Chapter 8
Capacity planning in healthcare: finding solutions for healthy planning in nursing home care

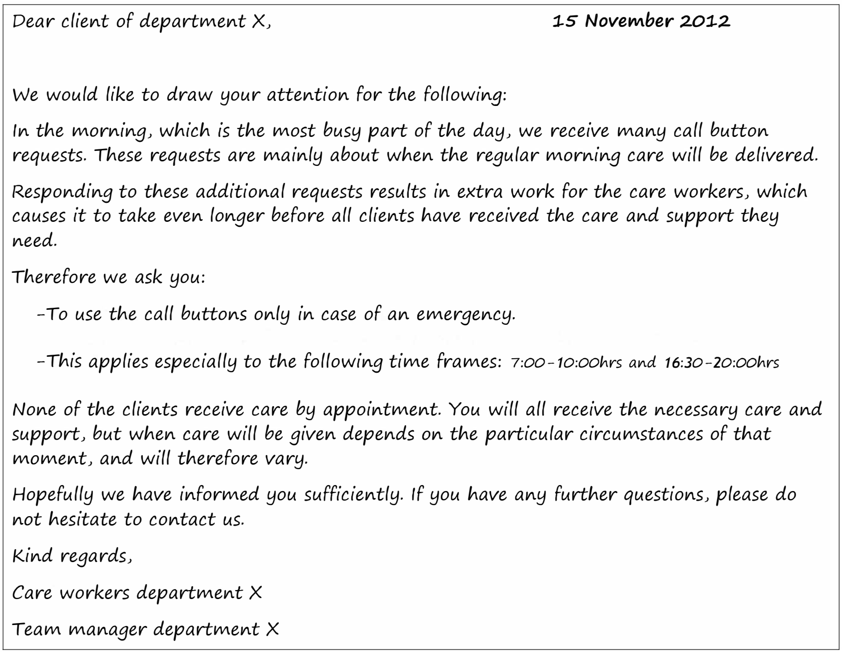
by Dr. D. Moeke & Dr. R. Bekker

8.1 Introduction

Context and background

Demographic projections reveal that the EU population is turning increasingly 'grey' in the coming decades. This change in demographics puts pressure on the financial sustainability of the European long-term healthcare systems because of two main reasons. Firstly, the prevalence of physical or mental disability, and thus dependency, increases with age (especially with very old age groups, 80+). As such the increasing share of elderly is likely to lead to an increase in the demand for Long Term Care (LTC). Secondly, chronic diseases such as cancer, heart disease and diabetes are more common among older people. In fact, many elderly have more than one chronic disease. Consequently, the ageing population has increased the pressure on the long-term healthcare systems and will continue to do so in the forthcoming decades. In order to ensure the long-term financial sustainability of the European healthcare systems, providers of long-term care are challenged to provide their care services in a more efficient manner.

Nursing homes play an important role in the provision of long-term care. A nursing home can be described as a facility with a domestic styled environment that provides 24-hour functional support and care for persons who require assistance with activities of daily living and who often have complex health needs and increased vulnerability [1, p. 183]. It should be noted that in this chapter no distinction is made between nursing homes and residential homes (less intensive care) as, in practice, the boundary between the two is diffuse [2]. Often, nursing home clients are in need of ongoing assistance with basic activities of daily living due to physical or psychological disabilities. Consequently, in order to live their lives according to their own daily routines, nursing home clients depend greatly on timely delivery of care and support. As such, quality of care in a nursing home largely depends on the coordination and timing of service delivery. In practice, however, nursing homes have to balance the goal of meeting clients' preferences with the efficient use of resources. The real-life letter reproduced in Figure 1.1 illustrates how some nursing homes struggle to meet these seemingly contradictory goals. It shows how nursing home department X has difficulties with meeting the time preferences of the clients during busy periods of the day with the resources available. It is even taken for granted by department X that clients have little or no influence on their daily routine as "none of the clients receive care by appointment" and "when care will be given depends on the particular circumstances of that moment, and will therefore vary".



Dear client of department X, **15 November 2012**

We would like to draw your attention for the following:

In the morning, which is the most busy part of the day, we receive many call button requests. These requests are mainly about when the regular morning care will be delivered.

Responding to these additional requests results in extra work for the care workers, which causes it to take even longer before all clients have received the care and support they need.

Therefore we ask you:

-To use the call buttons only in case of an emergency.

-This applies especially to the following time frames: 7:00-10:00hrs and 16:30-20:00hrs

None of the clients receive care by appointment. You will all receive the necessary care and support, but when care will be given depends on the particular circumstances of that moment, and will therefore vary.

Hopefully we have informed you sufficiently. If you have any further questions, please do not hesitate to contact us.

Kind regards,

Care workers department X

Team manager department X

Figure 8.1: Real-life letter

Capacity planning plays a key role in ensuring that an organization has the capability to respond sufficiently to the level of demand experienced, see e.g. [3]. For nursing homes, care workers are by far the most important resource. This is due to the fact that care workers are responsible for the daily care and supervision of the residents and their labor costs account for a significant proportion of the total healthcare expenditure [4]. The main focus of capacity planning should thus be on matching availability of care workers (*supply)* with the needs and preferences of clients (*demand*).

Matching supply and demand in nursing home practice is not straightforward, as it always takes place within a `client – care worker – organizational' triangle. Clearly, the needs and preferences of the nursing home clients are a vital element of capacity planning. More specifically, we find that their needs and preferences should dictate the timing of care and support activities and should be the starting point of any capacity plan. In turn, the role of the organization is to match the required care and support with appropriate nurse staffing levels (i.e. matching supply and demand), within the boundaries of human and financial resources. Finally, the professional responsibility of the care workers should also be considered. For example, care workers are confronted with an inherent tension between their desire to respect and foster the personal autonomy of the client and their responsibility to act in the best interest of the client [36]. Therefore, we like to stress that from a care workers' perspective, capacity planning is more than blindly obeying to time schedules, as it could undermine their professional dignity[37].

Thus, one could say that the main focus of capacity planning in a nursing home setting is on getting the right number of care workers with the right set of skills in the right job at the right time. The result of an adequate capacity planning process is a well-balanced workload, having major benefits for both the clients as well as the care workers. For the clients, a well-balanced workload will result in less waiting and care being delivered closer to their preferred time. From the perspective of the care worker, the literature shows that not having a balanced workload contributes to emotional exhaustion (e.g., [5, 38]) and potential cognitive impairment (e.g. [6]). Hence, capacity planning plays a crucial role in ensuring the physical and mental wellbeing of the care worker.

Furthermore, nursing homes are becoming more and more information-intensive enterprises. Nowadays, they have access to large amounts of clinical and operational data due to the increasing adoption of electronic health records and other IT systems. This development enhances data-driven decision making and accelerates digital transformation. Also, when it comes to capacity planning nursing homes are starting to recognize data-driven decision making as essential to improve their (future) operations.

Contribution and outline

This chapter illustrates how data-driven capacity planning can support nursing homes in their search for ways to further reduce their costs while maintaining an appropriate quality level of care. This is done using a framework based on a four-step approach adopted from workforce management in call centres. In addition, it is also shown that data-driven capacity planning allows for a more evenly spread workload for the care workers.

