# Towards High-Value(d) Nursing Home Care

Moeke, D.

2016

Chapter 1

# General introduction

The first section of this chapter is intended to place the work presented in this thesis in a broader societal context. Section 1.2 provides an overview of the main contributions of this thesis. Finally, this chapter concludes with a brief outline of the chapters that follow.

## 1.1 Background

Like many other Western countries, the Netherlands is plagued by increasing financial pressure on its healthcare system (Van der Horst et al., 2011). According to former finance minister Jan Kees de Jager, the rising cost of healthcare is the matter of most concern to the Dutch treasury:

> We will survive the European debt crisis. And the 18 billion euros in cutbacks that this cabinet wants to carry out will also succeed. But my biggest concern is ever-growing healthcare costs. The costs of healthcare have risen by four percent per annum over the past decades, where the average GDP growth has been two percent. This is mathematically unsustainable. (Netherlands Info Service News Bulletin, 2015).

Also the large and growing share of public spending on long-term care (LTC) has become a central issue for the Dutch government (Van der Horst et al., 2011). According to the World Health Organization (WHO), long-term care can be defined as:

> The system of activities undertaken by informal caregivers (family, friends and/or neighbors) and/or professionals (health and social services) to ensure that a person who is not fully capable of self-care can maintain the highest possible quality of life, according to his or her individual preferences, with the greatest possible degree of independence, autonomy, participation, personal fulfillment and human dignity. (World Health Organization, 2000, p. 6)

In 2010, Dutch public spending on LTC accounted for 3.7% of GDP[1], which was the highest among the OECD countries (OECD/ European Commission, 2013). Between 2009 and 2012, the spending on LTC rose by more than 20% and on average accounted for more than 40% of the total public healthcare spending (Dutch Healthcare Authority, 2013). Recent projection scenarios show a non-negligible increase in public spending on LTC over the forthcoming decades (Schut et al., 2013; Lipszyc et al., 2012; Van der Horst et al., 2011). It is estimated that, under current policies, long-term care expenditure in the Netherlands will reach 8.1% of GDP by 2060, which is more than three times the EU average (Schut et al., 2013). However, it should be noted that projections of demographic, economic and political developments are surrounded by considerable uncertainty and "the longer the projection period, the higher is the degree of uncertainty, especially in the domain of health and disability trends" (Lipszyc et al., 2012, p. 7).

Long-term care is provided both in institutions (residential care) and in communities (home care). In this thesis, the primary focus is on institutional care provided in nursing homes[2]. A nursing home is "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" (Sanford et al., 2015, p. 183). In 2013, Dutch public spending on nursing home care accounted for more than 40% of the total public spending on LTC (VWS, 2015).

In an attempt to reduce the growth rate of long-term care expenditures, without compromising quality, radical reforms were introduced by the Dutch government at the start of 2015 (VWS, 2014; Ministry of Economic Affairs, 2014). Under these reforms, the responsibility for most formal long-term care

[1] Gross Domestic Product

[2] In this thesis no distinction is made between nursing homes and residential homes (less intensive care) as, in practice, the boundary between the two is diffuse (Hamers, 2011).

services has been decentralized and taken over by local authorities. New clients requiring lighter types of care no longer receive an indication for admitted patient care. Instead, they will receive care at home. Consequently, on average, the care needs of nursing home clients are likely to become more severe as healthier people are drawn away from nursing home facilities.

Together with ongoing budget cuts, these reforms put serious pressure on Dutch nursing homes, which face the challenge of providing high-quality, cost-efficient care in a rapidly changing healthcare landscape. When it comes to nursing home care, the relevance of 'quality of care' differs from many other service settings. Nursing home clients are often 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 preferences, nursing home clients depend greatly on being provided with the opportunity to influence the delivery of their own care. In other words, an important measure of quality of care in a nursing home context is the extent to which the needs and preferences of the clients are being met. According to KPMG (2013, p. 18), "the design and delivery of care must focus on the needs of the individual rather than" as has often been the case "on the systems and procedures of the provider". This view on quality is in line with a client-centred[3] care approach, in which clients' wants, needs and preferences are respected and acted on and where clients are autonomous and able to decide for themselves (e.g., Kohn et al., 2001; Coulter, 2002). From this perspective, quality of care in a nursing home, largely depends on the coordination and timing of service delivery (Nies et al., 2010). In practice, when it comes to the coordination and timing of service delivery, 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 team 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". Fortunately there is still room for improvement. According to KPMG (2013) and Hamers (2011), the long-term care sector could benefit enormously from further research.

[3]In this thesis the term 'client' or 'resident' instead of 'patient' is used to designate the recipient of nursing home services.

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 1.1:** Real-life letter

It is commonly believed that insights, concepts and tools from the field of logistics can contribute to achieving high-value care (e.g., Bakker, 2004; Vissers and Beech, 2005; Nachtmann and Pohl, 2009; Verkooijen, 2006), where value refers to outcomes relative to costs (Porter, 2010). The modern interpretation of logistics has its origins in military operations, where it was used to describe the activities related to the efficient and effective distribution and storage of supplies and personnel. Dr William Müller, a public instructor of military science at the University of Gottingen, is believed to be the founder of modern logistics management[4]. He used the term 'logistics' in 1811 in his book titled *The Elements of the Science of War* (Müller, 1811), which is considered to be the first logistics textbook. In the decades that followed, the use of logistics gradually spread to cover business activities, often referred to as logistics management. Logistics management is defined by the Council of Supply Chain Management Professionals as:

[4]http://www.supplychainopz.com/2013/05/logistics.html

> That part of supply chain management that plans, implements, and controls the efficient, effective forward and reverse flow and storage of goods, services and related information between the point of origin and the point of consumption in order to meet customers' requirements. (Council of Supply Chain Management Professionals, 2015).

In the last decades logistics management has evolved into an increasingly important source of competitive advantage for organizations to ensure the availability of the right product (or service), in the right quality, and in the right condition, at the right place, at the right time, for the right customer, at the right cost (Shapiro and Heskett, 1985).

In line with this definition, logistic management in healthcare can be described as: ensuring the delivery of the right care, at the right time, in the right place, by the right person, where what is 'right' is based on a trade-off between the needs and preferences of the individual client and the efficient use of available resources. This description encompasses the above mentioned optimization challenge, namely meeting the needs and preferences of the individual client without compromising cost-efficiency. It is this optimization challenge that is addressed in this thesis.

## 1.2 Main contributions

The main objective of this thesis is to contribute to the optimization of daily nursing home operations, where optimal is defined as meeting the healthcare needs and time preferences of the nursing home clients in the most cost-efficient manner. To achieve this objective, knowledge and insights from the field of logistics management is used. More specifically, the thesis contributes to this aim by:

1. Exploring the principle of personal autonomy, which lies at the heart of client-centred care.

2. Presenting a conceptual framework for better understanding the fundamentals of healthcare logistics in a nursing home setting.

3. Providing insight into how demand patterns fluctuate over time and over the course of a day, using real-life data.

4. Providing insight into the consequences of these fluctuations in terms of workload, waiting time and staffing requirements, taking into account possible scale- and skill-mix efficiencies.
5. Providing a better understanding of the number of care workers required to meet sufficiently the needs of the nursing home clients regarding 'care on demand'.
6. Examining the performance of a small-scale living facilities, in terms of meeting the time preferences of their residents, under different assumptions and suggesting improvements regarding the allocation of care workers.

This thesis consists of two major parts. In the first part (Chapters 2 and 3), the emphasis is on providing a conceptual foundation for the chapters that follow. The conceptual foundation presented in this first part is the result of the exploration of existing literature and logical reasoning. In the second part of the thesis (Chapters 4, 5, 6 and 7) real-life nursing data are analyzed using quantitative methods stemming from Operations Research (OR). The focus of this second part of the thesis in on providing insight into how demand patterns fluctuate over time and over the course of a day and the consequences of these fluctuations in terms of workload, waiting time and staffing requirements, taking into account possible scale- and skill-mix efficiencies.

## 1.3 Outline of the thesis

Chapter 2 explores the principle of personal autonomy in healthcare by considering its historical-philosophical underpinnings. Amongst healthcare providers there exist many different views on what personal autonomy is and how it should be facilitated. To better understand the current struggle with the concept of personal autonomy in healthcare we examined the historical views related to this topic. The results presented in this chapter can be used by policy-makers and managers in their decision making on how to facilitate personal autonomy in healthcare. Chapter 2 is based on Moeke and Van Andel (2015).

Chapter 3 presents a conceptual framework for healthcare logistics in nursing homes. Existing literature on client-centred care and (healthcare) logistics is used to develop a framework which offers a structure for both future research and practice. Chapter 3 is based on Moeke and Verkooijen (2013).

Next, Chapter 4 investigates how demand patterns of scheduled care of five Dutch nursing home departments fluctuate over time and over the course of a day. Furthermore, we examined the consequences of these fluctuations in terms of workload and waiting time, taking into account possible advantages of pooling care workers from multiple departments. Chapter 4 is based on Moeke et al. (2015a).

Chapter 5 provides insight into how and why 'scale of scheduling' and the enlargement of care workers' jobs (blending tasks of different qualification levels) affect the number and type of staff required to meet the preferences (in terms of day and time) of nursing home clients. The scheduled care activities of three separate decision-making units within a single Dutch nursing home are analyzed. Chapter 5 is based on Moeke et al. (2014).

Chapter 6 examines the 'care on demand' process in a Belgian nursing home. Based on the analysis of real-life 'call button' data, a queueing model is presented which can be used by nursing home managers to determine the number of care workers required to meet a specific service level. Chapter 6 is based on Van Eeden et al. (2014).

Finally, Chapter 7 examines the performance of a nursing home department with four small scale living facilities (SSLF) under different assumptions. Furthermore, by using the results and insights of the previous chapters, improvements regarding the allocation of care workers are suggested. Chapter 7 is based on Moeke et al. (2015b).

Chapter **2**

# Back to the future: how the past shapes our present discussions on patient autonomy

**Abstract**

Today most healthcare providers have embraced the principle of personal autonomy as central to their strategic aims and objectives. However, amongst healthcare providers there exist many different views on what personal autonomy is and how it should be facilitated. To better understand the current struggle with the concept of personal autonomy in healthcare we examined the historical views related to this topic. Our investigation shows that there can be distinguished three major traditions, each of which defines preconditions for autonomous behavior. These preconditions are: (1) rationality and rational faculties, (2) individual rights and legislation and (3) free property rights, free market and free trade. We found that the three historical traditions still play a key role in current discussions on personal autonomy in healthcare. As such, a thorough understanding of these traditions is of vital importance for policy-makers and managers in healthcare.

This chapter is based on Moeke and Van Andel (2015).

## 2.1 Background

In the recent decades much attention has been given to personal autonomy in healthcare decision-making (Proot et al., 1998; Greaney et al., 2012). Hence, it has even become one of the fundamental principles of medical ethics (Beauchamp and Childress, 2009). This increased emphasis on personal autonomy is often seen as a social reaction to a paternalistic tradition, whereby the healthcare professional makes decisions based on what he or she finds to be in the clients' best interest, i.e. 'the healthcare professional knows best' (Proot et al., 1998). As such, the principle of personal autonomy respects the right of self-determination and non-interference of others when making decisions about themselves. In a healthcare context it is usually associated with allowing or enabling clients to make decisions about their own healthcare (Entwistle et al., 2010). Oshana (2003, p. 100) describes this as: *"the condition of being self-directed, of having authority over one's choices and actions whenever these are significant to the direction of one's life"*.

Today most healthcare providers have embraced the principle of personal autonomy as central to their strategic aims and objectives. However, amongst healthcare providers there are many different views on what personal autonomy is and how it should be facilitated. This often leads to suboptimal policies, both within the healthcare institution as well as within the supply chain, which can affect both the individual patient and the healthcare institution.

We therefore feel that policy makers are in need of guidance in their struggle with the concept personal autonomy in healthcare. As such, we believe the study presented in this chapter can support them in their decision making on how to facilitate personal autonomy within the healthcare supply chain.

The remainder of this chapter is structured as follows. The next section explores how the principle of personal autonomy has been regarded by scholars over time. It shows that three major traditions can be distinguished. Section 2.3 describes how these three traditions found their way into healthcare policy. Section 2.4 shows how the different traditions manifest themselves in current healthcare practice. Finally, Section 2.5 presents the conclusions of this chapter.

## 2.2 Personal autonomy: historical interpretations

Personal autonomy and related concepts such as individual liberty and individualism have been the topic of discussion for over 2000 years. In this section we aim to show how these concepts have been interpreted over the ages.

### Personal autonomy: early perspectives

The word autonomy stems from the Greek words *autos* [ατός] and *nomos* [νόμος] meaning: making one's own laws, or self-rule. However, in (pre-)Socratic Greece and the early Roman empire, individualism as known today was as good as non existent. The free citizens[1] of cities such as Athens or Rome were seen as part of an indivisible body of citizens who had plights rather than rights to take responsibility for the welfare of the city-state or *polis*. Nevertheless, within this context concepts such as personal autonomy and liberty were already topics of discussion. Plato and Aristotle, for instance, devoted part of their work to discussing concepts such as self-rule and freedom.

In his *Nicomachean Ethics* and *Politics*, Aristotle provides one of the first systematic analysis of what we would now call 'personal autonomy' and how it relates to 'state intervention'. Aristotle states that, in essence, happiness is the main goal of all self-regulated (i.e., autonomous) human endeavor and that this happiness can be achieved in different ways. He feels, however, that there are better and worse notions of happiness and better and worse ways of how to achieve it:

> Now of the chief good (i.e. of happiness) men seem to form their notions from the different forms of life, as we might naturally expect: the many and most low conceive it as pleasure, and hence they are content with the life of sensual enjoyment. For there are three lines of life which stand out prominently to view, that just mentioned, and the life of society, and thirdly the life of contemplation. (Aristotle, 1998, p. 4)

According to Aristotle, the highest form of happiness could only be achieved by reasonable and rational beings. Hence, only philosophers, those who were able to live a life of contemplation, would be able to reach this ultimate state of happiness. As such, personal autonomy was explained as self-rule in the

[1] Which excluded women, children and slaves Goff (2004).

sense that the individual could rule himself, and choose his individual path, only insofar as it was guided by reason.

A similar view is voiced by Plato (1998) who stated that the ideal state or society ought to be organized in such a way that those with the faculties to rule, the philosopher or philosopher-king, would rule those who could not and thus create a maximum of happiness for both rulers and ruled.

The subject of (individual) liberty was also studied by the Roman stoic Epictetus. Epictetus (2010) pursued the line of Plato and Aristotle by stating that true freedom is the greatest good of all and that it can be achieved only through a specific state of mind. Epictetus believed that as long as one is slave to desire, fear, ambition or tied to a friend, lover, spouse or sibling, one can never truly be free. To become truly autonomous one has to train one's mind in accepting God's will. That is, not to desire, prize or become attached to something or someone and accept the variable, unstable, unpredictable and unreliable nature of things. According to Epictetus, such a state of mind could be achieved through philosophical reflection:

> Start with things that are least valuable and most liable to be lost –things such as a jug or a glass– and proceed to apply the same ideas to clothes, pets, livestock, property; then to yourself, your body, the body's parts, your children, your sibling and your wife. (Epictetus, 2010, p. 71)

Autonomy or self-rule in Epictetus' view can be described as being able to detach oneself from pleasure, pain, earthly needs and wants and to accept things as they come.

In the Medieval era the teachings of Plato, Epictetus and, in particular, Aristotle remained very influential. Augustine (1991), for one, agreed with Aristotle that every human strives to be happy. He further believed that it was the task of philosophers to define happiness (i.e. this supreme good) and how it should be achieved. According to Augustine, the task of moral philosophy was to perform an inquiry into this supreme good. That is, the good that provides the standards for all our actions and which is sought for its own sake and not as a means to an end.

Thus, in the early perspectives on personal autonomy, reason and the rational faculties were considered as preconditions for self-rule. Seen from this perspective, when individuals are lacking in reason or rationality, someone with the capacity to act rationally should decide for them.

### Modernity and the birth of individualism

The Enlightenment was a great catalyst for a more egalitarian society, freedom of speech and press and individual rights. Or as Kant stated:

> Enlightenment is man's emergence from his self-incurred immaturity. Immaturity is one's inability to use one's own understanding without the guidance of another. This immaturity is self-incurred if its cause is not lack of understanding but lack of resolution and courage to it without the guidance of another. The motto of the enlightenment is therefore: Sapere aude! Have courage to use your own understanding! (Kant, 2010, p. 1)

Within this context Rousseau (1997, p. 49) wrote: "Man was born free and he is everywhere in chains". According to Rousseau, to renounce one's freedom is to renounce one's humanity, one's rights as a human and equally one's duties. The only reason that people would surrender their freedom is when they see advantage in doing so. According to Rousseau, the main difficulty concerning the topic of freedom may be expressed accordingly:

> How to find a form of association which will defend the person and goods of each member with the collective force of all, and under which each individual, while uniting himself with the others, obeys no one but himself, and remains as free as before. (Rousseau, 1997, p. 60)

The answer to the question posed by Rousseau, was seen in the promotion and acceptance of values such as representative democracy, basic human rights, individual liberty and freedom of expression. The pursuit of these ideals eventually gave way to the events that resulted in the French Revolution (Isreal, 2011).

### Mill and Berlin on liberty

The ideals which were central to the French Revolution, were further developed by John Stuart Mill and Isaiah Berlin. Both are often seen as the leading authors on the topic of (individual) liberty. According to Mill:

> The only freedom which deserves the name is that of pursuing your own goods in your own way, so long as we do not attempt to deprive others of theirs, or impede their efforts to obtain it. Each is the proper guardian of his own health whether bodily or mental and spiritual. (Mill, 2010, p. 21)

In addition, Berlin argues that individuals make choices based on the values that make them human:

> In the end, men choose between ultimate values; they choose as they do because their life and thought are determined by fundamental moral categories and concepts that are, at any rate over large stretches of time and space, a part of their being and thought and sense of their own identity; part of what makes them human. (Berlin, 1958, p. 31)

### Late modernity: neo-liberal autonomy

From the 1980s onwards the dominant view on personal autonomy could be described as neo-liberal. The essence of neo-liberalist freedom or neo-liberal autonomy was strongly voiced by Nobel laureate Milton Friedman (1912-2006). According to Friedman (1993), human well-being could be best advanced by liberating an institutional framework characterized by free property rights, free market and free trade:

> The great virtue of a free market system is that it does not care what color people are; it does not care what their religion is; it only cares whether they can produce something you want to buy. It is the most effective system we have discovered to enable people who hate one another to deal with one another and help one another. (Friedman, 1993, p. 19)

Not only does the free market system, according to Friedman (2009), display a form of 'color-blindness', but a free economy "gives people what they want instead of what a particular group thinks they ought to want" (p. 15).

### Personal autonomy throughout history: conclusions

Although painted with rather broad strokes we do feel the overview presented above shows the major perspectives on (the facilitation of) personal autonomy. In the Classical perspective rationality is seen as precondition for personal autonomy in the sense that only rational individuals are able to govern their own lives and those who are not should be governed by rational rulers. In the Modern perspective individual rights and legislation are seen as the precondition for personal autonomy in the sense that after the Enlightenment individual rights or the rights of man became a major topic for the common

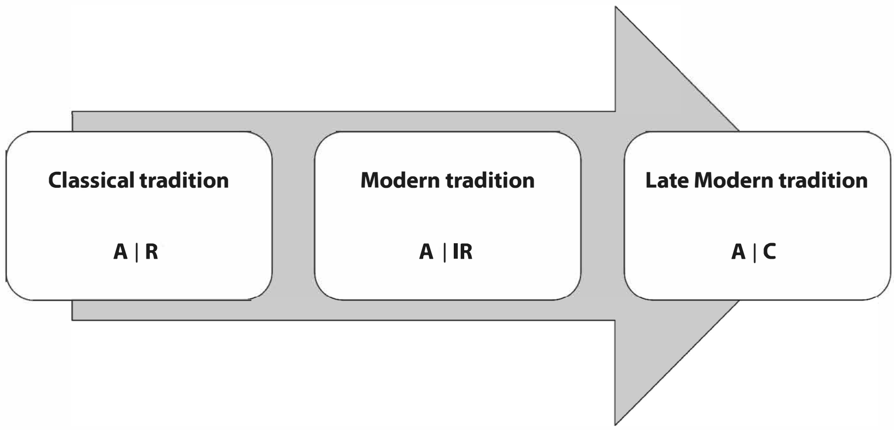


**Figure 2.1:** Interpretation of personal autonomy throughout history

people and, eventually, for governments as well. In essence the principle of individual rights and legislation gave individuals the opportunity to pursue their personal life goals in any way they saw fit. In the Late Modern perspective free market and free trade are seen as the precondition for personal autonomy in the sense that being able to act as a consumer is regarded as acting autonomously.

Thus, the three major traditions can be summarized as follows (see also Figure 2.1):

1. The Classical tradition in which rationality is seen as the precondition for personal autonomy (A|R)

2. The Modern tradition in which individual rights and legislation are seen as the precondition for personal autonomy (A|IR)

3. The Late Modern tradition in which free property rights, free market and free trade are seen as the precondition for personal autonomy (A|C)

The following section provides an overview of how the three major traditions found their way towards healthcare policy.

## 2.3 Personal autonomy in healthcare: historical development

### Doctor knows best (A|R)

From a healthcare perspective, the Classical tradition dates back to Hippocrates of Cos (c.460 - 370 BC). Hippocrates was a Greek physicist and is regarded as the father of medicine as a rational science. He advocated a beneficent paternalistic paradigm in which physicians have the right to decide what is in the best interest of the patient. This is also reflected in the so called Hippocratic oath, which is still taken today by students graduating from medical school. The oath states that: "I will use treatment to help the sick according to my ability and judgment". In other words, a physician should make decisions regarding the medical treatment of the patient based on his or her medical expertise. According to Will (2011):

> The language of the oath creates the impression that physicians (unlike others) have a superior knowledge that makes them capable of diagnosing illness, which carries with it the responsibility to offer treatments for the benefit of the sick. (Will, 2011, p. 670)

Hence, the healthcare professional alone had the knowledge and rationality to know what's best for the patient. This tradition continued up until the Enlightenment (Will, 2011). It should be mentioned that since September 2003 all medical faculties in the Netherlands have been using a new medical oath (Westerveld et al., 2005). An important new aspect of this oath is that "the view of the patient should be respected".

### Patient rights (A|IR)

Legal protection of personal autonomy in healthcare traces its origins back to 28 January 1891, when the Prussian minister of the interior circulated a memorandum regarding the use of tuberculin in the prison system. In this memorandum[2] the minister makes a clear reference to the autonomy of the patient:

[2] Ministerialblatt für die gesamte innere Verwaltung in den Königlich Preussischen Staaten. 1891. 52: 27

> Voraussetzung sei dabei, daß die Behandlung mit dem Koch'schen Mittel nur in frischen und sonst geeigneten Fällen auch nicht gegen den Willen der Kranken angewendet werde. [It is also necessary that Kochs' substance be used only in recent and appropriate cases and never against the will of the sick person].

This official public document can be regarded as a major turning point in the shift from paternalism towards more respect for the individual rights of a patient.

The first detailed regulations on informed consent in Western medicine came in 1900 after critical press reports and political debate in the Prussian parliament on the Neisser case (Vollmann and Winau, 1996). Albert Neisser, a German physician and bacteriologist, had injected serum from patients suffering from syphilis into a group of healthy patients, mostly prostitutes, who were admitted for other reasons. He had neither informed the women about the risks involved, nor had he obtained their consent. His experiment caused public debate as some of the women developed syphilis. The case led in 1900 to the adoption of the so-called Prussian Directive, in which the Prussian authorities advised medical directors of hospitals and clinics that research interventions should not go forward if "the human was a minor or not competent for other reasons" or if the subject had not given his or her "unambiguous consent" after a "proper explanation of the possible negative consequences" of the intervention. This was the first known governmental directive on clinical research practices (Vollmann and Winau, 1996).

### Patient as consumer (A|C)

To help raise the effectiveness and efficiency of their public services, at the end of the 1980s many western countries adopted a new, strongly market-orientated form of public management known as New Public Management (NPM). From a NPM perspective, a welfare state is conceived as a market-based delivery system, while the citizen is seen as a consumer (Olsen, 1988). Consequently, the adherence to NPM ideology resulted in the introduction of market and quasi market-type mechanisms, which were thought to raise the 'customer responsiveness' of professional bureaucrats (Pollitt, 1995). Eventually market principles, competition and choice found their way into healthcare sectors as well. In this new paradigm the patient was expected to act as a consumer (Calnan, 2010). Consequently, some policy makers and healthcare managers viewed personal autonomy as a market-based concept which defines healthcare users as individualistic actors striving to maximize their preferences. In this interpretation, the patient is regarded as a consumer in search of the best product (i.e. care, support, medicine, treatment, etc.). This rational-economic interpretation of patient autonomy resulted in the adoption of concepts such as hospitality (Severt et al., 2008; Aiello et al., 2010; Wu et al., 2013) and an increased focus on consumer satisfaction (Abusalem et al., 2013; Barr et al., 2006). Furthermore, there has been a growing focus on providing patients with the right information about the benefits of the products of healthcare providers.

## 2.4 Current affairs: personal autonomy in healthcare today

All of the three perspectives presented above still play a key role in current discussions on personal autonomy in healthcare. Below we describe, with the help of anonymized real-life examples, how the three major traditions shape current healthcare practice.

### Example I: A|R

> Hospital X is a university medical centre built to the highest standards of medical design and well equipped with 'state-of-the-art' technology. A highly skilled multidisciplinary team of specialists aims to provide the highest possible standards of clinical care in a research-informed environment. To ensure the best possible medical outcome for the patients, the hospital uses standardized, evidence-based multidisciplinary pathways. A pathway consists of an appropriate sequence of clinical interventions, timeframes and expected outcomes for a homogeneous patient group.

In this example, the knowledge and expertise of a team of specialists is dominant. Patients are seen as passive recipients of care and are considered incapable of making rational decisions about their own treatment. The focus lies on the physical aspect of the patient's disease with the purpose of ensuring the best possible clinical outcome according to the latest medical standards.

**Example II: A|IR**

> Hospital Y strives to support the individual needs and preferences of its patients as much as possible. After patients are referred to the hospital, they can make an appointment with the specialist of their own choice at a time that suits the patient best. Patients are actively involved in the decision-making process and facilitated in self-management. For example, for patients with a chronic illness which requires frequent monitoring, there is the possibility to make use of tele-monitoring equipment. This allows the patients to manage their own health.

In this example, the individual preferences and needs of the patient are taken as a starting point. The belief underlying this approach is that the patient is most knowledgeable about his or her own needs. The ultimate goal of this hospital is to facilitate and support the patient in such a way that the treatment meets the preferences of the client in terms of what, when, where and by whom.

**Example III: A|C**

> Hospital Z aims to make the hospital experience of patients and their visitors as pleasant as possible. Consequently, the hospital considers consumer service to be of the highest priority. Hospitality employees are available at the information desk in the main lobby to assist visitors as needed. The hospital has a wide range of cafés, restaurants, shops and other facilities to make the stay of the patients and visitors as pleasant as possible. The hospital offers the option to upgrade to a private or semi-private room at an extra charge. Furthermore, in return for extra payment, patients can make use of hotel-style room service.

In this example, patients and visitors are regarded as persons who buy goods and/ or services from the hospital. Hence, as consumers, they have the power to choose how to spend their money. The aim of this hospital is to create ultimate customer satisfaction by providing the patient with services and products which meet or exceed their expectations.

## 2.5 Conclusions and discussion

In this chapter we have considered the historical-philosophical underpinnings of personal autonomy in healthcare. Three main traditions were identified which still play a fundamental role in current discussions on personal autonomy in healthcare: the Classical, the Modern and the Late Modern tradition.

