# Home Healthcare Integrated Staffing and Scheduling

María I. Restrepo[a,b,1,*], Louis-Martin Rousseau[a,b], Jonathan Vallée[c]

[a] *Centre Interuniversitaire de Recherche sur les Réseaux d'Entreprise, la Logistique et le Transport, CIRRELT, Montréal, Québec H3T 1J4, Canada*

[b] *Département de Mathématiques et de Génie Industriel, Polytechnique Montréal, Montréal, Québec H3C 3A7, Canada*

[c] *AlayaCare, 4200 St Laurent Blvd Suite 800, Montréal, Québec H2W 2R2, Canada*

**Abstract**

Workforce planning for home healthcare represents an important and challenging task involving complex factors associated with labor regulations, caregivers' preferences, and demand uncertainties. This task is done manually by most home care agencies, resulting in long planning times and suboptimal decisions that usually fail to meet the health needs of the population, to minimize operating costs, and to retain current caregivers. Motivated by these challenges, we present a two-stage stochastic programming model for employee staffing and scheduling in home healthcare. In this model, first-stage decisions correspond to the staffing and scheduling of caregivers in geographic districts. Second-stage decisions are related to the temporary reallocation of caregivers to neighboring districts, to contact caregivers to work on a day-off, and to allow under- and over-covering of demand. The proposed model is tested on real-world instances, where we evaluate the impact on costs, caregiver utilization, and service level by using different recourse actions. Results show that when compared with a deterministic model, the two-stage stochastic model leads to significant cost savings as staff dimensioning and scheduling decisions are more robust to accommodate changes in demand. Moreover, these results suggest that flexibility in terms of use of recourse actions is highly valuable as it helps to further improve costs, service level, and caregiver utilization.

*Keywords:* Staffing and scheduling, Home healthcare, Two-stage stochastic programming, Context-free grammars

## 1. Introduction

Home healthcare refers to any type of care given to a patient at his own home rather than in a healthcare facility like a hospital or a clinic. Caregivers (e.g., personal support workers,

[*] Corresponding author

*Email addresses:* maria-isabel.restrepo-ruiz@polymtl.ca (María I. Restrepo), louis-martin.rousseau@cirrelt.net (Louis-Martin Rousseau), jonathan.vallee@alayacare.com (Jonathan Vallée)

[1] Present address: CIRRELT, Pavillon André-Aisenstadt, Montreal, QC H3T 1J4, Canada

nurses, and therapists) meet the patients' needs by bringing all necessary equipments at their homes and therein provide care. This activity increases the quality of life for the patients, as they are allowed to remain at home where they are most comfortable. Moreover, it yields relevant cost savings for the entire healthcare system as hospitalization costs are avoided (Lanzarone & Matta, 2014).

Home healthcare planning includes different decision levels that are usually classified in three main categories: *strategic planning*, *tactical planning*, and *operational planning* (Hulshof et al., 2012). Strategic planning relates to problems addressing structural decision making to design and to dimension the healthcare delivery process. This planning level often involves long planning horizons in which decisions are based on aggregate information and forecasts. Some applications include *districting* problems in which the geographic territory where home care agencies operate is partitioned in districts (i.e. smaller geographic zones). Tactical planning is related to medium-term decision making dealing with the implementation of strategic decisions. Examples of problems in this decision level include *personnel scheduling* problems, where work patterns are designed and allocated to caregivers to meet a forecasted and often uncertain demand for services. Operational planning includes short-term decision making related to the execution of the healthcare delivery process. Applications include *visit rescheduling* where visit schedules are updated a few days in advance or during the execution day, to respond to events such as caregiver absenteeism, incoming urgent care requests, and changes in visit requirements.

The spatial distribution of patients and the uncertainty in demands represent some important features found in home healthcare workforce planning. The incorporation of these aspects increases the complexity of the problems under study. However, including them in the modeling and solution process could have a positive impact on an efficient service delivery in terms of costs and quality. First, the integration of decisions in several districts usually generates flexible staffing and scheduling solutions that respond in a better way to fluctuating demand, since caregivers are allowed to work in a different district than the one they are dedicated to (Lahrichi et al., 2006). In a similar way, the incorporation of demand uncertainty provides solutions that will be more robust to accommodate changes in demand associated with the arrival of new patients and with changes in patients' conditions.

In this paper, we focus on the integration of two medium-term workforce planning problems: the *staff dimensioning problem* and the *caregiver scheduling problem*. This integration deals with the definition of the number of caregivers to recruit per district, as well as with the allocation of schedules to caregivers while considering demand uncertainty. Caregiver schedules are defined by sequences of *work stretches* and *rest stretches*. Work stretches contain a consecutive number of work days, where each work day contains exactly one *shift* (e.g., morning shift, night shift) executed in one district. Similarly, rest stretches represent a consecutive number of *days-off*. The composition of feasible schedules is subject to work regulations ensuring, among others, that there is a minimum rest time between consecutive shifts, that each work stretch includes a sequence of shifts between a minimum and a maximum value, and that each rest stretch contains a sequence of days-off between a minimum and a maximum value.

Our work is motivated by the challenges experienced in AlayaCare, a start-up company based in Canada developing software solutions for home healthcare agencies. Most of these agencies currently lack the tools to forecast future demands, to manage their labour resources, and to optimize work assignments. Hence, the staffing and scheduling planning is mostly done manually by experienced coordinators. Since this planning method often fails to include most of the rules for the composition of schedules, as well as accurate demand forecasts, it results in the inability to hire an adequate number of caregivers, to retain current caregivers, and to meet the needs of patients.

This paper has the following contributions. First, to the best of our knowledge, our work is the first to propose an optimization approach that integrates staffing and scheduling decisions in the context of home healthcare. To do so, we present a two-stage stochastic programming model where *first-stage decisions* correspond to the staffing and scheduling of caregivers at each geographic district, and *second-stage decisions* are related to the temporary reallocation of caregivers to neighboring districts, to contact caregivers to work on a day-off, and to allow under-covering and over-covering of demand. Second, although other authors have already benefit from the expressiveness of *context-free grammars* to build short-term schedules with a planning horizon of one day (see Restrepo et al. (2017); Côté et al. (2013)), we believe that our work is the first that uses context-free grammars to build schedules over long time horizons (i.e., one month or more) guaranteeing *horizontal work regulations* such as the minimum rest time between consecutive shifts and the allocation of a minimum and a maximum number of shifts to each work sequence. Context-free grammars allow to easily incorporate horizontal regulations as a set of recursive rewriting rules (or productions) to generate patterns of strings (Hopcroft et al., 2001), in our case, to generate caregiver schedules. Third, we discuss how to forecast the demand of home care services and how to integrate these forecasts in a two-stage stochastic programming model. Fourth, we perform an extensive computational study on real-based data to evaluate the impact in costs, caregiver utilization and service level, by using several recourse actions, various scheduling policies and different planning horizons.

The paper is organized as follows. In Section 2, we review related works on caregiver staffing and scheduling for healthcare. In Section 3, we present the methodology to solve the integrated caregiver staffing and scheduling for home healthcare. Computational experiments are presented and discussed in Section 4. Concluding remarks and future work follow in Section 5.

## 2. Related Work

Healthcare planning problems for hospitals have been extensively studied over the past years. In particular, nurse staffing and scheduling problems have attracted most of the attention from the operations research community since the generation of high-quality nurse schedules can lead to improvements in hospital resource efficiency, in patient safety and satisfaction, and in administrative workload (Burke et al., 2004). Recent approaches to this problem include the works presented in Maenhout & Vanhoucke (2013) and Kim & Mehrotra (2015). Maenhout & Vanhoucke (2013) present a branch-and-price procedure to solve an integrated nurse staffing and scheduling problem, where the number of nurses has to be determined for each profession in order to balance, over several months, the workforce costs and the coverage of patients in multiple hospital departments. Results indicate that staffing multiple departments simultaneously and including nurse skills into the staffing decisions lead to significant improvements in schedule quality in terms of cost, employees' job satisfaction, and effectiveness in providing high-quality care. Kim & Mehrotra (2015) present a two-stage stochastic integer program with mixed-integer recourse to integrated nurse staffing and scheduling. In the problem, first-stage decisions define initial staffing levels and schedules, while second-stage decisions adjust these schedules at a time epoch closer to the actual date of demand realization. Results show that, when compared with a deterministic model, the two-stage stochastic model leads to significant cost savings. The work of Kim & Mehrotra is similar to ours as the authors use a two-stage stochastic integer programming program with recourse to solve integrated staffing and scheduling problems in healthcare. The objective of both works is to find initial staffing levels and schedules to minimize overall labor costs by right-sizing the staff and by balancing understaffing and overstaffing costs. However, their work differs in some important spects from ours. First, as opposed to our work, the work of Kim & Mehrotra does not consider the spatial dimension in the planning, since the staffing and scheduling is done for nurses in a hospital and not for caregivers that need to visit patients in different geographic zones. Second, the authors assume that work patterns repeat from week to week during the planning horizon and that all possible weekly patterns are generated in advance. Instead, in our approach, caregiver schedules are allowed to be different from week to week, and weekly schedules are not generated in advance, as one of the objectives of our model is to build (with context-free grammars) caregiver schedules that guarantee several work regulations. Third, regarding the use of recourse actions, both works allow for calling in additional staff when needed. However, our work uses an additional recourse action corresponding to the reallocation of caregivers to neighbor areas and, contrary to Kim & Mehrotra, we do not allow to cancel shifts from the scheduled staff.

Problems related to the routing and scheduling of human resources involve the most important volume of existing investigations in home healthcare planning. These problems define the assignment of caregivers to patients, as well as the design of caregivers routes to reduce traveling distances, to decrease overtime costs, and to improve the *continuity of care*. Continuity of care guarantees that a patient is most of the time visited by the same caregiver in the whole duration of the care plan. Home healthcare routing and scheduling problems often require the incorporation of several constraints related to the management of caregivers' work regulations, to the matching of caregivers' skills and patients' requirements, and to the satisfaction of patients' and caregivers' preferences. Since the addition of these constraints often makes the modelling and solution of this problem intractable, different authors have proposed heuristic methods such as tabu search algorithms (Hertz & Lahrichi, 2009) and rolling horizon approaches (Bennett & Erera, 2011; Nickel et al., 2012) to efficiently solve practical instances of this problem. Exact approaches have also been developed in Bachouch et al. (2011) and Cappanera & Scutellà (2014) to deal (in an integrated way) with assignment, scheduling, and routing decisions.

Real applications of routing and scheduling of human resources in home healthcare often require the optimization of multiple objectives, as well as the incorporation of uncertainty in demands to obtain robust solutions that react better to changes in demand. In that order of ideas, Duque et al. (2015) and Braekers et al. (2016) propose bi-objective optimization approaches to maximize the quality of service and to minimize the distance travelled by the caregivers. Lanzarone et al. (2012) formulate different scenario-based stochastic programming models to solve the robust nurse-to-patient assignment problem that preserves the continuity of care and balances the operators' workloads. Lanzarone & Matta (2012) use analytical policies to address the nurse-to-patient assignment problem, in which both continuity of care and demand uncertainty are considered. Nguyen et al. (2015) present a variant of a home care problem in which the availability of nurses is uncertain (e.g., nurses might call sick on short notice). To address this problem, the authors propose to use a matheuristic optimization approach for robust nurse-to-patient assignment and nurse scheduling and routing. Carello & Lanzarone (2014) and Lanzarone & Matta (2014) present robust approaches for the nurse-to-patient assignment under continuity of care. In the former work, the authors apply the robust cardinality-constrained approach proposed in Bertsimas & Sim (2004) to incorporate the uncertainty in patients' demands. In the latter work, the authors propose an analytical policy that takes into account the stochasticity of new patient's demand and nurses' workloads. Hewitt et al. (2016) solve the nurse-to-patient assignment problem and develop a solution method to incorporate uncertainty in demand, as future patient requests are often unknown at the time of planning. Cappanera et al. (2018) extend the cardinality-constrained robust approach presented in Cappanera & Scutellà (2014) to include uncertainty in patients' demands in a home care problem integrating assignment, scheduling and routing decisions. The interested reader is referred to Fikar & Hirsch (2017) for a recent survey of current works in home healthcare routing and scheduling.