The remainder of this chapter is structured as follows. The next section introduces measures that can be used to evaluate the performance of capacity planning efforts. Section 3 provides insight into three concepts which are crucial for a thorough understanding of the domain of data-driven capacity planning. In Section 4, a four-step approach for data-driven capacity planning is introduced and substantiated. Finally, in Section 5, the added value and challenges of data-driven capacity planning are being discussed.

8.2 Data-driven performance measures

Before elaborating on performance measures associated with data-driven capacity planning, it is important to realize that there are two types of healthcare demand. The difference between the two is explained below.

Nursing homes are challenged to operate effectively and efficiently in an uncertain environment. A well-known definition of uncertainty is that of Galbraith [7, p. 5]. He defines it as: “the difference between the amount of information required to perform the task and the amount already possessed by the organization”. In other words: “if the task is well understood prior to performing it, much of the activity can be preplanned” [8, p. 28]. In practice, for some of the care activities it is possible, based on the individual needs and preferences of the clients, to make a fairly detailed schedule in advance. Examples of this type of activities are ‘giving medicine’ and ‘help with getting out of bed in the morning’. In this chapter, these activities will be referred to as deterministic care activities.

On the other hand, a nursing home should also have the responsiveness to handle what [8, p. 30] refers to as “non-routine, consequential events that cannot be anticipated and planned for in advance”. ‘Providing assistance with toileting’ is an example of a care activity which is carried out in response to random or unexpected demand. In this chapter, activities of this type will be referred to as random care activities.

Client-centered performance measures

As mentioned in the introduction, in order to live their lives according to their own daily routines, nursing home clients depend greatly on timely delivery of care and support. In this light, from a data-driven perspective, two prominent performance measures are:

- *The percentage of care tasks that should be conducted within Y minutes of the preferred delivery time (for deterministic care).* For example, a nursing home department may aim to conduct at least 95% of the deterministic care tasks within 15 minutes of the preferred delivery time.

- *The percentage of the care requests that should have a response time less than Y minutes (for random care).* For example, a nursing home department may aim to respond to at least 95% of the random care requests within 10 minutes.

Another common performance measure is the *average waiting time*. Although average waiting time can be a valuable measure, it should be used with care. The flaw of this performance measure is that possible fluctuations in waiting are not considered. For example, suppose that clients A, B and C have to wait respectively 19, 4 and 1 minutes in response to a random care request. In this case, the average waiting time is 8 minutes. Nevertheless, client A has to wait more than two times longer than the average! Hence, the average does not reveal the variability in the waiting time.

Capacity-centered performance measures

The nursing home clients are considered to be a prominent stakeholder, as they essentially are the customers of a nursing home organization. However, as mentioned in the introduction, nursing home care always takes place within a ‘client – care worker – organization’ triangle. As such, the perspectives of the care workers and organization are also of importance.

First of all, it should be realized that if clients have to wait a ‘backlog’ is created, where backlog is defined as: *the total amount of unfinished work due to not delivering the required care on time*. Large backlogs are undesirable because it increases the pressure on care workers, and leads to dejected clients as they lose grip on their daily routine (which may cause stress for the care worker as well). In addition, backlogs can lead to more work as a result of an increasing number of questions from clients.

From the perspective of the organization as well as the care worker, ’capacity utilization’ is considered to be an important performance measure. Capacity utilization indicates the relative amount of available capacity that is being used to supply the demand. Regarding the utilization of care workers, it is useful to divide the working time of care workers into four categories. First, care worker provide care; these are the client-related activities during which the care worker is present at the client’s site (or preparing or wrapping up). Second, there are administrative duties that are often related to the client. The third category concerns the time that care workers do not perform client-related care or administrative duties. This may include care workers waiting on purpose for ‘random demand’ of clients, or time that is left. Finally, the fourth category involves other moments when care workers are not performing any tasks, such as holidays, illness, training, meetings and paid breaks. In call centres, the fraction of time staff is unavailable to answer calls relative to the total capacity is called *shrinkage*. Koole [9], reports that a shrinkage of 40% is not exceptional. In practice, typically a fixed percentage for shrinkage is used. In this chapter, we restrict ourselves to the first and third category, i.e. the moments that care workers are present and available to perform care-related activities. Hence, the administrative tasks and shrinkage are not directly considered in this chapter. As such, we define the occupancy of a care worker as follows:

$$\text{Occupancy} = \frac{\textit{Total time spend on care tasks}}{\textit{Total time spend on care tasks and total idle time}} \text{ x } 100\%$$

In this context, we prefer the term occupancy over utilization to stress that the considered net capacity involves the time that care workers are present and available to perform care-related activities.

Optimization approaches

When it comes to data-driven capacity planning two type of optimization approaches can be applied. In the first approach, the available capacity is taken as starting point and the central question is: *how can we maximize the performance, given the available (amount of) capacity?* In the second approach, a pre-determined performance objective is taken as point of departure and the central question is: *how much capacity is (when) needed in order to achieve a pre-determined performance objective?* At the moment, the first approach is most prevalent in practice.

8.3 Elementary concepts

The literature on hospital logistics shows that the extent to which healthcare demand can be planned for is largely influenced by the following three interdependent concepts: (1) variability, (2) predictability and (3) scale [e.g. 10–13]. In the next subsections, it will be illustrated why and how these concepts are also of importance in relation to capacity planning in a nursing home context.

Variability

Nursing homes are challenged to meet the needs and preferences of their clients efficiently, despite variability (i.e., fluctuations) in demand over time. Without variability in demand, capacity planning would be a simple one-time exercise. However, in every-day nursing home practice, demand fluctuates from day to day and even from hour to hour, which makes capacity planning a challenging task.

Figure 1.2 shows the demand for care and/or support during a regular day (i.e., the number of clients in need of support) of a department within a Dutch nursing home facility. It can be seen that demand varies during the day. As most clients wake up between 7:00 and 10:00 hours and need help with getting out of bed and personal hygiene, a high demand can be observed during this time period. Furthermore, the figure also shows a peak in demand around 17:00 hours, due to an increased need for assistance with toileting. This increased need for assistance with toileting is caused by the regular afternoon tea at 16:00 hours.

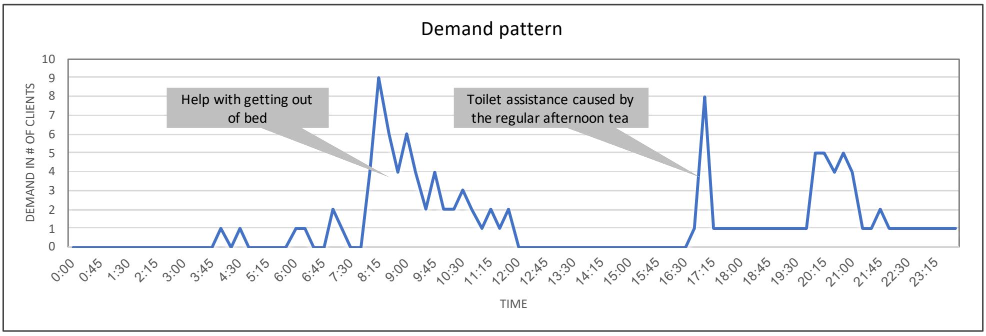


Figure 8.2: Variability in demand, based on [14]

According to [15], understanding and distinguishing between natural and artificial variability is key to improving healthcare processes. The high demand between 7:00 and 10:00 hours, due to help with getting out of bed and personal hygiene, is an example of so-called natural variability. Natural variability is inherent to the system and a direct result of the actual needs and preferences of clients.