In the Classical tradition, rationality is seen as the precondition for personal autonomy in the sense that only rational individuals are able to govern their own lives and those who are not should be governed by rational rulers. Today we see this classical perspective represented by the fact that healthcare professionals are regarded as rational with regard to patients' health and thus able to decide for patients what they should and should not do, eat, drink, etc.

In the Modern tradition, individual rights and legislation are seen as the precondition for personal autonomy in the sense that after the Enlightenment individual rights became a major topic for the common people and, eventually, for governments as well. In essence the principle of individual rights gave individuals the opportunity to pursue their personal life goals in any way they saw fit. Today this perspective is represented by laws, legal procedures and patients' rights organizations.

In the Late Modern tradition, free property rights, free market and free trade are seen as the precondition for personal autonomy in the sense that being able to act as a consumer is regarded as acting autonomously. Today users of publicly funded services are regarded more and more as customers who are in need of products provided by, for instance, government healthcare providers. When the patient is able to 'purchase' the product of his or her choice, he or she is then regarded as acting autonomously. Today this perspective is represented by organizations which have adopted concepts such as hospitality and customer service.

The three traditions presented in this chapter assist policy makers and managers to understand more fully the concept of patient autonomy and to help evaluate the trade-offs that their policy choices might entail. When facilitating personal autonomy in healthcare policy makers should realize that tension exists between these three traditions.

Although, clients may be quite knowledgeable about their health situation, in most cases they do not know all the relevant clinical details and its consequences. Furthermore, in a stressful and/ or emergency situation relying on the opinion of a medical specialist or other healthcare professionals can be a relief for the patient and the patient's family. A major pitfall of the classical perspective is a lack of respect for the ability of patients to make their own choices about their care or treatment. The opinion of the healthcare professional takes priority over the right to self-determination.

The main advantage of the modern perspective is that patients are able to make choices about their own healthcare based on their individual preferences and needs. In practice, this may imply that happiness and emotional welfare are put before physical health and safety. A possible pitfall of this approach is that it can lead to an organizational model in which healthcare professionals must comply with every patient request or demand. Furthermore, a fully demand-led model can result in cost-inefficiencies.

The third tradition allows patients to control their own healthcare costs, they become more conscientious about their spending on healthcare goods and services. It creates an incentive for healthcare consumers to make choices based on their individual needs, preferences and budget. Competition will encourage healthcare institutions to become more efficient and to allocate resources to their most profitable use. However, a healthcare system which is driven by competition and profit can lead to inequalities as it 'favors' wealthier healthcare consumers. Accordingly, medical care could become unaffordable for lower-income families. Furthermore, patients are not like regular customers as they are in a more vulnerable position and often lack the necessary information to make informed decisions.

The above shows that finding a suitable balance is a challenging quest. It raises interesting questions, for policy makers and managers in healthcare, about the extent to which it is possible to combine these traditions into a single organization concept. The findings presented in this chapter may hopefully aid policy-makers and managers in their decision making on personal autonomy in healthcare.

Chapter 3

# Healthcare logistics in a nursing home setting: doing more with less

**Abstract**

This chapter will outline the current status of research on healthcare logistics for nursing home care. By integrating related literature a conceptual framework is developed that provides a more structured and integral approach to the understanding of the aspects that are of importance when it comes to healthcare logistics in a nursing home setting. The framework offers guidance for future research and practice.

This chapter is based on Moeke and Verkooijen (2013).

## 3.1 Background

From the perspective of the Modern tradition (see Chapter 2), in which autonomy equals individual rights, it seems logical and necessary: healthcare that defines the needs and preferences of a client as starting point in shaping care delivery practices and all of the supporting services. Accordingly, one would expect that healthcare providers engage their clients in their own care and provide them with the opportunities to influence the organization and delivery of healthcare services. Although most healthcare providers have embraced respect for personal autonomy as central to their strategic aims and objectives (see also Chapter 2), in practice, the influence of clients on their own care often is still limited.

Several studies indicate that even healthcare clients who are highly dependent on assistance with basic activities in daily living, often have little influence on their daily routine (e.g., Hammar et al., 2014; Persson and Wästerfors, 2009; Simmons et al., 2011). When it comes to Dutch nursing home care, Hamers (2011) states that:

> In verpleeghuizen bepalen zorgverleners veelal wat goed voor ouderen is en nemen zij veel taken over van bewoners in het regime van de dagstructuur. [In nursing homes care workers often decide what is best for the client and are largely in control of the daily routine]. (Hamers, 2011, p. 15)

It should be stressed that often this is done with the best intentions of helping the client or under regulatory pressure (Hamers, 2011). Besides this lack of influence by clients, most nursing homes face ever tightening budget constraints. As such, finding the right balance between meeting the needs and preferences of clients and their consumption of scarce resources is paramount for successful nursing home operations.

As stated in Chapter 1, it is widely believed that insights, concepts and tools from the field of logistics can contribute to providing high-quality, client-centred care in a more cost-efficient manner (e.g., Bakker, 2004; Vissers and Beech, 2005; Nachtmann and Pohl, 2009; Verkooijen, 2006). This chapter outlines the current status of research on healthcare logistics for nursing home care and develops a conceptual framework. In this context, the term conceptual framework can be described as a visual or written product that “lays out the key factors, constructs, or variables, and presumes relationships among them” (Miles and Huberman, 1994, p. 440). As such, a conceptual framework provides understanding “rather than offering a theoretical explanation, as do quantitative models” (Jabareen, 2009, p. 51). The framework presented in this chapter is designed to provide a more structured and integral approach to the understanding of the aspects that are of importance when it comes to healthcare logistics in a nursing home setting.

The rest of this chapter is organized as follows. The next section presents the results of a systematic literature review on frameworks, models or concepts for ‘healthcare logistics’ which are designed (or can be used) for organizing healthcare activities in a nursing home setting. Next, Section 3.3 defines the key concepts of this study, i.e., ‘client-centred care’ and ‘healthcare logistics’. Section 3.4 then present the conceptual framework, using the results of sections 3.2 and 3.3 as a starting point. Finally, in Section 3.5, the conclusions of this chapter are presented.

## 3.2 Literature review

This section presents a literature review on existing frameworks, models or concepts for ‘healthcare logistics’ which are designed (or can be used) for organizing healthcare activities in a nursing home setting.

### Search strategy

Existing models for ‘healthcare logistics’ were identified through an electronic search of literature using the databases PubMed, ScienceDirect and Google Scholar from January 2003 to April 2013. The following Boolean combination was used for the search: “(health care logistics OR healthcare logistics OR patient logistics OR patients logistics) AND (framework OR model OR concept) AND (customer centred OR client centred OR patient centred OR customer centered OR client centered OR patient centered OR demand led OR customer led OR patient led OR patient driven OR user driven OR demand driven OR customer driven)”. We performed the literature search on April 8, 2013.

We then examined all the identified literature (including ‘grey’ literature). Derived from the background and purpose of this study, the inclusion criteria were set as follows:

- The models, frameworks or concepts should include both the needs/preferences of individual clients and efficiency as relevant aspects of healthcare logistics.

- The models, frameworks or concepts should be designed (or could be used) for organizing healthcare activities in a nursing home setting.

Finally, the references from the included literature were screened thoroughly for possibly relevant studies or literature using the same criteria cited above. This type of search strategy is often referred to as the 'snowball method' (e.g., Maso and Smaling, 1998).

### Search outcome and conclusions

The search method described in Section 3.2 resulted in a total of 125 'hits'. Next, we examined the titles and/ or abstracts of these hits to determine their match with the inclusion criteria. Only one publication met the inclusion criteria. We then analyzed the full text of that study and screened the references for possible relevant publications. This search strategy resulted in only two usable publications.

Based on this literature study one can conclude that there is very little available in terms of a conceptual framework for organizing nursing home care, let alone a framework which embodies a client-centred approach as well as the logistic aspects involved. The vast majority of the available literature on healthcare logistics focuses on care provided in a hospital setting. The so-called OERmodel (Verkooijen, 2006), which we also used in a previous study (Moeke and Verkooijen, 2010), was the only concept we found that uses a client-centred approach for organizing healthcare activities in a nursing home setting. Therefore, the knowledge gained from these two publications has been used as initial guidance for the development of the conceptual framework. The OERmodel is the result of a doctoral dissertation study and can be described as an action- and organization model which uses a client-centred approach for organizing (health)care activities. It provides guidance to care workers in both nursing homes and homecare organizations.

## 3.3 Client-centred care and healthcare logistics

This section provides more insights into the key concepts of the conceptual framework, namely: 'client-centred care' and 'healthcare logistics'.

## Client-centred care

Today client-centred care, also referred to as patient-centred care, is recognized by the Institute of Medicine (IOM) as one of the key components of healthcare quality, where client-centred care is defined as: "providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions" (Kohn et al., 2001, p. 6).

In other words, from a client-centred perspective, healthcare providers should engage clients in their own care and provide them with the opportunities to influence the organization and delivery of healthcare services. In our opinion, the amount of influence an individual client wants to have is strongly related to the need for 'self-directing', which can be defined as:

> het organiseren en/of coördineren van het eigen leven met als doel een goed leven in eigen ogen [the organization and /or coordination of one's own life with the objective of having a good life– in one's own opinion]. (Verkooijen, 2006, p. 70).

Based on a review of the literature, Rodriguez-Osorio and Dominguez-Cherit (2008) conclude that healthcare clients' preferences regarding their role in the decision-making process vary substantially. In line with this reasoning, Flynn et al. (2006) distinguish between 'autonomists' who wish to make decisions themselves versus 'delegators' who prefer a physician to make important decisions. And although individual preferences regarding their role in decision making differ from client to client, Chewning et al. (2012, p. 15) state that: "the number of patients who prefer participation has increased over the past three decades so that the majority of patients prefer to participate in decisions during the encounter".

Nursing home clients are often 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 preferences, nursing home clients largely depend on being provided with the opportunity to influence the delivery of their own care. From an organizational perspective, the (preferred) influence of a nursing home client in decision making can be divided into the following four categories (Verkooijen, 2006):

1. The moment (day and time) at which the care is delivered (*When?*)

2. The place where the care is delivered (*Where?*)

3. The person who provides the care (*Who?*)

4. The form and content of the delivered care (*What and how?*)

### Healthcare logistics

Healthcare logistics is a relatively new and multidisciplinary field of research. At an abstract level healthcare logistics can be defined as:

> Het zo beheersen van behandel-/zorg-/ondersteuningsprocessen en de daarmee verbonden inzet van medewerkers, informatie- en goederenstromen, dat tegen optimale kosten aan de wensen van cliënten kan worden voldaan [The control of treatment/care/support activities and the related staff planning, information, and flow of goods in such a way that the preferences of clients will be met cost effectively]. (Moeke and Verkooijen, 2010, p. 28).

Based on this definition and the four categories of influence presented in the previous subsection, healthcare logistics can be defined as: ensuring the delivery of the right care, at the right time, in the right place, by the right person, where what is 'right' is based on a trade-off between the needs and preferences of the individual client and the efficient use of available resources. This definition of healthcare logistics serves as point of departure for the conceptual framework presented in the next section. It should be noted that, regarding the delivery of nursing home care, a distinction should be made between two types of healthcare activities. For some of the care activities it is possible, based on the 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'. These activities are also referred to as 'care by appointment' or 'scheduled care' (see also Chapters 4 and 5). On the other hand, there are healthcare activities which are carried out in response to random or unexpected demand such as assistance with toileting. Activities of this type are also referred to as 'care on demand' or 'unscheduled care' (see also Chapter 6).

## 3.4 Conceptual framework

This section presents our conceptual framework, using the results of sections 3.2 and 3.3 as a starting point. In order to give further substance to the separate elements of the framework, we have made use of generally accepted and applied concepts/theories, experience from practice and logical reasoning.

### Boundaries to a client-centred approach

From a client-centred perspective nursing home clients are considered to be the most important stakeholder. However, they are not the only ones. Nursing home care always takes place within a 'client–healthcare professional–organization' triangle. Hence, in practice, the influence of a client is always bounded. In her dissertation Verkooijen (2006) states that there are four important boundaries regarding the influence of a client:

1. The professional responsibility of the care worker.
2. The organizational responsibility of the nursing home.
3. Generally accepted norms and values.
4. The healthcare budget received and/or clients' own financial resources.

#### Professional responsibility

In practice, 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 interests of the client (Rodriguez-Osorio and Dominguez-Cherit, 2008). In other words, clients' preferences may conflict with the *professional responsibility*, and thus with the moral autonomy, of the care worker (McParland et al., 2000). Hence, "promoting patient autonomy for its own sake does not necessarily constitute moral agency" (Woodward, 1998, p. 1050). Consequently, the influence of an individual client is bounded by the professional responsibility of the care worker. This professional responsibility, which is translated into professional guidelines and insights, mainly relates to the form and content (what and how) of the healthcare (see Figure 3.1).

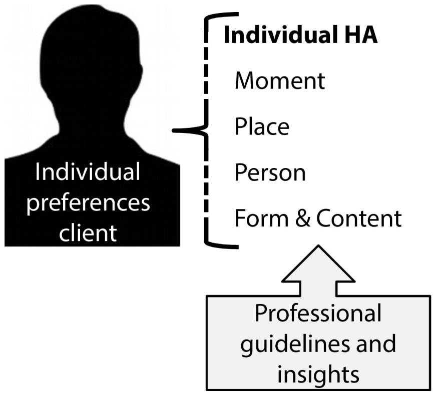


**Figure 3.1:** Influence of professional guidelines and insights on individual healthcare activities (HA)

### Strategic guiding principles

Furthermore, the influence of a nursing home client is also shaped by the *organizational responsibilities* of the nursing home. With regard to responsibilities and decision-making in a hospital setting, Brailsford and Vissers (2011, p. 223) state that: "decision making is carried out in more of a political arena in which the interests of different stakeholders need to be balanced". The same applies to nursing homes, as nursing home managers also have to take the interests of multiple stakeholders (e.g., employees, other organizations in the supply chain and society) into account. In addition, a nursing home client should adhere to *generally accepted norms and values* of fairness and equity. As such, these are regarded as an important boundary as well. Finally, the *budgetary boundaries* of the organization should not be overstretched. Organizational responsibilities, generally accepted norms and values and budgetary boundaries should be translated into strategic guiding principles, which act as a reference for decisions on healthcare logistics (see Figure 3.2). Based on our earlier study (Moeke and Verkooijen, 2010) we consider the mission statement, vision, logistics objectives, strategy and the logistics control philosophy to be important guiding principles.

The mission statement and vision, which are integrally tied to each other, are the cornerstones by which decisions on healthcare logistics are made. A mission statement should capture the unique and enduring purpose of the healthcare organization (Schwartz and Cohn, 2002). More specifically, it reflects why an organization exists, what it is trying to accomplish (Bart and Tabone, 1999) and how the organization intends to operate. A vision, on the other hand, can be defined as an aspirational description of what an organization would like to achieve or accomplish; an image of the future the organization seeks to create (Spallina, 2004; Bart and Hupfer, 2004; Levin, 2000). Based on research into the development of some of the most successful US corporations, Collins and Porras (1996) conclude that a well-conceived vision consists of two major components: a core ideology and an envisioned future state. Logistics objectives can be seen as a translation of the mission and vision into specific logistics performance goals. According to the definition by the Council of Supply Chain Management Professionals, both efficiency and effectiveness play a crucial role with respect to the logistics performance of an organization. Gleason and Barnum (1982, p. 380) define effectiveness as "the extent to which an objective has been achieved". From a client-centred care perspective, in which the clients' preferences are the starting point for the delivery of care (Bosman et al., 2008), effectiveness can be defined as the attainment of desired outcomes for clients (Arthur and James, 1994) or more specifically: the extent to which the preferences of the client are being met in terms of when, where, by whom, what and how. Efficiency, on the other hand, refers to "the degree to which resources have been used economically" (Gleason and Barnum, 1982, p. 380). As stated earlier, finding a balance of both effectiveness and efficiency is one of the essential tasks of logistics management.

Strategy can be described as a long-term approach to achieve the objectives of an organization. The essence of strategy in a competitive environment is to establish a sustainable competitive advantage i.e., combining activities in such a way that they deliver a unique mix of values (Porter, 1996). Treacy and Wiersema (1993) proposed the following three generic value disciplines:

1. Operational excellence: organizations pursuing operational excellence focus on making their operations lean and efficient.

2. Customer intimacy: organizations pursuing customer intimacy continually tailor and shape products and services to fit an increasingly fine definition of the customer.

3. Product leadership: companies pursuing product leadership strive to produce a continuous stream of state–of–the–art products and services.

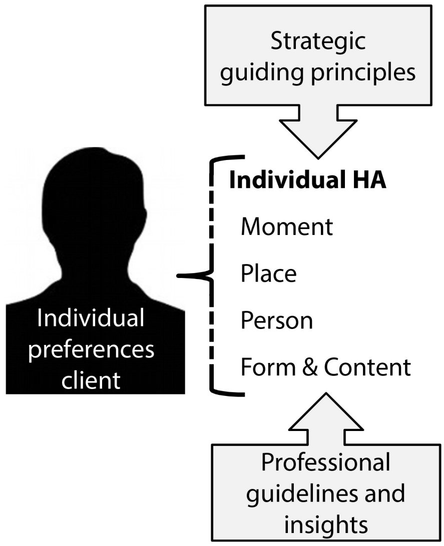


**Figure 3.2:** The influence of strategic guiding principles on the individual HA

According to Treacy and Wiersema (1993), an organization should focus on a specific value proposition, while meeting threshold standards in the other dimensions of values. The mission statement, vision, logistics objectives and strategy should be encapsulated into a logistics control philosophy. The control philosophy articulates the extent to which the preferences of the individual client should be determinative for the care delivery. There are roughly three different types of control philosophies: supply-driven, demand-driven and demand-led. Supply-driven means that the client has no (or very little) influence on the care delivery in terms of when, where, by whom, what and how. Demand-driven care can be positioned in between and the clients' influence is often limited to choosing from a set of predefined options. Demand-led is regarded as the most extreme form of client-centred care: the preferences of the individual client are determinative for the (health)care delivery. Because the logistics control philosophy should be consistent with the other strategic guiding principles, there is no 'one best choice' regarding the type of control philosophy.

## Balance between planning and reacting

Nursing homes are challenged to operate effectively and efficiently in an uncertain environment. A well-known definition of uncertainty is that of Galbraith (1973). He defines it as: "the difference between the amount of information required to perform the task and the amount already possessed by the organization" (p. 5). In other words: "if the task is well understood prior to performing it, much of the activity can be preplanned" (Galbraith, 1974, p. 28). When it comes to nursing home operations, planning can be described as: making informed and detailed decisions about future healthcare activities in terms of when, where, by whom, what and how, with the purpose of efficiently meeting the needs and preferences of the nursing home clients.

On the other hand a nursing home should also have the responsiveness to handle what Galbraith (1974, p. 30) refers to as "non-routine, consequential events that cannot be anticipated and planned for in advance". Here, we define responsiveness as: the actions or behavior of a system using a set of capabilities to address purposefully and timely unexpected changes triggered by external stimuli, like a change in demand. This definition is largely based on a definition presented by Davis Jr and Manrodt (1992).

To summarize, it could be said that nursing home facilities should preplan their healthcare activities as much as possible to ensure a 'basic level of effectiveness and efficiency'. In our framework we refer to this as *planning*. At the same time, a nursing home should develop a set of capabilities to deal adequately with uncertainty. The latter is translated in our framework as *reacting*.

In Section 3.4 provides insight into how (historical) information on aggregate demand, that is the sum of individual demands, can help to preplan healthcare activities better.

## Aggregate demand: variability, predictability and scale

The literature on hospital planning shows that the extent to which healthcare demand can be planned for is largely influenced by the following three interdependent variables, namely: (1) variability, (2) predictability and (3) scale (e.g., Joustra et al., 2010; De Bruin et al., 2007; Van Oostrum, 2009; Upshur et al., 2005). In the next subsections, each of these concepts will be discussed in relation to aggregate healthcare demand. With the help of examples we illustrate why and how these concepts are also of relevance with regard to the planning of healthcare activities in a nursing home context.

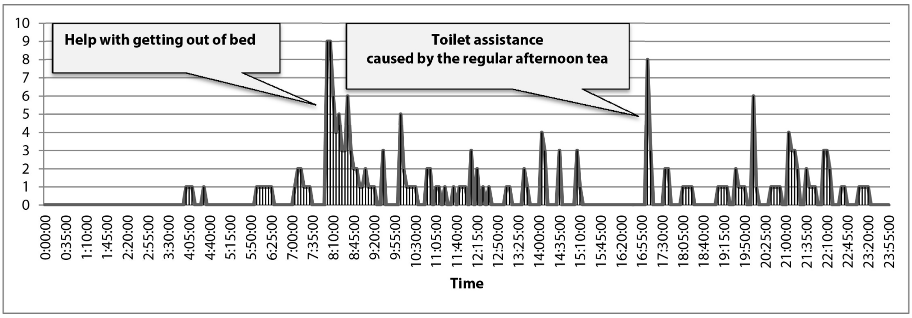


**Figure 3.3:** Variability in aggregate demand

### Variability

Nursing homes are challenged to meet the needs and preferences of their clients efficiently, despite variability (i.e., fluctuation) in demand over time. Without variability in demand, 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 planning a challenging task.

Figure 3.3 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 are in need of 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.

According to Litvak and Long (2000), 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 the 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**

In this subsection we look at how the predictability of demand (patterns) influences the extent to which healthcare activities can be planned for.

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 known as deterministic variability, can be easily planned for. Unfortunately, in practice, not all demand is fully known in advance. However, by analyzing historical data we can often identify regular patterns (i.e. systematic changes) like seasonality and trends. Such regular patterns can be planned for. For example, the peak in demand during the early morning in Figure 3.3 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. For instance, the exact number of call-button requests during a certain time-interval (see Chapter 6) 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. For example, in Section 6.3 we show that despite the fact that the exact number of call requests and corresponding delivery times (during the night) cannot be predicted, they behave stochastically and can be quantified with the help of a probability distribution. Thus, although the exact variability in demand cannot be predicted beforehand, it can be quantified.

**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 in the context of inventory pooling (e.g., Yang and Schrage, 2009; Benjaafar et al., 2005; Eppen, 1979). The so-called pooling principle suggests variability is reduced when demand is aggregated. This is because it becomes more likely that high demand from one client will be balanced out by low demand from another client. Statistically, the advantage of pooling is the consequence of a reduction in 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, Berg et al. (2005) argue that, generally, healthcare activities are better predictable at a more aggregate level:

> 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. (Berg et al., 2005, p. 79)

Furthermore, the enlargement of scale increases the flexibility in planning. Figure 3.4 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 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 healthcare workers available. Furthermore, we assume that a healthcare 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

| 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 3.4:** Increased flexibility in planning due to an increase in scale

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.

Figure 3.5 shows how variability, predictability and scale of healthcare activities are interrelated and how they influence the balance between planning and reacting.

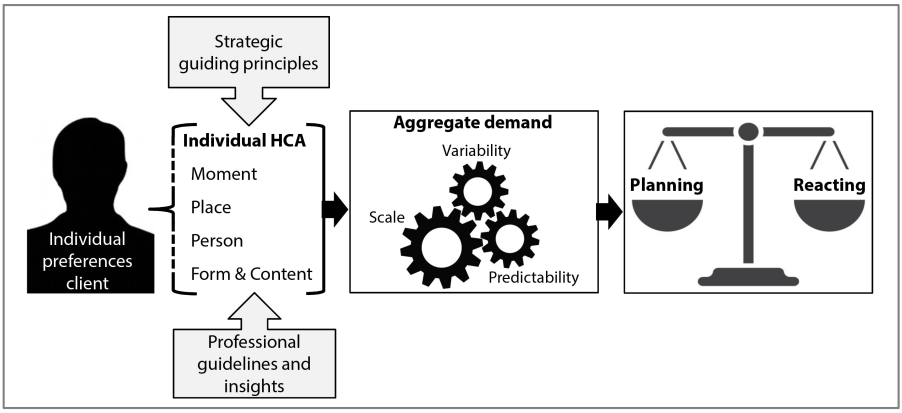


**Figure 3.5:** Variability, predictability and scale and the balance between planning and reacting

## Aspects of logistics design

The extent to which healthcare activities can be planned for has influence on the logistics design of a nursing home facility and vice versa. Therefore, the framework is extended by incorporating the following five aspects of logistics design: (1) *basic structure*, (2) *capacity planning*, (3) *information system*, (4) *organizational structure* and (5) *control system* (see Figure 3.6). Aspects (1), (3), (4), (5) are regarded as important by Verstegen (1989) and Van Goor et al. (2000). In addition, we included capacity planning (2) in our model as an important aspect of logistics design (e.g., Matta et al., 2014; Vissers et al., 2001). The following subsection discusses each of the design aspects in more detail. It should be noted that the emphasis of the discussion lies on short term (day-to-day) to medium term (up to 12 months) time horizons.

### Basic structure

Basic structure refers to how processes are structured at a comparatively aggregate level. When it comes to the delivery of care in a nursing home, two types of processes can be distinguished: (1) processes that focus on care activities which are to be carried out in a fixed sequence (so-called care routes) and (2) processes that focus on dealing with sudden or random demand.

### Capacity planning

In this context capacity planning can be defined as the process of determining the optimal level of resources required to meet future healthcare demand. Care workers are the most important resource of a nursing home. Therefore, capacity planning in a nursing home typically consist of: (1) estimating the minimum amount of care workers needed to fulfill the healthcare demand during a certain period of time (time horizon: months), (2) allocating care workers to shifts, whereby avoiding under- and over staffing as much as possible (time horizon: weeks) and (3) assigning care workers to specific activities or tasks with time specifications (time horizon: days-hours).

Depending on the amount of uncertainty in demand, more or less capacity flexibility is needed. Slack capacity, functional flexibility and/or numerical flexibility are methods commonly applied to deal adequately with (sudden) changes in demand. Slack capacity can be described as capacity that is in excess of the minimum necessary to achieve a certain level of organizational output, to handle unforeseen peaks in demand. Task or functional flexibility 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 (Atkinson, 1987). For example, in Chapter 5 we show that the enlargement of care workers' jobs can have a substantial positive effect on the number of care workers required to meet the time preferences of nursing home clients. Numerical flexibility can be defined as "the ability of firms to adjust the number of workers, or the level of worked hours, in line with changes in the level of demand for them" (Atkinson, 1987, p. 90). Common examples of numerical flexibility are: flexible contracts, hiring of temporary workers, and outsourcing or subcontracting.

### Information system

We define information system as a set of interrelated components that collect (or retrieve) information to support logistics decision making. According to Galbraith (1977), an important function of an information system is reducing the gap between the amount of information required for decision making and the amount of information available. When it comes to healthcare logistics, ideally, a nursing home information system should provide insight into (1) the needs and preferences of the nursing home clients, (2) the actual and historical demand, (3) the corresponding care durations, (4) the available capacity and (5) the logistics performance.

### Organizational structure

An organizational structure is a system used to define how the various tasks and responsibilities are distributed, coordinated and controlled within an organization. Centralization and decentralization are two opposite ways to transfer responsibilities within an organization. In a centralized structure, the decision-making authority is concentrated and often delegated to higher levels of an organizational hierarchy. In contrast, in a decentralized structure decision making has been disaggregated into a number of subunits, divisions, or teams each making its own decisions (Siggelkow and Levinthal, 2003). An advantage of decentralization is faster decision making and response time which can be desirable in urgent and/or uncertain situations. However, decentralized decision making can lead to sub-optimization, which occurs when different subunits, divisions and/or teams take decisions which are optimal for that specific unit, but may not be optimal for the organization as a whole.