Contrarily to the routing and scheduling of caregivers, integrated staffing and scheduling problems for home healthcare have been rarely studied in the literature. This problem is highly relevant, as human resources need to be properly managed in order to avoid inefficient visit schedules, treatment delays, and low quality of service (Matta et al., 2014). Two medium-term home healthcare nurse scheduling problems are addressed in Trautsamwieser & Hirsch (2014) and in Wirnitzer et al. (2016). In these works, a given set of nurses is allocated to schedules which are built by including work regulations associated with the allocation of days-off between work stretches, the allocation of rest times between consecutive working days, and the allocation of a maximum working time per day and per week. Trautsamwieser & Hirsch (2014) use a branch-and-price-and-cut solution approach to solve the problem over a one-week planning horizon. Experiments on real-world based instances show that the proposed method helps to significantly reduce the schedule planning time when compared to a manual planning process. Wirnitzer et al. (2016) present a mixed integer programming (MIP) model to address the nurse scheduling problem for longer planning horizons (e.g., one month). Experiments on real-world instances suggest that using the MIP model not only helps to reduce the time to generate the schedules, but also improves the solution quality from the patients and from the nurses point of view. A home healthcare nurse staffing problem with uncertain demands is studied in Rodriguez et al. (2015). The authors propose to use a two-stage stochastic programming approach where first-stage decisions correspond to a global staff dimensioning, while second-stage decisions are related to the allocation of schedules (that do not include work regulations or continuity of care) to nurses with different skills. Results indicate that the proposed approach helps decision-makers with staffing and scheduling decisions before opening a home healthcare service or before hiring a new nurse.

Forecasting patients' demands represents an important step in robust approaches for planning and managing resources in health care. These forecasts can create alerts for the management of patient overflows, they can enhance preventive health care, and when used as an input for planning human resources, they can significantly reduce the associated costs in overstaffing and understaffing (Soyiri & Reidpath, 2013). Several methods have been proposed in the literature to forecast demands and to support healthcare providers in human resource planning before the care execution. These forecasting methods include, among others, Markovian decision models (Lanzarone et al., 2010; Garg et al., 2010), Bayesian models (Argiento et al., 2016), and autoregressive moving average models (Jalalpour et al., 2015). In this paper, we use a *decomposable time series model* (Harvey & Peters, 1990) to forecast the demand since this type of model is relatively easy to implement and to explain to the end user.

The literature review in home healthcare planning reveals that no method has been proposed to integrate caregiver staffing and scheduling when demand is stochastic and when the composition of schedules includes complex work regulations, in particular, existing works show that when rules for the composition of schedules are included in the problem, staffing decisions are not considered since it is assumed that these decisions have been already taken in a previous step of the decision process (Defraeye & Van Nieuwenhuyse, 2016). In a similar way, when staffing decisions are included in the problem, the composition of caregivers' schedules does not consider important work rules such as the allocation of a minimum rest time between consecutive shifts. This paper addresses these gaps in the literature by proposing a model that integrates staff dimensioning with staff scheduling decisions for a medium-term home healthcare problem. Furthermore, the proposed model includes uncertainty in demands and the incorporation of several work rules for the generation of caregivers' schedules, providing solutions that are expected to react in a robust way to variations in demand and that comply with workplace agreements. We remark that although other works have already used context-free grammars to solve personnel scheduling problems under stochastic demand (see Restrepo et al. (2017)), our work is the first one that uses grammars to build schedules over time horizons longer than one day (i.e., a month or longer). Additionally, our work differs from the work in Restrepo et al. (2017) by three other aspects. First, this paper considers staffing decisions, while the work presented in Restrepo et al. (2017) assumes that the number of employees is already given. Second, in this paper, employees can work in different geographic areas, while in Restrepo et al. (2017) all employees are assumed to work in a single place. Third, while this paper uses the reallocation of caregivers to neighboring areas and the possibililty of calling caregivers to work during one of their days-off as the set of recourse actions, the work in Restrepo et al. (2017) uses the allocation of activities and breaks to daily shifts to protect against demand uncertainty.

Next section presents the definition and formulation of the problem studied in this paper.

## 3. Problem Definition and Formulation

The integrated caregiver staffing and scheduling problem for home healthcare considers a territory divided into $|C|$ *geographic areas* or *districts*, each one covering several patients. We assume that each patient is assigned to only one district. The *planning horizon* includes $|D|$ days, where each day $d \in D$ is covered by a set of *working shifts* $S$ characterized by a set of attributes, namely: a start time $b_s$, a day of the week $d_s$ (e.g. Monday, Tuesday,...), a length $l_s$, and a cost $c_s$ that depends on the shifts's length $l_s$ and the day of the week $d_s$. Each district $c \in C$ defines a different type of caregiver $e \in E$ ($E = C$) working in at most one shift $s \in S$ per day. To guarantee the continuity of care for patients, caregiver $e \in E$ should work most of the time in his district. However, caregivers might be temporarily reallocated (at the expense of an additional cost) to a *compatible* district $c \in C$ during shift $s \in S$ to meet unexpected demands. Campbell (2011) showed that schedule flexibility resulting from the reallocation of employees can be more valuable than the perfect information about demand, especially when demand uncertainty is high.

We assume that demands (expressed as the number of visits during day $d \in D$ in district $c \in C$ and shift $s \in S$) are uncertain. Hence, when solving the integrated caregiver staffing and scheduling problem for home healthcare we consider two types of decisions. The first type includes the first-stage decisions, which define the staffing levels (i.e. the number of caregivers to hire), as well as the allocation of individual schedules to each caregiver. The second type incorporates the second-stage decisions, which define the adjustment of caregivers' schedules few days before their execution. These adjustments include the caregivers reallocation to compatible districts, contacting caregivers to work during their day-off, and allowing demand over-covering and under-covering. Because schedules must be available to caregivers at least one month in advance to allow for choices, we assume that the planning horizon is larger than or equal to 4 weeks. At the beginning of this planning horizon staffing and scheduling decisions (first-stage decisions) are made to minimize the sum of the total staffing costs, the expected recourse costs, and the expected over-covering and under-covering costs. Since the actual demand is often revealed one week in advance, the planned schedules are adjusted at the beginning of each week for the following week. These adjustment decisions (second-stage decisions) are applied for each type of shift at each day of the week.

The methodology to solve the problem studied in this paper is divided in three steps. The first step is related to the demand forecasting and scenario generation. The second step involves the definition of caregivers' schedules by means of grammars. The third step uses a two-stage stochastic programming optimization model for caregiver staffing and schedule allocation. The description of these steps is presented next.

### *3.1. Demand Forecasting and Scenario Generation*

The ability of accurately forecast the demand for visits is a fundamental requirement for developing robust decision support tools in home healthcare resource planning. In fact, several strategic and tactical decisions in home healthcare are based on forecasts of demand for resources. For instance, recruitment decisions are mainly driven by forecasts on the amount of visits required by the patients in a given planning horizon. If this demand is accurately predicted, several operational problems such as under-utilization and over-utilization of caregivers can be avoided. On the contrary, inaccurate forecasts threatens the quality of the plans obtained leading to more expensive solutions that could be infeasible for some demand scenarios. In this section, we present a methodology for demand forecasting and scenario generation in home healthcare. We remark that the methods used to forecast and to generate scenarios for the demand are possible approaches, developing and evaluating different methods for these tasks is out of the scope of this work.

### *3.1.1. Demand forecasting*

To estimate the number of patients $b_{dcs}$ to visit during day $d \in D$ in district $c \in C$, and shift $s \in S$, we use a decomposable time series model with three main model components: *growth*, *seasonality*, and *holidays*. These components (included in equation (1)) represent the *growth* function ($g_{dsc}$) which models *non-periodic changes* in the value of the time series, the *periodic changes* function ($s_{dsc}$) modelling weekly or yearly seasonality, and the *effects of holidays* function ($h_{dsc}$) including effects from days such as christmas and new year's day. The error term $\epsilon_{dsc}$ represents irregular changes in demand, which are not accommodated by the time series model.

$$b_{dsc} = g_{dsc} + s_{dsc} + h_{dsc} + \epsilon_{dsc}, \text{ for each } s \in S, \, c \in C \tag{1}$$

Equation (1) is estimated with Facebook Prophet which is an open source library to create quick, accurate and completely automated time series forecasts. This tool uses an *additive regression model* with four components: i) a piecewise linear or logistic growth curve to detect changes in trends by selecting change points from the historical data; ii) a yearly seasonal component modeled using Fourier series; iii) a weekly seasonal component using dummy variables; iv) a user-provided list of relevant holidays. Unlike with ARIMA models, the time series measurements do not need to have a regular period. Hence, there is no need to interpolate missing values to fit. The reader is referred to (Taylor & Letham, 2018) for more information on how Facebook Prophet works.

### *3.1.2. Scenario generation*

In generating the different scenarios for our problem we only consider uncertainty in the number of visits per day, per shift, and per district. Therefore, we assume that the duration of patients' visits and travel times are deterministic parameters which are included in the caregiver capacities (i.e. the number of patients visited per shift). We allow these capacities to vary with the day of the week, with the type of shift, and with the district where the caregiver is working. For instance, the capacity of night shifts is generally lower than the capacity of morning shifts, as patients visited at night need care for longer periods than patients visited in the morning. We assume that the number of visits per day, per shift, and per district is a random variable with finite support. In addition, we define $\Omega_d$ as a set of scenarios for the demand at each day $d \in D$, and $p_d^{(w)} > 0$ as the probability of occurrence of scenario $w \in \Omega_d$. Note that $\sum_{w \in \Omega_d} p_d^{(w)} = 1, \forall \, d \in D$.

The scenarios for the demand are generated with Monte Carlo simulation. We assume that given the estimated values for the mean of demand ($\hat{b}_{dsc}$) and the estimated values for the upper bound ($\hat{b}^u_{dsc}$) of a $(1 - \alpha)$ confidence interval returned by Facebook Prophet after fitting model (1) to the historical data, the standard deviation $\hat{\sigma}_{dsc}$ can be computed with equation (2).

$$\hat{\sigma}_{dsc} = (\hat{b}^u_{dsc} - \hat{b}_{dsc}) \times \frac{\sqrt{n}}{Z_{1-\frac{\alpha}{2}}} \tag{2}$$

Where $Z_{1-\frac{\alpha}{2}}$ is the value for a standard normal variable with a $1 - \frac{\alpha}{2}$ probability to the right, and $n$ denotes the size of the training set used to estimate time series model (1). Once the values for $\hat{\sigma}_{dsc}$ are obtained, we can compute the demand for the number of visits in district $c \in C$ and shift $s \in S$ during day $d \in D$ under scenario $w \in \Omega_d$ as:

$$b^{(w)}_{dsc} = \max\left\{0, \left\lfloor \hat{b}_{dsc} + R * \frac{\hat{\sigma}_{dsc}}{\sqrt{n}} \right\rceil \right\} \tag{3}$$

Where $R$ represents the value of a random variable that follows a standard normal distribution and $\lfloor\ \rceil$ denotes the nearest integer function.