On the other hand, the large need for assistance with toileting around 17:00 hours is an example of artificial variability as it is created by the way the system is set-up and managed. To be more specific, the peak in demand is caused by the afternoon tea which is served around 16:00 hours and is driven by personal preferences and priorities of the care workers (rather than actual needs and preferences of the nursing home clients). In most cases artificial variability undermines the effectiveness and the efficiency of healthcare systems and should therefore be eliminated. However, artificial variability can also be the result of well-considered decision-making. For instance, most nursing homes make use of so-called ‘general medicine rounds’. During such a medicine round, a single care worker wheels around a trolley to provide nursing home clients with the necessary medication. Providing medicine at fixed times is a well-considered choice as it has two major advantages. First, making a single care worker responsible for administering medication reduces the risk of interruptions, which can lead to medication errors. Secondly, pooling the provision of medication leads to a more efficient allocation of resources.

Predictability

If one can predict the variability in demand, it can be planned for. Here, predictability is defined as *the degree to which a correct prediction or forecast can be made regarding the healthcare activities required to meet the demand of the nursing home clients*. Distinguishing between predictable and unpredictable variability is relevant for nursing home managers and policymakers as the need for reactive decision making (i.e. responsiveness) increases and the potential for efficient and effective planning decreases when healthcare activities become less predictable.

Some of the variability in demand is largely predictable as it can be determined in advance. This type of variability, also referred to as deterministic variability, can be easily planned for.

Unfortunately, in practice, not all demand is fully known in advance. However, by analyzing historical data often regular patterns can be identified (i.e. systematic changes) like seasonality and trends. Such regular patterns can be planned for. The peak in demand during the early morning is an example of predictable variability. Hence, it can be expected that during the early morning most clients need help with getting out of bed and personal hygiene.

On the other hand, demand can also be unpredictable. The exact number of random care requests during a certain time-interval is an example of unpredictable variability in demand. However, although single random events are by definition unpredictable, in many cases the frequency of different outcomes over a large number of events shows relatively less fluctuation and can be described using probability distribution functions.

Scale

The planning of healthcare activities is also influenced by scale, where scale is defined as the level of aggregation of demand. This phenomenon has been studied extensively, for instance, in the context of inventory pooling [e.g., 17– 19]. The so-called pooling principle suggests that the relative variability is reduced when demand is aggregated. Essentially, this is due to the possibility that high demand from one client is balanced out by low demand from another client. Statistically, the advantage of pooling is the consequence of a reduction in relative variability as the standard deviation of the sum of two random variables is smaller than the sum of the two standard deviations (if the coefficient of correlation is smaller than 1). When variability decreases, demand becomes more predictable. In other words, predictability and scale are interrelated concepts. For example, [20] argues that, in general, healthcare activities are better predictable at a more aggregate level. Furthermore, the enlargement of scale increases the flexibility in planning.

> Given adequate numbers, even the emergency consultation or admission is predictable at an aggregate level, and can thus be planned for. It can be predicted how many patients will visit an outpatient clinic without a scheduled appointment each day, or how many emergency surgeries come to the hospital daily [20, p. 79].

Small scale living facility 1 (1 care worker)		
Client	Time preference	Bottleneck
1	9:30	
2	7:00	X
3	7:30	
4	7:00	X
5	10:00	
6	8:00	

Small scale living facility 2 (1 care worker)		
Client	Time preference	Bottleneck
7	9:30	
8	8:30	X
9	9:00	
10	10:00	
11	8:30	X
12	6:30	

Scheduling living facilities 1 & 2 together (2 workers)		
Client	Time preference	Bottleneck
12	6:30	
2	7:00	
4	7:00	
3	7:30	
6	8:00	
8	8:30	
11	8:30	
9	9:00	
1	9:30	
7	9:30	
5	10:00	
10	10:00	

Figure 8.3: Increased flexibility in planning due to an increase in scale

JUPITER CASE: Effect of scale

Figure 8.3 provides an example of the effect of scale on flexibility in planning. In each of the two small-scale living facilities of a nursing home department called Jupiter there are six clients in need of care. Each client has his or her own time preference concerning the delivery of morning care and there are only two care workers available. Furthermore, we assume that a care worker spends 30 minutes per client. When each of the two care workers is assigned to a single living facility, it is not possible to meet the time preferences of all clients. The peaks in demand of facilities 1 and 2 between 7:00 and 7:30 hours and 8:30 and 9:00 hours, respectively, cannot be satisfied with the existing capacity. In other words, there are clients who have to wait until the care worker is available to provide the necessary care and/or support. Moreover, a single delay can cause a chain reaction. In this example, small scale planning will result not only in waiting time for clients 2 or 4 and 8 or 11, but also for clients 3, 6, 7, 9 and 10. By merging the planning (i.e. schedule) of the two living facilities, it becomes possible to meet all of the individual time preferences.
Hence, the peaks in demand are balanced out. Advantages that result from carrying out a process on a larger scale are also referred to as 'economies of scale'.

These principles with respect to scale are of great practical importance, as there is a trend towards small-scale living facilities. Policy makers and nursing home managers should not blindly focus on creating small-scale living facilities, without taking potential economies of scale into account. Hence, creating small-scale living facilities should not become an end in itself [22].

8.4 Data-driven capacity planning: a four-step approach

Capacity planning takes place at different organizational levels. Often a distinction is made between strategic, tactical and operational level, which corresponds to the long, middle-long and short term. On each planning level decisions are made that influence the next planning level. Here, the focus lies primarily on capacity planning at a tactical level, i.e., the aim is to determine the required capacity levels for the next weeks or months. This is precisely the time scale that involves making plans that strive to balance between available capacity and variable demand. At a strategic level, it is of primary importance to maintain a properly sized and well composed workforce that meets future demand scenarios. Typically, this is heavily influenced by political and societal developments and the situation at the labor market. Moreover, over the course of the day changes in the usage of capacity may be required due to unforeseen circumstances. This requires real-time control at an operational level.

As mentioned, in this chapter the emphasis lies on the tactical aspects of capacity planning. Following the lines of workforce management in the call centre domain, see e.g. [23], the care-related capacity planning process can be divided into four stages: (1) workload prediction, (2) staffing, (3) shift scheduling and (4) rostering & tasks assignment. In the first stage (i.e., workload prediction) historical demand data is analyzed, with the aim of predicting the expected workload over time. The workload predictions form the basis for the second stage (i.e., staffing) where the focus lies on determining the corresponding staffing levels over time in order to meet the demand, thereby taking uncertainty into account. Next, in the third stage, the aim is to develop a shift schedule to meet the staffing levels from the second stage, without over-stretching the available staffing hours. Sometimes, the second and third stages are combined, which will often turn out to be convenient for a nursing home setting. From a methodological perspective, shift scheduling deals with the problem of determining the working shifts (start and end times, breaks, etc.), together with the assignment of the number and type of care workers to each shift. Finally, in stage four the focus lies on assigning care workers to specific shifts and tasks. More specifically, it deals with the following two questions: Which of the available care workers should be assigned to which shift(s)? And, which of the available care workers should perform which care tasks at which time of the day in order to meet the demand of the nursing home clients as closely as possible? Such task schedules are generated for short periods, i.e., days. We like to note that a nursing home may decide not to assign tasks to care workers in advance. In that case, tasks are being distributed among the team of care workers on the spot (i.e., in real time). Still, the first three steps ensure that supply and demand are sufficiently balanced. Figure 1.4 shows the four stages of the care-related capacity planning process.