### Control system

In this context, we define control system as a system with the purpose of assuring the logistics performance. The control system should be embedded in the continues improvement cycle, also known as the quality cycle, of a nursing home. According to Kruis (2008, p. 4), "a management control system is expected to be effective when it can manage the specific control problems an organization or organizational unit faces". The control problem in a nursing home can be described as meeting the needs and preferences of the clients, without sacrificing too many resources. Examples of logistics performance measures are: (1) care worker-to-patient ratio, i.e. the number of clients divided by the number of care workers, (2) percent of hours of direct care, i.e. the ratio of direct care hours to total hours worked and (3) service level, e.g., $X\%$ of the clients have a response time at or below $Y$ minutes (for care on demand) and $X\%$ of the clients are being served within $Y$ minutes of their preferred delivery time (for care by appointment).

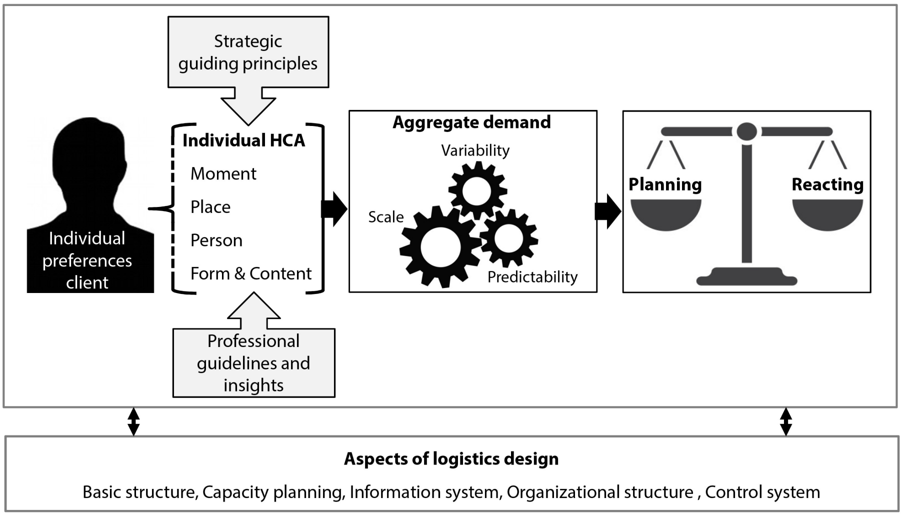


**Figure 3.6:** Five aspects of logistics design

## 3.5 Conclusions and discussion

Most nursing home clients are in need of ongoing assistance with activities in daily living, due to physical or psychological disabilities. This makes them vulnerable to and dependent upon the way in which nursing homes are organized and the healthcare workers caring for them. Unfortunately, recent studies show that nursing home clients still have little influence on their daily routine in terms of when, where, by whom, what and how this is carried out. Hence, nursing homes struggle to give substance to client-centred care. A complicating factor is that most nursing homes experience growing financial pressure. This increases the risk of an over-focus on efficiency and losing sight of the clients' needs and preferences.

To our opinion, the ultimate goal of a nursing home is to facilitate the need for self-directing of nursing home clients in such a way that it becomes possible for each resident to live the life they prefer to live. And although we realize that due to (increasing) resource limitations this goal can never be fully achieved, we do believe nursing homes have the social responsibility to try to achieve this goal as best they can. The conceptual framework presented in this chapter (see Figure 3.6) provides guidance in the search for a better balance between client-centredness and efficiency.

The literature study shows that there is very little available in terms of a framework for healthcare logistics which embodies both aspects. Therefore, we used existing insights on client-centred care and (healthcare) logistics as the starting point for the development of this framework. Generally accepted and applied concepts/theories, practical experience and logical reasoning were used to give further substance to the separate elements of the framework. As research on healthcare logistics in a nursing home setting is very scarce, we challenge researchers to collaborate on further research in this field. Finding usable data will be an important first step for future research in this promising field, as reliable and valid data is scarce and seldom collected.

However, the biggest challenge for further research will be not to lose sight of the needs and preferences of the client. Needs and preferences driven by the objective of having a good life– in one's own opinion. This is something that should always be kept in mind when conducting research in this area.

Chapter **4**

# Ebbs and flows in daily nursing home operations: a demand and workload analysis

**Abstract**

Nursing homes are challenged to develop staffing strategies that enable them to meet the needs of their clients efficiently in situations where the level of demand fluctuates. In this chapter we investigate how demand patterns of scheduled care of five Dutch nursing home departments fluctuate over time and over the course of a day. Furthermore, we examine the consequences of these fluctuations in terms of workload and waiting time, thereby taking into account possible advantages of pooling care workers from multiple departments. The results show that especially during the morning care (7:00-9:30 hours) the workload is high which leads to waiting times up to 40 minutes. We also show that pooling care work from multiple departments can have a substantial positive effect on the average waiting times.

This chapter is based on Moeke, Bekker and Schmidt (2015).

## 4.1 Background

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. Jack and Powers (2009). When it comes to nursing homes, care workers are by far the most important resource. They 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. Consequently, capacity management in a nursing home mainly consists of workforce planning. Effective workforce planning starts with a detailed insight and understanding of the fluctuation in demand (Higginson et al., 2011).

As stated in Chapter 3, concerning (health)care demand in a nursing home, a distinction should be made between 'scheduled care' (i.e. 'care by appointment') and 'unscheduled care' (i.e. 'care on demand'). This chapter focuses on scheduled care, for which it is possible, based on the needs and preferences of the residents, to make a fairly detailed planning in advance. In the sequel, the predefined activities of scheduled care with a preferred starting time are referred to as *preferred activity time* (PAT). More specifically, we investigate how the PATs fluctuate over time and over the course of a day, using data from five independent nursing home departments of a single Dutch nursing home. We also analyze the consequences of these fluctuations in terms of workload and waiting time, and take into account possible advantages of pooling care workers from multiple departments. The aim is to provide insight in the PATs in relation to workload and waiting time on an aggregate level.

This chapter yields primary insight in demand patterns and capacity requirements of scheduled care activities. The analysis of waiting time is based on *First Come First Served* (FCFS) assumptions, i.e. activities are carried out in the order of their PAT. As such, when it comes to the assignment of care workers to activities, durations of care activities and the exact location of clients are not taken into account. The analysis presented in this chapter is the first step towards balancing efficiency and meeting clients' preferences, and may serve as input for more advanced scheduling algorithms, as well as the use of more flexible staffing mechanisms.

The remainder of this chapter is structured as follows. Section 4.2 provides an overview of the existing literature and gives a brief description of the empirical context. In Section 4.3, we investigate how the PATs fluctuate over time and over the course of a day. Through the use of simulation, we analyze the consequences of these fluctuations in terms of workload and waiting time, thereby taking into account potential economies of scale by measuring the effect of pooling care workers from multiple departments. Finally, in Section 4.4, we present our conclusions and propose future research directions.

## 4.2 Context

The following subsections give an overview of related literature and provide information regarding the empirical context of the departments under study.

### Operations research in nursing home care

There is a growing body of Operations Research (OR) literature on capacity planning in healthcare. The vast majority of the 'OR in healthcare' literature is on capacity decisions in hospitals (e.g., Rais and Viana, 2011) and to date the area of nursing home care has received hardly any attention. This finding is in line with Hulshof et al. (2012), who provide an overview of studies in the field of OR and Management Science (MS) in healthcare and find that the body of OR/MS literature directed to residential care services is limited. One prominent reason for this is the lack of (reliable) data. According to Spruit et al. (2014), most of the available data is qualitative and largely unstructured. Furthermore, Spruit et al. (2014) point out that the available data is often not used for decision making.

Outside the field of OR there have been numerous studies that address the relationship between staffing levels and the quality of care in a nursing home setting. The literature study of Spilsbury et al. (2011) provides a systematic overview. In their study, the authors conclude that, although there is a growing body of literature, it is difficult to draw conclusions and offer recommendations based on existing studies. Spilsbury et al. (2011) also conclude that existing studies mainly focus on clinical outcomes as a measure of quality.

The allocation of care workers in a home care setting has been studied more extensively (e.g., Mankowska et al., 2014; Rasmussen et al., 2012; Eveborn et al., 2006). When it comes to home care, care workers are assigned to care and support activities, which should be carried out within a client-specific time window. As these activities are performed at the clients' homes, the spatial component is a crucial element. The nursing home scheduling problem is related to the vehicle routing problem with time windows, which is known to be NP-hard (e.g., Bredström and Rönnqvist, 2008). Since home care organizations

**Table 4.1:** Type of care per department

| Department | Type of care |
|---|---|
| A | Somatic care (assisted living) |
| B | Psychogeriatric care (assisted living) |
| C | Somatic care |
| D | Somatic care |
| E | Psychogeriatric care |

usually have to schedule tens or hundreds of activities, approximation methods (or heuristics) have been proposed to give satisfactory outcomes within an acceptable time frame.

### Nursing home context

The nursing home departments under study provide accommodation for people who require assistance with daily activities such as washing, dressing, eating, drinking and taking medication. Furthermore, medical attention is given as required. Each of the five departments is responsible for its own (health)care activity scheduling process. Table 4.1 shows that the type of care differs per department.

In order to make it possible for nursing home residents to live their lives according to their preferences, all departments aim to deliver the necessary care as close as possible to the time preferences of the residents, i.e. minimizing delay. The time preferences are inventoried on a regular basis, using a standardized, systematic method.

Maintaining the appropriate number and mix of care workers is critical to meeting the time preferences of the clients (see also Chapter 5). The care worker-to-client ratios currently applied by departments C, D and E are shown in Figure 4.1. Due to the lack of data, we assume that departments A and B use the same worker-to-client ratios. Furthermore, Figure 4.1 also shows worker-to-client ratios that are based on a recent study of The Netherlands Institute for Health Services Research (NIVEL) (Hingstman et al., 2012) among a large sample of Dutch nursing homes. There seems to be a reasonable fit between the ratios used by the nursing homes under study and the ratios presented by NIVEL. We note that the peak around 15:00 hours presented in the NIVEL study is due to information transfer between shifts.

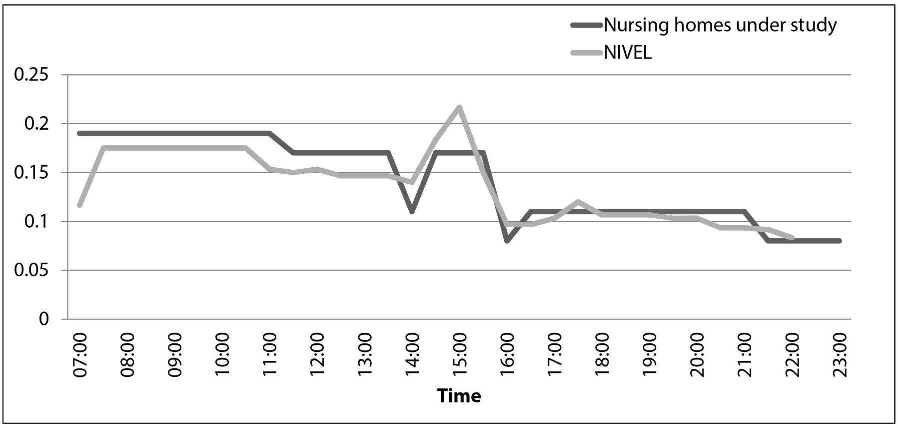


**Figure 4.1:** Care worker-to-client ratios

For the delivery of care and support, the departments under study 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. Qualification-level one (QL1) is required to perform the least complex tasks whereas QL4 is required to perform the most complex tasks (see also Table 5.1 in Chapter 5). As the aim is to provide insight on an aggregate level, task requirements in terms of qualification levels are not taken into account.

## 4.3 Analysis of demand, workload and waiting times

Five separate datasets were used (one for each department). Each dataset contains the care delivery data of 91 days, from January 1 until April 4 2014, and consists of the following variables:

- *Client ID* – the ID of a specific client.
- *Preferred Activity Time (PAT)* – the preferred starting time of the healthcare activity.
- *Date* – the date of the PAT.
- *Task description* – a brief description of the activity (i.e. healthcare task) entered as free text.

**Table 4.2:** Number of clients and size of dataset

| Dep. | Number of clients | Average rows of data per day | Average rows of data per day / Number of clients |
|---|---|---|---|
| A | 49 | 123 | 2.5 |
| B | 9 | 61 | 6.8 |
| C | 34 | 176 | 5.2 |
| D | 28 | 260 | 9.3 |
| E | 33 | 263 | 8.0 |
| Total | 153 | 884 | 5.8 |

- *Expected service time* – expected duration of the activity in minutes.

Table 4.2 shows the size of the datasets and the number of clients per department. The final column gives the average number of care activities per client. It can be observed that the number of activities differs substantially between departments, and that department A (assisted living for somatic clients) has fewer activities per client.

The analysis is structured as follows. First, for each department, the demand patterns (Section 4.3) and care delivery times (Section 4.3) are analyzed. In Section 4.3 the average workload during the course of a day is examined. Finally, we provide insight in the consequences of the fluctuation in workload in terms of waiting time, and take into account possible advantages by pooling care workers from multiple departments. In the interest of the readability, the visualization of the data is only for the two departments C and D. Figures and tables for the other departments can be found in the Appendix.

## PAT-analysis

To get an impression of the demand patterns, we divide the days into time periods (time buckets) of 5 (or 30) minutes and calculate the number of PATs within each time bucket during the course of a day. The goal of this PAT analysis is to create a vector with the estimated number of PATs per interval. This vector should visualize the demand over the course of a day for a department $Y$, where $Y \in \{A, ..., E\}$. We let $t$ denote the number of time intervals; for time buckets of 5 (or 30) minutes, the vector has length 288 (or 48). Let $X_{t,d}(y)$ be the number of PATs in interval $t$ at day $d \in \{1, ..., 91\}$, for department $y$, where $y \in \{A, ..., E\}$. For department $y$ this results in a matrix of the form:

**Table 4.3:** Results of the Friedman test

| Department | 5 minute intervals $p$-value | 30 minute intervals $p$-value |
|---|---|---|
| C | 0.9063 | 0.999 |
| D | <2.2e-11 | <2.2e-11 |

$$\begin{pmatrix} X_{1,1}(y) & X_{1,2}(y) & \cdots & X_{1,91}(y) \\ X_{2,1}(y) & X_{2,2}(y) & \cdots & X_{2,91}(y) \\ \vdots & \vdots & \ddots & \vdots \\ X_{T,1}(y) & X_{T,2}(y) & \cdots & X_{T,91}(y) \end{pmatrix}$$

The first step is to determine to what extent the PATs vary between days within each department. Hence, the following assumption is tested:
For each $y$ = A,...,E,

$$X_{t,1}(y) = X_{t,2}(y) = X_{t,3}(y) = ... = X_{t,91}(y), \qquad t = 1, 2, ..., T. \tag{4.1}$$

Because there is no well-known distribution for the number of PATs in a time interval, we use the non-parametric Friedman test. Table 4.3 shows the results of the Friedman test for both 5 and 30 minute intervals.

The test results show that only department D has a $p$-value $< 0.05$ (see Appendix 4.A for the results from the other departments). This allows a tentative conclusion to be drawn, indicating that only for department D the PATs vary between days. An interaction plot was used to analyze further the outcome of department D (see Appendix 4.F). The interaction plot shows a different pattern for the first ten days compared to the final 81 days. In fact, after day ten the interaction plot indicates that there are little differences between the days.

Next, we investigate the daily PAT patterns in more detail using the column vector $\overline{X}(y) = (\overline{X}_1(y), \overline{X}_2(y), ..., \overline{X}_{48}(y))'$, where $\overline{X}_t(y) = \frac{1}{91}\sum_{d=1}^{91} X_{t,d}(y)$ represents the average number of PATs at time $t$ for department $y$ (here $'$ denotes transpose). In order to make it possible to compare the demand patterns of the departments we plotted the relative PAT-values in relation to the total number of clients on the vertical axis (see Figures 4.2, 4.3 and Appendix B). To smooth out the inconsistencies in the first 10 days of department D, a 20% trimmed mean is used to determine the PAT-values per interval.

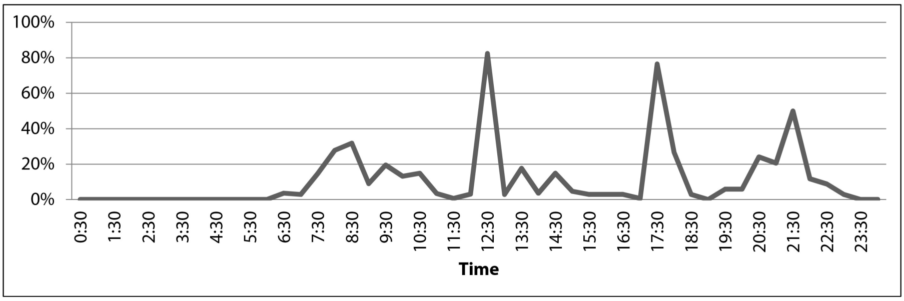


**Figure 4.2:** Relative number of PATs for department C

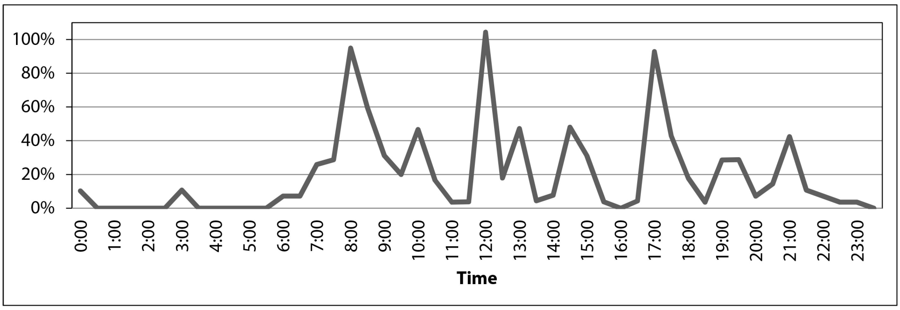


**Figure 4.3:** Relative number of PATs for department D

The plots show there is considerable fluctuation in PATs over the course of a day. Most of the fluctuation can be explained by the major activities of daily living. Between 0:00 and 6:00 hours there are hardly any activities (some clients have medication though). During the early morning, most residents need help with getting out of bed and/ or dressing. Around lunch and dinner time, there is need for assistance with feeding. Finally, at the end of the day, some of the residents need assistance with getting to bed. Furthermore, it can be observed that at some points in time the peaks exceed 100%. This is because some clients have more than one PAT during some time intervals. Department C has clearly fewer PATs, but the peaks are comparable to department D. Although peaks are dominated by the activities in daily living, we see that there are some differences in the exact times that peaks occur.

### Duration of care delivery

In addition to number of PATs, the workload is determined by the duration of care delivery. Figures 4.4, 4.5 and Appendix 4.C shows how the 'expected'

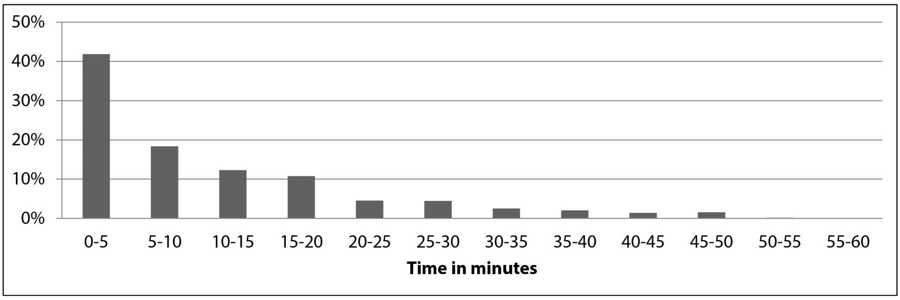


**Figure 4.4:** Distribution of care delivery times for department C

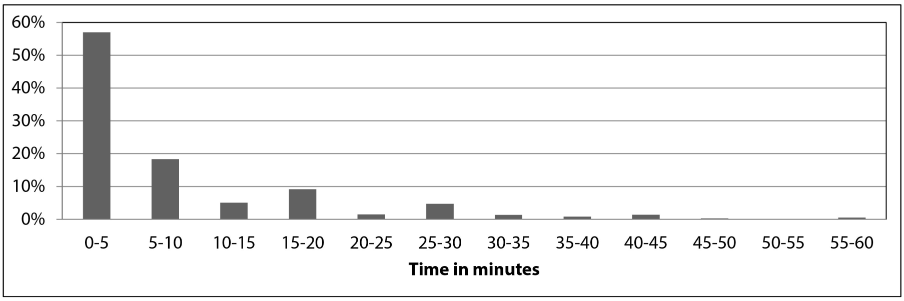


**Figure 4.5:** Distribution of care delivery times for department D

care delivery times (i.e. service times) are distributed for the departments under study.

The average 'expected' care delivery times for C and D are 13.2 and 10.0 minutes, respectively, with a standard deviation of 11.2 and 10.3 minutes. Most care delivery times are short and take 0–5 (or 5–10) minutes. However, there is considerable variation in care delivery times with activities that may take well over half an hour.

## Workload analysis

Based on the PATs and care delivery durations, this subsection examines the workload. The workload provides the aggregate demand for care. That is, the workload at time $t$ is the number of clients who need care at time $t$ ignoring capacity constraints. As such, it prescribes the required number of care workers at any time if demand would have been met directly (no waiting is allowed). In addition, the workload (i.e. the required number of care workers) is compared with the real-life staffing levels. As there is hardly any scheduled care during the night, we consider the time frame 7:00–22:00 hours. Now, we indicate how the workload and staffing levels are determined and then compare them to assess whether there is a mismatch.

**Workload:** Let $L_{t,d}(y)$ be the workload at day $d$ in interval t, for department y, where $t = 1$ denotes the first interval 07:00–07:05 hours and $y = A, ..., E$. For department $y$, $S_{t,d}(y)$ denotes the cumulative number of start times at day $d$ in interval $t$, i.e. the number of PATs between 7:00 hours and $t$, and $E_{t,d}(y)$ denotes the cumulative number of end times at day $d$ in interval $t$. Then, $L_{t,d}(y)$ is determined by

$$L_{t,d}(y) = S_{t,d}(y) - E_{t,d}(y) \quad t = 1, 2, ..., T \quad d = 1, ..., 91 \quad y = A, .., E$$

We then consider the average workload $\overline{L}(y) = (\overline{L}_1(y), \overline{L}_2(y), ..., \overline{L}_T(y))'$, where $\overline{L}_t(y) = \frac{1}{91}\sum_{d=1}^{91} L_{t,d}(y)$ is the averaged workload in time interval $t$ for department $y$.

**Staffing level:** The staffing levels are determined by the actual worker-to-client ratios, see Figure 4.1. Let $\overline{R}_t$ be the care worker-to-client ratio in interval $t$ and $I(y)$ the number of clients of department $y$. The available capacity in interval $t$ for department $y$ is then $\overline{C}_t(y) = \overline{R}_t I(y)$, rounded to the nearest integer.

The workload (i.e. the number of residents in need of care) and the available capacity are visualized together in Figures 4.6 and 4.7 and Appendix 4.D. The figures show that, at some moments during the day, the available capacity is insufficient to meet the time preferences of the residents. Hence, residents sometimes have to wait until a care worker is available to provide the necessary care and/ or support. Especially during morning care (7:00–9:00 hours) the workload is high. This finding is in line with the findings presented in Chapter 5 and the study of Sloane et al. (2007). Furthermore, analysis of the task descriptions corresponding to the (high) peaks in demand shows that some peaks are caused by 'serving coffee and tea' and 'giving medicines'. Such activities are sometimes combined, see Section 4.3 below.

## Waiting time analysis

In this section we assess the impact of the workload by determining the virtual waiting time during the course of an average day (7:00–22:00 hours). The virtual waiting at time $t$ is defined as the time that a client would have to

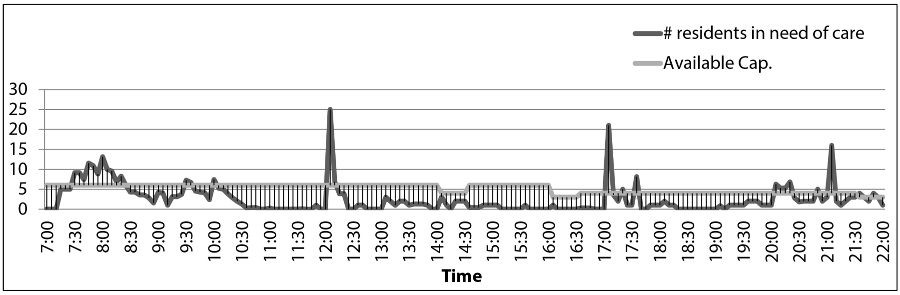


**Figure 4.6:** Workload and available capacity for department C

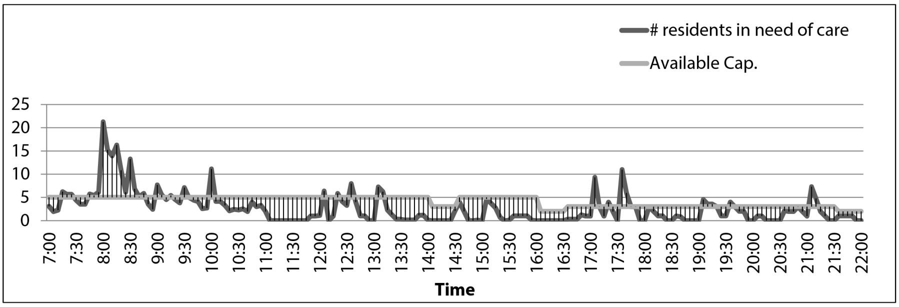


**Figure 4.7:** Workload and available capacity for department D

wait if the client were to have a (new) PAT at time $t$. The following simulation procedure is used to calculate the virtual waiting times[1]:

1. The day is divided into time buckets of 5 minutes ($t = 1, 2, ..., 180$).
2. Starting at 7:00 hours ($t = 1$), for every time step, care activities based on PATs are assigned to available care workers.
3. When all care workers are busy, a virtual queue is filled with the remaining activities.
4. PATs in the queue have priority over a 'new' PAT. Hence, a FCFS policy is used.

In order to make the virtual waiting times as realistic as possible, we made the following assumption with respect to 'serving out medicine'. During so-called 'medicine rounds' one care worker wheels around a medicine trolley to the individual residents. As such, during a medicine round, one of the care

[1] A Microsoft Excel VBA code was used to perform the simulation based on the available historic data.

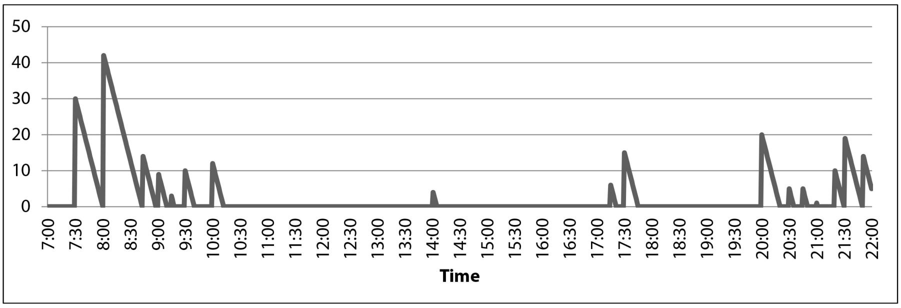


**Figure 4.8:** Waiting times (in minutes) for department C

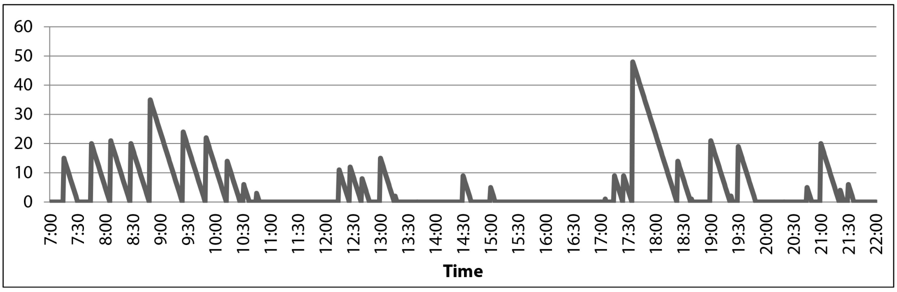


**Figure 4.9:** Waiting times (in minutes) for department D

workers is not available for other activities. Based on the data we used 5 minutes per resident as a norm to calculate the total duration of the 'medicine round'. PATs related to medicine are not taken into account for the (virtual) waiting times. The results of the simulations for the departments C and D are visualized in Figures 4.8 and 4.9, see Appendix 4.E for the virtual waiting times of departments A, B and E.