An example on the scenario generation for a given day $d \in D$ is shown in Tables 1 and 2. Table 1 presents for each combination of districts and shifts (denoted as $d_0$, $d_1$, $d_2$, and $d_3$ for the districts, and $a_4$, $m_4$, $m_8$, and $n_{10}$ for the shifts) the values for the forecasted mean demand ($\hat{b}$), the values for the lower bound and upper bound ($\hat{b}^l$, $\hat{b}^u$) of a 90% confidence interval for the forecasted demand, the values for the actual value of the demand ($b$), and the values for the possible values for the demand (list) with their corresponding frequency (count), after running 500 simulations. Table 2 shows a sample of 10 scenarios from the 500 scenarios generated. Each column from this table presents the demand values (number of visits) during day $d$ for each combination of districts and shifts.

| district_shift | $b$ | $\hat{b}^l$ | $\hat{b}^u$ | $\hat{b}$ | list | count |
|---|---|---|---|---|---|---|
| $d_0\_m_4$ | 1 | 1 | 2 | 1 | [1] | [500] |
| $d_0\_m_8$ | 2 | 1 | 3 | 1 | [1, 2, 0, 3] | [314, 157, 25, 4] |
| $d_0\_n_{10}$ | 2 | 1 | 2 | 1 | [1, 2, 0] | [466, 30, 4] |
| $d_1\_a_4$ | 7 | 5 | 9 | 7 | [7, 6, 5, 8, 9, 4, 10, 3] | [151, 134, 87, 77, 23, 22, 5, 1] |
| $d_1\_m_4$ | 5 | 2 | 8 | 4 | [4, 5, 6, 3, 2, 7, 1, 8, 0, 9] | [122, 110, 84, 78, 49, 31, 14, 9, 2, 1] |
| $d_1\_m_8$ | 14 | 12 | 17 | 14 | [13, 14, 15, 12, 16, 11, 17, 18, 10] | [126, 122, 106, 56, 47, 22, 13, 5, 3] |
| $d_1\_n_{10}$ | 9 | 6 | 10 | 7 | [8, 9, 7, 6, 10, 11, 5, 12] | [182, 128, 116, 33, 32, 5, 3, 1] |
| $d_2\_m_4$ | 2 | 1 | 3 | 1 | [1, 2, 0, 3] | [304, 154, 40, 2] |
| $d_2\_m_8$ | 2 | 1 | 2 | 1 | [1, 2, 0] | [460, 34, 6] |
| $d_2\_n_{10}$ | 2 | 1 | 2 | 1 | [1, 2, 0] | [420, 78, 2] |
| $d_3\_a_4$ | 3 | 1 | 5 | 2 | [3, 2, 4, 1, 5, 0] | [179, 176, 67, 59, 10, 9] |
| $d_3\_m_4$ | 4 | 3 | 6 | 4 | [4, 3, 5, 2, 6, 7, 1] | [204, 129, 116, 26, 23, 1, 1] |
| $d_3\_m_8$ | 4 | 2 | 6 | 4 | [3, 4, 2, 5, 1, 6, 7] | [192, 147, 88, 55, 9, 8, 1] |
| $d_3\_n_{10}$ | 4 | 3 | 6 | 4 | [4, 3, 5, 2, 6, 1] | [212, 151, 93, 30, 11, 3] |

Table 1: Results for the demand forecasting and Monte Carlo simulation.

| | Scenario | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| district_shift | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $d_0\_m_4$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $d_0\_m_8$ | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 1 |
| $d_0\_n_{10}$ | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 |
| $d_1\_a_4$ | 7 | 5 | 5 | 6 | 8 | 8 | 5 | 7 | 6 | 7 |
| $d_1\_m_4$ | 3 | 3 | 5 | 5 | 5 | 4 | 6 | 6 | 6 | 2 |
| $d_1\_m_8$ | 11 | 15 | 13 | 13 | 14 | 13 | 13 | 11 | 16 | 11 |
| $d_1\_n_{10}$ | 7 | 11 | 9 | 9 | 5 | 7 | 6 | 8 | 8 | 6 |
| $d_2\_m_4$ | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 |
| $d_2\_m_8$ | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| $d_2\_n_{10}$ | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 |
| $d_3\_a_4$ | 3 | 3 | 3 | 2 | 2 | 2 | 3 | 4 | 3 | 1 |
| $d_3\_m_4$ | 4 | 5 | 4 | 4 | 3 | 3 | 5 | 5 | 3 | 3 |
| $d_3\_m_8$ | 3 | 3 | 5 | 3 | 3 | 5 | 3 | 2 | 3 | 4 |
| $d_3\_n_{10}$ | 3 | 5 | 4 | 5 | 3 | 4 | 4 | 3 | 6 | 3 |

Table 2: Example of 10 scenarios for a given day in the planning horizon.

### *3.2. Grammars*

A *context-free grammar* is a set of recursive rewriting rules (or productions) used to generate patterns of strings, or (in the case of personnel scheduling) to generate schedules or daily shifts. Context-free grammars have been successfully used in the context of personnel scheduling. Applications include the solution of multi-activity and multi-task shift scheduling problems (Côté et al., 2013; Boyer et al., 2012) and multi-activity tour scheduling problems (Restrepo et al., 2017, 2016).

A context-free grammar consists of a four-tuple $G = \langle \Sigma, N, \mathcal{S}, P \rangle$, where $\Sigma$ is an alphabet of characters called the *terminal symbols*, $N$ is a set of *non-terminal symbols*, $\mathcal{S} \in N$ is the starting symbol, and $P$ is a set of *productions* represented as $A \rightarrow \alpha$, where $A \in N$ is a non-terminal symbol and $\alpha$ is a sequence of terminal and non-terminal symbols. The productions of a grammar are used to generate new symbol sequences until all non-terminal symbols have been replaced by terminal symbols. A *context-free language* is the set of sequences accepted by a context-free grammar.

A *parse tree* is a tree where each inner-node is labeled with a non-terminal symbol and each leaf is labeled with a terminal symbol. A grammar recognizes a sequence if and only if there exists a parse tree where the leaves, when listed from left to right, reproduce the sequence. A *DAG* $\Gamma$ is a *directed acyclic graph* that embeds all parse trees associated with words of a given length $n$ recognized by a grammar. The DAG $\Gamma$ has an and/or structure where the and-nodes represent productions from $P$ and or-nodes represent non-terminals from $N$ and letters from $\Sigma$. An and-node is true if all of its children are true. An or-node is true if one of its children is true. The root node is true if the grammar accepts the sequence encoded by the leaves. The DAG $\Gamma$ is built with a procedure proposed in Quimper & Walsh (2007) using bottom-up parsing and dynamic programming.

In employee scheduling, the use of grammars allows one to include work rules regarding the definition of *work stretches* and *rest stretches* in an easy way. Thus, feasible schedules can be represented as words in a context-free language. Specifically, for the problem addressed in this paper we use grammars to:

- Generate work stretches representing sequences of work spanning a minimum and a maximum number of days.
- Generate rest stretches denoting sequences of days-off spanning a minimum and a maximum number of days.
- Define a minimum and a maximum consecutive number of morning, afternoon, and night shifts within a work stretch. For instance, a given work stretch cannot have more than 3 night shifts in a row.
- Forbid infeasible transitions between shifts by associating costs to productions. For instance, a night shift cannot be followed by a morning shift.
- Allocate a rest stretch between two work stretches.

**Example 1**

*Consider the following grammar for an employee scheduling problem where the planning horizon consists of five days, work stretches have a length of three consecutive days, and days-off can be allocated in consecutive or nonconsecutive days:*

$G = (\Sigma = (w, r), N = (\mathcal{S}, F, Q, W, R), P, \mathcal{S}),$

*Where productions* $P$ *are:* $\mathcal{S} \rightarrow RF|FR|QR$, $F_{[3,3]} \rightarrow WW$, $W \rightarrow WW|w$, $Q \rightarrow RF$, $R \rightarrow RR|r$ *and symbol* $|$ *specifies the choice of production. Letter* $w$ *represents the allocation of a working shift and letter* $r$ *represents the allocation of a day-off.* $P_{[min,\ max]}$ *restricts the subsequences generated by production* $P$ *to a length between a minimum and maximum number of days.*

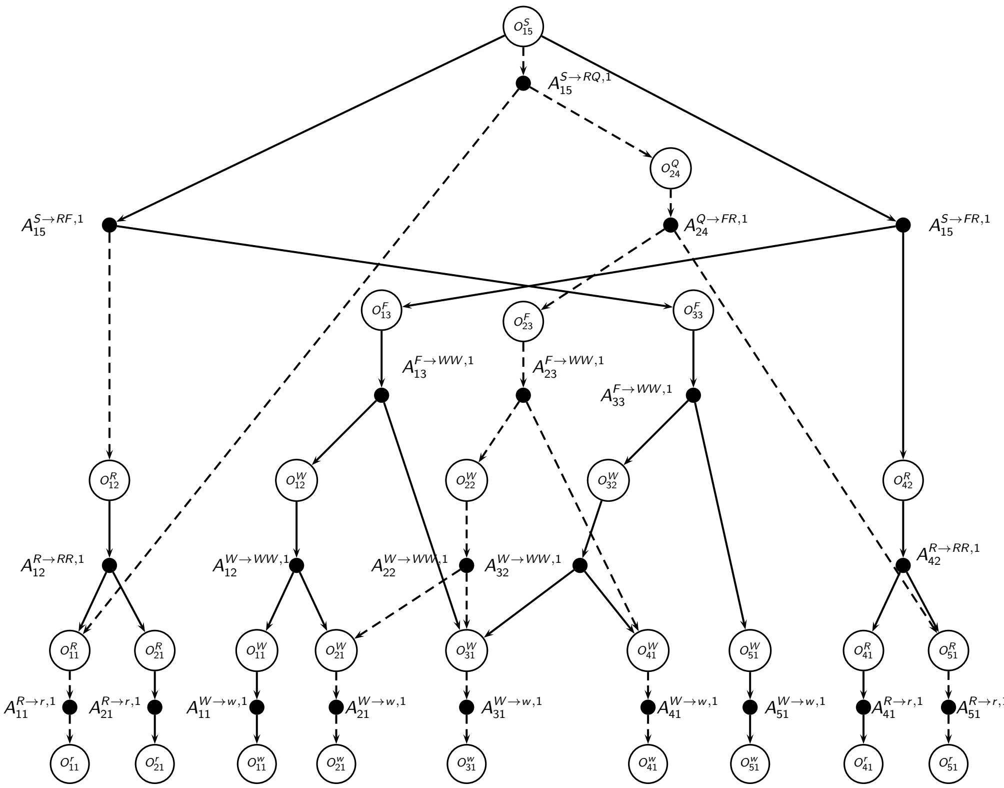
*In this grammar, production* $F_{[3,3]} \rightarrow WW$ *generates two non-terminal symbols* $W$*, meaning that the schedule will include a work stretch of exactly three days. Production* $Q \rightarrow RF$ *generates two non-terminal symbols* $R$ *and* $F$*, meaning that the schedule will start with a rest stretch and then it will include a work stretch of exactly three days. Production* $R \rightarrow RR$ *generates two non-terminal symbols* $R$*, meaning that the schedule will include a rest stretch. Productions* $W \rightarrow w$ *and* $R \rightarrow r$ *generate terminal symbols associated with the allocation of a shift and with the allocation of a day-off to the schedule of an employee, respectively. The* *last three productions are $\mathcal{S} \to RF$, $\mathcal{S} \to FR$, and $\mathcal{S} \to QR$. The first production generates a schedule starting with two days-off followed by a work stretch. The second production generates a schedule starting with a work stretch followed by two days-off. The last production generates a schedule starting with one day-off, followed by a work stretch, to finish with one day-off. The three words recognized as valid schedules by the grammar in this example are $rrwww$, $wwwrr$, and $rwwwr$.*

Let $O_{dl}^{\pi}$ be the or-nodes associated with $\pi \in N \cup \Sigma$ (i.e. with non-terminals from $N$ or letters from $\Sigma$) that generate a subsequence from day $d$ of length $l$. Note that if $\pi \in \Sigma$, the node is a leaf and $l$ is equal to one. On the contrary, if $\pi \in N$ the node represents a non-terminal symbol and $l \geq 1$. $A_{dl}^{\Pi,k}$ is the $k^{th}$ and-node representing production $\Pi \in P$ generating a subsequence from day $d$ of length $l$. There are as many $A_{dl}^{\Pi,k}$ nodes as there are ways of using $\Pi$ to generate a sequence of length $l$ from day $d$. As previously mentioned, undesired productions (i.e. transitions between a night shift and a morning shift) are penalized by a cost denoted as $c_{dl}^{\Pi,k}$. The sets of or-nodes, and-nodes, and leaves of DAG $\Gamma$ are denoted by $O$, $A$, and $L$, respectively. The root node is described by $O_{1n}^{\mathcal{S}}$ and its children by $A_{1n}^{\Pi,k}$. The children of or-node $O_{dl}^{\pi}$ are represented by $ch(O_{dl}^{\pi})$ and its parents by $par(O_{dl}^{\pi})$. Similarly, the children of and-node $A_{dl}^{\Pi,k}$ are represented by $ch(A_{dl}^{\Pi,k})$ and its parents by $par(A_{dl}^{\Pi,k})$. For more details on the use of grammars in employee scheduling we refer the reader to Côté et al. (2011).