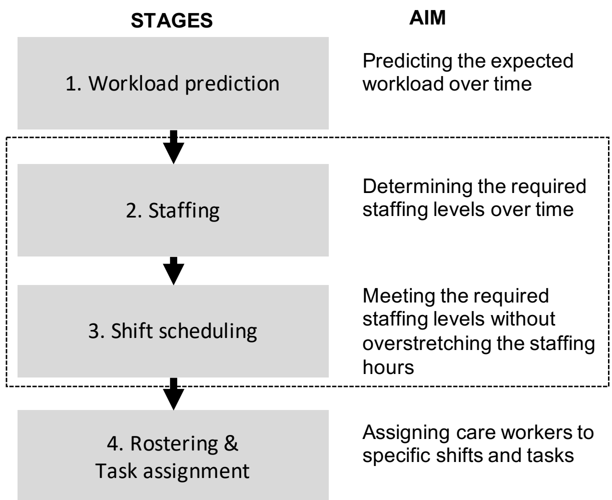


Figure 8.4: Stages of the tactical care-related capacity planning process

All stages are related as decisions at a higher-level stage form the input for the next stage. An optimal use of capacity requires decision making on all stages simultaneously. Due to the analytical complexity, such a fully integrated approach is not (yet) possible.

The result of the planning cycle above are rosters and task-schedules. This is, however, not yet the end of the story as unforeseen events might occur. Examples of such events are unavailability of care workers due to illness, holidays, and training, sudden changes in client needs, and administrative duties. This requires constant and real-time management of staffing capacity and assigning staff to activities. Also, the total available capacity should be sufficient to cope with shrinkage (see Section 8.2). As this involves a substantial amount of time, a good estimate of the fraction of time that care workers spend on non-care-related activities is crucial when determining the required amount of capacity. In addition, also the possibility of absenteeism should be paid attention to by creating some flexibility, e.g. by working with a flex pool. The organization of such kind of flexibility is out of scope of this chapter.

Finally, it should be noted that for the delivery of care and support, most nursing homes make use of the so-called differentiated practice. Based on their education and expertise, care workers are hierarchically divided into four distinct qualification levels (QLs). Depending on the required education and complexity of care, healthcare tasks are assigned to a healthcare worker with that specific qualification level. For comprehensibility reasons, in this chapter, qualification levels are not taken into account.

Workload prediction

As the role of capacity planning is to balance capacity with demand, it is crucial to have insight in how the demand behaves over time. Although this may seem evident, it should be envisaged that insight in healthcare demand is often lacking in nursing home settings. Availability of data (or the lack of it) is a key element here. Below, a distinction is made between 'deterministic demand' and 'random demand' (see also Subsection 1.2).

First, during the day, the majority of the demand is related to activities of daily living (ADLs). Although there may be strong fluctuations in ADL related demand, it is possible to collect the time preferences of each client regarding ADLs and estimate the corresponding care durations. Combining the time preference of an activity with the duration provides an estimate of the workload, i.e., *the number of care workers required to meet the demand*. In practice, regarding ADLs, we see that there are peaks in the workload especially during the morning and, to a smaller extent, in the evening (see e.g., [16, 21, 24]).

Client	Preferred	Duration	End
12	06:30	15	06:45
2	07:00	10	07:10
4	07:00	25	07:25
3	07:30	30	08:00
6	08:00	15	08:15
8	08:30	45	09:15
11	08:30	70	09:40
9	09:00	45	09:45
1	09:30	30	10:00
7	09:30	30	10:00
5	10:00	15	10:15
10	10:00	30	10:30

Figure 8.5: Care activities, time preferences and expected durations

JUPITER CASE: Workload prediction

The nursing home department Jupiter (see also Section 8.3) aims to deliver the necessary care and support as close as possible to the time preferences of the residents. Figure 8.5 provides an overview of the time preferences regarding ADL care activities for each of the 12 nursing home residents. In addition, for each of the required care activities the care duration is estimated. Based on these data, the manager of Jupiter has made a prediction of the workload. This workload prediction is shown in Figure 8.6. In this example the expected workload fluctuates between 1 and 4.

Second, there is ‘random demand’. Data involving random demand is scarce, as many organizations do not register such activities. However, some nursing homes make use of a call button system in which every call button request is registered automatically in a central database. In addition, all care workers are equipped with a key card. Every time a care worker enters or leaves the room of a resident, the key card is swiped along an electronic keypad, registering the timestamp and the location. If such a system is not available, then an estimate is the best that can be achieved, which should be calibrated based on practical experience.

Staffing

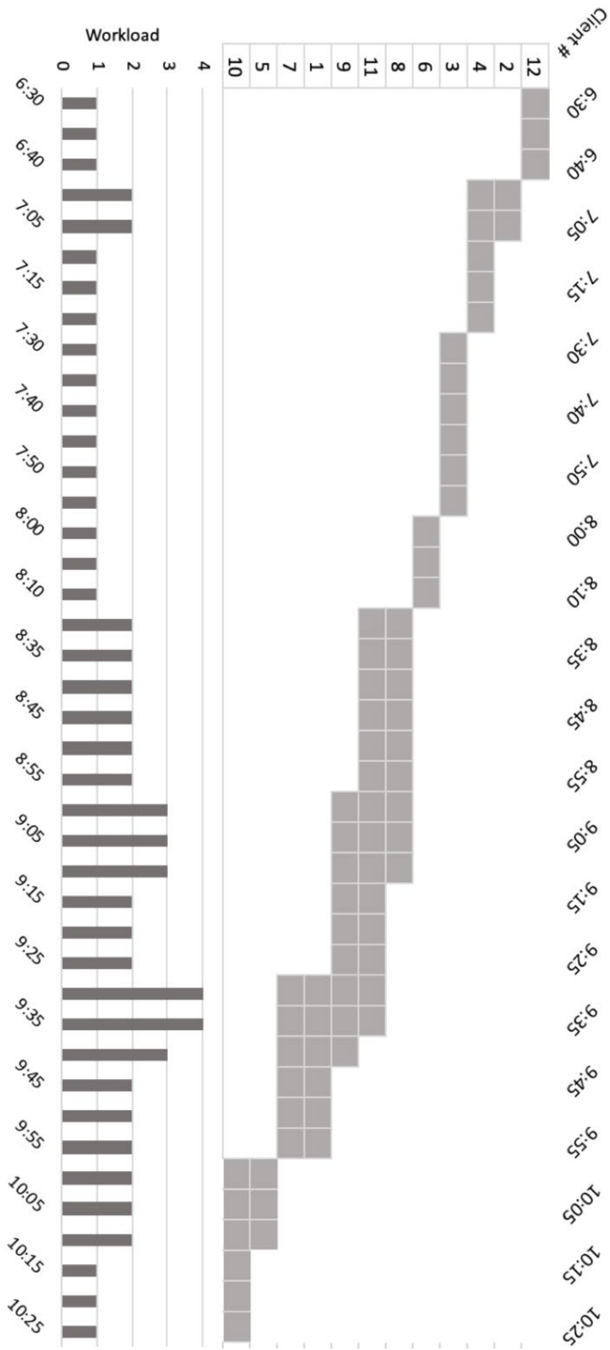
Given the workload predictions, the next step is to determine the corresponding staffing levels over time. Again, a difference should be made between deterministic and random demand. If demand is fully deterministic, the ideal staffing level is equal to the workload. Hence, in that idealized situation, staff is fully utilized (occupancy = 100%), whereas there is sufficient staff to meet demand. The problem is that in practice the workload fluctuates widely (see e.g. Figure 8.2), and the staffing capacity is not flexible enough to follow such a pattern. This issue can be partly covered by shift scheduling in the next stage. In this case, staffing and shift scheduling should be a one-step approach, see also Section 8.4.3.