The results show that the longest virtual waiting times can be observed during the morning, with waiting times up to 40 minutes for both departments C and D. Also during the evening (say between 19:00 and 21:30 hours), when clients go to bed, significant waiting occurs. Compared with the morning, the peak in the virtual waiting time is typically smaller, whereas there is more variation in the timing at which the peak occurs. Between 11:00 and 17:00 hours waiting times are typically short (there is some waiting for departments D and E), whereas just before or just at dinner time there is typically a peak in the virtual waiting time. Department A forms an exception. This can be explained by the fact that the residents of this department are less dependent and thus need less assistance with getting out of bed during the morning.

**Table 4.4:** Simulation results with the waiting time $W$ in minutes

| | Departments | |
|---|---|---|
| | C | D |
| # Clients | 34 | 28 |
| # PATs | 111 | 203 |
| $\overline{W}$ | 5.82 | 11.80 |
| $\sigma_{\overline{W}}$ | 7.00 | 6.42 |

However, it could be the case that in practice there are fewer care workers available for department A, as the care worker-to-client ratios we used for the analysis are based on C, D and E.

Table 4.4 and Appendix 4.E show both the average waiting time ($\overline{W}$) and the standard deviation of the waiting times ($\sigma_{\overline{W}}$). The average waiting time varies between 0.53 (department A) and 11.80 minutes (department D), whereas the standard deviation of the waiting times varies between 1.17 minutes for department A and 7.00 minutes for department C. This indicates that even in the absence of uncertainty, considerable waiting occurs before care activities are carried out compared with their PAT.

### Impact of scale

Finally, we investigate the effect of scale on the virtual waiting times. To be more specific, we examine the advantages of pooling care workers and PATs of two departments. The effect of pooling departments C and D is shown in Figure 4.10, see Appendix 4.E for the effect of combining departments A and C, and B and E, respectively. Table 4.5 and the table in Appendix 4.E show the effect of pooling in terms of total and average waiting times. As the total number of clients for departments C and D is 62, we assume that distances between different clients are not excessive and thus not influence waiting times. When aggregating on a larger scale, we think travel time should be taken into account.

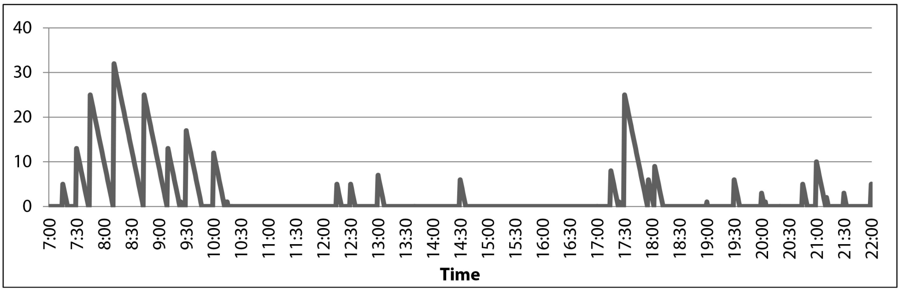


**Figure 4.10:** Waiting times (in minutes) for departments C & D

**Table 4.5:** Simulation results with W in minutes for departments C & D

| | | |
|---|---|---|
| # Clients | | 62 |
| # PATs | | 314 |
| Pooling C & D | $\overline{W}$ | 6.29 |
| | $\sigma_{\overline{W}}$ | 6.04 |
| Unpooled | *Weighted* $\overline{W}$ | 9.69 |

The final row of Table 4.5 shows the weighted average waiting time for the unpooled situation, which is based on Section 4.3. The weights are based on the number of PATs at each department and the weight for department $i$ in case of pooling a set $Y$ is determined by

$$\omega_i = \frac{\#\ \text{PATs}_i}{\sum_{j \in Y} \#\ \text{PATs}_j}.$$

The table shows that pooling departments B and C has a positive effect on the average waiting time. For department C, the waiting time slightly increases (from 5.82 to 6.29) whereas the standard deviation of the waiting time decreases. The average waiting time for department D clearly decreases. So, department D benefits from pooling in terms of average waiting time and for department C the waiting time becomes slightly more stable. For instance, the peak in the virtual waiting time process for the pooled situation is just above 30 minutes, whereas it is above 40 minutes for department C in the unpooled situation. The table in Appendix 4.E shows that pooling departments A and C, and B and E, respectively, has a strong positive effect on the average (virtual) waiting time. In those cases, the average waiting time decreases for all departments when they are pooled.

More generally, we observe that pooling departments has a positive impact on respecting PATs. Nonetheless, it seems that some peaks, in particular during morning care, cannot be avoided by adjusting scale.

## 4.4 Conclusions and discussion

This chapter investigated how the demand patterns of scheduled care of five Dutch nursing home departments fluctuate over time and over the course of a day. The analysis shows that most of the fluctuations can be explained by the major activities of daily living. We also examined the consequences of these fluctuations in terms of workload and waiting time. The results show that especially during the early morning the workload is high, which results in (virtual) waiting times up to 40 minutes. Furthermore, we have shown that increasing the scale of scheduling by pooling care workers from multiple departments has a substantial positive effect on the virtual average waiting times. The pooling effects, although considerable, are less pronounced during busy periods of the day. This is because, (1) these busy periods take place at roughly the same time for each of the departments and (2) during these periods most of the departments work near maximum capacity. Finally, the results of our analysis show that the staffing levels are not well balanced over the course of a day.

Based on the above, we see two primary areas of interest to meet client preferences in practice. First, organizing care on a (slightly) larger scale helps to satisfy clients PATs as peaks in PATs may differ between departments. For the major activities in daily living, organizing on a large scale decreases waiting times, but they remain considerable. To meet PATs during such 'busy' intervals, more flexibility is required. When it comes to increasing the flexibility, there are basically three strategies that can be applied (and can be mixed). A first possible strategy is to increase the demand-side flexibility, by using time windows during which the activities are supposed to be carried out. Another possible strategy is modifying staffing levels during the course of a day, which is referred to in Chapter 3 as numerical flexibility. 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. Finally, a third possible strategy is to increase functional flexibility (e.g., multi-tasking). For example, in Chapter 5 we show that considerable efficiency gains can be obtained by enlarging care workers' jobs (i.e. blending tasks of different qualification levels).

To further support the organization of nursing home care, we envisage the following subsequent steps. A primary element is that we only considered 'scheduled care' in the current chapter. It is however not clear to what extent nursing homes should mix 'scheduled' and 'unscheduled' care activities. In this chapter, the underlying assumption regarding waiting times is that care workers are dedicated to 'scheduled care'. For future research, we aim to study mixing 'scheduled' and 'unscheduled' activities. Obtaining good data on both type of activities from the same department and the same period seems crucial, as these type of data can be compromised. For example, long waits for scheduled activities may trigger unscheduled care requests.

Finally, for conciseness of presentation, we ignored differences in qualification levels (QLs) of care workers. Moreover, in our waiting time analysis we assumed a FCFS policy, where a preference is according to a time instant. Including QLs, time preferences entered as intervals instead of single time epochs, and travel time (in particular when care is organized on a larger scale) leads to interesting and advanced scheduling methods. The current dataset and analysis is an important first step, which may well serve as input for such optimization procedures.

## Acknowledgements

We would like to thank the anonymous nursing care organization, Verkooijen & Beima and CCmath for providing us with a dataset.

## 4.A Friedman test

Table 4.A.1: Results of the Friedman test

| Department | 5 minute intervals $p$-value | 30 minute intervals $p$-value |
|---|---|---|
| A | 0.7661 | 0.7117 |
| B | 1 | 1 |
| C | 0.9063 | 0.999 |
| D | $< 2.2e\text{-}11$ | $< 2.2e\text{-}11$ |
| E | 1 | 1 |

## 4.B Relative number of PATs

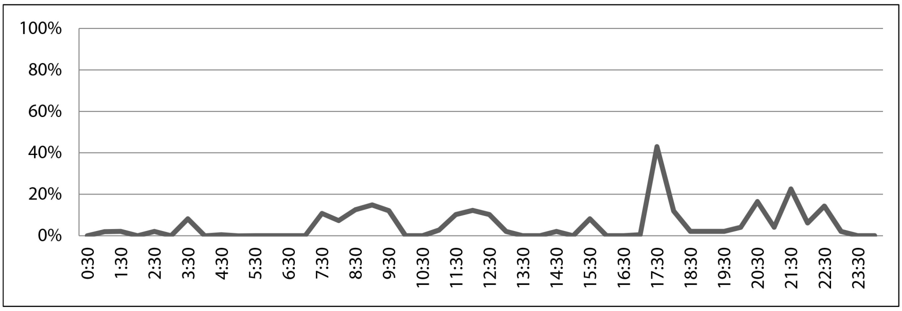


**Figure 4.B.1:** Relative number of PATs for department A

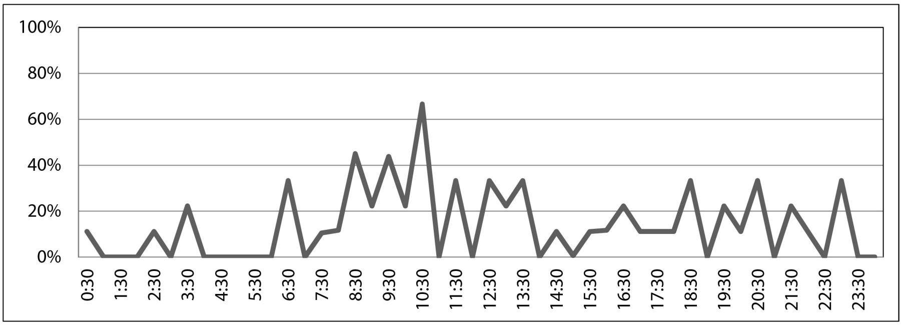


**Figure 4.B.2:** Relative number of PATs for department B

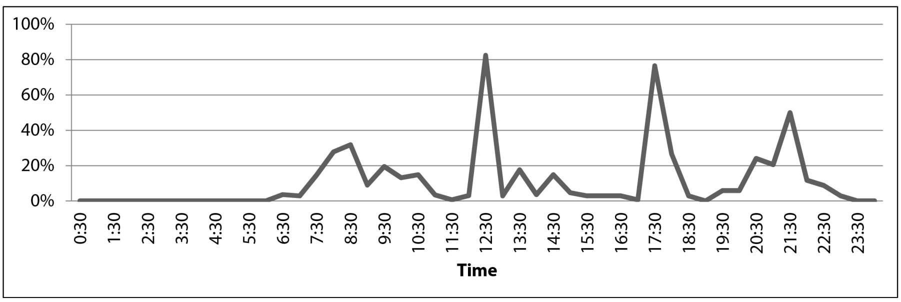


**Figure 4.B.3:** Relative number of PATs for department C

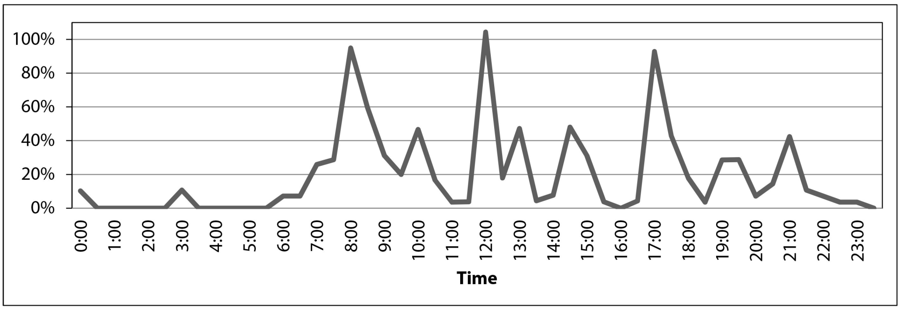


**Figure 4.B.4:** Relative number of PATs for department D

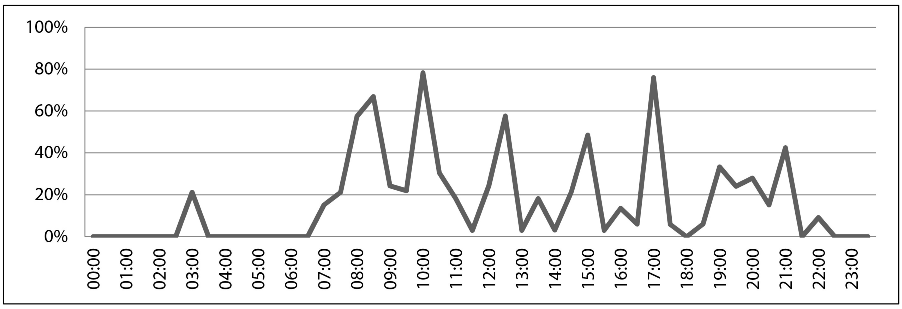


**Figure 4.B.5:** Relative number of PATs for department E

## 4.C Care delivery times

**Table 4.C.1:** Care delivery times

| Department | Average service time in minutes | Standard deviation service time in minutes |
|---|---|---|
| A | 12.5 | 9.0 |
| B | 11.5 | 11.8 |
| C | 13.2 | 11.2 |
| D | 10.0 | 10.3 |
| E | 12.4 | 8.9 |

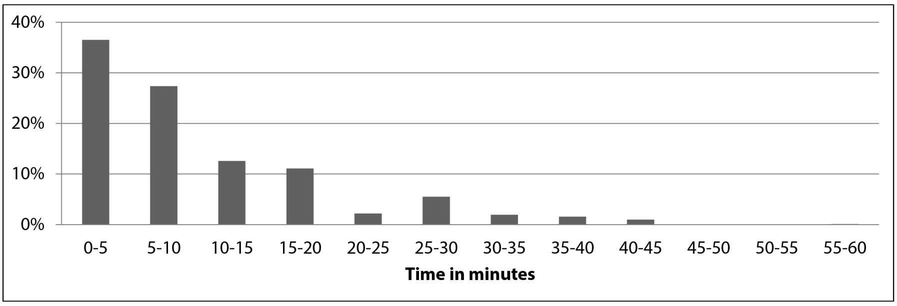


**Figure 4.C.1:** Distribution of care delivery times for department A

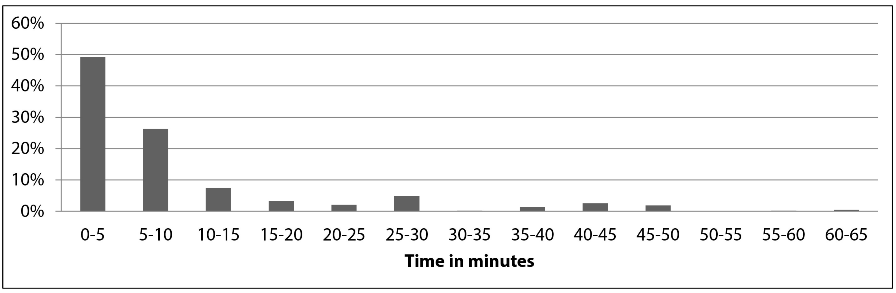


**Figure 4.C.2:** Distribution of care delivery times for department B

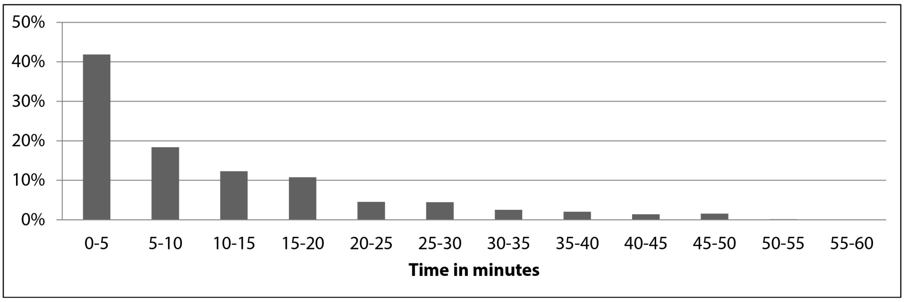


**Figure 4.C.3:** Distribution of care delivery times for department C

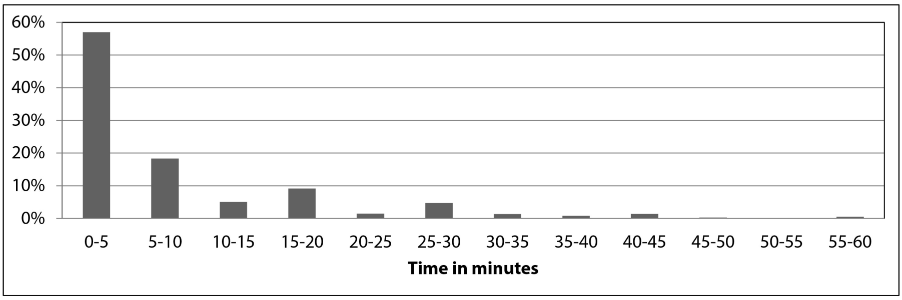


**Figure 4.C.4:** Distribution of care delivery times for department D

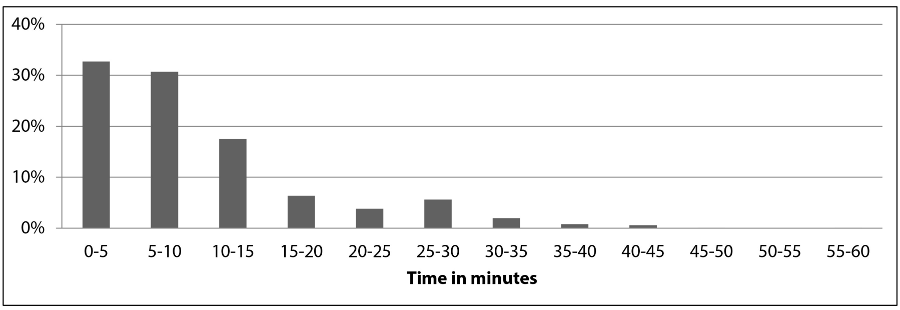


**Figure 4.C.5:** Distribution of care delivery times for department E

## 4.D Workload and available capacity

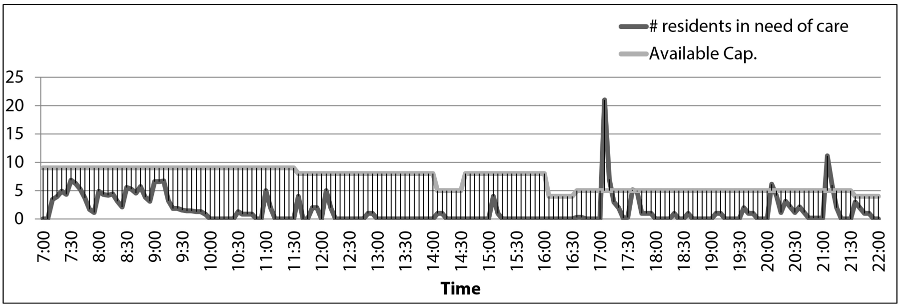


**Figure 4.D.1:** Workload for department A

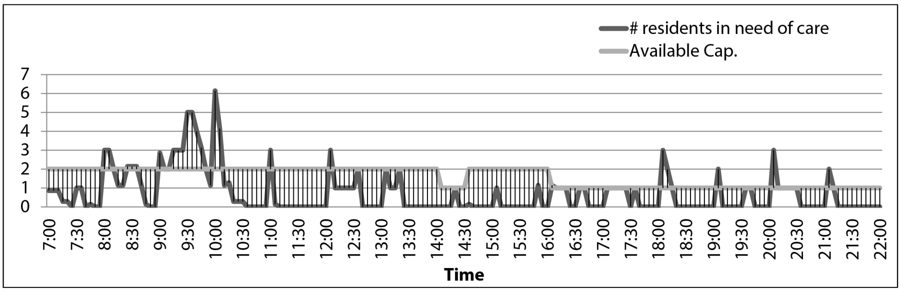


**Figure 4.D.2:** Workload for department B

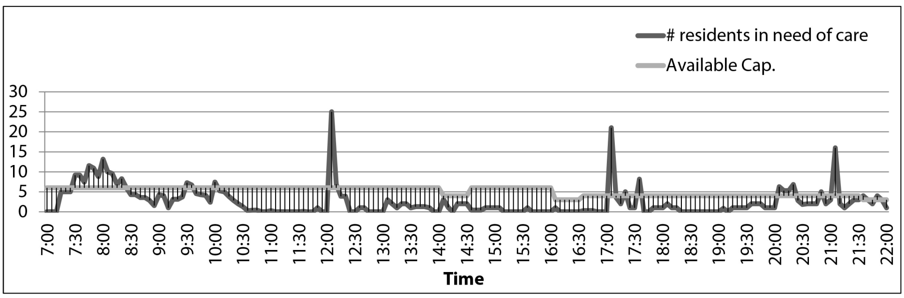


**Figure 4.D.3:** Workload for department C

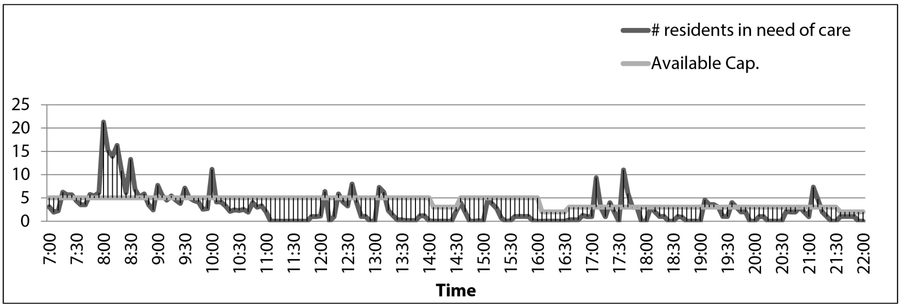


**Figure 4.D.4:** Workload for department D

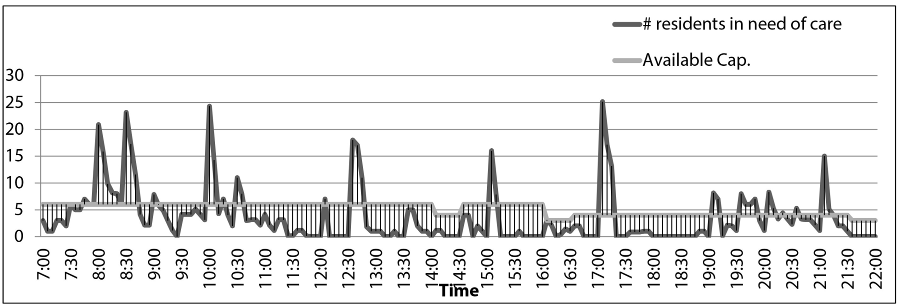


**Figure 4.D.5:** Workload for department E

## 4.E Overview waiting time analysis

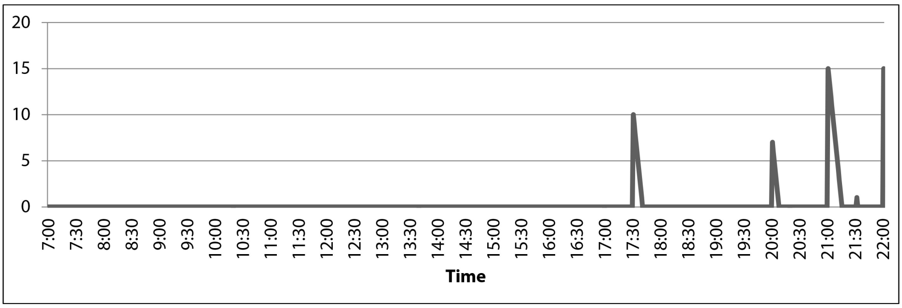


**Figure 4.E.1:** Waiting times for department A

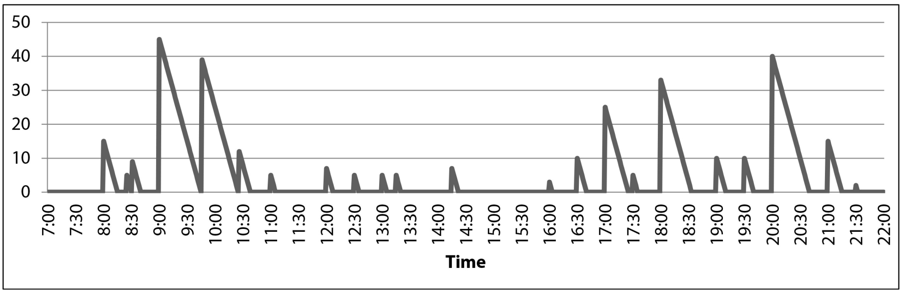


**Figure 4.E.2:** Waiting times for department B

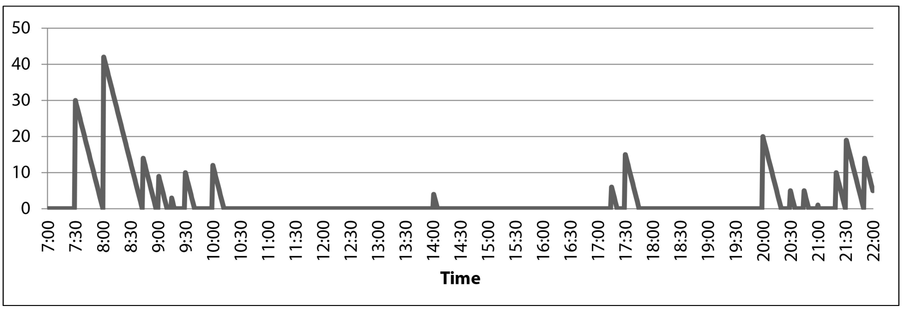


**Figure 4.E.3:** Waiting times for department C

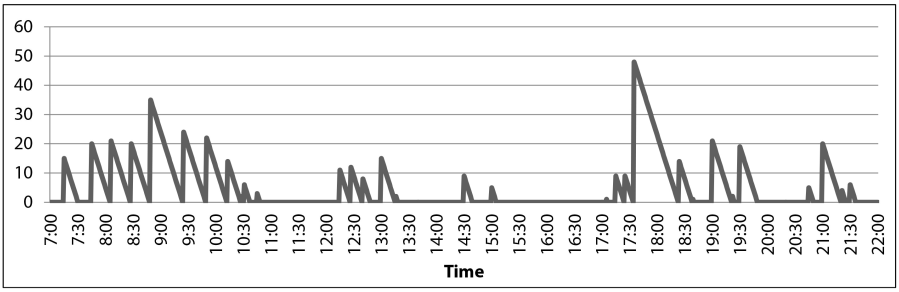


**Figure 4.E.4:** Waiting times for department D

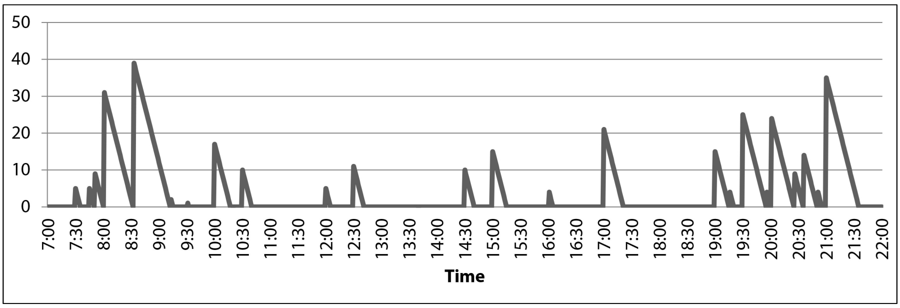


**Figure 4.E.5:** Waiting times for department E

**Table 4.E.1:** Results waiting time analysis departments

| | Departments | | | | |
|---|---|---|---|---|---|
| | A | B | C | D | E |
| # Clients | 49 | 9 | 34 | 28 | 33 |
| # PATs | 93 | 54 | 111 | 203 | 204 |
| $\overline{W}$ | 0.532 | 6.88 | 5.82 | 11.80 | 9.87 |
| $\sigma_{\overline{W}}$ | 1.17 | 5.51 | 7.00 | 6.42 | 5.15 |

**Table 4.E.2:** Results waiting time analysis pooled

| | | Departments | | |
|---|---|---|---|---|
| | | A & C | C & D | B & E |
| Pooled | # Clients | 83 | 62 | 42 |
| | # PATs | 204 | 314 | 259 |
| | $\overline{W}$ | 0.51 | 6.29 | 5.91 |
| | $\sigma_{\overline{W}}$ | 1.01 | 6.04 | 3.90 |
| Unpooled | $weighted\overline{W}$ | 3.41 | 9.69 | 9.21 |

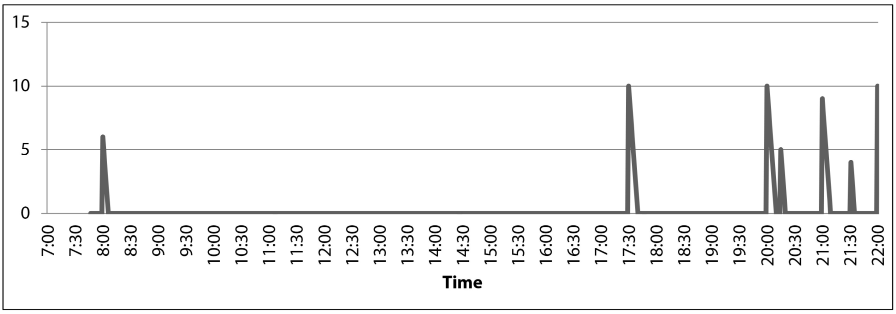


**Figure 4.E.6:** Waiting times for department A & C

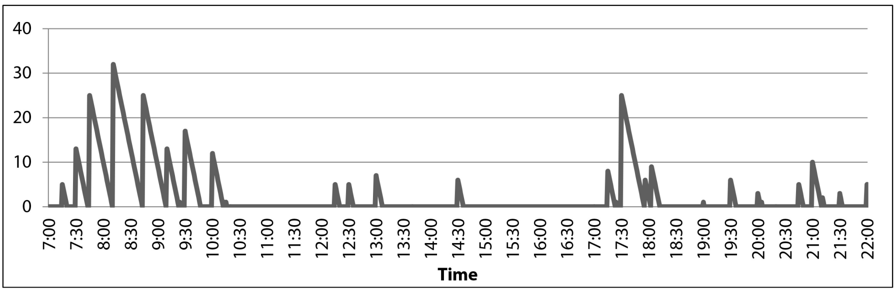


**Figure 4.E.7:** Waiting times for department C & D

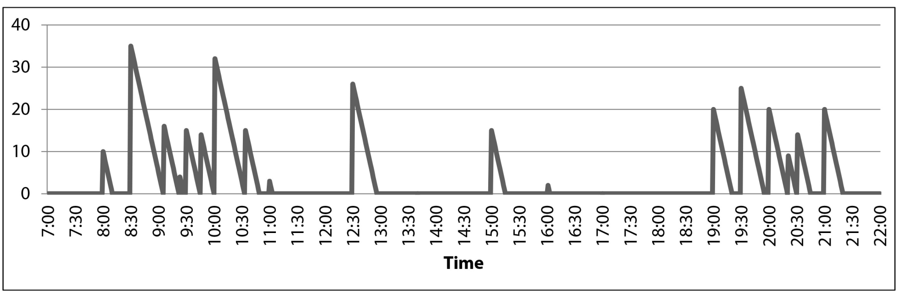


**Figure 4.E.8:** Waiting times for department B & E

## 4.F Interaction plot department D

The plot below displays PAT levels, which imply that, on some day $d$, a level $x$ is attained when at least one of the $X_{t,d}(y)$ is equal to $x$. Each dot in the interaction plot represents a PAT-level.

