MIP-CL, and COIN-CBC.

1.2.3 Constraint Programming

Constraint Programming (CP) has it origins in artificial intelligence and computer science, being first developed in the areas of constraint-based reasoning and logic programming [24]. Recent CP systems incorporate mathematical optimization techniques as well, and have been increasingly applied to solve optimization problems for operations research and other business applications. In most cases, discrete (finite domain) CP systems are used for this purpose. CP solvers are designed to be exact and provide a provably optimal solution. Like MIP solvers, they can be terminated within a given time limit and return a heuristic solution. Some CP solvers also provide the associated optimality gap, i.e., the guaranteed maximum distance from the optimum.

CP models allow variables to take any set of discrete values, and expressions can be of any algebraic form (linear, nonlinear, logical, and even relational). In addition, CP systems offer a library of special expressions, such as the AllDifferent constraint that requires a set of variables to take distinct values. Most systems also offer dedicated syntax to represent scheduling problems, for example using the concepts of activities and resources [25], or interval variables [26]. Similar to mixed integer programming, CP technology can scale easily to thousands of variables and constraints; CP solvers can be especially effective for applications that contain a scheduling component.
The most powerful CP solvers to date include IBM ILOG CPLEX CP Optimizer, Google OR-tools, Gecode, and Choco. The solver SCIP is a hybrid solver that can handle both MIP and CP models [27].

1.2.4 Heuristics and Dedicated Methods

In some cases, the requirements and scale of the problem can be too complex to be solved with a single generic MIP or CP model. In such cases, one typically develops a heuristic approach, or a dedicated method that is tailored to the application at hand. Heuristics are usually designed to find good solutions relatively quickly, but cannot prove optimality or provide a bound on the solution quality.

Heuristics come in different forms. As mentioned above, one can use MIP or CP models with a solving time limit to turn these exact methods into heuristic methods. Another, often used, approach is to decompose the problem into smaller sub-problems that are easier to handle. For example, a one-year planning problem may be decomposed over time into twelve monthly planning problems. As another example, a large regional staff planning problem may be decomposed into smaller district staffing problems.

It is also possible to develop heuristic methods that do not require a MIP or CP model, but instead work directly on problem-specific data structures. There exist principled approaches such as local search that can provide a systematic framework for developing heuristics [28]. These are often combined with meta-heuristics to further refine the solutions [29]. Popular heuristic methods include Tabu search [30] and simulated annealing [31].

1.2.5 Technology Comparison

When choosing a specific optimization technology, one needs to balance 1) the expectations or requirements of the delivered solution, 2) the solver capabilities in terms of solving time and memory requirements, 3) the time to develop the solution, and 4) maintenance and flexibility of the solution.

Solution expectations and solver capabilities

When the solution is expected to coarsely balance multiple resources, for example when planning global staffing levels over a long time horizon, linear programming may be suitable, since no detailed integer decisions are required. Since linear programming can be solved very efficiently in practice, there are usually no solver limitations in this case.

If instead we need to develop a dispatching tool that generates operational nurse schedules, a MIP or CP-based approach would be needed to handle the discrete de-

2) In some cases, it is possible to apply a formal decomposition method that provides provably optimal solutions, for example using Benders decomposition or column generation.
cisions. As a rule of thumb, most modelers will first attempt a MIP model, as this technology can provide good results on a wide range of problem classes. For more constrained scheduling and routing applications, however, CP models can perform better than MIP and may be the preferred technology.

In some cases, the requirements and scale of the problem can be too complex to be solved with a monolithic MIP or CP model. In such cases, one typically develops a decomposition approach (in which individual components can still be solved with MIP or CP), or a dedicated method. Popular decomposition methods that have been used in the home health care literature include column generation (or branch-and-price) and Benders decomposition (particularly Logic-based Benders). Specific examples and references of these decomposition approaches as well as dedicated heuristic methods will be provided in Section 1.5.

**Development time and maintenance**

Home care solutions often contain a core set of problem characteristics, for which models have been developed in the literature, and additional provider-specific requirements. When the additional requirements are limited, or can be handled in a post-processing step, it is often desirable to deploy an off-the-shelf software package that is designed to solve a specific problem class (e.g., employee scheduling or routing). However, many medium to large size organizations have additional needs that are not addressed by commodity solutions, for example due to their organizational structure, or because of physical or legal requirements. In such cases, a tailored solution is often developed by a software consulting firm.