For random fluctuations in demand, some slack capacity is required. Using estimates for the average number of care activities and their average durations, it is possible to estimate the performance (waiting time and occupancy) using queueing models. A particularly useful model that is applied in a variety of service settings is the celebrated M/M/s (or Erlang delay) model. This model requires the arrival rate (i.e. rate at which new care activities occur) and mean duration of an activity and gives, for a given number of care workers, the waiting time (both in terms of its average and the service level). The performance is reasonably robust in its assumptions and easy to apply, e.g. due to online calculators. The underlying assumptions involve a Poisson arrival process of random care activities and exponential durations of such activities. The Poison arrival assumption is shown to hold in many cases for uncontrolled arrival requests. Also, the model is fairly insensitive to the exponential distribution of care durations, as long as the coefficient of variation of these durations is close to one. Regarding a nursing home setting, [16] showed that the M/M/s model is applicable to determine the staffing during the night.

In Figure 8.7 the impact of the occupancy on the waiting time is visualized. Specifically, it shows the fraction of clients that wait more than 10 minutes. Based on [16], it is assumed that random care activities take 3 minutes on average. The number of care workers is 2, representing staffing overnight. It can be observed that a fast drop in performance occurs when the occupancy is getting close to 100%. This reveals that in situations with random demand, high occupancy is undesirable, as it will lead to excessive waiting. Indirectly, this also influences the wellbeing of care workers and the associated increase in risk of mistakes. In this example, an occupancy of 80% will lead to about 20% of the clients waiting more than 10 minutes, which seems a reasonable choice. Of course, this concerns a managerial trade off that is also highly affected by the available budget. From a different angle, it could be argued that the results of the queueing analysis should provide the starting point for decisions about available budgets. The shape of the curve in Figure 8.7 is common for this type of queueing systems. However, it should be stressed that the actual values depend on the assumptions and parameters involved. Hence, it is not possible to give a single occupancy level that is appropriate for all systems.

The influence of parameters is visualized in Figure 8.8, where the scale of the system is being varied. Again, it is assumed that random care activities take 3 minutes. Furthermore, an 80/10 service level is used as `golden standard'; the aim is to respond to at least 80% of the random care requests within 10 minutes. The figure shows that if the scale increases, in terms of the number of care workers, occupancy also increases. This reflects economies of scale, where some fluctuations cancel out and the system becomes more efficient (see also 8.3.3). This gain in efficiency is most pronounced for small systems. This implies that organizing care at a small scale may have serious consequences in terms of efficiency. Unless some form of flexibility in the use of capacity is organized, it is recommended to avoid such small-scale settings as much as possible.

Figure 8.6: Workload prediction based on Figures 8.3 and 8.5 with individual (top) and aggregate (bottom) workloads



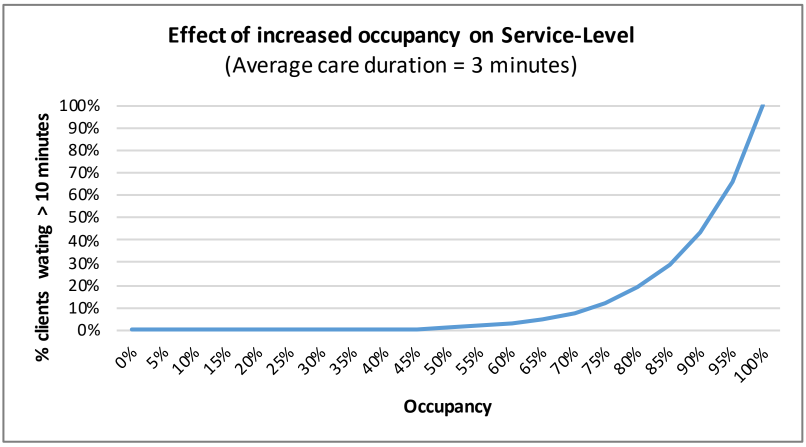


Figure 8.7: Fraction of clients waiting at least 10 minutes as a function of occupancy

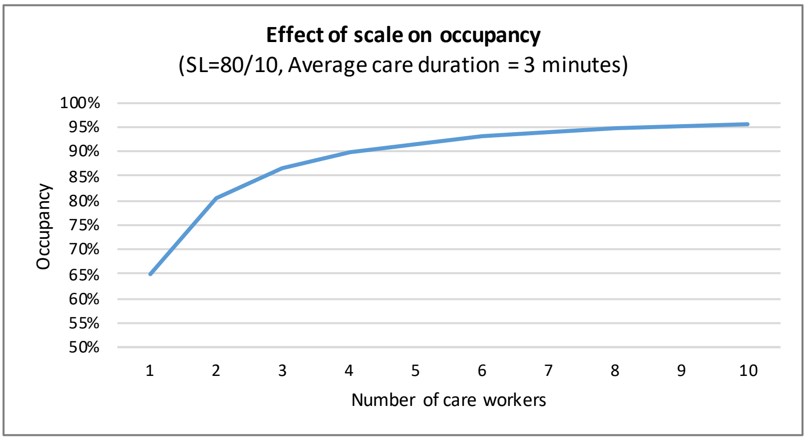


Figure 8.8: Impact of scale for a SL where 80% of the clients wait at most 10 minutes

Shift scheduling

In the stage of shift scheduling, the goal is to translate the staffing levels over the course of a day, that are determined at the previous stage, into shifts that should be carried out by care workers. A shift here specifies the start and end of the working time during a day, and possibly the breaks (which we neglect here). Again, there is a dichotomy between deterministic and random demand. For deterministic demand, it will typically not be desirable to be able to instantly meet all demand; this will lead to inefficient use of staffing capacity due to the rather strong fluctuations in demand compared to the length of a shift. As such, the staffing and scheduling stages should be integrated. Random demand usually has a clearly more stable pattern over the course of a day (including the night). For such demand, it is desirable to have slack capacity at (almost) any moment in time. This setting is much more comparable to traditional service systems, see e.g. [25–27] for some shift scheduling approaches.

Shift scheduling for deterministic demand will be discussed first. More specifically, the following question is considered: which shifts should be carried out to optimize performance? From the perspective of a nursing home client, this means waiting should be avoided as much as possible, where waiting is defined as: *the difference between actual starting time of a care activity and the preferred starting time*. As the waiting time is difficult to determine in an analytical fashion, 'backlog' is used as a proxy (see Section 1.2). In this light, it should be pointed out that the backlog may also affect the wellbeing of the nursing staff, further supporting this choice.