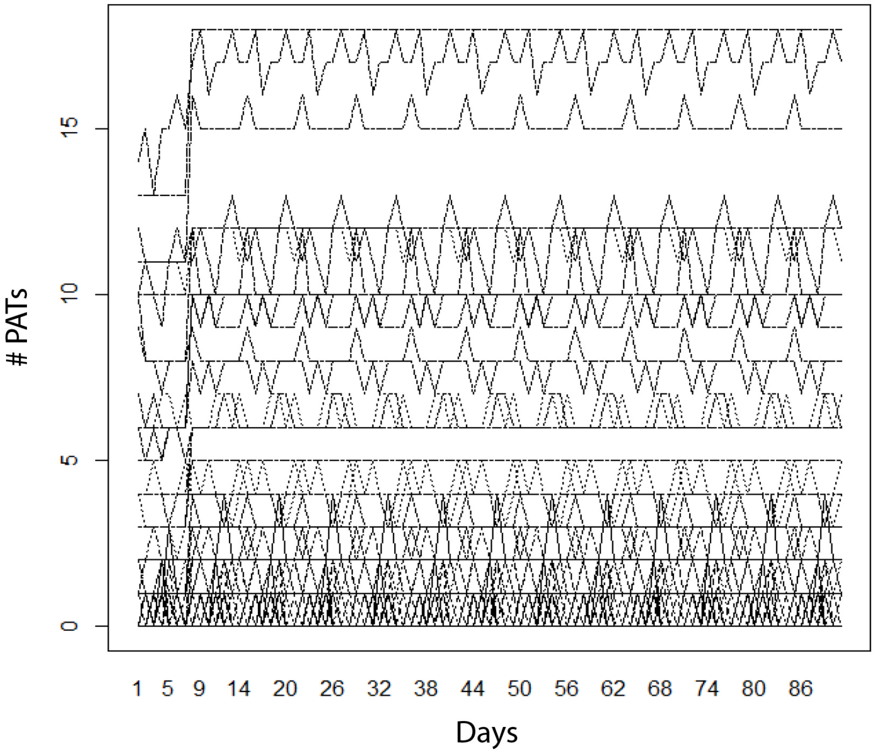


**Figure 4.F.1:** Interaction plot for department D (5 minute intervals)

Chapter 5

# Scale and skill-mix efficiencies in nursing home staffing: inside the black box


**Abstract**

This chapter provides insight in how and why 'scale of scheduling' and the enlargement of care workers' jobs (blending tasks of different qualification levels) affect the number and type of staff required to meet the preferences (in terms of day and time) of nursing home residents. We examined the scheduled care activities of three separate decision-making units within a single Dutch nursing home. The results show that in most of the examined cases, substantial scale and skill-mix economies can be achieved. Furthermore, we also found that the correlation between the demand patterns of different types of care tasks is of considerable importance when it comes to possible scale and skill-mix efficiencies.

This chapter is based on Moeke, Koole and Verkooijen (2014).

## 5.1 Background

To make it possible for nursing home clients to live the lives they prefer, it should be possible for them to have influence on the moment (day and time) at which care will be delivered. According to Bosman et al. (2008), nursing home clients “do not want to adjust their lives to the carers schedule” (p. 523). As such, nursing homes should aim to deliver the necessary care as close as possible to the time preferences of their clients. Unfortunately, in current practice, many nursing homes struggle to meet this expectation. Chapter 4 showed that, especially during the morning care, considerable waiting times may occur due to resource constraints. We concluded Chapter 4 with saying that in order to meet the time preferences during ‘busy’ intervals, more flexibility is required. In this chapter we examine the effectiveness of functional flexibility during the morning care (7:00–11:00 hours). That is, we investigate how and why the enlargement of care workers’ job tasks (blending tasks of different qualification levels) affect the number and type of staff required to meet the preferences (in terms of day and time) of nursing home clients. To do so, we make use of an advanced scheduling algorithm (see Section 5.3) which incorporates time windows during which the activities are supposed to be carried out. Furthermore, we also more closely examine the effects of ‘scale of scheduling’ in terms of efficiency. Questions with respect to scale are prominent, as there is an ongoing tendency towards small-scale living facilities when it comes to nursing home care. These small-scale living facilities “aim at providing nursing care in small groups (6-10 residents per house) emphasizing normalization of daily life and encouraging residents to participate in meaningful activities” (Verbeek et al., 2010, p. 663).

The rest of this chapter is organized as follows. In the next section, we justify the study by looking at related literature. Section 5.3 provides an overview of the dataset used for this study and gives a brief description of the empirical context. The results in terms of possible scale and skill-mix efficiencies are presented in Section 5.4. Next, in Section 5.5 we propose a correlation measure to quantify the correlation between the demand patterns of different types of care tasks. In Section 5.6 we validate our correlation measure with help of a simulation experiment. Finally, in Section 5.7, we present our conclusions and propose future research directions.

## 5.2 Previous research

As we know from other areas of scheduling research (e.g., call centre scheduling) in general the number of resources needed decreases relative to size (Gans et al., 2002). These economic advantages that result from carrying out a process on a larger scale are often referred to as 'economies of scale'. Economies of scale in long-term care facilities have been the subject of several international studies. The most commonly used methods for measuring scale efficiencies are data envelopment analysis (DEA) (e.g., Kooreman, 1994; Chattopadhyay and Ray, 1996), stochastic frontier analysis (SFA) (e.g., Farsi et al., 2008; Knox et al., 2007) and regression analysis (e.g., Christensen, 2004; Blank and Eggink, 2001; Gertler and Waldman, 1992). However, these studies:

1. Mainly focus on the relative efficiency of health-care facilities, but give little or no insight into how structural and process-related factors interrelate to contribute to economies of scale on a more operational level.
2. Do not use the degree to which the clients' preferences are being met as a measure of effectiveness.

Another important potential contributor in making more efficient and effective use of care workers is changing the 'skill-mix' or 'scope of practice' of the workforce (e.g., Dubois and Singh, 2009; Schluter et al., 2011). According to Buchan and Calman (2005, p.4): "Skill mix is a relatively broad term which can refer to the mix of staff in the workforce or the demarcation of roles and activities among different categories of staff". Based on the existing literature on skill-mix in health care we can conclude that:

1. Current studies provide insufficient practical guidance regarding the most efficient and/-or effective skill-mix. Most studies focus on the impact of poor staffing levels or skill-mixes, rather than seeking to develop or evaluate innovative ways of working (Flynn and Mckeown, 2009). According to Masterson (2004), much of the literature states that there is a need to define an appropriate skill-mix but contains very little on how this could be done. Therefore future research should focus on what combination of skill level contributes to quality in the most efficient manner (Spilsbury et al., 2011).
2. Reliable and valid information on skill-mix and staffing levels is scarce and data is seldom collected (Masterson, 2004). The CNA (2004) states

that: "there is a general lack of uniform, reliable and available data on nurse staffing, preventing the effects of skill-mix from being understood" (p. 11).

3. There is no "one size fits all" (Masterson, 2004). Therefore, a skill-mix cannot be considered in isolation from its (organizational) context (Dubois and Singh, 2009). This is important to bear in mind, because without a robust process for calculating appropriate nurse staffing levels for particular settings it is likely that care workers will lack confidence in such evaluation exercises (Flynn and Mckeown, 2009).

4. Current studies do not incorporate the degree to which the clients' preferences are being met as a measure of output. Instead, they mainly focus on 'clinical outcomes' for clients which are easier to measure (Spilsbury et al., 2011).

Thus, from a direct usability perspective, the concepts of 'economies of scale' and 'skill-mix' largely remain a black box. Based on our literature review we also came to the conclusion that, for the purpose of this study, a more precise definition of the concept of 'skill-mix' is needed. Therefore, in this chapter a distinction is made between 'grade mix' and 'skill-mix' (Gibbs et al., 1991). Grade mix, as defined in this chapter, refers to the number of care workers in each category that is required. Skill-mix, on the other hand, refers to the demarcation of roles and activities among the different categories of care workers. By 'category' we mean the distinct levels of nursing practice based on educational preparation and defined competencies. In this chapter we intend to pry open the black box by finding an answer to the following research question: when looking at the scheduled of care activities in a nursing home, how and why do staffing decisions with respect to scale and skill-mix affect the number and type of staff required to meet the preferences (in terms of day and time) of nursing home residents?

## 5.3 Dataset

To find an answer to the research question a six-day dataset of three separate locations within a single Dutch nursing home facility has been analyzed. The dataset (one for each location) consists of the following variables:

- *Client ID* - the ID of a specific client.

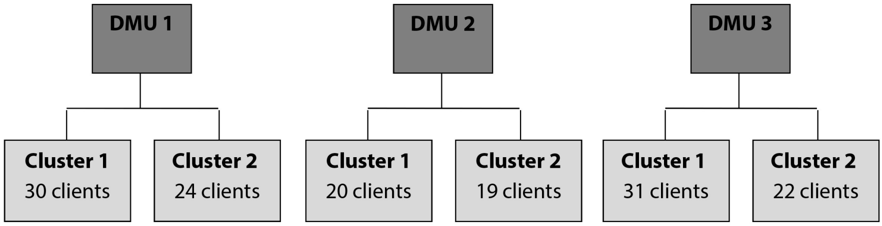


**Figure 5.1:** Number of clients living in each cluster

- *Preferred Activity Time (PAT)* - the preferred starting time of the healthcare activity.
- *Task description* - a brief description of the activity (i.e. healthcare task) entered as free text.
- *Qualification level* - the qualification level required to perform the healthcare task.
- *Date* - the date of the PAT.
- *Expected service time* - expected duration of the activity in minutes.

More detailed information about the characteristics of the dataset can be found in the Appendix in Tables 5.A.2 and 5.A.1. Because each of the three locations is responsible for its own (health)care activity scheduling process, we use the term decision-making unit (DMU). Each DMU consists of two clusters with separate rooms for each client (see Figure 5.1). The current situation (i.e. base case setting) can be described as follows. All three DMUs provide somatic care and work with a fixed pool of care workers. When it comes to the daily scheduling of (health)care activities, care workers are allocated to a specific cluster. In the remainder of this chapter, this is referred to as 'small scale scheduling'. New incoming clients are assigned to a cluster based on both room availability and expected intensity of care.

In their study, Sloane et al. (2007) state that the need for assistance with activities of daily living is most concentrated during the morning. This is consistent with our finding: Figure 5.2 shows that the morning care between 7:00 and 11:00 hours accounts for more than 50% of the total daily workload[1]. Because of this extra workload burden most nursing homes struggle to meet he

[1]Based on a 6-day average.

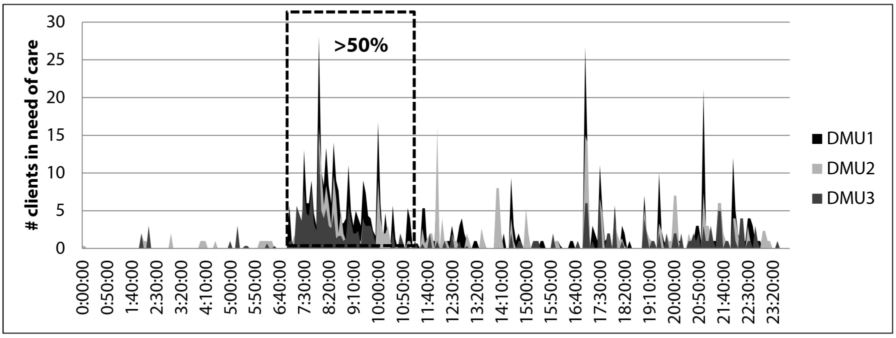


**Figure 5.2:** Demand pattern for care activities

individual preferences and needs of their clients efficiently during this part of the day. In this chapter we therefore focused on the plannable care activities during the morning shifts (7:00–11:00 hours).

## Scheduling support tool

To ensure adequate scheduling of (health)care activities, nursing homes require a method to identify clients' needs which enables an appropriate allocation of care workers. For their activity scheduling all three DMUs make use of a scheduling support tool called OERplanner. The OERplanner is based on the OERmodel developed by Verkooijen (2006). Consistent with the OERmodel, the OERplanner embodies a client-centred perspective and uses the clients' preferences, in terms of moment (day and time), as a starting point for determining the required number of care routes. Every care route consists of a number of activities, which are to be carried out in a fixed sequence, and is dedicated to a single care worker with the required qualification level.

As mentioned in Chapter 4, in Dutch nursing homes, care workers are often hierarchically divided into distinct qualification levels (QLs). This so called differentiated practice is based on a distinction in education, responsibility and complexity of care (Jansen et al., 1997). The idea behind the concept of differentiated practice is to make optimal use of the available care workers. The division of (health)care into distinct levels of nursing required by the client population and setting, should make it possible for nursing homes to achieve a more efficient and effective skill-mix. In the current situation all three DMUs use differentiated practice for the assignment of healthcare tasks to their care workers. Table 5.1 shows the QLs which are relevant for this study.

**Table 5.1:** Qualification Levels

| Qualification level | Tasks to be carried out |
|---|---|
| QL1 | Bringing food and drinks, cleaning, transferring, bed cleaning |
| QL2 | Getting in/out of bed, eating, toileting, making the beds, washing |
| QL3 | Giving medication, simple medical check ups |

### Algorithm of the OERplanner

The OERplanner uses an advanced algorithm to generate the care routes for the care workers. Starting with the highest qualification level it plans the activities one by one, in already existing care routes if possible within 15 minutes from the time desired by the client and with a nurse who has a qualification level that is close to that needed for the task to be scheduled. When this is not possible a new care route is created. This decision is based on a set of parameters that indicate to what extent the rules can be violated before a new care route is opened. By changing these parameters more emphasis can be given to filling up care routes completely, to adhering to the desired time, or to giving nurses tasks at their own qualification level.

## 5.4 Scale and skill-mix efficiencies

In order to answer the research question a six-day dataset has been used. For each day we examined the potential scale and skill-mix efficiencies by measuring the effect of:

1. Dedicating care workers to the DMU as a whole (large scale scheduling), without changing the skill-mix.
2. A change in skill-mix through (1) enlargement of QL3 workers' jobs with QL2 and QL1 tasks and (2) enlargement of QL2 workers' jobs with QL1 tasks (without enlarging the scale of scheduling).

To measure the potential efficiency gain, the number of care routes required according to the OERplanner was counted. As a reference we used the number of care routes which are required without applying scale and job enlargement (i.e. the current situation). Table 5.2 shows that in >85% of the examined cases substantial scale and skill-mix economies can be achieved. On average, the potential economies of scale for DMU1, DMU2 and DMU3 are 10%, 17%

and 21%, respectively. With respect to the potential skill-mix efficiencies the results show an average gain of respectively 9%, 26% and 23%. The potential efficiency gain is measured by counting the total number of care routes required. As reference we used the number of care routes which are required without applying scale and job enlargement.

Next, we examined the impact of a change in scale and skill-mix on the required grade-mix for each of the examined days (see Table 5.3). When applying large scale scheduling, in >65% of the cases fewer QL3 workers are needed. Job enlargement seems to have the opposite effect on the number of QL3 workers needed. In >70% of the cases, job enlargement causes an increase in the number of QL3 workers needed. Potential skill-mix efficiencies are mainly achieved because fewer QL2 workers are required (in >70% of the cases). Because highly qualified care workers are more expensive, differences in hourly labor costs should be taken into account in order to determine whether or not the skill-mix efficiencies found outweigh the cost of the additional QL3 workers. The average hourly costs for a QL1, 2, or 3 care worker are €17.50, €19 and €21 , respectively. For each scenario Table 5.4 shows the labor costs per hour, assuming that each of the care workers is assigned from 7:00 to 11:00 hours. In >80% of the cases substantial efficiency gains in terms of labor costs can be achieved. On average, the potential scale efficiencies for DMU1, DMU2 and DMU3 are 10%, 17% and 20%, respectively. When it comes to the potential skill-mix efficiencies in terms of labor costs the results show an average gain of 6%, 23% and 20%, respectively.

**Table 5.2:** Potential scale and skill-mix efficiencies (in number of routes)

| DMU 1 | | 01/04/11 | 02/04/11 | 03/04/11 | 04/04/11 | 05/04/11 | 06/04/11 | Average |
|---|---|---|---|---|---|---|---|---|
| | Without | 14 | 15 | 15 | 16 | 15 | 16 | |
| # Routes | Large scale | 14 | 14 | 13 | 14 | 13 | 14 | |
| | Job enlargement | 14 | 13 | 14 | 15 | 13 | 14 | |
| Scale efficiency | | 0% | 7% | 13% | 13% | 13% | 13% | 10% |
| Skill-mix efficiency | | 0% | 13% | 7% | 6% | 13% | 13% | 9% |
| DMU 2 | | 01/04/11 | 02/04/11 | 03/04/11 | 04/04/11 | 05/04/11 | 06/04/11 | Average |
| | Without | 9 | 11 | 12 | 10 | 12 | 12 | |
| # Routes | Large scale | 9 | 9 | 9 | 9 | 9 | 9 | |
| | Job enlargement | 6 | 7 | 10 | 9 | 8 | 9 | |
| Scale efficiency | | 0% | 18% | 25% | 10% | 25% | 25% | 17% |
| Skill-mix efficiency | | 33% | 36% | 17% | 10% | 33% | 25% | 26% |
| DMU 3 | | 01/04/11 | 02/04/11 | 03/04/11 | 04/04/11 | 05/04/11 | 06/04/11 | Average |
| | Without | 9 | 10 | 9 | 9 | 10 | 9 | |
| # Routes | Large scale | 7 | 7 | 7 | 8 | 8 | 7 | |
| | Job enlargement | 7 | 7 | 7 | 7 | 8 | 7 | |
| Scale efficiency | | 22% | 30% | 22% | 11% | 20% | 22% | 21% |
| Skill-mix efficiency | | 22% | 30% | 22% | 22% | 20% | 22% | 23% |

**Table 5.3:** Required grade-mix

| DMU 1 | | **01/04/11** | | | **02/04/11** | | | **03/04/11** | | | **04/04/11** | | | **05/04/11** | | | **06/04/11** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Routes (without) | | 14 | | | 15 | | | 15 | | | 16 | | | 15 | | | 16 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 2 | 8 | 4 | 3 | 8 | 4 | 3 | 8 | 4 | 3 | 9 | 4 | 3 | 8 | 4 | 3 | 9 | 4 |
| # Routes (large scale) | | 14 | | | 14 | | | 13 | | | 14 | | | 13 | | | 14 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 2 | 8 | 4 | 3 | 7 | 4 | 3 | 7 | 3 | 3 | 8 | 3 | 3 | 7 | 3 | 3 | 7 | 4 |
| # Routes (job enlargement) | | 14 | | | 13 | | | 14 | | | 15 | | | 13 | | | 14 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 2 | 6 | 6 | 1 | 6 | 6 | 1 | 6 | 7 | 1 | 7 | 7 | 1 | 5 | 7 | 0 | 8 | 6 |
| DMU 2 | | **01/04/11** | | | **02/04/11** | | | **03/04/11** | | | **04/04/11** | | | **05/04/11** | | | **06/04/11** | |
| # Routes (without) | | 9 | | | 11 | | | 12 | | | 10 | | | 12 | | | 12 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 2 | 4 | 3 | 2 | 6 | 3 | 2 | 7 | 3 | 2 | 5 | 3 | 2 | 7 | 3 | 2 | 7 | 3 |
| # Routes (large scale) | | 9 | | | 9 | | | 9 | | | 9 | | | 9 | | | 9 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 1 | 6 | 2 | 1 | 6 | 2 | 1 | 6 | 2 | 1 | 5 | 3 | 1 | 6 | 2 | 1 | 6 | 2 |
| # Routes (job enlargement) | | 6 | | | 7 | | | 10 | | | 9 | | | 8 | | | 9 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 0 | 2 | 4 | 0 | 4 | 3 | 0 | 5 | 5 | 0 | 5 | 4 | 0 | 5 | 3 | 0 | 5 | 4 |
| DMU 3 | | **01/04/11** | | | **02/04/11** | | | **03/04/11** | | | **04/04/11** | | | **05/04/11** | | | **06/04/11** | |
| # Routes (without) | | 9 | | | 10 | | | 9 | | | 9 | | | 10 | | | 9 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 2 | 3 | 4 | 2 | 5 | 3 | 2 | 3 | 4 | 4 | 2 | 3 | 2 | 5 | 3 | 2 | 4 | 3 |
| # Routes (large scale) | | 7 | | | 7 | | | 7 | | | 8 | | | 8 | | | 7 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 1 | 3 | 3 | 1 | 4 | 2 | 1 | 3 | 3 | 1 | 4 | 3 | 1 | 4 | 3 | 1 | 4 | 2 |
| # Routes (job enlargement) | | 7 | | | 7 | | | 7 | | | 7 | | | 8 | | | 7 | |
| QL | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| # Routes | 0 | 3 | 4 | 0 | 3 | 4 | 0 | 3 | 4 | 0 | 4 | 3 | 0 | 4 | 4 | 0 | 3 | 4 |

**Table 5.4:** Potential scale and skill-mix efficiencies (costs/hr)

| DMU 1 | | 01/04/11 | 02/04/11 | 03/04/11 | 04/04/11 | 05/04/11 | 06/04/11 | Average |
|---|---|---|---|---|---|---|---|---|
| | Without | 271 | 289 | 289 | 308 | 289 | 308 | |
| Costs (€/hr) | Large scale | 271 | 270 | 249 | 268 | 249 | 270 | |
| | Job enlargement | 275 | 258 | 279 | 298 | 260 | 278 | |
| Scale efficiency | | 0% | 7% | 14% | 13% | 14% | 12% | 10% |
| Skill-mix efficiency | | 1% | 11% | 3% | 3% | 10% | 10% | 6% |
| **DMU 2** | | **01/04/11** | **02/04/11** | **03/04/11** | **04/04/11** | **05/04/11** | **06/04/11** | **Average** |
| | Without | 174 | 212 | 231 | 193 | 231 | 231 | |
| Costs (€/hr) | Large scale | 174 | 174 | 174 | 176 | 174 | 174 | |
| | Job enlargement | 122 | 139 | 200 | 179 | 158 | 179 | |
| Scale efficiency | | 0% | 18% | 25% | 9% | 25% | 25% | 17% |
| Skill-mix efficiency | | 30% | 34% | 13% | 7% | 32% | 23% | 23% |
| **DMU 3** | | **01/04/11** | **02/04/11** | **03/04/11** | **04/04/11** | **05/04/11** | **06/04/11** | **Average** |
| | Without | 176 | 193 | 176 | 171 | 193 | 174 | |
| Costs (€/hr) | Large scale | 138 | 136 | 138 | 157 | 157 | 136 | |
| | Job enlargement | 141 | 141 | 141 | 139 | 160 | 141 | |
| Scale efficiency | | 22% | 30% | 22% | 8% | 19% | 22% | 20% |
| Skill-mix efficiency | | 20% | 27% | 20% | 19% | 17% | 19% | 20% |

## 5.5 Measuring the correlations

Our initial theoretical statement was that potential scale and/or skill-mix efficiencies, when it comes to the scheduling of care activities, largely depend on the degree of correlation between the different types of demand patterns within and between the separate clusters of a DMU. This phenomenon has been researched extensively in the context of inventory pooling. Eppen (1979) shows in his study that when demand is aggregated across different locations, it becomes more likely that high demand from one customer will be balanced out by low demand from another customer. Thus, when the correlation between demands from two separate inventory locations becomes higher, the benefits from inventory pooling decreases. With respect to staffing in nursing homes we expect that:

1. If the demand patterns of two separate clusters are weakly correlated, care workers can be exchanged more effectively between the clusters. Therefore, a weak correlation would suggest a higher potential for improving efficiency by increasing the scale of scheduling from cluster level to DMU level.

2. If the different types of care tasks are weakly correlated then job enlargement of care workers' jobs, through extending the range of tasks, is expected to have a positive effect on efficiency because of an increase in flexibility of allocation.