When developing a tailored solution, one typically starts with a baseline MIP (or CP) model. These can usually be based on similar models in the literature, and can be quickly adapted to the organization’s needs. This allows for a relatively fast delivery of a prototype solution, which is followed by an iterative process of testing and refinement. The overall development process may take several months or more and can result in scalable and detailed solutions. If the solution uses a commercial optimization solver, one should also budget for the solver license.

The alternative would be to develop a new dedicated solution method, perhaps from scratch, or by re-using optimization solvers for sub-parts of the problem. For some problems, for example certain employee scheduling problems, a dedicated heuristic approach may work well. The development time may be similar to a MIP or CP based solution, and does not require the use of a commercial solver. That said, it may be challenging to adapt, extend, or maintain dedicated solutions, especially over a longer time period (e.g., several years). Instead, optimization models can be extended or adapted much more easily, by simply changing the model description. Moreover, as optimization models are algebraic, model maintenance and development by multiple people is much easier as well. Lastly, the performance of optimization solvers is continuously improving, which means that the same optimization model can provide better solutions, or scale to larger instances, when updating the solver to a new version.

In the following sections, we present the most common operational challenges in
home care, and their respective optimization-based solutions.

1.3 Territory Districting

When applicable, the districting is the first problem to solve to tackle the HHCP. In the province of Quebec for example where local authorities offer the home care services, it is a very common approach to divide the territory into districts. Each district will then be staffed (nurses, social workers, etc).

A districting problem consists of grouping small geographic areas, called basic territorial units, into larger geographic clusters, called districts. These districts should be balanced (for example of equitable size), contiguous, and compact. Typical examples for basic units are census codes, zip code areas or even aggregated demand points of patients. In home care applications, the districts should have a good accessibility with respect to public transportation and have an equitable workload based on service and travel time [32, 33].

Let $J = \{1, \ldots, n\}$ be the set of basic units and $d_{ij}$ the distance between units $i$ and $j$. In the home care context it usually represents the road distance (or travel time). Activity measures are defined for each unit: demand (such as the number of patients and frequencies of service), service time (total duration of visits), etc. The objective is to aggregate these basic units into $p$ districts; $p$ could be either equal to the number of providers (in this case we seek to define one district per nurse), or a smaller number which we can refer to as teams. These $p$ districts should satisfy the planning criteria of balance, compactness, and contiguity. As highlighted in [34], “the” mathematical model for districting problems does not exist, due to the various design decisions, although there are many commonalities.

Three measures of balance are commonly used. Let $w_k$ be the workload of district $k$. The first measure of balance $B_1$ is based on the relative deviation of the district workload $w_k$ from the mean workload $\mu = \frac{\text{Total workload}}{\text{Number of districts}(p)}$:

$$B_1 = \left| \frac{w_k - \mu}{\mu} \right| , 1 \leq k \leq p$$

If this balance measure is equal to 0, it ensures that all districts have the exact same workload, which in practice may be difficult to achieve.

The second measure will instead concede a priori a certain relative deviation $\alpha > 0$ from perfect balance and only measures the imbalance exceeding this threshold:

$$B_2 = \frac{1}{\mu} \max \{ w_k - (1 + \alpha)\mu, (1 - \alpha)\mu - w_k, 0 \}$$

The districts are balanced if the workload is between the lower bound and the upper bound. The value of $\alpha$ should be fixed by management.

A third measure will instead minimize the deviation between the maximum and the minimum workload of districts.

$$B_3 = \max \{ w_1, \ldots, w_p \} - \min \{ w_1, \ldots, w_p \}$$
The choice of the fairness measure usually depends on the context, and the decision is to be made by the manager.

Districting approaches are commonly based on mixed integer programming or heuristic methods. To reduce the day-to-day traveling distance, one aims to establish contiguity and compactness of the districts. Different measures for contiguity and compactness exist; we refer the reader to [34] for more details. A typical districting model is location-allocation model, which assigns basic units to the “seed” of a district, ideally in order to maximize the compactness of the districts.