Figure 1 shows the DAG $\Gamma$ associated with the grammar from Example 1. Observe that this figure includes three parse trees, each one representing one word (schedule) recognized by the grammar. As an example we present in dashed lines the parse tree generating schedule $rwwwr$.

The works of Restrepo et al. (2017) and Côté et al. (2011) on anonymous tour scheduling problems with multiple activities are examples of the use of context-free grammars to represent the work rules involved in the composition of shifts. In both works, the authors present implicit grammar-based integer programming models where the word length $n$ corresponds to the number of periods in the planning horizon, the set of work activities corresponds to letters in the alphabet $\Sigma$, and each employee is allowed to work in any work activity. In the model, the logical clauses associated with $\Gamma$ are translated into linear constraints on integer variables. Each and-node $A$ and each leaf $L$ in $\Gamma$ are represented by an integer variable denoting the number of employees assigned to a specific subsequence of work. Since this grammar-based model efficiently encapsulates the constraints for the generation of the schedules, it is used as a component in the formulation of the two-stage stochastic problem presented next.

Figure 1: DAG Γ on schedules of length five.




### *3.3. Two-Stage Stochastic Optimization Model*

The formulation of the two-stage stochastic programming model requires a previous definition of the grammars and DAGs Γ containing specific work regulations for the composition of valid caregiver schedules. Since work regulations could vary depending on the type of caregiver, we define a different grammar and a different DAG $\Gamma^e$ for each $e \in E$. The notation used for the formulation of the problem is as follows:

*Parameters:*

– $\kappa^e_{dsc}$: number of visits a caregiver of type $e \in E$ working on shift $s \in S$ can perform in district $c \in C$ during day $d \in D$;

– $c^e_{ds}$: non-negative cost associated with one caregiver of type $e \in E$ working on shift $s \in S$ during day $d \in D$;

– $c^{\Pi,k,e}_{dl}$: non-negative cost associated with the $k^{th}$ and-node representing production Π from $\Gamma^e$, producing a sequence from day $d \in D$ of length $l$ for caregiver $e \in E$;

– $\hat{b}_{dsc}$: mean demand for the number of visits in district $c \in C$ and shift $s \in S$ during day $d \in D$;

– $b_{dsc}^{(w)}$: demand for the number of visits in district $c \in C$ and shift $s \in S$ during day $d \in D$ under scenario $w \in \Omega_d$;

– $c_{dsc}^{+}$, $c_{dcs}^{-}$: non-negative demand over-covering and under-covering costs for district $c \in C$ and shift $s \in S$ during day $d \in D$, respectively;

– $t_{dsc}^{e}$: non-negative transition cost associated with the reallocation one caregiver of type $e \in E$ to district $c \in C$ during day $d$ and shift $s \in S$;

– $r_{ds}^{e}$: non-negative cost associated with assigning shift $s \in S$ to a caregiver of type $e \in E$ during its rest day $d \in D$;

– $\delta_{sc}^{e}$: binary parameter that takes value 1 if caregiver $e \in E$ admits a reallocation to district $c \in C$ during shift $s \in S$, and it assumes value 0 otherwise.

*Decision variables:*

– $u^e$: variable that denotes the number of caregivers of type $e \in E$ to hire;

– $v_{dl}^{\Pi,k,e}$: variable that denotes the number of caregivers of type $e \in E$ assigned to the $k^{th}$ and-node representing production $\Pi$ from $\Gamma^e$ producing a sequence from day $d \in D$ of length $l$;

– $y_{ds}^{e}$: variable that denotes the number of caregivers of type $e \in E$ working on shift $s \in S$ during day $d \in D$ (equivalent to the number of caregivers of type $e \in E$ assigned to leaf $O_{d1}^{s,e}$);

– $y_{dr}^{e}$: variable that denotes the number of caregivers of type $e \in E$ having rest during day $d \in D$ (equivalent to the number of caregivers of type $e \in E$ assigned to leaf $O_{d1}^{r,e}$);

– $x_{dsc}^{e(w)}$: variable that denotes the number of caregivers of type $e \in E$ assigned to work in district $c \in C$ and shift $s \in S$ during day $d \in D$ under scenario $w \in \Omega_d$;

– $z_{ds}^{e(w)}$: variable that denotes the number of caregivers of type $e \in E$ assigned to work during a day-off on shift $s \in S$ during day $d$ and scenario $w \in \Omega_d$;

– $s_{dsc}^{+(w)}$ and $s_{dsc}^{-(w)}$: slack variables denoting demand over-covering and under-covering in district $c \in C$ and shift $s \in S$ during day $d \in D$ under scenario $w \in \Omega_d$, respectively.

The formulation for the stochastic caregiver staffing and scheduling problem, is as follows.

$$\min \sum_{d\in D}\sum_{s\in S}\sum_{e\in E} c^{e}_{ds} y^{e}_{ds} + \sum_{d\in D}\sum_{e\in E}\sum_{A^{\Pi,k,e}_{dl}\in A^{e}} c^{\Pi,k,e}_{dl} v^{\Pi,k,e}_{dl} + \mathcal{Q}(\mathbf{y}) \tag{4}$$

$$y^{e}_{ds} = \sum_{A^{\Pi,1,e}_{d1}\in par(O^{s,e}_{d1})} v^{\Pi,1,e}_{d1},\ \forall\, d\in D,\ e\in E,\ s\in S, \tag{5}$$

$$y^{e}_{dr} = \sum_{A^{\Pi,1,e}_{d1}\in par(O^{r,e}_{d1})} v^{\Pi,1,e}_{d1},\ \forall\, d\in D,\ e\in E, \tag{6}$$

$$u^{e} = \sum_{A^{\Pi,k,e}_{1n}\in ch(O^{\mathcal{S},e}_{1n})} v^{\Pi,k,e}_{1n},\ \forall\, e\in E, \tag{7}$$

$$\sum_{A^{\Pi,k,e}_{dl}\in ch(O^{\pi,e}_{dl})} v^{\Pi,k,e}_{dl} = \sum_{A^{\Pi,k,e}_{dl}\in par(O^{\pi,e}_{dl})} v^{\Pi,k,e}_{dl},\quad \forall\, e\in E,\ O^{\pi,e}_{dl}\in O^{e}\setminus\{O^{\mathcal{S},e}_{1n}\cup L^{e}\}, \tag{8}$$

$$u^{e}\geq 0 \text{ and integer},\ \forall\, e\in E, \tag{9}$$

$$v^{\Pi,k,e}_{dl}\geq 0 \text{ and integer},\ \forall\, d\in D,\ e\in E,\ A^{\Pi,k,e}_{dl}\in A^{e}, \tag{10}$$

$$y^{e}_{ds}\geq 0 \text{ and integer},\ \forall\, d\in D,\ e\in E,\ s\in S, \tag{11}$$

$$y^{e}_{dr}\geq 0 \text{ and integer},\ \forall\, d\in D,\ e\in E. \tag{12}$$

The objective of model (4)-(12) is to minimize the total staffing cost (i.e. allocation of working shifts to caregivers), the penalization for certain transitions between shifts (i.e. transition from night shifts to morning shifts), and the *expected recourse function* $\mathcal{Q}(\mathbf{y})$. Constraints (5)-(6) set the value of variables $y^{e}_{ds}$ and $y^{e}_{dr}$ as the summation of the value of the parents of leaf nodes $O^{s,e}_{d1}$ and $O^{r,e}_{d1}$, respectively. Constraints (7) define the number of caregivers of type $e\in E$ to hire. Constraints (8) guarantee, for every or-node in $\Gamma^{e}$, $e\in E$ excluding the root node $O^{\mathcal{S},e}_{1n}$ and the leaves $L^{e}$, that the summation of the value of its children is the same as the summation of the value of its parents. Constraints (8) can be seen as flow conservation equations where or-nodes $O^{\pi,e}_{dl}$ represent "transition nodes". The constraints for those transition nodes guarantee that if $m$ caregivers of type $e$ are allocated to the productions generating the subsequence associated with node $O^{\pi,e}_{dl}$, those $m$ caregivers have to be distributed along all the possible ways to use $\pi$ to generate a sequence of length $l$ from position $d$ ($ch(O^{\pi,e}_{dl})$). Consider the following example using the DAG $\Gamma$ from Figure 1. Assume that three caregivers are assigned to and-node $A^{W\to WW,1}_{22}$ (represented by variable $v^{W\to WW,1}_{22}$) and that one employee is assigned to and-node $A^{W\to WW,1}_{12}$ (represented by variable $v^{W\to WW,1}_{12}$). Since these two and-nodes have one child in common (i.e. or-node $O^{W}_{21}$) the number of employees allocated to $O^{W}_{21}$ is four. Now, since or-node $O^{W}_{21}$ has one child ($A^{W\to w,1}_{21}$) these four employees must be allocated to a working shift during day 2.

The expected recourse function $\mathcal{Q}(\mathbf{y})$ is denoted by $\mathcal{Q}(\mathbf{y}) \equiv \mathbb{E}_{\xi}[\mathcal{Q}(\mathbf{y}, \xi)]$. The *recourse function* $\mathcal{Q}(\mathbf{y}, \xi_d(w))$ for a given realization $w$ of $\xi$ and fixed values for the allocation of caregivers to shifts and days-off ($\bar{y}^e_{ds}$, $\bar{y}^e_{d-1r}$, $\bar{y}^e_{dr}$, and $\bar{y}^e_{d+1r}$) is represented by:

$$\min \sum_{e \in E} \sum_{s \in S} \sum_{c \in C} t^e_{dsc} x^{e(w)}_{dsc} + \sum_{e \in E} \sum_{s \in S} r^e_{ds} z^{e(w)}_{ds} + \sum_{s \in S} \sum_{c \in C} \left( c^+_{dsc} s^{+(w)}_{dsc} + c^-_{dsc} s^{-(w)}_{dsc} \right) \tag{13}$$

$$\sum_{c \in C} \delta^e_{sc} x^{e(w)}_{dsc} = (z^{e(w)}_{ds} + \bar{y}^e_{ds}), \ \forall\, s \in S, \, e \in E, \tag{14}$$

$$\sum_{s \in S} z^{e(w)}_{ds} \leq \bar{y}^e_{d-1r}, \ \forall\, e \in E, \tag{15}$$

$$\sum_{s \in S} z^{e(w)}_{ds} \leq \bar{y}^e_{dr}, \ \forall\, e \in E, \tag{16}$$

$$\sum_{s \in S} z^{e(w)}_{ds} \leq \bar{y}^e_{d+1r}, \ \forall\, e \in E, \tag{17}$$

$$\sum_{e \in E} \kappa^e_{dsc} x^{e(w)}_{dsc} - s^{+(w)}_{dsc} + s^{-(w)}_{dsc} = b^{(w)}_{dsc}, \ \forall\, s \in S, \, c \in C, \tag{18}$$

$$x^{e(w)}_{dsc} \geq 0 \text{ and integer}, \ \forall\, e \in E, \, s \in S, \, c \in C, \tag{19}$$

$$z^{e(w)}_{ds} \geq 0 \text{ and integer}, \ \forall\, e \in E, \, s \in S, \tag{20}$$

$$s^{+(w)}_{dsc}, \, s^{-(w)}_{dsc} \geq 0, \ \forall\, s \in S, \, c \in C. \tag{21}$$

The objective of model (13)-(21) is to minimize the reallocation costs, the costs of contacting caregivers to work on a day-off, and the penalization for demand over-covering and under-covering. Constraints (14) define the reallocation of caregivers of type $e \in E$ working on shift $s \in S$ to compatible districts. Constraints (15)-(17) set the valid conditions to contact caregivers to work on a day-off. That is, if an employee is having three days-off in a row only the day-off in the middle of the rest stretch can be assigned to a working shift. Constraints (18) ensure that the total number of caregivers working on day $d \in D$, shift $s \in S$, and district $c \in C$ is equal to the demand subject to some adjustments related to demand under-covering and over-covering. Constraints (19)-(21) set the non-negativity and integrality of variables $x^{e(w)}_{dsc}$ and $z^{e(w)}_{ds}$, and the non-negativity of variables $s^{+(w)}_{dsc}$ and $s^{-(w)}_{dsc}$.