JUPITER CASE: Constant staffing

As many nursing homes, Jupiter tries to stabilize the available capacity. During the period between 6:30 and 10:30, Jupiter works with 4 shifts of 2 hours each. Stabilizing capacity then implies that there are two shifts working from 6:30 – 8:30 and two shifts from 8:30 – 10:30. From the workload prediction in Figure 8.6, it may be observed that during the early shifts there is quite some overcapacity; in fact, the occupancy rate will only be 40% if they adhere to the time preferences of the clients. During the last two shifts, things become tight. In fact, even though both care workers will be constantly busy with providing care and support, they will run 20 minutes short in time. Moreover, clients 9, 7, 1, 5, and 10 will have to wait yielding a total waiting time of 90 minutes.

In the case of Jupiter, it can be observed that constant staffing gives rise to high working pressure and long waiting times during the day.

Next, a mathematical optimization model is used to determine the required shifts. To get some intuition for the approach, first the optimization problem is presented; a more general version of the model can be found in [24]. In this example, it is allowed to start the 2-hour shifts at every multiple of 15 minutes. That means that potential starting times are 6:30, 6:45, 7:00, . . ., 8:30. This provides 9 types of shifts. Let x_k be the number of care workers that are scheduled for shift type k, with $k = 1, 2, \ldots, 9$, representing the different starting times; note that x_k will be zero when this shift is not carried out. To determine the number of care workers that are working at any moment, a_{tk} is needed, where $a_{tk} = 1$ if shift type k works during interval t, and $a_{tk} = 0$

otherwise. Here, $t = 1, \ldots, 48$ represents all five-minute intervals between 6:30 and 10:30. The staffing level during interval t is then $c_t = \sum_{k=1}^{9} x_k a_{tk}$, for $t = 1, 2, \ldots, 48$.

To determine the backlog, it is assumed that a team of care workers can all jointly work when there is any workload (this may not always be the case when the workload is due to a single client). The backlog during interval t, denoted by q_t, can then be recursively defined, as follows:

$$q_{t+1} = \max\{q_t + L_t - c_t, 0\},$$

where L_t is the new amount of work in interval t that follows from the workload prediction of the first stage, and c_t is the staffing capacity following from the shift schedule x_k. Observe that there is an almost linear relation between q_{t+1} and q_t (up to the $\max\{\cdot, 0\}$ operator); this can be exploited in the formulation. Specifically, we can now formulate the shift scheduling problem as a Mixed Integer Linear Programming problem, as follows:

$$\begin{aligned}
\text{Minimise} \quad & \sum_{t=1}^{48} q_t & & \quad (1.1)\\
\text{subject to} \quad & c_t = \sum_{k=1}^{9} x_k a_{tk}, & t = 1, \ldots, 48 & \quad (1.2)\\
& q_{t+1} \geq q_t + L_t - c_t, & t = 1, \ldots, 47 & \quad (1.3)\\
& \sum_{k=1}^{9} x_k = 4 & & \quad (1.4)\\
& c_t \geq 1, & t = 1, \ldots, 48 & \quad (1.5)\\
& q_t \geq 0, & t = 1, \ldots, 48 & \quad (1.6)\\
& x_k \in \mathbb{N}_0, & k = 1, \ldots, 9 & \quad (1.7)
\end{aligned}$$

Equation (1.1) reflects that the goal is to minimize the total backlog. Equation (1.2) gives the staffing capacity at time t in terms of the shifts. Equation (1.3), jointly with the minimization, provides the recursive relation between the backlogs at successive intervals. Equation (1.4) provides that there are only 4 shifts available, whereas (1.5) guarantees that at any moment at least 1 care worker is available. Finally, (1.6) provides that backlogs are non-negative and (1.7) makes sure that an integer number of each shift type is scheduled (including 0). From this formulation, it may be observed that alternative choices in the formulation are possible.

Finally, it should be noted that the performance of shifts in terms of waiting times can only be determined after the activities are assigned to care workers. Doing this in an efficient way is the goal of the next stage.

JUPITER CASE: Shift scheduling

Using the shift scheduling formulation of Equations (1.1)–(1.7), it turns out that it is best to use the shifts as presented in the table below:

Number of shifts	Start time	End time
1	6:30	8:30
1	8:00	10:00
2	8:30	10:30

Using these shifts, the peak in demand between 9:00 and 9:40 can be handled much better. As there is only one care worker between 6:30 and 8:00, clients 4 and 3 now have to wait (which was previously not the case). However, the total waiting time decreases to 35 minutes, and there is no work remaining at all at 10:30.

Rostering & task assignment

In the fourth step of the capacity-planning process, care workers need to be assigned to shifts such that each care worker has a roster. In addition, if desired, care workers should be assigned to specific activities. So far, the steps in the framework of the capacity planning process (Figure 8.4) can be independently carried out for each day separately. This also applies for assigning activities to staff members. Rostering of staff, on the other hand, should be carried out for consecutive days, weeks, or sometimes even months. In that sense, the time horizon of the rostering process deviates from the other steps.

Composing a roster or timetable for the personnel is a classical topic in workforce management that appears in many industries and organizations. There are different ways to organize this (see also [9]). First, it is possible that staff provides preferences and unavailability beforehand. The rostering is then a complex scheduling or timetabling problem, requiring an advanced optimization algorithm to determine a timetable that attempts to satisfy all constraints and preferences of the staff as much as possible. We refer to [28] for an overview of scientific literature in this area. A disadvantage of this approach is that it is often difficult to quantify the perceived quality of a roster, see e.g. [29]. A second option is that staff can choose their shifts on some auction system (see e.g. [30]). Third, assignment of nurses to shifts can be done on the basis of a self-rostering system. This approach has the advantage that the staff has direct control over their schedule and can therefore balance their personal and professional lives. This approach is of course more of an organizational nature than a quantitative approach. See e.g. [31] for a reported pilot on self-rostering.

The assignment of tasks to care workers is more specific for a nursing home setting. Specifically, the 'route' of each care worker should be decided upon, meaning that each activity should be assigned to one care worker next to the starting time of the activity. This problem can be viewed as an unrelated parallel machine scheduling problem, but the performance measure crucially differs. In machine scheduling problems, the performance is typically in terms of the make-span (i.e. time the last activity finishes), whereas we are interested in the waiting time of clients (i.e. delay in starting activity compared to the preferred activity time). Another related area is home care scheduling, where care workers visit clients at home preferably within a time window indicated by the client. An interesting example of a home care task assignment problem involving time windows is [32]. From the literature it becomes evident that the task assignment problem is NP-hard, implying that large (real life) instances cannot be solved to optimality. Compared to home care, an advantage of nursing homes is that distances, and thus travel times, between clients are much smaller. As such, it seems easier to develop a fast task assignment heuristic with reasonable performance.