To formulate the location-allocation model, we introduce the following decision variables:

- $x_{ij}$ equals 1 if basic unit $j$ is assigned to basic unit $i$ and 0 otherwise, and;
- $y_i$ equals 1 if basic unit $i$ is selected as district seed and 0 otherwise.

The model is then formulated as:

$$\text{min} \sum_{i,j \in J} w_i d_{ij}^2 x_{ij} \quad (1.1)$$

subject to:

$$\sum_{j \in J} x_{ij} = 1 \quad \forall j \in J \quad (1.2)$$

$$\sum_{j \in J} w_j x_{ij} \geq (1 - \alpha) \mu y_i \quad \forall i \in J \quad (1.3)$$

$$\sum_{j \in J} w_j x_{ij} \leq (1 + \alpha) \mu y_i \quad \forall i \in J \quad (1.4)$$

$$\sum_{i \in J} y_i = p \quad (1.5)$$

$$y_i, x_{ij} \in \{0, 1\} \quad \forall i, j \in J \quad (1.6)$$

The constraints ensure that each basic unit is assigned to one district only (1.2), that the workload of each district is within the boundaries (1.3)–(1.4) and that $p$ district "seeds" are selected (1.5). No explicit measure of compactness is used in this model, however, as the objective is to minimize the total distance.

In [33], the compactness is integrated as a hard constraint in two ways: by limiting the maximum distance between two basic units that would be assigned to the same district and by minimizing a compactness measure which is the maximum distance between two basic units assigned to the same district. With this approach, they are able to solve problems where a maximum of 100 units are to divide into 4 districts.

While this type of formulation is useful for few hundred basic units, it becomes intractable when the problem size grows. Heuristics approaches then prove to be more efficient, while being flexible to include almost any practical criterion and measure for the design of districts.

For example, Blais et al. [32] use a tabu search algorithm to solve the districting problem for a local community health center in Montreal. The objective is to divide the territory in six districts in order to balance the workload. A tabu search is an
iterative method that locally explores the space of solutions. It starts from an initial solution and moves to the others through one or multiple movements. In this context, each solution is a feasible districting and is evaluated using a measure (i.e., an objective function).

Blais et al. [32] also propose a mobility level measure for each solution $s$:

$$f(s) = \frac{\sum_{k=1}^{p} \sum_{i,j \in D_k | i \leq j} v_i v_j d_{ij}}{n_k (n_k - 1) / 2 (\sum_{k \in D_k} v_i)^2}$$

where $D_k$ represents the set of units in district $k$, $v_i$ is the number of visits in basic unit $i$, and $n_k$ is the number of basic units included in district $k$. If the value of $f(s)$ is low, this represents that the ease of travel is high. Their criteria to measure the workload equilibrium is very similar to $B_2$.

1.4 Provider-to-Patient Assignment

We next consider the provider-to-patient assignment problem. First we introduce workload measures, followed by fairness evaluation, generalized assignment models that allow to model this problem, and we finish with dynamic approaches to consider re-assignments of new patients.

1.4.1 Workload Measures

Most papers in the literature only use service time (duration of visits) and traveling time to determine provider’s workload [16, 17, 32, 35]. In [15], a new workload measure is introduced. The rationale is that all indirect duty should be considered, especially for hospice and chronic care. This workload is derived from the notion of case management. Since providers (case managers, pivot nurse, or reference nurse) have to coordinate care and need to perform work outside the "visit", this work should be accounted for. Four categories of patients are considered:

- Short term (post-surgery, post-hospitalization, ...);
- Long term needing punctual care;
- Long term needing continuous care;
- Palliative patients.

To scale the workload between these patients, a reference visit to short term patients (typically of 20 minutes) is used and all other visits are then aligned to this one. For example, a visit to a palliative patient usually requires 4 times more work. This value was determined with the managers and the care team and may be adapted to adjust to the context.

We note that estimating the traveling distances is important when considering assignment models. Yalçındağ et al. [16] discuss the use of different methods to estimate travel times of care givers. They compare the Kernel regression technique...
(based on historical data) to the average (average distance from one patient to all the others) and $k$-nearest neighborhood search methods. The Kernel regression technique can be successfully used in a two-stage approach that first assigns patients to caregivers and then defines their routes. However it needs historical data. The average (which is the simplest to use) performs well when the territory to cover is not large.