Since we assumed that the number of visits per day, per shift, and per district is a random variable with finite support, where $\Omega_d$ is the set of scenarios for the demand at each day and $p^{(w)}_d > 0$ is the probability of occurrence of scenario $w \in \Omega_d$, the expected recourse function $\mathcal{Q}(\mathbf{y})$ can be expressed as:

$$\mathcal{Q}(\mathbf{y}) \equiv \sum_{d \in D} \mathbb{E}_{\xi}[\mathcal{Q}(\mathbf{y}, \xi_d)] \equiv \sum_{d \in D} \sum_{w \in \Omega_d} p^{(w)}_d \mathcal{Q}(\mathbf{y}, \xi_d(w)) \tag{22}$$

With this result, recourse functions (13)-(21) can be incorporated in (4)-(12) to obtain an *deterministic equivalent problem* given by:

$$
\begin{aligned}
f(\mathcal{Z}) = \min & \sum_{d \in D} \sum_{s \in S} \sum_{e \in E} c^e_{ds} y^e_{ds} + \sum_{d \in D} \sum_{e \in E} \sum_{A^{\Pi,k,e}_{dl} \in A^e} c^{\Pi,k,e}_{dl} v^{\Pi,k,e}_{dl} + \\
& \sum_{d \in D} \sum_{w \in \Omega_d} p^{(w)}_d \Big( \sum_{e \in E} \sum_{s \in S} \sum_{c \in C} t^e_{dsc} x^{e(w)}_{dsc} + \sum_{e \in E} \sum_{s \in S} r^e_{ds} z^{e(w)}_{ds} \Big) \\
& \sum_{d \in D} \sum_{w \in \Omega_d} p^{(w)}_d \Big( \sum_{s \in S} \sum_{c \in C} c^+_{dsc} s^{+(w)}_{dsc} + c^-_{dsc} s^{-(w)}_{dsc} \Big) \\
& (5) - (12) \text{ and} \\
& (14) - (21), \forall\, d \in D, w \in \Omega_d.
\end{aligned}
$$

Observe that model $\mathcal{Z}$ could involve a large number of variables and constraints, especially when the number of days in the planning horizon is large. However, since context-free grammars allow to handle multiple shift types and to represent complex work regulations in an implicit (compact) way, and since the size of the model does not depend on the number of caregivers to hire at each district, model $\mathcal{Z}$ can be efficiently solved for large instances without the need of decomposition methods.

## 4. Computational Experiments

In this section, we test the proposed approach on real-world instances from a home healthcare agency working with AlayaCare. First, we present information related to the agency's operations and to the rules for schedule generation. Second, we describe the procedure adopted for the generation of the instances and present the size of these instances. Third, we report and analyze the computational results and present a discussion on the practical aspects and managerial insights of the proposed approach.

The computational experiments were performed on a Linux operating system, 16 GB of RAM and 1 processor Intel Xeon X5675 running at 3.07GHz. The algorithm to solve the problem was implemented in C++. The deterministic equivalent problem $\mathcal{Z}$ was solved with CPLEX version 12.7.0.0. The time limit to solve each instance is proportional to the length of the planning horizon. For example, if a given instance is defined over 4 weeks, the time limit is set to 2 hours. Similarly, if a given instance is defined over 12 weeks, the time limit is set to 6 hours. A relative gap tolerance of 0.01 was set as a stopping criterion for solving the MILPs with CPLEX.

### *4.1. Operations and Schedule Generation*

- Operations: The test instances are generated based on 8-month historical data from operations of one private agency operating in Greater Toronto Area. This region is divided in four districts (i.e. $|C| = 4$). The agency operates in these districts 24 hours per day from Monday to Sunday. We only consider the staffing and scheduling of personal support workers, as they represent the largest portion of employees in the agency (70% of the total number of caregivers). Based on the agency's operations we defined four types of shifts: morning shifts of type 1 (denoted as $m_8$) starting at 7:00 with an 8-hour length; morning shifts of type 2 (denoted as $m_4$) starting at 10:00 with a 4-hour length; afternoon shifts (denoted as $a_4$) starting at 14:00 with a 4-hour length; and night shifts (denoted as $n_{10}$) starting at 18:00 with a 10-hour length. We assume that the base cost of each working time interval is 1\$ and that the shift allocation cost depends on the shift length, as well as on the day covered (weekend shifts are more expensive than weekday shifts). Because one of the objectives of the agency is to increase the service level, demand under-covering costs are set to a large value equal to the cost of each visit ($c^e_{ds}/\kappa^e_{ds}$) multiplied by 10. Similarly, the costs for the demand over-covering are equivalent to the cost of each visit ($c^e_{ds}/\kappa^e_{ds}$) multiplied by 0.5. The values for these costs, for the capacities of shifts, as well as other parameters characterizing each type of shift are presented in Table 3. Observe that the costs presented in this table do not consider a 20% surcharge for weekend days. In addition, the cost of contacting a caregiver to work on a day-off is $r^e_{ds} = c^e_{ds} * 2$, the surcharge for allowing transitions between districts is $t^e_{dsc} = 10\%$, and the transition costs between forbidden shifts is equal to $c^{\Pi,k,e}_{dl} = 1000\$$.

| Parameter | Shift | | | |
|---|---|---|---|---|
| | $m_8$ | $m_4$ | $a_4$ | $n_{10}$ |
| Shift allocation cost $c^e_{ds}$ (\$) | 8 | 4 | 4 | 10 |
| Under-covering cost $c^-_{dsc}$ (\$) | 40 | 40 | 20 | 100 |
| Over-covering cost $c^+_{dsc}$ (\$) | 2 | 2 | 1 | 5 |
| Capacity $\kappa^e_{ds}$ (number of visits) | 2 | 1 | 2 | 1 |
| Max_days | 6 | 4 | 4 | 3 |

Table 3: Costs and capacity values for each type of shift.

| District | District | | | |
|---|---|---|---|---|
| | $d_0$ | $d_1$ | $d_2$ | $d_3$ |
| $d_0$ | 1 | 0 | 0 | 0 |
| $d_1$ | 0 | 1 | 1 | 0 |
| $d_2$ | 0 | 1 | 1 | 1 |
| $d_3$ | 0 | 0 | 1 | 1 |

Table 4: District compatibilities.

- Schedule composition: The work regulations for the schedule composition are the following

  1. The minimum and maximum number of days in each work stretch are 4 and 6, respectively.
  2. The minimum and maximum number of days in each rest stretch are 1 and 3, respectively.
  3. A rest stretch is necessary between two work stretches.
  4. Each shift has a maximum number of consecutive times it can appear in a work sequence. These values are presented in row Max_days of Table 3. For instance, a work stretch cannot contain more than 3 night shifts in a row.

- Grammar: Let $w_s$ be a terminal symbol that defines working on shift $s \in S$. Let $r$ be a terminal symbol that represents a rest period. Let $F$ and $R$ be non-terminal symbols representing work and rest stretches, respectively. Let $s_u$ be the maximum number of consecutive times shift $s$ can appear in a work sequence. In productions $\Pi \in P$, $\Pi \xrightarrow{ctr}_{[\text{min, max}]}$ restricts the subsequences generated by a given production to a length between a minimum and maximum number of days, and $ctr$ denotes a cost associated with the production. The grammar and the productions that define valid schedules for caregiver of type $e \in E$ during a planning horizon of four weeks are as follows:

$$
\begin{aligned}
G^e =& (\Sigma = (w_s \;\; \forall s \in S, r),\\
& N = (\mathcal{S}, F, H, J_s, J_s', J_s^2, J_s^3 \, \forall s \in S, R), P, \mathcal{S}),\\
& \mathcal{S}_{[28,28]} \to RHR|RH|HR,\\
& H \to FRFRFRF,\\
& F_{[4,6]} \overset{c_{tr}}{\to} J_s J_s',\ \forall s \in S; F_{[4,6]} \to J_s^2 J_s^3,\ \forall s \in S,\\
& J_s' \overset{c_{tr}}{\to} J_{s'} J_{s'}',\ \forall s \in S,\ \forall s' \in S \backslash \{s\}; J_s' \to J_{s'}^2 J_{s'}^3,\ \forall s \in S,\ \forall s' \in S \backslash \{s\};\\
& J_{s[0,s_u]} \to J_s^2 J_s^3,\ \forall s \in S,\\
& J_s^2 \to J_s^2 J_s^3,\ \forall s \in S; J_s^3 \to w_s,\ \forall s \in S;\\
& R_{[1,3]} \to rR; R \to r.
\end{aligned}
$$

### *4.2. Instances Generation and Size of Problems*

Three instances spanning planning horizons from 4 to 12 weeks and including 500 demand scenarios were generated to test our model. These instances were built with the procedures presented in Sections 3.1.1 and 3.1.2. Table 5 presents for each instance (denoted as I1, I2, and I3), the number of or-nodes, the number of and-nodes, and the number of leaves in each DAG $\Gamma^e, \forall\, e \in E$. This table also presents the number of variables and the number of constraints for the first-stage and second-stage components of model $\mathcal{Z}$. Note that the size of the model is not proportional to the number of caregivers, as the employee dimension is included in the model in an implicit way.

| | Instance | | |
|---|---|---|---|
| | I1 (4 weeks) | I2 (8 weeks) | I3 (12 weeks) |
| Or-nodes | 1,551 | 3,102 | 6,204 |
| And-nodes | 3,614 | 7,228 | 14,456 |
| Leaves | 140 | 280 | 560 |
| First-stage constraints | 6,208 | 12,416 | 24,832 |
| Second-stage constraints | 305,000 | 610,000 | 1,120,000 |
| First-stage integer variables | 15,016 | 30,032 | 60,064 |
| Second-stage integer variables | 560,000 | 1,120,000 | 2,240,000 |
| Second-stage continuos variables | 224,000 | 448,000 | 896,000 |

Table 5: Instances size.