Let us now specify a simple and greedy task assignment heuristic. With such a heuristic it is also possible to determine waiting times for activities (see the end of Algorithm 1). Moreover, it may provide an initial solution for an optimization procedure. The greedy heuristic, shown in Algorithm 1, starts at 6:30 and iterates through time, here in time steps of 5 minutes. At every time instant, the algorithm first updates which activities finish, then updates the status of the care workers (including possible starting and ending of shifts), and finally determines which activities are allowed to start and by which care worker this is carried out. This final step is the most interesting. The activities will be inspected in a particular order indicating their priority. This priority can be influenced by sorting the preferred activity schedule. An appealing way to sort activities is by:

1. Preferred starting time: activities with an earlier preferred starting time should start first, i.e., First Come First Served (FCFS).
2. Short activities first, as this will result in less overall waiting, i.e., Shortest Job First (SJF).

Next, if multiple care workers are available, it needs to be determined which care workers will take care of the next activity. In Algorithm 1, this is indicated by a *Select care worker()* procedure. A simple rule would be to assign an activity to an arbitrary care worker or to the care worker that is idle for the longest time. However, it may involve more complex rules taking e.g. the remaining length of the shift, or the amount of work carried out so far into account. For instance, next to availability, a possible requirement is that a care worker may only start an activity if the activity will finish before the end of a shift (or by the care worker that exceeds the end of the shift by the smallest amount). Another assignment rule could be to assign an activity to the care worker with the smallest occupancy so far. Of course, mixtures of such rules can also be constructed, depending on desired criteria.

Another option for improvement is to start some activities before their preferred activity time, in order to avoid delays later on. To optimize the task assignment, the preferences between waiting and earliness should be quantified; for example, all deviations from the preferred activity time may be valued equally bad. In that case, an optimal task assignment should minimize the deviations from the preferred activity time, by assigning activities to care workers and determining their starting time. Another performance measure might be to balance the workload among care workers (or a mixture of these performance measures). Due to complexity of the optimization problem, we typically have to rely on heuristic solution methods for real life instances; see [33] for an example of an advanced exact solution method for home care scheduling and routing. Another interesting study in the home care scheduling area is [32], which develop a local search heuristic that tries to improve solutions by searching for better solutions in the neighbourhood of the current one. See e.g. [34] for a review of routing and scheduling in the home care setting. In particular, it becomes clear that there is a variety of (heuristic) solution procedures for related problems.

Algorithm 1 Greedy task scheduling algorithm

```
initialize   care workers status, waiting list
sort preferred activity schedule                ▷ use heuristic for smart sorting
for t = 6:30, 6:35,...,10:30 do
    if activities finish at t then finish these activities
    update care worker status                  ▷ incl. starts and ends of shifts
    if there are activities that may start at t then
        if there are idle care workers then
            Select care worker()               ▷ may involve different heuristic rules
        else
            add activity to waiting list
    update care worker status & waiting list
calculate performance measures
```

JUPITER CASE: Task assignment

To get a good feeling for the impact of the planning process, Jupiter uses three different scenarios. The first two scenarios use a greedy task assignment procedure, whereas the third scenario is based on an optimized approach. Moreover, the first scenario uses constant staffing, whereas the second and third scenario use the optimized staffing levels.

Scenario	Staffing	Task assignment	Waiting time (min.)	Earliness (min.)	Overtime (min.)
1	Constant	Greedy	90	0	20
1	Optimized	Greedy	35	0	0
2	Optimized	Optimized	5	15	0

From the results, we see that the largest gain in performance is achieved by determining the shifts appropriately. There is potential for optimizing task assignment, however the potential gains (in particular in practice) are smaller.

Finally, it should be mentioned that the task assignment procedure assigns activities to care workers in advance. This is possible when there is hardly any random demand during the considered time window or when there are separate care workers that handle random demand. If this do not apply, then the nature of the optimization problem changes. Moreover, it is then questionable what the added value of a predetermined task assignment actually is. The random demand will lead to changes in the assignment that was constructed in advance, and will require real-time updating of plans. Although the real-time control of such a system may give rise to interesting optimization problems, embedding real-time optimization approaches in software for nursing homes is considered to be a solution for the more distant future.

8.5 Value and challenges of data-driven capacity planning

Recent studies show that nursing homes could greatly benefit from applying data-driven capacity planning approaches. A good example is the study of [24] in which the potential gain of data-driven shift scheduling for three independent nursing home departments of single Dutch nursing home has been investigated. Figure 8.9 shows the main numerical results of this study. For all three departments substantial improvements can be observed, both in terms of average waiting time as well as in service level. The waiting time across the day, for each of the three departments, is visualized in Figure 8.10. As expected, waiting reduces significantly during rush hours, whereas there is only a slight increase in waiting time during non-rush hours. For example, during the morning rush hour the maximum average waiting times for departments C, D and E drop from 35, 45 and 35 minutes to around 12, 15 and 8 minutes, respectively. Furthermore, as the proposed staffing pattern is more balanced, it created a more evenly spread workload for the care workers. The added value of a well-balanced workload should by no means be underestimated in the light of care workers' wellbeing.

An important first step towards data-driven capacity planning would be to introduce objective performance measures regarding timely delivery of care and support. Consequently, nursing homes should have a sufficient information system to make it possible to work with those types of performance measures. However, in practice, there is a lack of reliable and valid data. Fortunately, due to developments in technology (e.g., ICT support for domestic tasks, robotics and registration systems), data generation is increasing rapidly. As such, an important future challenge will be to transform these data into tools that support decision making. This will be a challenging task as nursing home processes have many complex characteristics and research on nursing home operations from an applied mathematical perspective is still in its infancy.

Furthermore, in order to implement data-driven capacity planning, sufficient staffing flexibility must be ensured. In this case, a distinction should be between numerical and functional flexibility. Numerical flexibility can be defined as the ability of teams to adjust the number of workers, or the level of worked hours, in line with changes in the level of demand for them [35]. Numerical flexibility could, for example, be achieved by creating a flex pool. A flex pool consists of care workers who are 'on call' and available for work as and when required. Supplementing a core team of full-time care workers with flex pool workers allows nursing home managers to balance their staffing levels better over the course of a day. Functional flexibility can be defined as internal flexibility and refers to the ability of care workers to perform a broader range of tasks, which makes it possible to assign them to different tasks and jobs [35].

Current	Dep. C	Dep. D	Dep. E
Av. waiting time	9.59 minutes	14.49 minutes	10.34 minutes
15 min. SL	78.5%	70.9%	77.4%
Optimal	**Dep. C**	**Dep. D**	**Dep. E**
Av. waiting time	2.89 minutes	7.20 minutes	3.37 minutes
15 min. SL	96.0%	83.5%	95.0%

Figure 8.9: Overview main results study of [24]

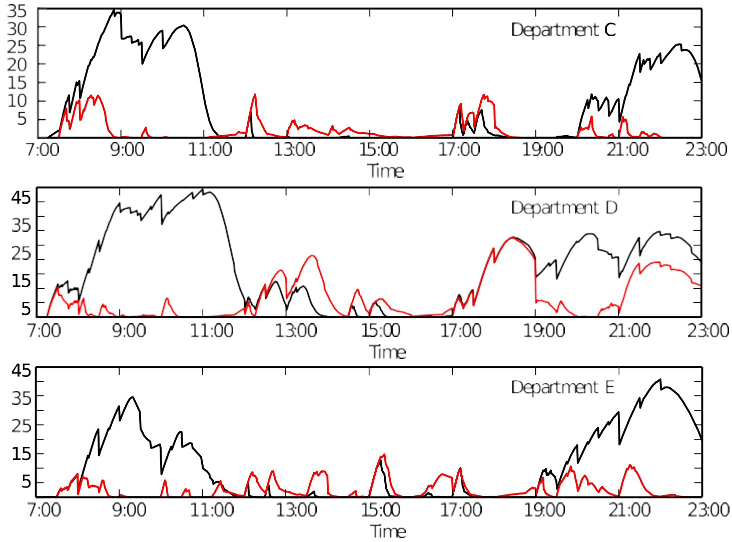


Figure 8.10: Waiting time (in minutes) during the day: current (black) versus optimal (red) for departments C, D and E

Bibliography

[1] A.M. Sanford, M. Orrell, D. Tolson, A.M. Abbatecola, H. Arai, J.M. Bauer, A.J. Cruz-Jentoft, B. Dong, H. Ga, A. Goel, et al. An international definition for "nursing home". *Journal of the American Medical Directors Association*, 16(3):181–184, 2015.