A limitation to this approach is however highlighted in [36]. While the approach seems very promising when distance are small (e.g., in dense urban territory) it does need seem to perform as well when distances are larger. Therefore, the authors in this study recommend the use of routing first and then assign.

1.4.2 Workload Balance

To measure workload balance, let’s introduce $w_i$ the workload of provider $i$, $w_{\text{max}}$ and $w_{\text{min}}$ the maximum and the minimum workloads respectively. Fairness can be evaluated by minimizing the:

- Deviation between max and min: $(w_{\text{max}} - w_{\text{min}})$
- Maximum workload: $w_{\text{max}}$
- Total workload: $\sum_i w_i$
- Total squared workload: $\sum_i w_i^2$

In other cases, instead of directly optimizing the workload, some authors focus on the deviation between maximum workload and the average. A target value (if available) could be substituted to the average. For example, for each solution $s$, the total overload of each provider $O_i$ is equal to the sum of three components:

- Visit load: which corresponds to the deviation between the current load of visits $V_i$ of provider $i$ and the average $\bar{V}$:
  $$ O_{i1}(s) = \max\{0, V_i(s) - \bar{V}\} $$

- Case load: which corresponds to the deviation between the current load of cases in each category (total number of assignment of patients $\sum_c x_{ic}$ so that patient $c \in C_j$, where $C_j$ is the set of patients of category $j$) and the average (total number of cases $n_j$ over the number of providers $|I|$). We consider that each additional patient receives the average number of visits $\bar{v}_{ij}$ of its category. Indirect work (through a penalty $p_{ji}$) is also derived from the category $j$ of the patient. This additional overload is:
  $$ O_{i2}(s) = \sum_{j \in J} \max \left\{ 0, \sum_{c \in C_j} x_{ic} - \left\lceil \frac{n_j}{|I|} \right\rceil \right\} \cdot \bar{v}_{ij}(s) \cdot p_{ji} $$

- Travel load: which corresponds to the deviation between the current travel $T_i$ of provider $i$ and the average $\bar{T}$:
In practice, we use multi-criteria optimization where the objective is to maximize (or minimize) a set of criteria. Since our optimization model can optimize only a single objective function, we introduce weights $\omega$ to represent the importance of each criterion. The appropriate values of $\omega$ result from discussion with the managers and staff. This type of multi-criteria functions also provide an efficient tool to test different scenarios: “what-if” one criterion is more important than another for example.

1.4.3 Assignment Models

We next combine all the above elements in a single optimization model. Here we provide the assignment model of [15] of operators to patients that includes the most complete workload measure introduced earlier. $C$ is the set of patients to assign, $v_c$ the number of visits required by patient $c$ and $h_{jc}$ the heaviness of a patient of category $c$ (this helps to differentiate between palliative patients and short term patients for example). We also define $t_{ic}$ as the travel load of patient $c$ to provider $i$. It is proportional to the number of visits $c$ requires.

\[
O_{i3}(s) = \max\{0, T_i(s) - T(s)\}
\]

Minimize $\sum_{i \in I} \omega_1 \cdot (O_{i1})^2 + \sum_{i \in I} \omega_2 \cdot (O_{i2})^2 + \omega_3 \cdot \left( \sum_{i \in I} (O_{i3})^2 + (T)^2 \right)$

subject to:

\[
\sum_{i \in I} x_{ic} = 1 \quad \forall c \in C \quad (1.7)
\]
\[
\sum_{c \in C} v_c \cdot h_{jc} \cdot x_{ic} - \nabla \leq O_{i1} \quad \forall i \in I \quad (1.8)
\]
\[
\sum_{c \in C_j} v_c \cdot x_{ic} = \bar{v}_{ij} \cdot \sum_{c \in C_j} x_{ic} \quad \forall i \in I, \forall j \in J \quad (1.9)
\]
\[
\sum_{c \in C_j} x_{ic} - \left\lceil \frac{n_j}{|I|} \right\rceil \leq s_{ij} \quad \forall i \in I, \forall j \in J \quad (1.10)
\]
\[
\sum_{j \in J} s_{ij} \cdot \bar{v}_{ij} \cdot h_{j} \leq O_{i2} \quad \forall i \in I \quad (1.11)
\]
\[
\sum_{i \in I} \sum_{c \in C} t_{ic} \cdot x_{ic} = |I| \cdot T \quad (1.12)
\]
\[
\sum_{c \in C} t_{ic} \cdot x_{ic} - \bar{T} \leq O_{i3} \quad \forall i \in I \quad (1.13)
\]
\[ x_{ic} \in \{0, 1\} \quad \forall i \in I, \forall c \in C \]  
(1.14)