We analyzed the statistical relationship between demand patterns using a weighted average of the Pearson product-moment correlation coefficient (Pearson's $r$) (Lee Rodgers and Nicewander, 1988). To measure the demand we divided the days into time periods (time buckets) of 5 minutes and calculated the number of clients in need of care within each time bucket ($k$). The Pearson product-moment correlation coefficient is defined as follows:

$$r(X,Y) = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{\sqrt{\sum_{k=1}^{n}(X_k - \overline{X})^2 \sum_{k=1}^{n}(Y_k - \overline{Y})^2}}$$

We define the following variables:

$$X_{ci} = (X_{ci1}, ..., X_{cin})$$
$$W_{ci} = \sum_{t=1}^{m} W_{cit}$$

where

$X_{cik} =$ demand (number of clients) cluster $c$, qualification level $i$, and interval $k$

$W_{cit} =$ workload (in minutes) cluster $c$, qualification level $i$, and task $t$

To give more insight into the scale efficiencies found we used the following formula, which gives a weighted average of the coefficient for different DMUs:

$$\overline{\rho}_{\text{Scale}} = \frac{\sum_{i=1}^{3}(W_{ci} + W_{di})r(X_{ci}, X_{di})}{\sum_{i=1}^{3}(W_{ci} + W_{di})}$$

with $c \neq d$.

The following formulas were used to give more insight into the efficiency effects of a change in skill-mix:

$$\overline{\rho}_{\text{Skill-mix}}(c) =$$

$$\frac{(W_{c1} + W_{c2})r(X_{c1}, X_{c2}) + (W_{c1} + W_{c3})r(X_{c1}, X_{c3}) + (W_{c2} + W_{c3})r(X_{c2}, X_{c3})}{2(W_{c1} + W_{c2} + W_{c3})}$$

$$\overline{\rho}_{\text{Skill-mix}} = \frac{(W_{c1} + W_{c2} + W_{c3})\overline{\rho}_{\text{Skill-mix}}(c) + (W_{d1} + W_{d2} + W_{d3})\overline{\rho}_{\text{Skill-mix}}(d)}{(W_{c1} + W_{c2} + W_{c3} + W_{d1} + W_{d2} + W_{d3})}$$

The results of the weighted average correlations ($\overline{\rho}_{\text{Scale}}$ and $\overline{\rho}_{\text{Skill-mix}}$) are presented in Table 5.5.

**Table 5.5:** Weighted average correlations

| DMU 1 | 01/04/11 | 02/04/11 | 03/04/11 | 04/04/11 | 05/04/11 | 06/04/11 | Average |
|---|---|---|---|---|---|---|---|
| Scale efficiency | 0% | 7% | 13% | 13% | 13% | 13% | 10% |
| $\overline{\rho}_{\text{Scale}}$ | 0.5 | 0.48 | 0.47 | 0.36 | 0.45 | 0.44 | 0.45 |
| Skill-mix efficiency | 0% | 13% | 7% | 6% | 13% | 13% | 9% |
| $\overline{\rho}_{\text{Skill-mix}}$ | 0.27 | 0.33 | 0.34 | 0.3 | 0.43 | 0.34 | 0.34 |
| **DMU 2** | **01/04/11** | **02/04/11** | **03/04/11** | **04/04/11** | **05/04/11** | **06/04/11** | **Average** |
| Scale efficiency | 0% | 18% | 25% | 10% | 25% | 25% | 17% |
| $\overline{\rho}_{\text{Scale}}$ | 0.52 | 0.52 | 0.53 | 0.48 | 0.56 | 0.54 | 0.53 |
| Skill-mix efficiency | 33% | 36% | 17% | 10% | 33% | 25% | 26% |
| $\overline{\rho}_{\text{Skill-mix}}$ | 0.18 | 0.2 | 0.23 | 0.14 | 0.15 | 0.17 | 0.18 |
| **DMU 3** | **01/04/11** | **02/04/11** | **03/04/11** | **04/04/11** | **05/04/11** | **06/04/11** | **Average** |
| Scale efficiency | 22% | 30% | 22% | 11% | 20% | 22% | 21% |
| $\overline{\rho}_{\text{Scale}}$ | 0.26 | 0.26 | 0.38 | 0.22 | 0.58 | 0.3 | 0.33 |
| Skill-mix efficiency | 22% | 30% | 22% | 22% | 20% | 22% | 23% |
| $\overline{\rho}_{\text{Skill-mix}}$ | 0.21 | 0.19 | 0.25 | 0.25 | 0.16 | 0.23 | 0.22 |

On average the $\overline{\rho}_{\text{Skill-mix}}$ is relatively low compared with the $\overline{\rho}_{\text{Scale}}$. This might explain why considerable skill-mix efficiencies can be achieved, even when care workers are dedicated to a specific cluster (=small scale scheduling). Based on results presented in Table 5.5 it is not possible to prove a direct relationship between the weighted average correlations and the potential efficiency gains. This is due to the fact that the variability between the weighted average correlations is limited. In order to examine this relationship further, we made use of simulation.

## 5.6 Simulation

To validate our method further, we quantified the relationship between the weighted average correlation and the required number of care routes with the help of a simulation performed with the OERplanner. A dataset with 40 shifts of a fictitious DMU (consisting of two clusters) was used as input for the simulation analysis. We made the following assumptions for the simulation experiment:

1. Each shift consists of 100 tasks. Based on the empirical data they were divided as follows: 26 QL1 tasks, 38 QL2 tasks and 36 QL3 tasks (see Appendix 5.A.1).
2. The tasks were equally divided over the two clusters.
3. Each wish time (preferred delivery time) was randomly generated.
4. The duration of each task (task size) was randomly chosen from the empirical distribution found. For each QL a separate empirical distribution was used.
5. Each task was planned within 10 minutes from the wish time.
6. The duration of the shifts varied from 1 to 4 hours (full hours only).

In order to quantify the relationship between the weighted average correlation and the required number of care routes (for each day) we calculated: (1) the weighted average correlations ($\overline{\rho}_{\text{Scale}}$ and $\overline{\rho}_{\text{Skill-mix}}$) and (2) the potential efficiency gain. The potential efficiency gain was measured by comparing the total number of care routes required, with the number of care routes required without applying scale and job enlargement. The results of the simulation experiment are presented in two scatter plots (see Figure 5.3 and 5.4), which show

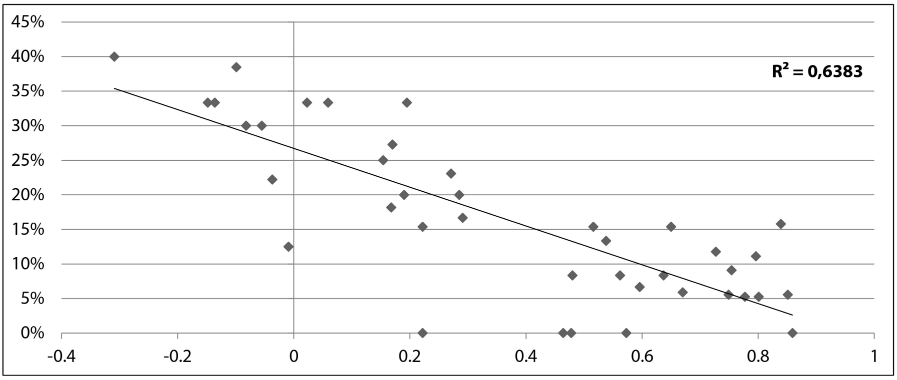


**Figure 5.3:** Quantification of the relationship between $\overline{\rho}_{\text{Scale}}$ and the potential efficiency gain

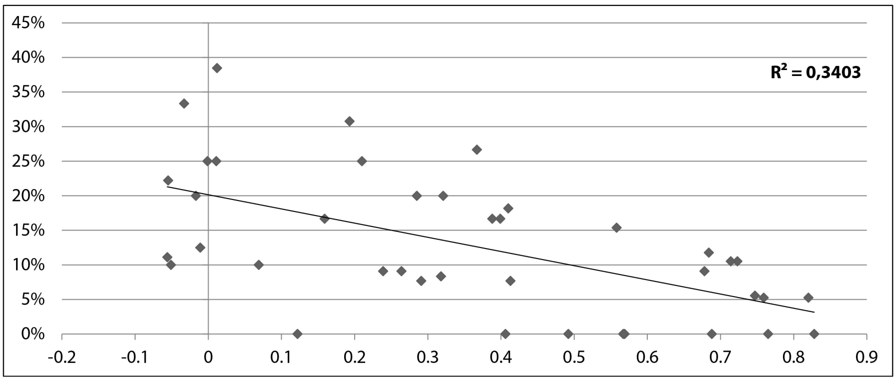


**Figure 5.4:** Quantification of the relationship between $\overline{\rho}_{\text{Skill-mix}}$ and the potential efficiency gain

the relationship between the weighted average correlation ($\overline{\rho}_{\text{Scale}}$ and $\overline{\rho}_{\text{Skill-mix}}$) and the potential efficiency gain. The $R^2$ values (0.6383 and 0.3404) suggest that there is a linear relationship between, and the potential efficiency gain. The exact results of the simulation can be found in the Appendix in table G3.

## 5.7 Conclusion and discussion

In this chapter we made a first step towards gaining more insight into the concepts of 'economies of scale' and 'skill-mix efficiency' in a nursing home setting, by using the scheduling data for care activities of three separate Decision Making Units (DMUs) within a single Dutch nursing home facility. The results show that, when taking the preferences of the nursing home residents as starting point (in terms of day and time), in >85% of the examined cases substantial scale and skill-mix economies can be achieved. The findings from this chapter have several implications for practice. First of all, Dutch nursing homes are struggling to survive on the tight budgets provided to them, let alone giving substance to a more client-centred approach. Therefore, we believe nursing homes should seriously investigate the benefits of (1) planning healthcare activities on a more larger scale and (2) the enlargement of care workers' jobs (i.e. blending tasks of different qualification levels). Secondly, the Dutch government's policy is increasingly directed towards small-scale, homelike care settings. With the results of this chapter in mind, policy makers and nursing home managers should not focus blindly on creating small-scale living facilities, without taking potential scale and skill-mix economies into account. In our opinion, the most important goal of a nursing home should be to facilitate its clients in such a way that it becomes possible for them to live the life they prefer at a reasonable cost. Hence, a small-scale care setting should not be an end in itself. Thirdly, the assumption that differentiated practices provide nursing homes with the most efficient and effective use of scarce resources is open to question, as this chapter shows that considerable skill-mix efficiencies can be achieved. A more theoretical contribution of this chapter is that it shows that the correlation between the demand patterns of different types of care tasks is relatively low. This explains why considerable skill-mix efficiencies can be achieved, even when care workers are dedicated to a specific cluster (=small scale scheduling). We quantified this assumption by performing a simulation experiment. The simulation results show a linear relationship between the weighted average correlation ($\overline{\rho}_{\text{Scale}}$ and $\overline{\rho}_{\text{Skill-mix}}$ ) and the potential efficiency gain. However, this chapter has a few limitations. As potential skill and/-or scale mix efficiencies are influenced by the (organizational) context one has to be careful with the generalization of the results presented in this chapter. Furthermore, this chapter is solely focused on the plannable care activities and does not take random healthcare demand into account.

## Acknowledgements

We would like to thank the anonymous nursing care organization for the dataset and Verkooijen & Beima and CCmath for letting us use the OERplanner.

## 5.A Information on dataset

**Table 5.A.1:** Distribution number of tasks

| QL | QL1 | QL2 | QL3 | Total |
|---|---|---|---|---|
| # Tasks | 385 | 561 | 553 | 1499 |
| % | 26% | 37% | 37% | 100% |

**Table 5.A.2:** Additional information on dataset

| DMU 1 | # Tasks | Average tasksize | SD tasksize | Workload distribution (Qualification Levels) | | |
|---|---|---|---|---|---|---|
| Date | | Minutes | Minutes | QL1 | QL2 | QL3 |
| 01/04/11 | 131 | 11.1 | 9.8 | 15.0% | 59.3% | 25.7% |
| 02/04/11 | 127 | 9.7 | 8.7 | 16.3% | 53.7% | 30.0% |
| 03/04/11 | 127 | 9.6 | 8.7 | 16.4% | 53.1% | 30.4% |
| 04/04/11 | 131 | 10.4 | 9.6 | 16.2% | 53.1% | 30.7% |
| 05/04/11 | 130 | 10.1 | 8.9 | 16.4% | 55.3% | 28.3% |
| 06/04/11 | 131 | 10.2 | 9.2 | 16.5% | 53.7% | 29.8% |

| DMU 2 | # Tasks | Average tasksize | SD tasksize | Workload distribution (Qualification Levels) | | |
|---|---|---|---|---|---|---|
| Date | | Minutes | Minutes | QL1 | QL2 | QL3 |
| 01/04/11 | 61 | 11.4 | 9.8 | 5.1% | 58.4% | 36.5% |
| 02/04/11 | 57 | 11.2 | 10.1 | 4.2% | 57.4% | 38.4% |
| 03/04/11 | 58 | 11.2 | 10.0 | 4.1% | 56.6% | 39.2% |
| 04/04/11 | 67 | 12.2 | 10.4 | 6.8% | 55.4% | 37.8% |
| 05/04/11 | 61 | 11.8 | 10.6 | 6.5% | 58.1% | 35.4% |
| 06/04/11 | 62 | 11.6 | 10.5 | 4.0% | 60.6% | 35.3% |

| DMU 3 | # Tasks | Average tasksize | SD tasksize | Workload distribution (Qualification Levels) | | |
|---|---|---|---|---|---|---|
| Date | | Minutes | Minutes | QL1 | QL2 | QL3 |
| 01/04/11 | 62 | 11.9 | 8.2 | 11.9% | 49.1% | 39.0% |
| 02/04/11 | 54 | 11.9 | 8.5 | 8.7% | 54.3% | 37.0% |
| 03/04/11 | 50 | 11.1 | 7.4 | 9.9% | 47.9% | 42.1% |
| 04/04/11 | 67 | 10.1 | 7.4 | 15.1% | 52.7% | 32.2% |
| 05/04/11 | 63 | 11.1 | 9.2 | 12.2% | 57.4% | 30.4% |
| 06/04/11 | 60 | 10.8 | 7.4 | 12.1% | 52.9% | 35.0% |

**Table 5.A.3:** Results simulation (Care routes with a total duration of 10 minutes or less were excluded from the count)

| Day | $\overline{\rho}_{\text{Scale}}$ | Scale efficiency | $\overline{\rho}_{\text{Skill-mix}}$ | Skill-mix efficiency | Day | $\overline{\rho}_{\text{Scale}}$ | Scale efficiency | $\overline{\rho}_{\text{Skill-mix}}$ | Skill-mix efficiency |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.222 | 0% | 0.285 | 20% | 21 | 0.516 | 15% | 0.406 | 0% |
| 2 | 0.464 | 0% | 0.410 | 18% | 22 | 0.650 | 15% | 0.558 | 15% |
| 3 | 0.573 | 0% | 0.569 | 0% | 23 | 0.839 | 16% | 0.759 | 5% |
| 4 | 0.478 | 0% | 0.413 | 8% | 24 | 0.291 | 17% | 0.318 | 8% |
| 5 | 0.859 | 0% | 0.820 | 5% | 25 | 0.168 | 18% | 0.264 | 9% |
| 6 | 0.801 | 5% | 0.723 | 11% | 26 | 0.285 | 20% | 0.321 | 20% |
| 7 | 0.777 | 5% | 0.714 | 11% | 27 | 0.190 | 20% | 0.122 | 0% |
| 8 | 0.851 | 6% | 0.828 | 0% | 28 | -0.037 | 22% | -0.055 | 22% |
| 9 | 0.749 | 6% | 0.747 | 6% | 29 | 0.271 | 23% | 0.193 | 31% |
| 10 | 0.670 | 6% | 0.688 | 0% | 30 | 0.154 | 25% | 0.159 | 17% |
| 11 | 0.596 | 7% | 0.567 | 0% | 31 | 0.170 | 27% | 0.239 | 9% |
| 12 | 0.562 | 8% | 0.388 | 17% | 32 | -0.055 | 30% | -0.017 | 20% |
| 13 | 0.637 | 8% | 0.492 | 0% | 33 | -0.082 | 30% | 0.069 | 10% |
| 14 | 0.480 | 8% | 0.399 | 17% | 34 | -0.136 | 33% | -0.056 | 11% |
| 15 | 0.754 | 9% | 0.678 | 9% | 35 | 0.059 | 33% | -0.033 | 33% |
| 16 | 0.796 | 11% | 0.765 | 0% | 36 | -0.148 | 33% | -0.001 | 25% |
| 17 | 0.727 | 12% | 0.684 | 12% | 37 | 0.023 | 33% | 0.011 | 25% |
| 18 | -0.009 | 13% | -0.011 | 13% | 38 | 0.195 | 33% | 0.210 | 25% |
| 19 | 0.538 | 13% | 0.367 | 27% | 39 | -0.099 | 38% | 0.012 | 38% |
| 20 | 0.222 | 15% | 0.291 | 8% | 40 | -0.309 | 40% | -0.051 | 10% |

Chapter 6

# Care on demand in nursing homes: a queueing theoretic approach


**Abstract**

The focus in this chapter lies on the 'care on demand' process in a Belgian nursing home. Based on the analysis of real-life 'call button' data, a queueing model is presented which can be used by nursing home managers to determine the number of care workers required to meet a specific service level. An 80/10 service level is proposed for this nursing home, meaning that at least 80% of the clients should receive care within 10 minutes after a call button request. To the best of our knowledge, this is the first attempt to develop a quantitative model for the 'care on demand' process in a nursing home.

This chapter is based on Van Eeden, Moeke and Bekker (2014).

## 6.1 Background

In Chapters 4 and 5 the focus was on the analysis of 'care by appointment' (i.e. scheduled care activities). The focus in this chapter lies on healthcare activities which are carried out in response to random, unexpected demand. This type of activity is also referred to as 'care on demand' or 'unscheduled care'. A considerable part of the 'care on demand' activities in nursing homes consists of responding to the requests of nursing home residents made through the use of call buttons. Most nursing homes struggle with determining the appropriate number of care workers needed to respond to these call button requests. The main reason for this is that a well-founded quantitative approach is generally lacking. Also staffing decisions concerning 'care on demand' are often made without a sound rational basis. In the ideal situation staffing decisions are based on a quantification of the needs and preferences of the nursing home residents, in other words the demand, and the duration of the healthcare tasks associated with these needs and preferences. Unfortunately, in most nursing homes this type of information is not available. Even basic staffing information such as information about actual staffing hours is often of poor quality (Havig et al., 2011) or not available at all (Harrington et al., 2012). Our experience is that the lack of reliable data is a common problem in healthcare facilities. However, it is a more pronounced problem in nursing home facilities as they are often low-tech, paper based organizations. In this chapter a queueing model is developed, using data from a Belgian nursing home, to gain more insight into the 'care on demand' process and its performance. Thereby, this study provides better understanding of the number of resources (i.e. the number of care workers) required to sufficiently meet the needs of the nursing home residents regarding 'care on demand'. We analyze demand patterns over the course of a day, whereas our main focus is on the night shift. The reason for this is that the 'care by appointment' activities are scarce during the night, which minimizes the risk that the 'care on demand' data is compromised by 'care by appointment' tasks.

Clearly, the study presented in this chapter addresses an issue of great societal relevance. More specifically, in order to increase the efficiency without losing sight of the needs of residents, it should be possible for nursing home managers to analyze and monitor the performance of healthcare processes. From a scientific point of view there is hardly any insight into demand processes in nursing homes. To the best of our knowledge this is the first endeavor to study ‘care on demand’ activities in a nursing home setting using a queueing theoretic approach.

This chapter is structured as follows. In the next section, we outline and justify the study by looking at related literature. In addition, we describe the nursing home context and its relation to queueing theory and propose a performance measure for ‘care on demand’ in nursing homes. In Section 6.3, we analyze the call-buton data of a single Belgian nursing home. Based on this analysis we present a queueing model in Section 6.4. In Section 6.5 the constructed queueing model is used to analyze different scenarios. Finally, in Section 6.6, we present our conclusions and directions for future work.

## 6.2 Nursing home context

In this section the ‘care on demand’ process is described in more detail. First, we justify the use of a queueing theoretic approach. Next, we provide more insight into the empirical context regarding the ‘care on demand’ activities in a Belgian nursing home. In the final subsection we define a performance measure which can be used for assessing the ‘care on demand’ process.

### A queueing theoretic approach

Waiting lines or queues occur whenever the demand for a service exceeds the systems capacity to provide that service. From a client perspective, long queues have a detrimental effect on the perceived quality of service. Unfortunately, congestion is commonplace in many areas of healthcare and has become an important issue in the provision of healthcare services. In addition to diminished patient satisfaction, waiting can have a serious impact on the well-being of patients or clients. In the case of nursing home residents, excessive waiting for care and/or support limits their freedom of living the lives they prefer as they are often in need of ongoing assistance with activities of daily living. Queueing theory is the mathematical study of waiting lines or queues and can be useful in describing and analyzing healthcare processes (Hall, 2012), where ‘healthcare process’ refers to a set of activities and/or procedures that a client takes part in to receive the necessary care. A growing number of studies shows that queueing models can be helpful in assessing the performance of healthcare processes in terms of waiting time and utilization of critical resources (Fomundam and Herrmann, 2007; Lakshmi and Iyer, 2013). The most common resources in a healthcare process are physicians, care workers, beds, and (specialized) equipment. When it comes to the delivery of healthcare in a nursing home setting, care workers can be regarded as the most important resource as they are responsible for the daily care and supervision of clients. Furthermore, they account for a significant proportion of the total costs of nursing home care.

By now, there is a considerable amount of literature on queueing models for capacity decisions in hospitals. The vast majority of publications address issues related to medium and long-term capacity decisions, with a strong emphasis on bed occupancy, see e.g., Cochran and Bharti (2006); De Bruin et al. (2007); El-Darzi et al. (1998); Gorunescu et al. (2002); Fomundam and Hermann (2007); Hulshof et al. (2012); Lakshmi and Iyer (2013) for an overview. The central issue typically is to determine the capacity (either in number of beds or staff) required to accommodate the randomly arriving demand. In the short-term, meeting the randomly occurring needs of patients lead to waiting for care delivery. The literature on short-term performance for clinical care is much more limited. Most related to our setting are two studies proposing models for nurse staffing levels in hospital wards. Yankovic and Green (2011) use a two-dimensional Markov model to describe nursing workload due to admissions and discharges in addition to the fluctuations in needs of patients that arise while a patient occupies a bed. They determine nurse staffing levels by evaluating the system performance numerically. Furthermore, they demonstrate that admission or discharge blocking caused by nurse shortage can have a significant impact on system performances and show that prespecified nurse-to-patient ratio policies cannot achieve a consistently high service level. Véricourt and Jennings (2011) study a similar model, but consider a fixed number of patients. Their analysis is based on many-server asymptotics. Their results also suggest that nurse-to-patient ratio policies cannot achieve a consistently high service level. Our study differs in that we validate the model for short-term performance in a nursing home setting. Also, the model assumptions in Véricourt and Jennings (2011) and Yankovic and Green (2011) differ slightly from ours. The performance analysis in Yankovic and Green (2011) requires a more involved numerical procedure, whereas our formulas can be easily implemented in a decision support tool. The study of Véricourt and Jennings (2011) considers an interesting asymptotic regime, which is however less relevant for our practical setting. Although queueing theory has been shown useful in assessing staffing levels in hospitals, the use of queueing models in guiding staffing decisions in a nursing home setting is still very limited.

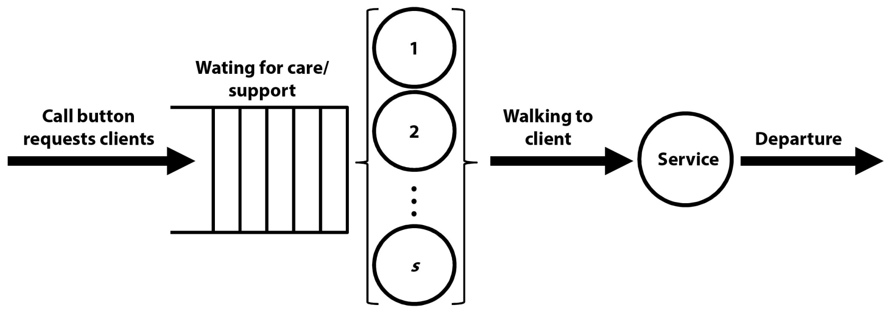


**Figure 6.1:** The 'care on demand' queueing system

## Empirical context

From a queueing theoretic perspective, healthcare processes can be viewed as a system in which clients have to wait for the care they need, receive the necessary care and then depart (Fomundam and Herrmann, 2007). The 'care on demand' process in a nursing home can be described as follows: when a nursing home resident needs care or support he/she pushes the call button in his/her room and waits until a care worker is available. The available care worker then moves to the room of the nursing home resident concerned. When the required care or assistance has been delivered, the resident leaves the 'care on demand' process (see Figure 6.1).

Queueing theory is an appropriate and useful method for modeling and analysing this 'care on demand' process because it can handle the random character of call button requests and the variability in duration of healthcare tasks.

The Belgian nursing home under study provides long-term residential care for up to 180 clients who are aged 65 and over. Although all residents need some assistance with activities of daily living, most of them are still largely self-sufficient. There are six care-providing departments, each of which is responsible for the care and support of a fixed number of residents. This nursing home uses a high-tech registration system for 'care on demand'. In particular, every call button request is registered automatically in a central data base. In addition, all care workers are equipped with a keycard. Every time a care worker enters or leaves the room of a resident, the keycard is swiped along an electronic keypad, registering the timestamp and the location. In this study we focus on the 'care on demand' activities during the night shift. The care is provided by a small number of care workers, as the total need for care or support is limited during this period of the day. The assigned care workers only have to respond to call button requests which are received in a single call centre.

### Performance measure

In order to make it possible for nursing home managers to monitor the performance of the 'care on demand' process they need a performance measure and an objective. A performance measure, as defined in this chapter, refers to a metric used to quantify the efficiency and/or effectiveness of a process (Neely et al., 2005). For 'care on demand' we find that from the standpoint of the client, waiting for care and/or support should be avoided as much as possible. Here, waiting refers to the time between call request and the moment a care worker is present, to be called response time. The response time should be below some threshold for most of the clients. In line with other service sectors, such as call centres, we propose to measure response times in terms of service levels. Specifically, we define a targeted time window $Y$ during which a care worker should be present at the client. The service level is defined as follows.

**Definition:** The *service level* $X/Y$ denotes that $X\%$ of the clients have a response time at or below $Y$ minutes.

Based on Pareto's principle, a typical value for $X$ is 80. Using a time window $Y$ of 10 minutes then yields a service-level of 80/10, meaning that in 80% of the requests a care worker is present at the client within 10 minutes after the request is generated. The queueing model presented in this chapter can be used as a tool to measure the performance of the 'care on demand' process and to determine the number of care workers required to meet a specific service level. Although performance management is widely used in the field of healthcare as a means to improve quality and efficiency, this type of performance management for 'care on demand' processes in a nursing home context has received hardly any attention.

### Outline of results

In this subsection we outline the results obtained in sections 6.3-6.5. First, from the data analysis we conclude that the average demand patterns, i.e. call requests and care delivery durations, are stable during the night. During the day, there are some distinct peaks that are caused by 'care by appointment' activities. Different days of the week show similar demand patterns. For call requests we conclude that the interarrival times may be approximated well by an exponential distribution. For care delivery durations, the conclusions are less affirmative. Based on the data analysis, we propose an $M/G/s$ queueing model to determine service levels, which also include travel times. This results in Equation (5) which expresses the tail probability of the response time. We note that such an equation may be readily implemented in a decision support tool. Moreover, we validated the queueing model with the actual waiting time data. Some numerical experiments show that an 80/10 service level would work well for this nursing home facility.

## 6.3 Data analysis

To provide insight into the 'care on demand' activities, we analyze the arrival process of call button requests and the actual care delivery process.