Since the size and complexity of problem $\mathcal{Z}$ increase with the number of scenarios, we decided to perform an analysis to evaluate how staffing and scheduling decisions (including a fraction of the scenarios) accommodate the real demand, and how these decisions react when they are evaluated on all generated scenarios (500). Specifically, for each instance we first solve problem $\mathcal{Z}$ with a fraction of the scenarios (e.g., 50 out of 500) to get the optimal solution for variables $y^e_{ds}$ and $y^e_{dr}$. These optimal values are fixed in second-stage problems (13)-(21), which are solved with the actual demand information and with all 500 scenarios. Table 6 presents the results for this evaluation on problem $\mathcal{Z}$ including reallocation of caregivers (Realloc. = 1) and contacting caregivers to work on a day-off (RestToW = 1). For each type of instance (Instance) and number of scenarios (Scen.), we present the status of the solution (Status), the recourse cost when the schedule is evaluated with the real demand (Real.C), and the recourse cost when the schedule is evaluated with 500 scenarios (Recour.C). The percentage increase in these two costs (Real.C and Recour.C) by using a fraction of the scenarios is presented in columns %I.Real.C and %I.Recour.C. This percentage is computed as: $\%I = 100 \times \frac{Cost - base\_cost}{base\_cost}$, where $Cost$ represents the value for Real.C and Recour.C, and $base\_cost$ denotes the recourse cost obtained after solving problem $\mathcal{Z}$ with the largest possible number of scenarios (300 for I1 instances, 125 for I2 instances, and 25 for I3 instances).

| Instance | Scen. | Realloc. | RestToW | Status | Real.C ($) | Recour.C ($) | %I.Real.C | %I.Recour.C |
|---|---|---|---|---|---|---|---|---|
| I1 | 5 | 1 | 1 | Optimal | 1,679.6 | 2,297.34 | 14.63% | 28.45% |
| I1 | 25 | 1 | 1 | Optimal | 1,505.8 | 1,807.29 | 2.77% | 1.05% |
| I1 | 100 | 1 | 1 | Optimal | 1,464.6 | 1,801.97 | -0.04% | 0.75% |
| I1 | 150 | 1 | 1 | Optimal | 1,508.6 | 1,800.15 | 2.96% | 0.65% |
| I1 | 200 | 1 | 1 | Optimal | 1,638.8 | 1,827.56 | 11.85% | 2.18% |
| I1 | 250 | 1 | 1 | Optimal | 1,506.8 | 1,824.26 | 2.84% | 2.0% |
| I1 | 300 | 1 | 1 | Optimal | 1,428 | 1,794.67 | -2.54% | 0.34% |
| I1 | 350 | 1 | 1 | Optimal | 1,567 | 1,835.46 | 6.95% | 2.62% |
| I1 | 400 | 1 | 1 | Optimal | 1,503.8 | 1,827.55 | 2.63% | 2.18% |
| I1 | 450 | 1 | 1 | Optimal | 1,465.2 | 1,788.53 | 0.0% | 0.0% |
| I2 | 5 | 1 | 1 | Optimal | 3,829 | 4,048.22 | 24.77% | 11.92% |
| I2 | 25 | 1 | 1 | Optimal | 3,106 | 3,832.19 | 1.21% | 5.95% |
| I2 | 50 | 1 | 1 | Optimal | 3,212 | 3,786.17 | 4.67% | 4.67% |
| I2 | 100 | 1 | 1 | Optimal | 3,100.8 | 3,685.55 | 1.04% | 1.89% |
| I2 | 150 | 1 | 1 | Optimal | 3,062.2 | 3,618.67 | -0.22% | 0.04% |
| I2 | 200 | 1 | 1 | Optimal | 3,068.8 | 3,617.14 | 0% | 0.0% |
| I3 | 5 | 1 | 1 | Optimal | 6,187.4 | 6,315.87 | 6.68% | 9.18% |
| I3 | 25 | 1 | 1 | Optimal | 5,904.4 | 5,827.59 | 1.8% | 0.74% |
| I3 | 50 | 1 | 1 | Optimal | 5,800 | 5,784.86 | 0% | 0% |

Table 6: Costs on stochastic instances for different number of scenarios.

To choose the number of scenarios that will be used in each instance we observed the values for the percentage differences in the recourse costs (%I.Recour.C). Since these differences are smaller than 0.5% for 300 scenarios for instances I1 and for 150 scenarios for instances I2, we decided to set $|\Omega_d| = 300$ for I1 and to set $|\Omega_d| = 150$ for I2. Regarding instances I3, we set $|\Omega_d|$ to 50 as the expected recourse cost (Recour.C) was smaller than the value of Recour.C for the other number of scenarios (5 and 25), and as the model was not able to solve instances with a larger number of scenarios.

*4.3. Computational Results*

In this section, we present the computational results after testing our model on real-world instances. First, we present the performance of the proposed model for different planning horizons. Second, we introduce an example to illustrate a typical output of the problem. Third, we analyze the impact of the type of recourse actions used in the costs and number of caregivers staffed. An analysis of the impact of schedule flexibility in the costs and number of caregivers staffed is presented at the end of this section.

Table 7 presents for each instance and each combination of recourse actions allowing caregiver reallocation (Realloc.) and working on a day-off (RestToW), the CPU time in seconds to solve the problem (Time), the status of the solution (Status), the total cost (Total.C), and the total number of caregivers to hire.

| | | | | | | | Caregivers | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Instance | Scen. | Realloc. | RestToW | Time (s) | Status | Total.C ($) | $d_0$ | $d_1$ | $d_2$ | $d_3$ | Total |
| I1 | 300 | 0 | 0 | 8.04 | Optimal | 12,858.8 | 4 | 43 | 6 | 22 | 75 |
| I1 | 300 | 1 | 0 | 418.45 | Optimal | 11,773.5 | 4 | 39 | 13 | 15 | 71 |
| I1 | 300 | 0 | 1 | 166.78 | Optimal | 11,528.5 | 4 | 39 | 5 | 20 | 68 |
| I1 | 300 | 1 | 1 | 1,121.53 | Optimal | 10,843 | 4 | 37 | 11 | 17 | 69 |
| I2 | 150 | 0 | 0 | 53.54 | Optimal | 24,990.7 | 4 | 43 | 6 | 21 | 74 |
| I2 | 150 | 1 | 0 | 1,876.03 | Optimal | 22,620.7 | 4 | 38 | 13 | 14 | 69 |
| I2 | 150 | 0 | 1 | 675.45 | Optimal | 21,441.6 | 3 | 38 | 5 | 17 | 63 |
| I2 | 150 | 1 | 1 | 5,212.85 | Optimal | 20,116.4 | 4 | 35 | 11 | 14 | 64 |
| I3 | 50 | 0 | 0 | 1,139.62 | Optimal | 35,995.5 | 4 | 34 | 6 | 17 | 61 |
| I3 | 50 | 1 | 0 | 10,522.7 | Optimal | 32,341.8 | 4 | 30 | 11 | 13 | 58 |
| I3 | 50 | 0 | 1 | 2,799.27 | Optimal | 30,033.5 | 3 | 30 | 6 | 16 | 55 |
| I3 | 50 | 1 | 1 | 11,983.2 | Optimal | 27,805.5 | 3 | 27 | 10 | 12 | 52 |

Table 7: Computational effort and results on stochastic instances.

Results from Table 7 indicate that the computational effort increase with the length of the planning horizon, as well as with the flexibility related to the recourse actions. Observe that it was possible to find an optimal solution for all instances. We can conclude that the recourse action that contributes the most to an increase in the CPU time is allowing the reallocation of caregivers (Realloc. = 1). Specifically, for instances I1, I2, and I3 and when Realloc. = 1 CPLEX was respectively 52, 35 and 10 times slower to solve the model when compared to solving the model with simple recourse, i.e. Realloc = 0 and RestToW = 0. When recourse RestToW is included in the model (contact caregivers to work on a day-off), these values increase to 140, 97, and 10 for instances I1, I2, and I3, respectively.

Results on staff dimensioning suggest that the number of caregivers to hire in districts $d_0$, $d_2$, and $d_3$ is very similar for instances spanning different planning horizons. However, for districts $d_1$ and $d_3$ we can observe some significative differences in the number of caregivers to hire (e.g., 27 caregivers for $d_0$ in instance I3 when Realloc = 1 and RestToW = 1 versus 37 caregivers for $d_0$ in instance I1 when Realloc = 1 and RestToW = 1). We remark that this result might be due to forecasting errors and changes in the magnitude of demands from one month to the other one.

### *4.4. Example 2: Output Illustration*

Tables 8 and 9 present an example of the schedules and the use of recourse actions after solving the two-stage stochastic programming model on an instance including a 4-week planning horizon and 10 scenarios. This example incorporates the use of recourse actions associated with under-covering, with over-covering, with the reallocation of caregivers to neighbor districts, and with contacting caregivers to work on a day-off. Table 8 shows four schedules (one per district) including the shift and day-off allocation at each day in the planning horizon. Recall that r represents the allocation of a day-off, and that $m_4$, $m_8$, $a_4$, and $n_{10}$ denote different types of shifts. For instance, a caregiver hired to work in district 3 ($d_3$) will be allocated in his last week to: afternoon shifts ($a_4$) in the first 2 days of the week; then he will work in the next 2 days in night shifts ($n_{10}$); the caregiver will finish the week with 3 consecutive days-off (r).

| | Date | | | | | | |
|---|---|---|---|---|---|---|---|
| district | 2017-07-03 | 2017-07-04 | 2017-07-05 | 2017-07-06 | 2017-07-07 | 2017-07-08 | 2017-07-09 |
| $d_0$ | $m_8\_d_0$ | $m_8\_d_0$ | $m_8\_d_0$ | $m_4\_d_0$ | $m_4\_d_0$ | r | $m_8\_d_0$ |
| $d_1$ | $m_4\_d_1$ | $m_4\_d_1$ | $m_4\_d_1$ | $m_8\_d_1$ | $m_8\_d_1$ | r | r |
| $d_2$ | r | r | r | $m_4\_d_2$ | $m_4\_d_2$ | $n_{10}\_d_2$ | $n_{10}\_d_2$ |
| $d_3$ | $m_4\_d_3$ | $m_4\_d_3$ | $a_4\_d_3$ | $a_4\_d_3$ | r | r | r |
| | 2017-07-10 | 2017-07-11 | 2017-07-12 | 2017-07-13 | 2017-07-14 | 2017-07-15 | 2017-07-16 |
| $d_0$ | $m_8\_d_0$ | $n_{10}\_d_0$ | $n_{10}\_d_0$ | r | r | r | $m_8\_d_0$ |
| $d_1$ | r | $m_4\_d_1$ | $m_4\_d_1$ | $m_8\_d_1$ | $m_8\_d_1$ | r | r |
| $d_2$ | r | r | r | $m_8\_d_2$ | $m_8\_d_2$ | $n_{10}\_d_2$ | $n_{10}\_d_2$ |
| $d_3$ | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| | 2017-07-17 | 2017-07-18 | 2017-07-19 | 2017-07-20 | 2017-07-21 | 2017-07-22 | 2017-07-23 |
| $d_0$ | $m_8\_d_0$ | $n_{10}\_d_0$ | $n_{10}\_d_0$ | r | r | r | $m_8\_d_0$ |
| $d_1$ | r | $m_4\_d_1$ | $m_4\_d_1$ | $m_8\_d_1$ | $m_8\_d_1$ | r | r |
| $d_2$ | $n_{10}\_d_2$ | r | r | r | $m_4\_d_2$ | $m_4\_d_2$ | $n_{10}\_d_2$ |
| $d_3$ | $m_8\_d_3$ | $m_8\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| | 2017-07-24 | 2017-07-25 | 2017-07-26 | 2017-07-27 | 2017-07-28 | 2017-07-29 | 2017-07-30 |
| $d_0$ | $m_8\_d_0$ | $m_8\_d_0$ | $n_{10}\_d_0$ | $n_{10}\_d_0$ | r | r | r |
| $d_1$ | r | $m_8\_d_1$ | $m_8\_d_1$ | $m_8\_d_1$ | $a_4\_d_1$ | $a_4\_d_1$ | r |
| $d_2$ | $n_{10}\_d_2$ | r | $m_4\_d_2$ | $m_4\_d_2$ | $m_4\_d_2$ | $m_8\_d_2$ | $m_8\_d_2$ |
| $d_3$ | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |

Table 8: Example of the schedules obtained with the two-stage stochastic programming model.

The shift and day-off allocation of the schedule for $d_3$ is used as an example to show the use of recourse actions related to the reallocation of caregivers to neighbor districts, and with contacting caregivers to work on a day-off. Table 9 shows for each day of the week from 2017-07-24 to 2017-07-30 the changes in the schedules due to the recourse actions used for 10 demand scenarios. Values in bold indicate that a recourse action was used to protect against uncertainty. For instance, during day 2017-07-24 and under scenario 10 the model decided to include a district reallocation (a caregiver from district $d_3$ is reallocated to district $d_1$). In a similar way, during day 2017-07-29 the model chose to use the recourse work on a day-off for scenarios 2, 7, 9, and 10 (e.g., in scenario 2, a caregiver is called to work on his day-day in a morning shift in district 3 ($m_4\_d_3$)).

| | Date | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2017-07-24 | 2017-07-25 | 2017-07-26 | 2017-07-27 | 2017-07-28 | 2017-07-29 | 2017-07-30 |
| Master schedule | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 1 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 2 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | $\mathbf{m_4\_d_3}$ | r |
| Scen. 3 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 4 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 5 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 6 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 7 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | $\mathbf{a_4\_d_3}$ | r |
| Scen. 8 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | r | r |
| Scen. 9 | $a_4\_d_3$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | $\mathbf{a_4\_d_3}$ | r |
| Scen. 10 | $\mathbf{a_4\_d_1}$ | $a_4\_d_3$ | $n_{10}\_d_3$ | $n_{10}\_d_3$ | r | $\mathbf{n_{10}\_d_3}$ | r |

Table 9: Illustration on the use of recourse actions in a schedule of a caregiver working in $d_3$.