\[ O_{i1}, O_{i2}, O_{i3} \geq 0 \quad \forall i \in I \]  
(1.15)

\[ \overline{v}_{ij}, s_{ij} \geq 0 \quad \forall i \in I, \forall j \in J \]  
(1.16)

\[ T \geq 0 \]  
(1.17)

Constraints (1.7) impose that each patient \( c \) is assigned to a provider \( i \). Constraints (1.8), (1.9), (1.11) and (1.12)-(1.13) define respectively the visit load, the case load and the travel load. Variable \( s_{ij} \) is used to define the positive deviation between the number of cases obtained by each provider in each category \( j \) and the average. Finally constraints (1.14)-(1.16) are domain definition.

This model is non-linear (constraints 1.11) and cannot be handled directly by commercial MIP solvers. Therefore, a heuristic approach is developed in [15]. In terms of problem size, they solve problems with more than 1400 patients, 26 staff members and 36 basic units.

1.4.4 Assignment of New Patients

Carell and Lanzarone [35] consider a robust assignment of new patients to providers. They use a rolling horizon of 26 periods and each decision made in a period is fixed for the following ones. Three sets of patients are considered: those not requiring continuity of care, those requiring continuity of care (with a provider already assigned or to be assigned) and those requiring a partial continuity of care.

To deal with the dynamics of the arrival of new patients, using a flexible assignment as described in [15] helps the stability of the re-assignments. In particular, the authors test 1-day, 3-day, and 7-day assignment periods. Patients from the previous period are fixed and only new patients are assigned in order to balance the workload. They show that if boundaries are flexible for patient assignment (i.e., providers are already covering for others outside their territory to alleviate for demand variation), then this time interval becomes less critical.

1.5 Task Scheduling and Routing

Task scheduling and nurse routing is most relevant when done in conjunction with the patient-to-nurse allocation. Otherwise, in case the tasks and patients have been allocated to nurses already, determining the associated routes is immediate. From an optimization modeling perspective, the core scheduling/routing of nurses for a single day corresponds to vehicle routing problem with time windows (VRPTW), in which each of the patient visits are supposed to occur within a given time window. Additionally, we typically need to respect nurse-patient compatibility constraints, workload balancing constraints, and continuity of care constraints for longer time horizons. The VRPTW is known to be a challenging problem class, and the addi-
tional constraints typically make it even harder to find provably optimal solutions. Therefore, most of the optimization approaches described in the literature present heuristic methods, although some dedicated exact approaches have been presented as well, as we will see.

One of the first optimization approaches for this problem was presented by Begur et al. [37], who compute daily nurse schedules/routes based on a master schedule of patient visits. They adapt well-known heuristic methods including the Clarke-Wright savings heuristics [38], as well as route improvement heuristics that were developed for the Traveling Salesman Problem (TSP). However, constraints to ensure continuity of care are not considered explicitly. Another early work, by Cheng and Rich [39], introduces formal MIP models for the scheduling and routing problem, but use a problem-specific heuristic for instances of larger scale.

Bertels and Fahle [40] introduce a heuristic approach that combines linear programming, constraint programming, and metaheuristics. They use CP to generate roster sequences, for which the associated optimal start times are computed by LP. CP is also used to find good initial routes. Route improvements are obtained by simulated annealing and tabu search.

Eveborn et al. [41] describe an integer programming model for the nurse scheduling/routing problem, based on a set covering formulation. Their model includes patient-nurse compatibility requirements and allows for visits that require multiple staff members. The set covering formulation explicitly lists the possible routes, which are generated to reflect the constraints. Since several other approaches in the literature follow a similar structure, we next present this set covering model in more detail.