[2] J.P.H. Hamers. *De intramurale ouderenzorg: nieuwe leiders, nieuwe kennis, nieuwe kansen; achtergrondstudie geschreven op verzoek van de Raad voor de Volksgezondheid en Zorg*. Maastricht University, Maastricht, 2011.

[3] E.P Jack and T.L. Powers. A review and synthesis of demand management, capacity management and performance in health-care services. *International Journal of Management Reviews*, 11(2):149–174, 2009.

[4] L. Di Giorgio, M. Filippini, and G. Masiero. Implications of global budget payment system on nursing home costs. *Health policy*, 115(2):237–248, 2014.

[5] G.S. Rai. Burnout among long-term care staff. *Administration in Social Work*, 34(3):225–240, 2010.

[6] L. Kuntz, R. Mennicken, and S. Scholtes. Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science*, 61(4):754–771, 2014.

[7] J.R. Galbraith. *Designing complex organizations*. Addison-Wesley Long-man Publishing Co., Boston, MA, 1973.

[8] J.R. Galbraith. Organization design: an information processing view. *Interfaces*, 4(3):28–36, 1974.

[9] G. Koole. *Call center optimization*. Lulu. com, 2013.

[10] P. Joustra, E. Van der Sluis, and N.M. Van Dijk. To pool or not to pool in hospitals: a theoretical and practical comparison for a radiotherapy outpatient department. *Annals of Operations Research*, 178(1):77–89, 2010.

[11] A.M De Bruin, A.C. Van Rossum, M.C. Visser, and G.M. Koole. Modeling the emergency cardiac in-patient flow: an application of queuing theory. *Health Care Management Science*, 10(2):125–137, 2007.

[12] J. Van Oostrum. *Appplying mathematica models to surgical patient planning*. PhD thesis, Erasmus University Rotterdam, 2009.

[13] R.E.G. Upshur, R. Moineddin, E. Crighton, L. Kiefer, and M. Mamdani. Simplicity within complexity: seasonality and predictability of hospital ad-missions in the province of Ontario 1988–2001, a population-based analysis. *BMC Health Services Research*, 5(1):13–19, 2005.

[14] D. Moeke and L. Verkooijen. Doing more with less: a client-centred approach to healthcare logistics in a nursing home setting. *Journal of Social Intervention: Theory and Practice*, 22(2):167–187, 2013.

[15] E. Litvak and M.C. Long. Cost and quality under managed care: irreconcilable differences. *The American Journal of Managed Care*, 6(3):305–312, 2000.

[16] K. Van Eeden, D. Moeke, and R. Bekker. Care on demand in nursing homes: a queueing theoretic approach. *Health Care Management Science [Advance online publication]*, 2014.

[17] H. Yang and L. Schrage. Conditions that cause risk pooling to increase inventory. *European Journal of Operational Research*, 192(3):837–851, 2009.

[18] S. Benjaafar, W.L. Cooper, and J.S. Kim. On the benefits of pooling in production-inventory systems. *Management Science*, 51(4):548–565, 2005.

[19] G.D. Eppen. Note-effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, 25(5):498–501, 1979.

[20] M. Berg, W. Schellekens, and C. Bergen. Bridging the quality chasm: integrating professional and organizational approaches to quality. *International Journal for Quality in Health Care*, 17(1):75–82, 2005.

[21] D. Moeke, R. Van de Geer, G. Koole, and R. Bekker. On the performance of small-scale living facilities in nursing homes: a simulation approach. *Operations research for health care*, 11:20–34, 2016.

[22] D. Moeke. Towards high-value (d) nursing home care: Providing client-centred care in a more efficient manner. *Available at SSRN 3116326*, 2016.

[23] N. Gans, G. Koole, and A. Mandelbaum. Telephone call centers: a tutorial and literature review. *Manufacturing and Service Operations Management*, 5(2):79–141, 2002.

[24] R, Bekker, D. Moeke, and B. Schmidt. Keeping pace with the ebbs and flows in daily nursing home operations. *Health care management science*, 22(2):350–363, 2019.

[25] A.T. Ernst, H. Jiang, M. Krishnamoorthy, and D. Sier. Staff scheduling and rostering: A review of applications, methods and models. *European journal of operational research*, 153(1):3–27, 2004.

[26] T. Aykin. Optimal shift scheduling with multiple break windows. *Management Science*, 42(4):591–602, 1996.

[27] S. Bhulai, G. Koole, and A. Pot. Simple methods for shift scheduling in multiskill call centers. *Manufacturing & Service Operations Manage-ment*, 10(3):411–420, 2008.

[28] E. K. Burke, P. De Causmaecker, G. V. Berghe, and H. Van Landeghem. The state of the art of nurse rostering. *Journal of scheduling*, 7(6):441–499, 2004.

[29] D. Ouelhadj, S. Martin, P. Smet, E. Ozcan, and G. Van den Berghe. Fair-ness in nurse rostering. 2012.

[30] M.L. De Grano, D.J. Medeiros, and D. Eitel. Accommodating individual preferences in nurse scheduling via auctions and optimization. *Health Care Management Science*, 12(3):228, 2009.

[31] L. Bailyn, R. Collins, and Y. Song. Self-scheduling for hospital nurses: an attempt and its difficulties. *Journal of Nursing Management*, 15(1):72–77, 2007.

[32] D.S. Mankowska, F. Meisel, and C. Bierwirth. The home health care routing and scheduling problem with interdependent services. *Health Care Management Science*, 17(1):15–30, 2014.

[33] M.S. Rasmussen, T. Justesen, A. Dohn, and J. Larsen. The home care crew scheduling problem: preference-based visit clustering and temporal dependencies. *European Journal of Operational Research*, 219(3):598–610, 2012.

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

[35] J. Atkinson. Flexibility or fragmentation? The United Kingdom labour market in the eighties. *Labour and Society*, 12(1):87–105, 1987.

[36] C.A. Rodriquez-Osorio and G. Dominguez-Cherit. Medical decision making: paternalism versus patient-centered (autonomous) care. *Current opinion in critical care*, 14(6): 708-713, 2008.

[37] S. Tønnessen, P. Nortvedt, and R. Førde. Rationing home-based nursing care: professional ethical implications. *Nursing ethics*, 18(3): 386-396, 2011.

[38] E.R. Greenglass, R.J. Burke, and L. Fiksenbaum. Workload and burnout in nurses. *Journal of community & applied social psychology*, 11(3): 211-215, 2001.