### The arrival process of call requests

Call button requests made during the period of three months, from February 1, 2013 to May 1, 2013, were analyzed. During this period in total 19,996 requests arrived. As a first indication, Figure 6.2 shows the number of arrivals per day. Observe that the number of arrivals fluctuates over time: during March more call requests were made than during February and April. Unfortunately, the time window of the data set is restricted to three months excluding the option to observe seasonal patterns. Given the state and mobility of residents, it seems likely that these fluctuations occur naturally; this idea is further supported by the huge variability in requests between residents and for requests over time.

Figure 6.3 shows the average number of arrivals per quarter hour. For instance, the first data point means that on average 1.6 call requests were made between 00:00 and 00:15 hours. From this figure it can be seen that around 8:30, 12:00 and 17:00 hours, on average, less call requests were made than during surrounding periods. These are moments in which the residents enjoy a joint activity like breakfast, lunch or dinner, whereby they generate fewer calls.

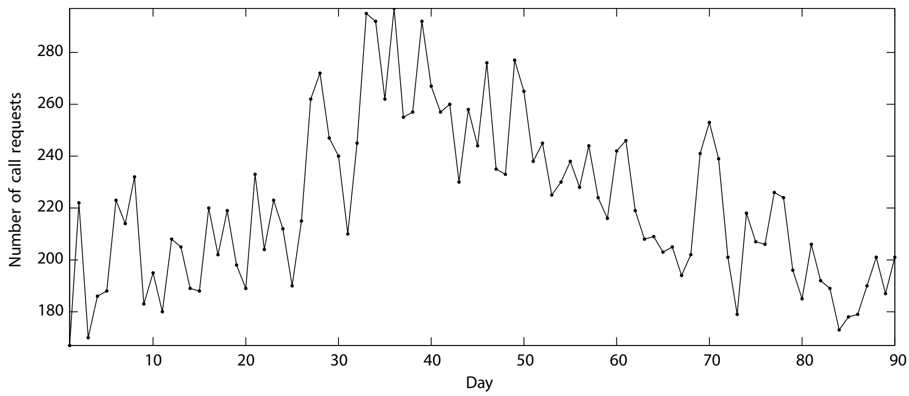


**Figure 6.2:** Number of call requests per day

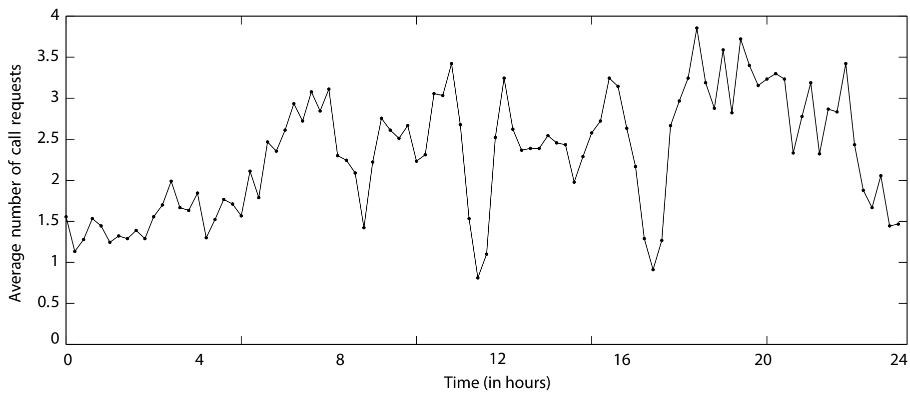


**Figure 6.3:** Average number of call requests per quarter hour

Moreover, it can be seen that the number of arrivals during the night[1] is fairly constant. In Figure 6.4 a boxplot is given for the number of arrivals during each 15-minute interval. A data point that exceeds 1.5 times the interquartile range is defined as an outlier, and is drawn as a circle. These boxplots confirm that the arrival rate over the course of a full day is inhomogeneous. For each day of the week a similar arrival pattern has been observed. As shown in Figure 6.5 each line represents another weekday; it can be seen that the arrival patterns correspond with the overall arrival pattern, as shown in Figure 6.3. This suggests that there is no structural weekly pattern.

[1]The term 'night' refers to the time period between 22:45 and 5:45 hours. This is the largest possible range in which the average number of arrivals per quarter hour does not exceed 2.2

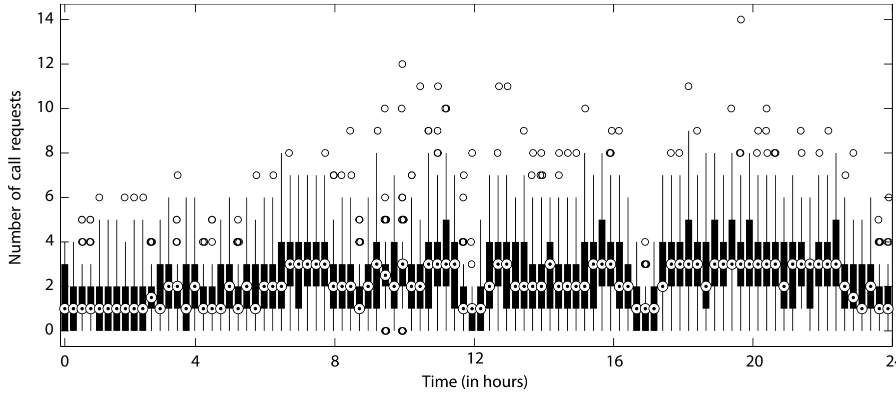


**Figure 6.4:** Boxplots of the number of call requests per quarter hour

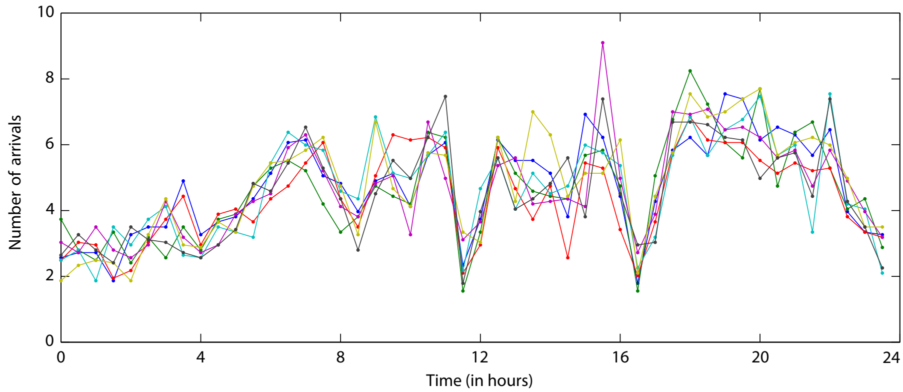


**Figure 6.5:** Average number of call requests per weekday per half hour. Each line represents a different weekday

A common way to deal with the daily cycle of call requests is to consider intervals for which the number of arrivals is relatively stable. A prominent example of such a method is the stationary independent period-by-period (SIPP) approach, where the arrival rate is averaged over the staffing interval (e.g., Green et al., 2001, 2007). We followed this method, but restricted the analysis to staffing decisions for the night period. The two main reasons are that (i) the data are not compromised by 'care by appointment' related data, and (ii) in the day time, staff may not be solely dedicated to 'care on demand'.

To confirm statistically whether the arrivals are constant during the night, i.e., between 22:45 and 5:45 hours, we used the the Kolmogorov-Smirnov test. This is done by testing for each combination of two 15-minute intervals the null hypothesis that the number of arrivals originate from the same underlying distribution. By using a significance level of 0.05 and applying the Bonferroni correction for the many tests that are done, this results in 0 out of 378 null hypotheses being rejected. However, the Bonferroni correction is known to be conservative. A more powerful testing procedure is the positive false discovery rate (pFDR), introduced by Storey (2002, 2003). This procedure also results in 0 null hypotheses being rejected. Hence, it can be assumed that the arrival rate is constant during the night. In the remaining part of this chapter we consider call requests that take place during this period, i.e. between 22:45 and 5:45 hours. Other time intervals of the day can be analyzed in a similar way.

During the night, in total 3,891 call requests were made with an average interarrival time of 9.37 minutes. The interarrival time is defined here as the difference between the arrival times of two consecutive call requests. The standard deviation of the interarrival times is 9.83 minutes and the coefficient of variation equals 1.05. In Figure 6.6 (left) a histogram of the interarrival times is shown. From this figure it can be seen that the underlying distribution of the interarrival times corresponds with a righttailed distribution. The exponential distribution, Gamma distribution and hyperexponential distribution have similar properties and are fitted to the data. The parameters of the exponential distribution and Gamma distribution are obtained by minimizing the mean squared error between the empirical and theoretical distribution, and the parameters of the hyperexponential distribution are obtained by using a three-moment fit, as given by Tijms (2003).

For each fitted distribution, we used the Kolmogorov-Smirnov (KS) test to test the null hypothesis that the underlying distribution of the interarrival times is equal to the specified distribution. The estimated parameters and the $p$-values of the KS tests are given in Table 6.1. These results show that all of the fitted distributions are rejected, using a significance level of 0.05. This is not surprising given the considerable number of data points. To consider smaller sample sizes, we also conducted the KS test for the 15-minute intervals between 22:45 and 5:45 hours separately, resulting in 28 tests. On average, a 15-minute interval has 139 call button requests during this 3 month period, which is a more suitable sample size for statistical testing. Using the KS test again and using a significance level of 0.05, the null hypothesis that the underlying distribution of the interarrival times is equal to the specified distribution is rejected for 11 intervals for the exponential and hyperexponential distribution and for all 28 intervals for the Gamma distribution. Based on the above, we may conclude

**Table 6.1:** Parameters and $p$-values for different distributions fitted with the interarrival times. The unit of time is minutes

| Distribution | Parameters | $p$-value |
|---|---|---|
| Exponential | $\hat{\lambda} = 0.11$ | 1.47e-11 |
| Gamma | $\hat{k} = 0.91$, $\hat{\theta} = 9.92$ | 3.02e-13 |
| Hyperexponential | $\hat{p}_1 = 0.14$, $\hat{p}_2 = 0.86$, $\hat{\lambda}_1 = 0.07$, $\hat{\lambda}_2 = 0.12$ | 1.47e-11 |

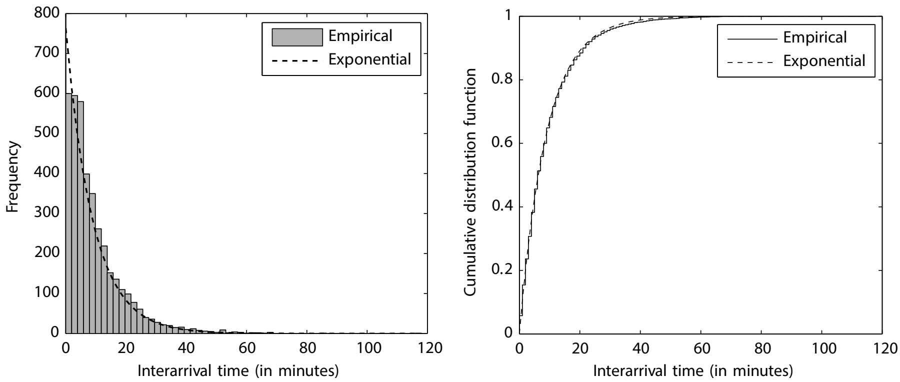


**Figure 6.6:** Scaled probability density function (left) and cumulative distribution function (right) of the interarrival times fitted with the exponential distribution

that the interarrival times closely resemble an exponential distribution, but the match is not perfect. This is also confirmed by the exponential QQ-plot (which is omitted here).

Figure 6.6 showes a scaled probability density function (left) and cumulative distribution function (right) of the interarrival times fitted with the exponential distribution. The plots of the Gamma distribution and hyperexponential distribution are similar. Figure 6.6 visually shows a good fit with the exponential distribution. Despite the fact that each of the fitted distributions was rejected by some of the KS tests, it seems practically useful to assume that the interarrival times are exponentially distributed.

## Duration of care delivery

Upon a call button request, delivery of care is assumed to take place between the arrival time of a care worker at the room of the client who made the request and the departure of the care worker from that room. In Figure 6.7 the average time for care delivery per quarter hour is plotted. Clearly, between 6:45 and

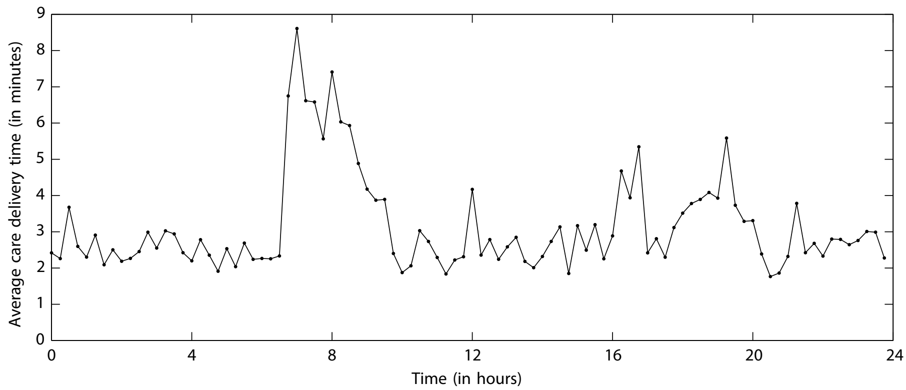


**Figure 6.7:** Average care delivery time per quarter hour

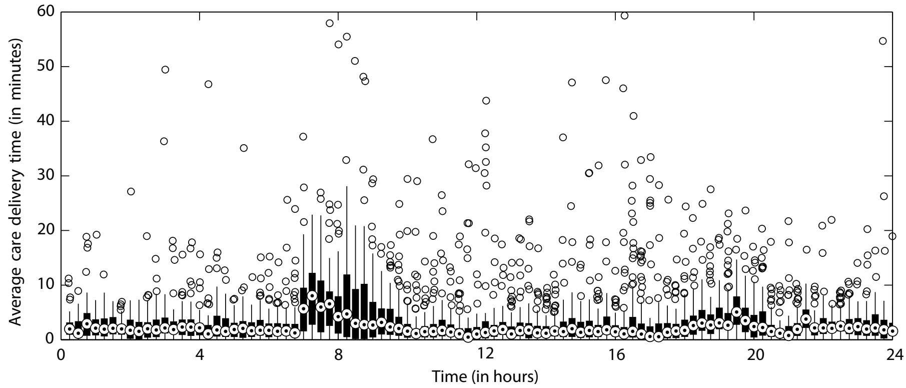


**Figure 6.8:** Boxplot of the average care delivery time per quarter hour

9:45 hours these times are longer than during other periods of the day. Around this time the residents wake up and mainly receive 'care by appointment'. In practice, call requests around this time are often related to the scheduled activities, which causes longer care delivery times. For each 15-minute interval the average care delivery time is determined. Figure 6.8 shows, for each 15-minute interval, a boxplot of the average care delivery times, each based on roughly 90 data points. These boxplots show that the durations of care delivery are highly variable during certain periods of the day. Moreover, the average duration of care delivery is fairly stable across the night.

To estimate the distribution of the care delivery times, the same approach is used as in Section 6.3. The KS test combined with the Bonferroni correction confirms that in each combination of the 15-minute intervals between 22:45 and

**Table 6.2:** Parameters and $p$-values for different distributions fitted to the care delivery times. The unit of time is minutes

| Distribution | Parameters | $p$-value |
|---|---|---|
| Exponential | $\hat{\mu} = 0.43$ | 2.24e-125 |
| Gamma | $\hat{k} = 0.53$, $\hat{\theta} = 4.78$ | 1.56e-88 |
| Hyperexponential | $\hat{p}_1 = 0.10$, $\hat{p}_2 = 0.90$, $\hat{\mu}_1 = 0.11$, $\hat{\mu}_2 = 0.56$ | 1.59e-45 |

5:45 hours the care delivery times originate from the same underlying distribution. When using a significance level of 0.05, 0 out of 378 null hypotheses are rejected. The same results are obtained when using positive false discovery rates.

During the night, in total 3,863 care delivery times were registered, with an average care delivery time of 2.56 minutes. The standard deviation of the care delivery times is 4.12 minutes and the coefficient of variation is equal to 1.61. The histogram in Figure 6.9 shows that the underlying distribution of the care delivery times is right-tailed. Obviously, a large number of call requests have a care delivery time of less then one minute. Probably, these are either 'false' requests that do not require assistance or are short questions. The coefficient of variation of 1.61 indicates that the underlying distribution of the care delivery times shows considerable variation, more than, for instance, the exponential distribution. Nonetheless, the exponential distribution, Gamma distribution and hyperexponential distribution are fitted to the data; the estimated parameters and the $p$-values of the KS tests are given in Table 6.2. These parameters are estimated in the same way as in Section 6.3. The null hypothesis is tested that the underlying distribution of the care delivery times is equal to the specified distribution. At a significance level of 0.05, the null hypothesis is rejected. As in Section 6.3, we carried out the KS test for the 15-minute intervals separately resulting in 28 rejections out of 28 tests for all three distributions.

Based on the tests above, it is clear that the care delivery times do not match well to any of the proposed distributions from a statistical viewpoint. The hyperexponential distribution has the highest $p$-value, which is an indication that this distribution gives the best fit with the data compared with the other fitted distributions. This is also confirmed by plots of the scaled probability density function, cumulative distribution function and QQ-plot made for each of the distributions. Plots of the former two are given for the hyperexponential

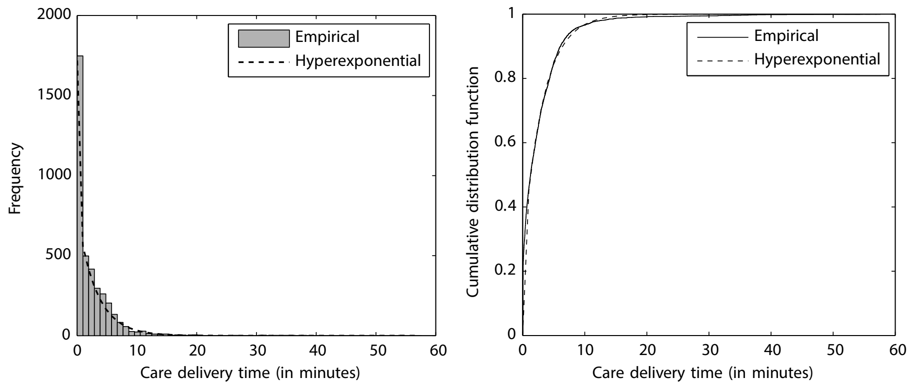


**Figure 6.9:** Scaled probability density function (left) and cumulative distribution function (right) of the care delivery times fitted with the hyperexponential distribution

distribution in Figure 6.9. These plots indicate that the hyperexponential distribution might be of some practical value. In that case, the parameters of the hyperexponential distribution, as given in Table 6.2, can be interpreted as follows: with probability $\hat{p}_2 = 0.90$, the client has a minor request taken on average $1/\hat{\mu}_2 = 1.79$ minutes, whereas with probability $\hat{p}_1 = 0.10$ the client has a large request taking $1/\hat{\mu}_1 = 9.28$ minutes on average.

## 6.4 Queueing model

The number of call requests and the care delivery times are key ingredients for a model that may be applied to determine staffing levels overnight. Such a model should be useful for management at a strategic or tactical level. We approximate the service level of the proposed model in Section 6.4 and validate the model in Section 6.4.

### Model and performance analysis

We identify three important features that the model should obey:

(i) The random nature of call requests and service times should be reflected in the model.

(ii) The travel time for a care worker to reach a client should be taken into account.

(iii) The model should be sufficiently simple to be useful in practice.

Queueing models are the natural candidate in view of feature (i). We note that detailed simulation models may also capture the stochastic nature, but due to the limited availability of data and process information we opt for simple models that reflect the key characteristics of the health delivery process. Such simple models demonstrate the important principles for supporting staffing decisions on a strategic or tactical level and are sufficiently simple to implement.

Below, we discuss the elements of the queueing model.

**Arrivals** As discussed in Section 6.3, the interarrival times overnight are approximated well by an exponential distribution. The arrival of call requests are therefore assumed to follow a Poisson process with rate $\lambda$. We note that the number of residents is bounded, suggesting that a finite-source queueing model may be applicable when we assume that a resident does not generate any new calls while he/she is waiting. However, closed queueing models are much more difficult to analyze. Moreover, the number of residents is large enough (180 residents) for the difference between open and closed models to be negligible. For small-scale living facilities, having in the order of 10 residents, some further analysis is required though.

**Service times** The time for care delivery is yet not entirely clear from Section 6.3. Moreover, the time required for care workers to react to call requests and travel to the corresponding room may be considerable. We refer to this combination as *travel time*. The service time $S$ is then defined as $S = S_1 + S_2$, with $S_1$ the overall travel time and $S_2$ the time for care delivery. Unfortunately, information about travel times can not be derived from the data; such durations also highly depend on the local situation. In the light of (iii), we propose to use a two-phase hypoexponential distribution with parameters $\gamma_1 \neq \gamma_2$ for $S_1$, which is the sum of two exponential durations. The first phase may be interpreted as time to react (e.g.,notice the call, finish current task) and the second phase as actual traveling time. The two-phase hypoexponential distribution has coefficient of variation between 0.5 and 1, and the peak in probability mass is at a point larger than zero if $\gamma_1, \gamma_2 < \infty$. We consider this approximation to be reasonable. Note that the above implies that the full service time $S$ is general, as we did not yet make an assumption for $S_2$.

**Servers** Let $s$ be the number of servers, representing the care workers. We assume that the care workers are dedicated to 'care on demand' tasks. This is reasonable for the overnight period due to the limited 'care by appointment' activities. By day, it depends on how the care process is organized whether 'care on demand' and 'care by appointment' are mixed or separated.

The arguments above support the use of the $M/G/s$ queueing model. At this point we would like to make two relevant remarks. First, the $M/G/s$ model is intractable and for its analysis we rely on approximations available in the literature. Second, the waiting time in the $M/G/s$ queue corresponds to the time when a care worker is available to visit a client. In line with the queueing literature, we refer to this as $W_Q$. The time that a client is actually waiting for a care worker includes traveling time, i.e., the performance measure of interest is $R = W_Q + S_1$, with $R$ referring to response time.

Let us now first consider approximations for the waiting time $WQ$ in the $M/G/s$ queue. An important point of reference is the $M/M/s$ queue. As the coefficient of variation of $S_2$ is 1.61 (see Subsection 6.3) and the coefficient of variation of $S_1$ is between 0.5 and 1, approximating the service time $S = S_1+S_2$ by an exponential random variable may not be that bad (the coefficient of variation of $S$ is likely not too far from 1).

We follow the approximation proposed by Whitt (1992, 1993) and consider the probability of delay $\mathbb{P}(W_Q > 0)$ and the waiting time distribution separately. The probability of waiting in the $M/G/s$ queue has been relatively well studied in the literature, see e.g., Kimura (1994) and Whitt (1992, 1993) and references therein. As noted in Whitt (1993, p.134) and Kimura (1994, p.371), the probability of waiting in the $M/M/s$ model is usually an excellent approximation for the probability of waiting in the corresponding $M/G/s$ queue. Hence we have

$$\mathbb{P}(W_Q > 0) \approx \frac{a^s}{(s-1)!(s-a)} \left[ \sum_{i=0}^{s-1} \frac{a^i}{i!} + \frac{a^s}{(s-1)!(s-a)} \right]^{-1}, \tag{6.1}$$

where $a = \lambda \mathbb{E} S$ is the offered load. Let $\rho = a/s$ denote the load per care worker and assume that $\rho < 1$. In heavy traffic, the conditional waiting time $(W_Q | W_Q > 0)$ has an exponential distribution. Moreover, for the $M/M/s$ queue the waiting time is also exponential. In line with Abate et al. (1995), Kimura (1994) and Whitt (1992) we suggest a simple exponential approximation of the waiting time given that the waiting time is positive. It remains to specify the parameter of this exponential distribution.

We follow Whitt (1992) and let the parameter coincide with heavy-traffic analysis, yielding

$$\mathbb{P}(W_Q > t) \approx \mathbb{P}(W_Q > 0)e^{-\beta t}, \tag{6.2}$$

with

$$\beta = \frac{2}{c_A^2 + c_S^2}(1-\rho)s. \tag{6.3}$$

Here the squared coefficient of variation of the interarrival times equals 1, because arrivals are assumed to follow a Poisson process; the squared coefficient of variation of the overall service time $S$ is given by

$$c_S^2 = \frac{\mathbb{V}\mathrm{ar}S}{(\mathbb{E}S)^2} = \frac{\mathbb{V}\mathrm{ar}S_1 + \mathbb{V}\mathrm{ar}S_2}{(\mathbb{E}S_1 + \mathbb{E}S_2)^2} = \frac{1/\gamma_1^2 + 1/\gamma_2^2 + 16.94}{(1/\gamma_1 + 1/\gamma_2 + 2.56)^2},$$

assuming that $S_1$ and $S_2$ are independent and $\mathbb{E}S_2$ and $\mathbb{V}\mathrm{ar}S_2$ follow from Section 6.3.

Now, we turn to the response time, which is the convolution of the waiting time with the travel time. Let $F_{S_1}(t)$ be the distribution function of $S_1$, which for $t \geq 0$, is given by,

$$F_{S_1}(t) = 1 - \frac{1}{\gamma_2 - \gamma_1}\left(\gamma_2 e^{-\gamma_1 t} - \gamma_1 e^{-\gamma_2 t}\right). \tag{6.4}$$

Conditioning on the waiting time, and combining the results above, provides an approximation for the quantity of interest: for $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(R \leq t) &= \mathbb{P}(W_Q = 0)F_{S_1}(t) + \int_0^t F_{S_1}(t-u)\mathrm{d}\mathbb{P}(W_Q \leq u) \\
&\approx (1 - \mathbb{P}(W_Q > 0))\left[1 - \frac{1}{\gamma_2 - \gamma_1}\left(\gamma_2 e^{-\gamma_1 t} - \gamma_1 e^{-\gamma_2 t}\right)\right] + \mathbb{P}(W_Q > 0) \\
&\quad \times \int_0^t \left(\beta e^{-\beta u} - \frac{\beta\gamma_2}{\gamma_2 - \gamma_1}e^{-\gamma_1 t}e^{-u(\beta - \gamma_1)} + \frac{\beta\gamma_1}{\gamma_2 - \gamma_1}e^{-\gamma_2 t}e^{-u(\beta-\gamma_2)}\right) du \\
&= 1 - (1 - \mathbb{P}(W_Q > 0))\frac{1}{\gamma_2 - \gamma_1}\left(\gamma_2 e^{-\gamma_1 t} - \gamma_1 e^{-\gamma_2 t}\right) \\
&\quad + \mathbb{P}(W_Q > 0)\Bigg(\frac{-\gamma_1\gamma_2}{(\beta-\gamma_1)(\beta-\gamma_2)}e^{-\beta t} - \frac{\beta\gamma_2}{(\gamma_2-\gamma_1)(\beta-\gamma_1)}e^{-\gamma_1 t} \\
&\qquad + \frac{\beta\gamma_1}{(\gamma_2-\gamma_1)(\beta-\gamma_2)}e^{-\gamma_2 t}\Bigg),
\end{aligned} \tag{6.5}$$

where we used Equations (6.4) and (6.2) in the second step and the third step follows from basic calculus. Here $\mathbb{P}(W_Q > 0)$ and $\beta$ are given by (6.1) and

(6.3), respectively. This simple formula thus provides the approximate service level, i.e., the fraction of clients who wait no longer than $t$ minutes for receiving care.

**Remark 1.** The most involved part in the performance analysis may be $\mathbb{P}(W_Q > 0)$. A more elementary approximation for this is due to Sakasegawa (1977)

$$\mathbb{P}(W_Q > 0) \approx \rho^{\sqrt{2(s+1)}-1}.$$

As this approximation is mainly accurate for high loads, we advocate the use of Equation (6.1).