Let $I$ be the set of nurses (caregivers), let $V$ be the set of patient visits, and let $J$ be the set of schedules/routes that can be allocated to a nurse. The set $J$ is pre-computed and contains, in principle, all possible allowed schedules. (Since there are exponentially many schedules we usually restrict ourselves to a subset of interesting schedules.) Each schedule $j \in J$ has an associated cost vector $c_{ij}$ which reflects the penalty or weight if nurse $i \in I$ follows that schedule, for example the travel length, possible violation of time windows, patient-to-nurse preference, etc. We introduce a binary parameter $a_{ijv}$ to indicate whether nurse $i \in I$ performs visit $v \in V$ in schedule $j \in J$ ($a_{ijv} = 1$) or not ($a_{ijv} = 0$).

We next introduce a binary variable $x_{ij}$ that denotes whether nurse $i \in I$ is assigned to schedule $j \in J$ ($x_{ij} = 1$) or not ($x_{ij} = 0$). The set covering formulation then ensures that each nurse is allocated to one schedule (1.19), and each patient visit is performed once (1.20), while minimizing the total cost:

$$\min \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \tag{1.18}$$

subject to

$$\sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{1.19}$$

$$\sum_{i \in I} \sum_{j \in J} a_{ijv} x_{ij} = 1 \quad \forall v \in V \tag{1.20}$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J. \tag{1.21}$$
Given the complexity of the resulting model, it is solved heuristically by the repeated matching approach in [41].

Chahed et al. [42] present an optimization approach for delivering chemotherapy at home. In this context, the sample lifetime of the product is particularly important, as it varies from 2 hours to several days. The optimization approach considers both production and distribution; the distribution problem corresponds to the home care scheduling and routing problem, for which Cahed et al. present an exact branch-and-bound approach using a MIP model.

In [43], Rasmussen et al. propose an exact solution method based on branch-and-price, which is a well-known approach for solving vehicle routing problems. Similar to [41] they introduce a set covering model, which is extended with side constraints to represent temporal dependencies. Branch-and-price decomposes the problem into a master problem and a subproblem. The master problem iteratively solves the set covering problem with a restricted set of variables, after which the subproblem determines whether improving schedules exists. If so, these are added to the master problem and the process repeats. The resulting column generation approach is then embedded in a systematic branch-and-price search to find provably optimal solutions.

Cappanera et al. [18] present a MIP model to jointly handle the assignment, scheduling, and routing of home care visits. The model allows multiple patients visits per week and respects patient-nurse compatibilities. Their solution approach relies on weekly patterns of patient visits, which are then allocated to the caregivers.

A different exact approach, based on logic-based Benders decomposition, was proposed by Heching and Hooker [44]. Like branch-and-price, Benders decomposition partitions the problem into a master and a subproblem, but performs constraint-generation rather than variable-generation. In [44], the master problem assigns caregivers to patients (one patient may have multiple visits), which is modeled and solved using MIP. The subproblem then computes the optimal route for each caregiver for each day of the week, which is modeled and solved with CP. If the route/schedule is not feasible, the subproblem returns a Benders cut that forbids that particular assignment.

Both Cappanera et al [18] and Heching and Hooker [44] ensure continuity of care within the one-week planning horizon of their respective models. While longer-term continuity of care can be handled by fixing previous patient-to-nurse assignment in a rolling horizon framework, this may lead to myopic sub-optimal solutions. For this reason, Güven-Koçak et al. [45] present an optimization approach that aims to better handle long-term continuity of care. Updates to the schedule are processed on a daily basis (usually because of new patients, or existing patients no longer need care), in which case a new schedule is computed that aims to be as consistent with the previous schedule as possible. Thus, the standard objective function to minimize travel time or total cost is now augmented with a penalty for inconsistent nurse-to-patient allocations. They present a MIP formulation as well as a heuristic solution method to solve the consistent home care delivery problem.

While most optimization methods for home care delivery consider static travel times, Rest et al. [46] consider time-dependent travel times, which can provide more
accurate solutions. They present a detailed MIP formulation for the problem as well as a heuristic method based on tabu search.

In the above, we have sketched some of the most common approaches to solving the scheduling/routing problem for home care delivery. We refer to the recent surveys by Fikar and Hirsch [47] and Cissé et al. [48] for a more detailed comparison of various approaches and different variants of the rich scheduling/routing literature in this problem domain.