### *4.5. Assessing the Impact of Different Recourse Actions*

In this section, we perform a comparison among the different types of recourse actions used in the two-stage stochastic programming model. The impact of allowing caregiver reallocation and working on a day-off is evaluated. Table 10 reports the percentage difference in the total cost (%D.Total.C), the percentage difference in the scheduling cost (%D.Sched.C), the percentage difference in the recourse cost (%D.Recour.C), and the percentage difference in the total number of caregivers staffed (%D.Staff), when flexibility regarding the use of different recourse actions is introduced in the model. These percentage differences are computed as $\%D = 100 \times (Value - base\_value)/base\_value$. $Value$ represents the final value for the total cost, for the staffing cost, for the recourse cost, and for the total number of caregivers staffed, and $base\_value$ denotes the value for the same attribute obtained after solving problem $\mathcal{Z}$ (on each instance I1, I2, and I3) with the base scenario. Since the base scenario corresponds to the use of simple recourse in the second-stage model (i.e. only allowing demand under-covering and over-covering) the differences in the recourse costs are mainly due to the reduction in demand under-covering and over-covering costs.

| Instance | Scen. | Realloc. | RestToW | %D.Total.C | %D.Sched.C | %D.Recour.C | %D.Staff |
|---|---|---|---|---|---|---|---|
| I1 | 300 | 1 | 0 | -8.44% | -5.35% | -21.17% | -5.33% |
| I1 | 300 | 0 | 1 | -10.35% | -12.7% | -0.65% | -9.33% |
| I1 | 300 | 1 | 1 | -15.68% | -12.87% | -27.22% | -8% |
| I2 | 150 | 1 | 0 | -9.48% | -6% | -23.03% | -6.76% |
| I2 | 150 | 0 | 1 | -14.2% | -17.84% | -0.03% | -14.86% |
| I2 | 150 | 1 | 1 | -19.5% | -17.36% | -27.85% | -13.51% |
| I3 | 50 | 1 | 0 | -10.15% | -6.16% | -24.73% | -4.92% |
| I3 | 50 | 0 | 1 | -16.56% | -18.91% | -7.99% | -9.84% |
| I3 | 50 | 1 | 1 | -22.75% | -20.5% | -25.26% | -14.75% |

Table 10: Impact of the type of recourse action used in the costs and number of caregivers staffed.

Results from Table 10 suggest that the introduction of flexibility in the use of recourse actions significantly reduces the total costs, as well as the number of caregivers staffed. These reductions appear to be larger for instances spanning planning horizons of 8 weeks or longer than for instances spanning 4 weeks. The recourse action with larger impact is contact caregivers to work on a day-off, and when this recourse action is integrated with the reallocation of caregivers, the reductions in costs become even larger. Solving an integrated problem including all districts instead of solving independent problems for each district generates a supplementary cost reduction, as well as an improvement in caregivers' utilization. In particular, allowing reallocation of caregivers to neighbor districts gives planners the flexibility to occasionally use resources from other districts to respond to changes in demands.

### *4.6. Assessing the Impact of Schedule Flexibility*

Since the two-stage stochastic programming problem becomes harder to solve with the length of the planning horizon, we perform an analysis on the impact of reducing schedule flexibility. Specifically, for instances including more that 4 weeks (I2 and I3), we solve the two-stage stochastic programming problem by imposing schedules starting at week 5 to be exactly the same as schedules from the previous 4 weeks. For instance, in a problem with a 8-week planning horizon, the schedules in week 5 must be the same as the schedules for week 1, the schedules in week 6 must be the same as the schedules for week 2, and so on.

Table 11 reports an analysis on the impact of schedule flexibility in the computational effort and results of model $\mathcal{Z}$. In particular, this table presents a comparison of the CPU times in seconds (Time (s)), of the total costs (Total.C), and of the number of caregivers staffed (Total.Staff) when schedules are completely flexible (Flex.) and when the scheduling flexibility is reduced (No.Flex) as explained above.

| | | | | Time (s) | | Total.C ($) | | Total.Staff | |
|---|---|---|---|---|---|---|---|---|---|
| Instance | Scen. | Realloc. | RestToW | Flex | No.Flex | Flex | No.Flex | Flex | No.Flex |
| I2 | 150 | 0 | 0 | 53.54 | 12.36 | 24,990.7 | 25,879.5 | 74 | 71 |
| I2 | 150 | 1 | 0 | 1,876.03 | 1,397.98 | 22,620.7 | 23,583 | 69 | 67 |
| I2 | 150 | 0 | 1 | 675.45 | 547.97 | 21,441.6 | 22,344.3 | 63 | 62 |
| I2 | 150 | 1 | 1 | 5,212.85 | 3,442.57 | 20,116.4 | 20,858.4 | 64 | 58 |
| I3 | 50 | 0 | 0 | 1,139.62 | 14.6 | 35,995.5 | 40,120.3 | 61 | 70 |
| I3 | 50 | 1 | 0 | 10,522.7 | 704.88 | 32,341.8 | 36,155 | 58 | 69 |
| I3 | 50 | 0 | 1 | 2,799.27 | 533.79 | 30,033.5 | 33,006.6 | 55 | 57 |
| I3 | 50 | 1 | 1 | 11,983.2 | 2,668.15 | 27,805.5 | 30,678.8 | 52 | 54 |

Table 11: Impact of schedule flexibility in the costs and number of caregivers staffed.

Results from Table 11 indicate that the method is in average 14 times faster when flexibility in the allocation of schedules is limited. This speed-up is more substantial for instances with type I3 as a longer time horizon is being considered. Observe that the total cost presents an increase when there is less flexibility associated with the allocation of schedules, as the two-stage model has less freedom to use recourse actions when needed. However, the number of caregivers to hire shows a different behavior for instances I2 and I3. Specifically, in I2 instances the value of Total.Staff becomes smaller when the schedule flexibility is reduced. On the contrary, when the schedule flexibility is reduced, the value of Total.Staff becomes larger for instances I3. This might be explained by the fact that for short time horizons (8 weeks), the model with less schedule flexibility (No.Flex) decides to hire less employees (even if this means to have some extra under-covering) in order to reduce the employee underutilization (visits over-covering). On the contrary, for longer time horizons (12 weeks) the No.Flex model decides to hire more employees as this restriction in the schedule allocation might significantly increase the visits under-covering and hence the total costs.

### *4.7. Value of the Stochastic Solution*

The VSS is a standard measure that indicates the expected gain from solving a stochastic model rather than its deterministic counterpart, the *expected value problem* (EV). The value of the stochastic solution is defined as $VSS = EEV - RP$, where $RP$ corresponds to the optimal value of problem (4)-(12) and $EEV$ corresponds to the expected value of using the EV solution. EV is problem (4)-(12) evaluated using the mean scenario $\bar{\xi}_d = \hat{\mathbf{b}}_d$ for each day $d \in D$. Given an EV solution ($\bar{\mathbf{y}}^*$), EEV corresponds to: $EEV = \sum_{d \in D} \sum_{w \in \Omega_d} p_d^{(w)} \mathcal{Q}\left(\bar{\mathbf{y}}^*, \xi_{\mathbf{d}}(\mathbf{w})\right)$. A large VSS means that uncertainty is important for the quality of the resulting optimal solution. On the contrary, a small VSS means that a deterministic approach based on the expected values of the random variables might be sufficiently good to take a decision. The reader is referred to Birge & Louveaux (2011) for an overview of stochastic programming.

Table 12 presents a comparison of the computational effort between the two-stage stochastic programming model (denoted as Stochastic) and the mean value problem (denoted as Deterministic). This effort is measured by the CPU time in seconds to solve the problem. This table also reports the total cost when the schedules obtained with the stochastic model and with the deterministic model are evaluated with the actual values for the demand (Real.C).

Table 13 presents an evaluation of the values of the stochastic solution. In particular, this table reports the expected gains in the total cost ($VSS_{Cost}$), in the scheduling cost ($VSS_{Scheduling}$), in the recourse cost ($VSS_{Recourse}$), and in the quantity of caregivers staffed ($VSS_{Staff}$) from solving the stochastic model rather than its deterministic counterpart. This evaluation is computed as: $VSS_i = 100 \times (EEV_i - RP_i)/EEV_i$, for all $i = \{Cost; Scheduling; Recourse; Staff\}$.

| | | | | Stochastic | | Deterministic | |
|---|---|---|---|---|---|---|---|
| Instance | Scen. | Realloc. | RestToW | Time(s) | Real.C ($) | Time (s) | Real.C ($) |
| I1 | 300 | 0 | 0 | 8.04 | 13,102.2 | 2.73 | 12,765.6 |
| I1 | 300 | 1 | 0 | 418.45 | 11,831 | 8.78 | 12,174.68 |
| I1 | 300 | 0 | 1 | 166.78 | 11,158.4 | 2.32 | 11,431.8 |
| I1 | 300 | 1 | 1 | 1,121.53 | 10,478.68 | 9.77 | 10,981.48 |
| I2 | 150 | 0 | 0 | 53.54 | 24,947.6 | 34.59 | 28,810.4 |
| I2 | 150 | 1 | 0 | 1,876.03 | 22,464.72 | 218.16 | 25,952.68 |
| I2 | 150 | 0 | 1 | 675.45 | 20,635 | 48.18 | 22,148.8 |
| I2 | 150 | 1 | 1 | 5,212.85 | 19,587.2 | 217.29 | 21,157.92 |
| I3 | 50 | 0 | 0 | 1,139.62 | 37,509.2 | 575.7 | 45,786.8 |
| I3 | 50 | 1 | 0 | 10,522.7 | 32,884.12 | 8706.94 | 40,853 |
| I3 | 50 | 0 | 1 | 2,799.27 | 30,655.2 | 577.08 | 31,873 |
| I3 | 50 | 1 | 1 | 11,983.2 | 28,262.8 | 5144.16 | 30,706.6 |

Table 12: Computational effort and results for the stochastic and deterministic models.

| Instance | Scen. | Realloc. | RestToW | $VSS_{Cost}$ | $VSS_{Scheduling}$ | $VSS_{Recourse}$ | $VSS_{Staff}$ |
|---|---|---|---|---|---|---|---|
| I1 | 300 | 0 | 0 | 15.53% | -17.62% | 60.89% | -13.64% |
| I1 | 300 | 1 | 0 | 14.12% | -11.34% | 59.68% | -5.97% |
| I1 | 300 | 0 | 1 | 5.61% | -2.04% | 25.75% | -1.49% |
| I1 | 300 | 1 | 1 | 5.12% | -1.74% | 28.77% | -1.47% |
| I2 | 150 | 0 | 0 | 16.33% | -20.5% | 61.8% | -17.46% |
| I2 | 150 | 1 | 0 | 14.62% | -13.8% | 60.97% | -9.52% |
| I2 | 150 | 0 | 1 | 3.35% | 0.86% | 10.56% | 0.0% |
| I2 | 150 | 1 | 1 | 4.27% | -0.2% | 20.15% | -3.23% |
| I3 | 50 | 0 | 0 | 18.67% | -22.81% | 63.58% | -19.61% |
| I3 | 50 | 1 | 0 | 17.25% | -15.61% | 63.92% | -13.73% |
| I3 | 50 | 0 | 1 | 5.34% | 0.26% | 18.68% | -7.84% |
| I3 | 50 | 1 | 1 | 6.56% | 0.64% | 24.13% | -4% |

Table 13: Value of the stochastic solution.