1.6 Perspectives

As we have seen in the previous sections, systematic quantitative models based on optimization technology has proven to be a powerful tool for decision-support in home care delivery. There are, however, still several avenues to improve the efficiency of home health care services. All fields of health care can benefit from the continuously improving mobile technologies, and home health care is no exception as show by a recent book and survey [49, 50]. This technological shift will enable new possibilities in the use of advanced predictive and prescriptive analytics tools. In the last part of this chapter we look at several new perspectives that we have seen emerging and that rely on data-driven decision making.

1.6.1 Integrated Decision Making Under a New Business Model

Over the last years the home care industry has witnessed the emergence of a new business model. Revenues of private agencies are evolving from a model where they were able to bill a fee for every service provided, to a model where each patient condition prescribes a fixed amount. Providing care of high quality, while remaining profitable, becomes more challenging as this model is more widely adopted. Most importantly, we can recognize that the individual planning problems, mentioned earlier, are now optimized separately while there are clearly interactions between them. For example, admitting certain patients, or certain patient/agency assignment, may require the need to hire additional staff, or lead to highly inefficient routes, which can significantly affect productivity.

These interactions can be made explicit by designing optimization models that span two or more of these decision problems. For example, it is possible to combine the dynamic acceptation of patients with the ongoing employee scheduling and routing problems. The resulting online, and stochastic model may be more challenging to solve, but as the algorithmic power of optimization improves, and machines get faster, solving these larger models may become routine in the near future.
1.6.2 Home Telemetering Forcasting Adverse Events

Since about 20% of Medicare patients are readmitted shortly after discharge\[51\], insurance companies have established financial penalties to hospital with high readmission rates 30 days after discharge\[52\]. Prediction of adverse events such as hospital readmissions, in the context of home health care patients, can thus have a significant economical impact.

Patients that benefit from Home Health Care are initially evaluated by a medical team to assess the level of assistance they will require. Despite the many advantages of home health care, one primary medical challenge of home caring for patients is the lack of constant supervision, as one would get within an institution. Therefore, there is always a risk that the expected medical outcome of a care plan is not achieved, that the patient falls and is not rescued promptly, or that he or she gets confused with medication leading to deterioration of his or her health condition and eventually to re-hospitalization.

When patients are admitted to a home care agency, they generally are visited by a nurse who will perform an initial needs assessment. Patients may be assigned to a Home Telemetering (HT) program, in case the agency offers such program. While on a HT program, patients answer a periodic questionnaire during which they will be asked to take some vital signs readings. This information is then transmitted to the HHC agency where a nurse monitors a HT case load. In some experimental cases, patients can wear sensory devices that capture and transmit vital signs in a continuous fashion.

Based on the patient diagnosed conditions and initial assessment, the care workers create alerts based on acceptable ranges of each measured vital signs, as shown in \[53\].

Sometimes, with more advanced systems, complex rules can be developed to get alerted based on combinations of suspect readings. In all cases, care workers bear the weight of setting up patients with the right set of alerts based on their conditions. The manually engineered rules then need to evolve with the patient’s condition in order to remain reliable.

When a vital sign reading is out of the acceptable range, the monitoring nurse can perform one or two of the following actions: (1) call the patient to determine next steps, and/or (2) schedule an in-person visit. The challenge is to prevent costly hospital readmissions and emergency room visits, but there is also a cost to each intervention. To add complexity, most of the alarms are false positives, not leading to adverse events.

Early detection of these events serves the purpose of the triple aim of improving outcomes: (1) quality of health services, (2) improving health of populations and (3) reducing costs \[54\].

Linear models such as multivariate logistic regression and Cox Proportional Hazard \[55\] \[56\] \[57\] \[58\] are often used because of their understandable nature. Indeed, most of the work so far has been interested in understanding the significant factors that lead to adverse events. Modern machine learning techniques, such as Neural
Networks have not been adopted yet, although have demonstrated success in many industries, from computer vision to market finance. One of the challenges of such approaches resides in their interoperability [59]. There are however new companies which have run case studies, reported in a white paper [60], that indicated that machine learning approaches could do significantly better than human-defined alerts, both in terms sensitivity and specificity.

1.6.3 Forecasting the Wound Healing Process

Today, wound care costs the Canadian health system at least $3.9 billion annually or 3% of total health expenditures [61] and is believed to be around $50 billions in the United States [62]. The portion of these costs dedicated to home care can be significant, as in the case of diabetic foot ulcers where direct-care cost to the Canadian health-care system of $547 million (2011 dollars) out of which $125M (around 23%) was spent in home caring. Public and private agencies must thus be able to assess the necessary effort required to manage each type of wound and each individual patient in order to plan the usage of their resources efficiently.

Although there has been some effort to forecast the healing duration of a wound we have not seen today any paper reporting the use of modern machine learning techniques. To enable the use of such methods, one would require access to a large body of high-quality data in a numerical format. Information on the nature of the wound (type, acuity, area, odor, grade, etc.) would need to be diligently recorded at every visit. Detailed information on the patient would also be quite useful, as many physiological factors, such as obesity, hypertension, and diabetes have a significant influence on wound healing. Table 1.2 gives the major factors that influence the healing of a wound, as stated in several studies [63, 64, 65, 66, 67].

<table>
<thead>
<tr>
<th>Systemic Factors</th>
<th>Local Factors</th>
<th>Organizational Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deficient eating and hydration</td>
<td>Infection</td>
<td>Absence of specialized nurse</td>
</tr>
<tr>
<td>Oxygen deficit</td>
<td>Chronic Wound</td>
<td>Multidisciplinary team</td>
</tr>
<tr>
<td>Stress</td>
<td>Wound area</td>
<td>Continuity of care</td>
</tr>
<tr>
<td>Bad sensorial perception</td>
<td>Presence of foreign body</td>
<td>Complete evaluation</td>
</tr>
<tr>
<td>Age and sexe</td>
<td>Hematoma</td>
<td>Following protocol</td>
</tr>
<tr>
<td>Obesity</td>
<td>Wound Location</td>
<td></td>
</tr>
<tr>
<td>Diabetes, HBP</td>
<td>Necrotic tissues</td>
<td></td>
</tr>
<tr>
<td>Tabagism, alcoolism</td>
<td>Wound pressure</td>
<td></td>
</tr>
<tr>
<td>Auto-immune disease</td>
<td>Wound hydration</td>
<td></td>
</tr>
<tr>
<td>Cultural beliefs</td>
<td>Wound vascularization</td>
<td></td>
</tr>
<tr>
<td>Sexual hormones level</td>
<td>Wound type</td>
<td></td>
</tr>
</tbody>
</table>
1.6.4
Adjustment of Capacity and Demand

One of important challenges faced by home care agencies in many markets is the alignment between their service capacity, that is the number of productive hours their workforce can provide in a week or a month, and the demand for service that comes from their customers. In particular it is quite common to have high peak of demand in the morning, because less mobile patients need help to get their day started while more autonomous customers prefer an early visit so that they can proceed with their daily schedule. Data gathered by the VHA Healthcare\footnote{www.vha.ca} shows the spread of thousands of daily visits over each hour of the day, as illustrated in Figure 1.1.

![Client Visits – Daily Trend](image)

\textbf{Figure 1.1} Average number of clients visited per daily hour by the VHA Healthcare agency

If most of the work is to be scheduled in the morning, then many of the caregivers are not needed during the afternoon. This can result in low motivation by the staff and high turnover of the caregivers, which may aim to find a full time position instead. This can be addressed by two possible approaches: (1) hiring part-time caregivers that will willingly work reduced hours, and (2) trying to shift demand to a later period during the day. Both of these approaches come with their own challenges.

It is therefore a challenging task to design and recruit an optimal workforce. Agencies would need to forecast accurately the demand per time period and day of the week, and using this forecast they would then need to design a set of work shifts that fit this forecast as best as possible. Such approaches are used widely in retail stores and call centers\footnote{68}, however in these service industries there is no notion of continuity of care between the employees and the customers. In fact, the assignment of customers to employees is performed each time a new customer requires a service. Dynamic pricing\footnote{69}, which is widely used in the airline and hospitality industry, may come as an option.
References


research and practice, 24 (1), 10–17.