Results from Table 12 indicate that the CPU time to solve the mean value problem is significantly smaller than the CPU time to solve the stochastic problem. However, when the schedules obtained after solving the deterministic problem are evaluated on the real demand these schedules perform worse (as the Real.C is larger in most instances), when compared to the performance of the schedules obtained with the stochastic problem. The differences in the real cost between the deterministic model and the stochastic model are of great importance in practice, since Real.C indicates how well the caregivers schedules react to the actual demand. Since the majority of values for Real.C are lower when demand uncertainty is included in the model, we can conclude that the schedules obtained with the stochastic model are more robust than the schedules obtained with the deterministic model (EV problem).

We remark that since the schedules obtained with the stochastic model are usually more robust than the schedules obtained with a deterministic model, Real.C is expected to be lower when evaluated with the stochastic schedules than when evaluated with the deterministic schedules. However, it may happen that in some cases this is not true. For example, in instance I1 with Realloc.=0 and RestToW=0. In this case, what could have happened was that the actual demand was very similar to the mean demand. Hence, when the dimensioning and scheduling decisions obtained with the deterministic model are evaluated on a single instance corresponding to the actual (observed) demand, Real.C is lower than the cost obtained with the stochastic model.

Results from Table 13 suggest that the two-stage stochastic model can lead to significant reductions in the total cost when compared to the mean value program, since all the VSSs associated with the total cost are positive values ranging from 3.35% to 18.67%. This result is mainly due to a reduction in the recourse costs associated with demand under-covering and over-covering. Observe that some instances have negative VSS for the scheduling costs ($VSS_{Sched.}$) and for the the staffing decisions ($VSS_{Staff}$). This means that the two-stage stochastic model selects a larger workforce than the deterministic model, resulting in more robust staffing and scheduling decisions that accommodate better to changes in demands.

### *4.8. Practical Aspects and Managerial Insights*

The methodology developed in this paper represents an important and general decision support tool for home care agencies interested in staff dimensioning and caregiver scheduling. Specifically, our computational experiments indicate that:

- The design of robust staffing and scheduling decisions require the incorporation of uncertainty in demands, as expected costs are smaller when uncertainty is included. This is explained by the fact that opposite to deterministic models, the strength of stochastic programming arises from the ability to represent solutions that protect against multiple possible future outcomes (Birge, 1995). Hence, the aptitude to identify solutions that handle or adapt best to the set of potential outcomes, relative to their probability of occurring, is expected to generate costs that are smaller when compared to a deterministic model when evaluated on several possible demand realizations.

- Including recourse actions such as allowing caregiver reallocation to neighbor districts and working on a day-off significantly improves the costs associated with the dimensioning decisions (staffing), as well as with demand under-covering and over-covering, resulting in the improvement of caregiver utilization and quality of service.

- Solving an integrated problem including all districts instead of solving independent problems for each district, generates supplementary cost reductions. In particular, allowing reallocation of caregivers to neighbor districts gives planners the flexibility to occasionally use resources from other districts to respond to changes in the demand or in caregivers' availabilities.

Even though the case study was done for a specific agency from AlayaCare, this agency was selected because it includes most of the key features of the consider problem (e.g., stochastic demands, several geographic areas, different types of shifts, several work regulations for the composition of schedules). Therefore, we believe that our study is general and that the conclusions drawn form the computational experiments can be similar if the methodology is tested in other practical cases.

The proposed model could be useful to evaluate the impact in costs and in the quality of solutions by using different recourse actions. Specifically, recourse actions including the allocation of overtime and the use of part-time caregivers could be tested to evaluate if an increase in recourse flexibility helps to decrease the scheduling costs and demand under-covering and over-covering costs. The model could also be used as a tool to detect the lack/excess of caregivers due to changes in demand. For instance, given a fix number of caregivers, the model will incur large under-covering costs if the size of the workforce is inadequate to satisfy all patient visits when demand increases. On the contrary, the solutions of the model will return large over-staffing costs if the size of the permanent workforce is too large for the demand. Moreover, the two-stage stochastic programming model could be extended to incorporate multiple types of caregivers with different skills, and to include information about current employees with their preferences and availabilities.

Regarding the computational effort and limits of the two-stage stochastic programming model, computational experiments indicate that the CPU time increases with the length of the planning horizon, with the number of scenarios, and with the flexibility in recourse actions. For each type of instance tested, we observed that the most important factor in this computational time increase was the value of RestToW (i.e. contact caregivers to work on their day-off) since problem $\mathcal{Z}$ was in average 100 times slower when RestToW was set to 1. We also observed that the computational time required to solve the problems can be reduced by 5 times in average by limiting the schedule allocation flexibility. One idea to deal with the computational limits of the method on larger planning horizons could be to use a rolling horizon approach. In this way, the complexity of the problem will be reduced as this method will gradually move along the planning horizon to incorporate stochastic information of the demand.

The work presented in this paper has some limitations that could be addressed in future work. These limitations are mainly related to the assumptions adopted to facilitate the modeling and solution of the problem under study. For instance, assuming that the duration of patients' visits and travel times are deterministic parameters could lead to suboptimal solutions, especially if caregivers perform several short visits within one day and the variability in these times is large. In the case of AlayaCare, this variability does not affect significantly the solution of the problem, as most of the caregivers are personal support workers that perform long visits during their shift. In addition, the practical use of the work presented in this paper can be affected by assuming that caregivers will accept to work when called during their day-off, since from time to time caregivers are free to reject this type of request from their employer. Moreover, in a home care setting where caregiver absenteeism rates are high, assuming that the workforce capacity is deterministic could lead to problems in the implementation of the solutions obtained. The last limitation of this work is related to the demand forecasting methods used, as other techniques could be explored to predict the demand in a more accurate way.

## 5. Concluding Remarks

We presented a two-stage stochastic programming model for integrated staffing and scheduling in home healthcare. In this model first-stage decisions correspond to staff dimensioning and to the allocation caregivers to schedules. Second-stage decisions are related to the temporary reallocation of caregivers to neighbor districts, to contact caregivers to work on their day-off, and to allow under-covering and over-covering. Results on real-world instances show that the use of the two-stage stochastic programming model helps to reduce demand under-covering and over-covering costs when compared to a deterministic approach using the mean demand. Moreover, computational results indicate that the use of flexible recourse actions significantly reduces the total costs, improves caregiver utilization, and increases the level of service.

An interesting avenue for future research is related to the development of specialized solution methods to tackle larger instances commonly found in practice. Future research could also include the use of different techniques for demand forecasting and for scenario generation to assess the impact of demand estimation accuracy in the solutions obtained with the two-stage stochastic programming problem.

## Acknowledgments

The authors would like to thank MEDTEQ which supported this work a grant. The authors would also like to thank to the data team at AlayaCare for their support in getting access to anonymized data. We also thank the two anonymous referees and the Associate Editor for their insightful comments and suggestions which allowed us to improve the paper.

## References

Argiento, R., Guglielmi, A., Lanzarone, E., & Nawajah, I. (2016). A bayesian framework for describing and predicting the stochastic demand of home care patients. *Flexible Services and Manufacturing Journal*, *28*, 254–279.

Bachouch, R. B., Guinet, A., & Hajri-Gabouj, S. (2011). A decision-making tool for home health care nurses' planning. In *Supply Chain Forum: an International Journal* (pp. 14–20). Taylor & Francis volume 12.

Bennett, A. R., & Erera, A. L. (2011). Dynamic periodic fixed appointment scheduling for home health. *IIE Transactions on Healthcare Systems Engineering*, *1*, 6–19.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, *52*, 35–53.

Birge, J. R. (1995). Models and model value in stochastic programming. *Annals of Operations Research*, *59*, 1–18.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Boyer, V., Gendron, B., & Rousseau, L.-M. (2012). A branch-and-price algorithm for the multi-activity multi-task shift scheduling problem. *Journal of Scheduling*, *17*, 185–197.

Braekers, K., Hartl, R. F., Parragh, S. N., & Tricoire, F. (2016). A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience. *European Journal of Operational Research*, *248*, 428–443.

Burke, E. K., De Causmaecker, P., Berghe, G. V., & Van Landeghem, H. (2004). The state of the art of nurse rostering. *Journal of scheduling*, *7*, 441–499.

Campbell, G. M. (2011). A two-stage stochastic program for scheduling and allocating cross-trained workers. *Journal of the Operational Research Society*, *62*, 1038–1047.

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

Cappanera, P., Scutellà, M. G., Nervi, F., & Galli, L. (2018). Demand uncertainty in robust home care optimization. *Omega*, *80*, 95–110.

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

Côté, M.-C., Gendron, B., & Rousseau, L.-M. (2011). Grammar-based integer programming models for multiactivity shift scheduling. *Management Science*, *57*, 151–163.

Côté, M.-C., Gendron, B., & Rousseau, L.-M. (2013). Grammar-based column generation for personalized multi-activity shift scheduling. *INFORMS Journal on computing*, *25*, 461–474.

Defraeye, M., & Van Nieuwenhuyse, I. (2016). Staffing and scheduling under nonstationary demand for service: A literature review. *Omega*, *58*, 4–25.

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

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

Garg, L., McClean, S., Meenan, B., & Millard, P. (2010). A non-homogeneous discrete time markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system. *Health care management science*, *13*, 155–169.

Harvey, A. C., & Peters, S. (1990). Estimation procedures for structural time series models. *Journal of Forecasting*, *9*, 89–108.

Hertz, A., & Lahrichi, N. (2009). A patient assignment algorithm for home care services. *Journal of the Operational Research Society*, *60*, 481–495.

Hewitt, M., Nowak, M., & Nataraj, N. (2016). Planning strategies for home health care delivery. *Asia-Pacific Journal of Operational Research*, *33*, 1650041.

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2001). Introduction to automata theory, languages, and computation. *ACM SIGACT News*, *32*, 60–65.

Hulshof, P. J., Kortbeek, N., Boucherie, R. J., Hans, E. W., & Bakker, P. J. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, *1*, 129–175.

Jalalpour, M., Gel, Y., & Levin, S. (2015). Forecasting demand for health services: Development of a publicly available toolbox. *Operations Research for Health Care*, *5*, 1–9.

Kim, K., & Mehrotra, S. (2015). A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research*, *63*, 1431–1451.

Lahrichi, N., Lapierre, S., Hertz, A., Talib, A., & Bouvier, L. (2006). Analysis of a territorial approach to the delivery of nursing home care services based on historical data. *Journal of medical systems*, *30*, 283–291.

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

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

Lanzarone, E., Matta, A., & Sahin, E. (2012). Operations management applied to home care services: the problem of assigning human resources to patients. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *42*, 1346–1363.

Lanzarone, E., Matta, A., & Scaccabarozzi, G. (2010). A patient stochastic model to support human resource planning in home care. *Production Planning and Control*, *21*, 3–25.

Maenhout, B., & Vanhoucke, M. (2013). An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems. *Omega*, *41*, 485–499.

Matta, A., Chahed, S., Sahin, E., & Dallery, Y. (2014). Modelling home care organisations from an operations management perspective. *Flexible Services and Manufacturing Journal*, *26*, 295–319.

Nguyen, T. V. L., Toklu, N. E., & Montemanni, R. (2015). Matheuristic optimization for robust home health care services. In *Proc. ICAOR* (p. 2).

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

Quimper, C.-G., & Walsh, T. (2007). Decomposing global grammar constraints. In *Principles and Practice of Constraint Programming–CP 2007* (pp. 590–604). Springer.

Restrepo, M. I., Gendron, B., & Rousseau, L.-M. (2016). Branch-and-price for multi-activity tour scheduling. *INFORMS Journal on Computing*, *28*, 1–17.

Restrepo, M. I., Gendron, B., & Rousseau, L.-M. (2017). A two-stage stochastic programming approach for multi-activity tour scheduling. *European Journal of Operational Research*, *262*, 620–635.

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

Soyiri, I. N., & Reidpath, D. D. (2013). An overview of health forecasting. *Environmental health and preventive medicine*, *18*, 1—9.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*, 37–45.

Trautsamwieser, A., & Hirsch, P. (2014). A branch-price-and-cut approach for solving the medium-term home health care planning problem. *Networks*, *64*, 143–159.

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