**Remark 2.** Kimura (1994) also suggests an exponential distribution for the conditional waiting time, but proposes a refined parameter $\beta$ that is also exact in light traffic, see Equation (5.12) of Kimura (1994). Since the more elementary $\beta$ above suffices, see also the next subsection; we advocate the use of the simpler one here.

**Remark 3.** Assuming that the number of residents is fixed at $N$ and that they do not generate any new calls during waiting, the relevant model is in fact the finite-source $M/G/s//N$ system. For exponential service times the number of customers in such a system is a birth-and-death process from which the stationary distribution of the number of waiting customers $\pi_i$ is easily derived see, for instance, Kleinrock (1975). The waiting time distribution is then

$$\mathbb{P}(W_Q > t) = e^{-s\mu t} \sum_{k=s}^{n} \frac{(n-k)\pi_k}{\sum_{i=0}^{n}(n-i)\pi_i} \sum_{j=0}^{k-s} \frac{(s\mu t)^j}{j!}.$$

This also leads to closed-form results, but the convolution with the travel time is now more involved. For small-scale living facilities, the M/M/s//N model is more appropriate. Caution is required when the coefficient of variation is far from 1. In that case, a viable option is to rely on relations between open and closed queueing systems, as in e.g., Satyam et al. (2013) and Whitt (1984).

### Model validation

We validated the model using the time frame between 22:45 and 5:45 hours again. Most parameters can be derived from Section 6.3. For this nursing home we assumed the number of care workers was fixed at three during the night. In practice, the number of care workers is seldom fixed throughout the year

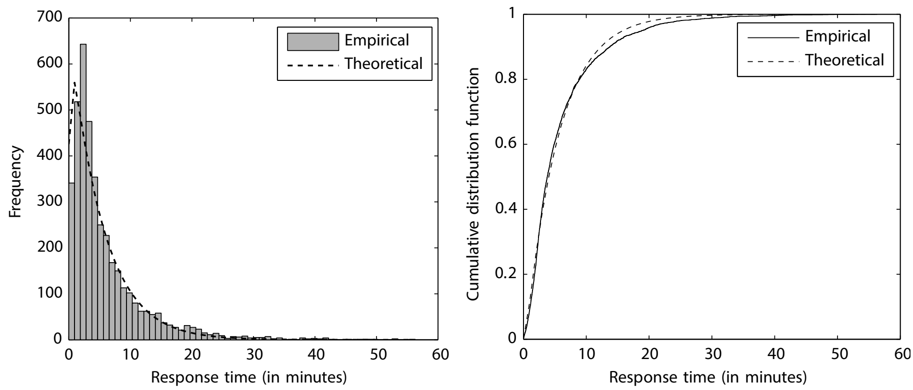


**Figure 6.10:** Scaled probability density function (left) and cumulative distribution function (right) of the empirical response times with the response time distribution based on the $M/G/3$ model

due to illnesses, deficient scheduling during staff leave, and other care activities such as 'care by appointment'. As mentioned, the parameters $\gamma_1$ and $\gamma_2$ can not be estimated directly as there is no data available for travel times. We estimated the parameters $\gamma_1$ and $\gamma_2$ by minimizing the mean squared error between the empirical cumulative distribution function of the response times and $\mathbb{P}(R \leq t)$ as presented in (6.5). For $s = 3$, this yields $\hat{\gamma}_1 = 2.87$ and $\hat{\gamma}_2 = 0.20$. From practical experience, these values seem to resemble the total *travel time* reasonably well.

The probability density function (pdf) and the cumulative distribution function (cdf) of the response time $R$ are displayed in Figure 6.10 along with the empirical versions. Because the response time is the convolution of $W_Q$ with the travel time $S_1$, it has no probability mass in zero. In other words, every client has to wait at least until a care worker is present in the room. Overall, the theoretical model fits fairly well to the empirical distribution. For very short response times, there is a difference in pdf. In particular, the peak of the empirical distribution is shifted a little to the right compared with our approximate queueing model. This indicates that the two-phase hypoexponential assumption for $S_1$ may not be perfect. For the cdf, the largest difference occurs for response time of about 20 minutes, although the differences are modest. For the empirical distribution also a small peak seems to emerge around this time window.

To verify the approximation, we also simulated the finite source $M/G/s//N$ model with $N = 180$ residents. For the travel time $S_1$, we used the two-phase hypoexponential distribution and for the care delivery time $S_2$, we randomly draw from the empirical data. The simulation results (5,000,000 care requests and a warm-up period of 100,000 care requests) are very similar to the response-time approximation. We compared the empirical distribution from the simulation with the approximate theoretical distribution function from Section 6.4. The mean relative error between these distributions is 0,1495%. We omitted the simulation results in Figure 6.10 as the lines for the simulation and the approximation are indistinguishable.

Our experience is that nursing homes are not managed based on agreements over service level. However, a typical quantity of interest is the fraction of clients who wait no longer than 10 minutes, which may be defined as the service level (see Section 6.2). For both the approximate queueing model and the data, this is slightly over 80%. For such service levels, the approximation is rather accurate. Our general conclusion is that for any practical purposes the $M/G/s$ model seems to suffice. More specifically, the service levels based on our $M/G/s$ approximation are not far from the realized service levels in the data. In the next section, we exploit the queueing model to evaluate the impact of different practical scenarios.

## 6.5 Numerical scenarios

In this section we use the approximate queueing model to obtain insight into different nursing home scenarios. Specifically, we investigate the impact of care delivery times, call requests and scale on the response time $R$. As introduced in Section 6.2, we focus on the service level (SL) as our performance measure. A SL of $X/Y$ denotes that

$$\mathbb{P}(R \leq Y) = \frac{X}{100},$$

that is, the response time is less then $Y$ minutes for $X$ percent of the clients. In the current situation a SL of 80/10 is met as the probability that the response time will be longer than 10 minutes is slightly below 0.2.

In view of the changing landscape for long-term care, nursing homes will increasingly face questions related to capacity decisions and retaining nursing staff. As noted in Brazil et al. (2012) and McGilton et al. (2013), and references therein, elderly people living in long-term care facilities have increasingly complex care needs. This may manifest in longer durations of care delivery, an increased number of call button requests, or both. In turn, this will affect appropriate staffing levels. Below, we briefly investigate the relations between service levels and the intensity of care.

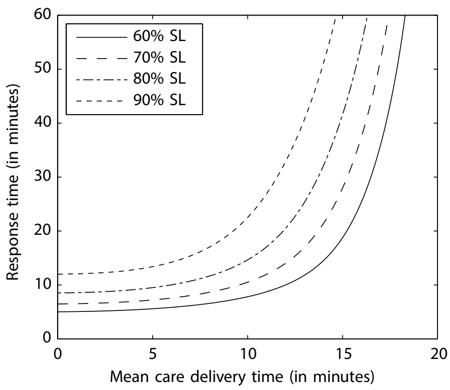


**Figure 6.11:** The impact of mean care delivery time (left) and intensity of call requests per quarter hour (right) on the response time $Y$ to achieve a fixed service level $X \in \{60, 70, 80, 90\}$

For all figures, we use the situation and parameters as described in Section 6.4 as our basic scenario. In Figure 6.11, we plotted the response time $Y$ for different service levels as the mean duration of care delivery $\mathbb{E}S_2$ varies (left) or the number of call requests $\lambda$ varies (right). For example, for the current situation the mean care delivery time is 2.56 minutes and on average 1.6 call requests arrive per quarter hour. We can read from the vertical axis that a 60% SL is achieved at roughly 5.2 minutes, i.e. $\mathbb{P}(R \leq 5.2) \simeq 0.6$. The 70, 80 and 90% service levels are achieved at approximately 6.7, 8.8, and 12.4 minutes. As such, for every mean care delivery time (left) and arrivals of call requests (right), the impact of different choices for the SL target may be read along the vertical axis. For the SL displayed, the difference between a SL of 80% and 90% is the largest, which indicates that the target time is increasing faster as the fraction of clients from whom the service should meet the response target increases. In other words, the amount of *extra* capacity requirements increases as the SL becomes more tight.

Reading Figure 6.11 along the horizontal axis, one may see the impact of increasing the mean time for care delivery (left) and intensity of call requests (right). The impact on the SL is rather modest as the parameters change slightly (which is due to the relatively small load of the system). Even if $\mathbb{E}S_2$ is 10 minutes, then in 80% of the cases a care worker is present within 15 minutes. Note that the impact is largest for the 90% SL.

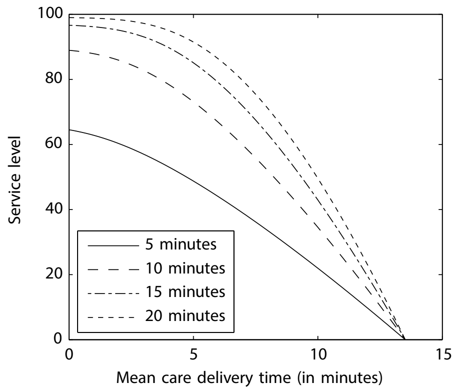


**Figure 6.12:** The impact of mean care delivery time on the service level $X$ for different values of the response target $Y \in \{5, 10, 15, 20\}$

Figure 6.12 shows the relation between the SL (for four different target response times $Y$) and mean care delivery time. For instance, for the current mean care delivery time of 2.56 minutes and a target time of 5 minutes the SL is only 59%, whereas the SL is 84, 94 and 98% for target times $Y$ of 10, 15, and 20 minutes, respectively. From the above it follows directly that a target time of 5 minutes is not very useful. Even if the mean care delivery time is negligible, the SL is then still just 60%. This is due to the traveling time. Again, for a fixed mean care delivery time, the SL for different response time targets may be read along the vertical axis. Using these curves a target in the range of 10–15 minutes seems appropriate. Reading Figure 6.12 along the horizontal axis, it can be seen which mean care delivery times can be handled while maintaining a specified SL target. As an example, the mean care delivery time may rise up to about 6 minutes to maintain an 80/10 SL. Based on the insights from the figures above, we advocate using a 80/10 SL for this particular situation.

Finally, we investigate the effect of scale. The mean care delivery time is fixed as in the basic scenario. As a starting point, we assume that an 80/10 SL is desirable. We vary the scale in terms of number of care workers (employees) and let the arrival rate of call requests vary accordingly, such that the 80/10 SL is met exactly. A difficult issue is the impact of scale on traveling times. It seems natural that traveling times increase as scale increases. To what extent this holds strongly depends on the local condition as design of the building and relative positions of different units. Moreover, for longer distances other types of transport (e.g., scooters) may be profitable, such that drawbacks related to

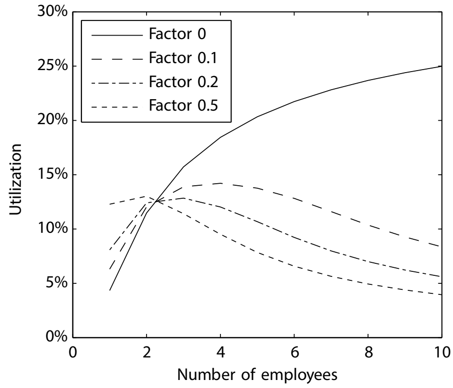


**Figure 6.13:** Impact of scale (number of employees) on utilization excluding traveling to achieve an 80/10 SL

scale may be circumvented. Here, we consider four scenarios where the mean travel time $\mathbb{E}S_1$ increases with 0, 10, 20, and 50% of the mean rate of call requests.

In Figure 6.13 we illustrate the impact of scale on the utilization for the four traveling times. The utilization is based on times of care delivery and excludes travel times. The low utilization follows from considerable travel times, requiring overcapacity. When mean travel times remain constant (factor is 0), we see that the utilization increases in a concave manner. This is in line with the general concept of economies of scale. At larger scale, the relative variability in required care decreases and the utilization increases to maintain the same SL. This effect however becomes weaker when scale increases. Hence, 'care on demand' should not be organized at too small a scale.

In case travel times increase with scale (factor larger than 0), there is a trade-off between traditional economies of scale and impact of travel times. From Figure 6.13, we see that the optimal size depends on the specific factor and, therefore, on the local situation. In any case, the optimal number of employees is at least two, confirming that organizational units for 'care on demand' should not be too small.

## 6.6 Conclusions and discussion

In this chapter we made a first step in trying to understand the real-life performance of the 'care on demand' process in a Belgian nursing home facility using a queueing theoretic approach. From a methodological point of view, the contribution of this study is twofold. First, by using real-life data we obtain insight into the number of call requests and care delivery times for 'care on demand' activities. Secondly, we developed a queueing model to support capacity decisions. Based on numerical experiments, we propose an 80/10 service level for this specific nursing home facility, meaning that at least 80% of the clients should receive care within 10 minutes after a call button request. Although we think that an 80/10 service level will be suitable in many situations, this may depend on the specific context.

From a practical perspective, this study provides a basis on which it is possible to develop a staffing support tool for 'care on demand' activities which would allow nursing home managers (1) to determine the number of care workers required to meet sufficiently the needs and preferences of the nursing home residents when it comes to 'care on demand' and (2) to better understand the implications of their decisions (i.e. what-if scenarios). We think that such a tool has the potential to make an important contribution in the quest for more efficiency, without losing sight of the needs of residents.

A model is never a complete representation of reality and the queueing model presented in this chapter is no exception. First of all, this study is limited in scope because it only addresses night time care. A similar approach could be taken for the 'care on demand' process during the day. During day time, the amount of 'care by appointment' activities is much larger than at night, which may lead to compromised 'care on demand' data. Moreover, data on traveling times is lacking. Although travel times may vary depending on the local situation, it would be of interest to model this in more detail. Finally, we used an approximation for the queueing model. This approximation is expected to work well in most nursing home situations, but the accuracy may decrease when, for example, the number of clients is getting very small. Despite the fact that long-term elderly care will become increasingly important in the next decades, the body of operations research (OR) literature directed to this topic is still very limited. Therefore we would like to challenge researchers in the field of OR to put more emphasis on research in long-term elderly care.

## Acknowledgements

The authors would like to thank Niko Projects for providing us with a dataset and Jan-Pieter Dorsman for the simulation of the finite-source queueing model.

Chapter **7**

# On the performance of small-scale living facilities: a simulation approach

**Abstract**

Scientific evidence on the impact of small-scale living facilities (SSLFs) on quality of life remains scarce. In this chapter a simulation model is developed to examine the performance of SSLFs, in terms of meeting the time preferences of their residents. The model is used to explore the impact of a change in demand characteristics, duration of care delivery, travel time, allocation flexibility, shifts, staffing levels, number of clients and allocation policy. The results show that to further improve the performance, the focus should lie on: increasing (1) the allocation flexibility of care workers and (2) the number of clients per SSLF. We also consider the problem of minimizing the total number of care workers, while maintaining a 95/15 service-level throughout the day. The results show that allocation flexibility together with numerical flexibility has a substantial positive effect on the performance. Furthermore, this study shows that simulation is a useful tool for assessing and improving daily nursing home operations. The presented simulation model provides a basis for building a decision support tool for nursing home managers.

This chapter is based on Moeke et al. (2015b).

## 7.1 Introduction

In the last decades, many Dutch nursing homes have transformed their care environments into more client-centred, small-scale care settings. These small-scale settings, often referred to as small-scale living facilities (SSLFs) or small-scale care facilities, provide care and/or support to a small group of residents in a homelike environment. This approach aims to enable residents to live their lives according to their own needs and preferences (Verbeek, 2011) or as stated by Willemse et al. (2014, p. 804) “daily life and care provision are adjusted to resident’s lifestyle and preferences as much as possible”. Despite the fact that most nursing homes have embraced the concept of SSLFs as a way to give substance to client-centred care, scientific evidence on the impact of SSLFs on quality of life (QoL) is still scarce (De Rooij et al., 2012; Verbeek et al., 2014).

As mentioned in Chapter 4, in order to make it possible for nursing home residents to live their lives according to their needs and preferences, the necessary care and support should be delivered as close as possible to the time preferences of the residents. Hence, with regard to the delivery of care and support earliness and waiting should be avoided, without overstretching the available budget. From this perspective, QoL largely depends on the coordination and timing of service delivery.

In this chapter a simulation model is developed to examine the performance of SSLFs, in terms of meeting the time preferences of their residents, under different assumptions. Furthermore, improvements in the allocation of care workers are suggested. Simulation is a commonly used method to study the effectiveness or efficiency of larger and/or complex systems which do not lend themselves for analytic approaches. Regarding the purpose of this study, the main advantage of a simulation approach is its flexibility, as parameters and assumptions of the underlying model can be adjusted relatively easy. As such, simulation (1) is well suited to answer ‘what-if’ questions and (2) allows for a detailed analysis of how components interact and the trade-offs involved.

The studies of Hamrock et al. (2013); Brailsford and Vissers (2011); Günal and Pidd (2010); Fone et al. (2003) show that, during the past two decades, simulation has been extensively used for modelling healthcare systems. The vast majority of existing studies focus on supporting better operational decision-making and planning in a hospital setting, with an emphasis on specific subsystems. Examples of those subsystems are: operating theatres (Saadouli et al., 2015; Zhang and Xie, 2015), emergency departments (Konrad et al., 2013; Kuo, 2014) and intensive care units (Zhu et al., 2012; Griffiths et al., 2010). As mentioned in the previous chapters (see e.g., chapters 4 and 6), to date the area of nursing home care received hardly any attention in the OR literature. To the best of our knowledge this is the first study examining the daily operations of SSLFs using simulation.

The remainder of this chapter is structured as follows. In the next section, the study is outlined and justified by providing an overview of the existing literature. Then, in Section 7.3 of this chapter, we present the analysis of the input data and in Section 7.4 the model is described. The scenarios together with their performance are presented in Section 7.5. In Section 7.6, for each half-hour interval, we determine the number of care workers required to meet a predetermined service-level. Finally, in the last section, conclusions are drawn and the results are discussed.

## 7.2 Background

Today, the concept of SSLFs is applied in both somatic and psychogeriatric care settings. However, most research on the effects of SSLFs has been conducted in psychogeriatric settings. This is not surprising considering the fact that SSLFs were originally developed for residents suffering from dementia. The term 'dementia' is a catch-all term for a group of symptoms caused by gradual death of brain cells (Thackery and Harris, 2003). Common symptoms are problems with concentration, memory, thinking, behavior and the ability to perform everyday activities. The often progressive nature of the disease makes that carrying out activities of daily living (ADL) becomes more and more of a challenge and causes an increase in dependency. Therefore, as the disease progresses, institutional nursing is often inevitable. Until the 80s of the last century the behavior of dementia clients was mainly interpreted from a disease perspective, whereby little attention was paid to the experiences and perceptions of the dementing person (De Lange, 2004). From the 90s onwards, more and more emphasis was put on how people with dementia cope with the consequences of their illness and on how they experience and value their personal situation (Finnema et al., 2000). The aim to make dementia care fit the feelings and emotional needs of each individual client has led to development of so-called 'integrated emotion-oriented care', in which "the offered care and activities are person-centered and well attuned to the abilities, experiences and preferences of the person with dementia" (Dröes et al., 2010, p. 153). This shift towards more holistic dementia care was accompanied by the emergence of small-scale and homelike nursing home settings. In the Netherlands, the first SSLF was introduced in the early 80s. In the decades that followed, encouraged by government policies and programs, the number of SSLFs increased. In 2010, roughly 25% of the residents with dementia who received nursing home care lived in SSLFs (Smit et al., 2012). Despite the lack of a uniform definition of a SSLF, the following common characteristics regarding SSLFs can be identified (Verbeek, 2011; De Rooij, 2012):

1. Providing care and/or support to 6-8 residents.
2. Home-like environment.
3. Small fixed team of care workers.
4. Care workers perform a broad range of tasks.
5. More individual decision making by care workers.
6. Residents should have influence on their daily routine.

Today it is generally assumed that living in a SSLFs will add to the QoL, however scientific evidence to support this assumption is sparse and mixed in its results.

In a comparative study Te Boekhorst et al. (2009) examined the effects of small-scale living for people with dementia compared with living in a traditional nursing home setting. The results of this study show that the residents living in a SSLF (1) needed less assistance with ADLs, (2) had more social engagement, (3) had greater sense of aesthetics, (4) had more to do during the day and (5) were prescribed less physical restraints. However, the authors did not find differences in cognitive status, behavioral problems and the prescription of psychotropic drugs. Verbeek et al. (2010) also investigated the effects of small-scale living compared with traditional nursing home care. The purpose of this study was to evaluate the effects of small-scale living facilities in dementia care on residents (QoL and behavior), family caregivers (experienced burden, involvement with care and satisfaction), and staff (job satisfaction and motivation). No convincing overall effects of small-scale living facilities were found. In their summary the authors state:

> Because governmental policies and, in some countries, financial support, are increasingly aimed at providing small-scale, homelike care, it is suggested that this may not be a final solution to accomplish high-quality dementia care and that other options should be considered.
> (Verbeek et al., 2010, p. 662)

Using a quasi-experiment, De Rooij et al. (2012) examined the benefits of small-scale living for residents with dementia, compared to traditional long-term care in the Netherlands and Belgium. Their findings indicate that both small-scale and traditional settings appear to have beneficial effects on different domains. The Dutch sample showed higher scores on 'social relations', 'positive affect' and 'having something to do' for small-scale settings compared to residents in traditional settings. Moreover, mean scores on 'caregiver relation' and 'negative affect' remained more stable over time among residents in small-scale settings compared to traditional settings. In the Belgian sample, the differences found between traditional and small-scale settings were less evident. The authors also found that residents 'felt more at home' in traditional settings. Moreover, the mean QoL scores on 'restless behavior', 'having something to do' and 'social relations' remained stable in the traditional setting but decreased in the small-scale settings.

In chapters 4 and 5 it is shown that, when it comes to efficiently meeting the time preferences of nursing home clients, scale of scheduling plays a prominent role. Applying small-scale scheduling can lead to less efficiency in terms of number of care workers needed to meet the demand. These findings are supported by our recent follow-up study (Lieder et al., 2015). The results of this study show that using large-scale scheduling is beneficial to the overall quality of service in terms of total tardiness in the obtained task schedules.

We believe the disadvantages of small scale organizing can be compensated for by applying job enlargement, but only up to a certain extent (see Chapter 5). Hence, our preliminary assumption is that *when the scale of scheduling becomes too small, it will become impossible to meet the preferences of the nursing home residents (in terms of moment and time) without increasing the number of care workers*. As care and support activities in SSLFs are organized on a relatively small-scale, we expect the quality of care in terms of meeting the time preferences of the residents to be low (i.e., we expect waiting times to be relatively long).

The main contributions of this chapter can be summarized as follows. Using simulation and real-life data we:

1. Examine the performance of SSLFs, in terms of meeting the time preferences of their residents, under different assumptions.

2. Suggest improvements (i.e., alternative approaches) regarding the allocation of care workers, using the simulation results and insights gained from the previous chapters.

3. Show that simulation is a useful tool for assessing and improving daily nursing home operations.

It should be stressed that, in addition to the previous chapters, in this chapter both scheduled and unscheduled care activities are taken into account. As reliable data is hard to obtain, we use the same data as used in Chapters 4 (scheduled care) and 6 (unscheduled care).

In the following section we present the analysis of the input data for the simulation model.

## 7.3 Data analysis

The main emphasis of this section is on the analysis of the main characteristics of the unscheduled care process during daytime (7:00-23:00 hours). See Chapter 4 for an extensive analysis of the care delivery process of scheduled care.

To make assumptions about the demand and the service times of unscheduled healthcare tasks during daytime, we use the call-button data that we collected for the study presented in Chapter 6. The characteristics of the care delivery process during the night (i.e., between 22:45 and 5:45 hours) are taken as point of departure. The two main reasons for this are that (i) the data are not compromised by 'care by appointment' related data, and (ii) during day time, unscheduled care is not solely initiated by call button requests.

Since the arrival pattern of each day of the week (see Figure 6.5) corresponds with the overall arrival pattern (see Figure 6.3) we assume there is no structural weekly pattern. In Chapter 6 it was shown that, for the night, it can be assumed the arrivals follow a homogeneous Poisson process (see Table 6.1 and Figure 6.6). However, the boxplots in Figure 6.4 and Table 7.1 indicate that the arrival rates during daytime are inhomogeneous. To confirm statistically whether or not the arrival rate is constant during the day, we conduct a $\chi^2$-test. Lunch-and dinner time (11:30-12:30 and 16:30-17:30 hours) are excluded from the analysis, because the limited number of call requests during lunch and dinner give a distorted view. For each of the 90-days, the $\chi^2$-test is used to determine whether the empirical arrival frequencies originate from a Poisson distribution. We use the following property of the Poisson process: given a fixed number of arrivals in an interval, the numbers of arrivals in non-overlapping subintervals have jointly a multinomial distribution. If the Poisson process is homogeneous between 7:00-23:00 hours and the subintervals are of an equal length of 30 minutes (i.e., 28 subintervals), then the multinomial probabilities are equal (i.e. 1/28). The $\chi^2$-test is performed in order to compare empirical arrival frequencies with the multinomial distribution. Using a significance level of 0.05, the null hypothesis is rejected for 18 (20%) of the 90 days. In addition, we also performed a KS-test to test the null hypothesis that the underlying distribution follows a Poisson distribution. The value of the parameter of the Poisson distribution is estimated using maximum-likelihood estimation (MLE). The test results show that the null hypothesis is rejected at 5% significance level, with a $p$-value of 0.00.

Based on the test results we assume the arrival process to be inhomogeneous with the arrival rate depending on the time of day. To account for this non-stationarity we divide the day into three time periods: (1) morning (7:00-12:00 hours), (2) afternoon (12:00-18:00 hours) and (3) evening (18:00-23:00 hours) as these time periods coincide with the repetitive cycle of daily routines. Next we repeat the $\chi^2$ test for each of the time periods separately [1]. The null hypothesis is now rejected for 13% (morning), 14% (afternoon) and 6% (evening) of the 90 days. Hence, within each time period the arrivals show a more stable pattern. Based on the test results, we assume the arrival process to be homogeneous within each time period.

However, it is observed that within each time period the number of requests fluctuate more than we would expect from a Poisson process. Under Poisson assumptions, the variance and the mean should be approximately the same. For the three time periods, Table 7.2 shows the 90-day average of the number of call requests, the corresponding variance and the variance-to-mean ratio. It can be observed that for each of the time periods the variance-to-mean ratio is larger than 1.

[1] lunch and dinner time are again excluded from the analysis

**Table 7.1:** Average number of call requests per hour (7:00-23:00 hours)

| | Hours | Average # call requests |
|---|---|---|
| Morning | 7:00-7:30 | 5.9 |
| | 7:30-8:00 | 5.2 |
| | 8:00-8:30 | 4.3 |
| | 8:30-9:00 | 3.6 |
| | 9:00-9:30 | 5.2 |
| | 9:30-10:00 | 5.2 |
| | 10:00-10:30 | 4.5 |
| | 10:30-11:00 | 5.9 |
| | 11:00-11:30 | 6.1 |
| | 11:30-12:00 | 2.3 |
| Afternoon | 12:00-12:30 | 3.6 |
| | 12:30-13:00 | 5.9 |
| | 13:00-13:30 | 4.8 |
| | 13:30-14:00 | 4.8 |
| | 14:00-14:30 | 4.9 |
| | 14:30-15:00 | 4.3 |
| | 15:00-15:30 | 5.1 |
| | 15:30-16:00 | 6.4 |
| | 16:00-16:30 | 4.8 |
| | 16:30-17:00 | 2.0 |
| | 17:00-17:30 | 3.9 |
| | 17:30-18:00 | 6.2 |
| Evening | 18:00-18:30 | 6.9 |
| | 18:30-19:00 | 6.5 |
| | 19:00-19:30 | 6.5 |
| | 19:30-20:00 | 6.4 |
| | 20:00-20:30 | 6.5 |
| | 20:30-21:00 | 5.6 |
| | 21:00-21:30 | 5.8 |
| | 21:30-22:00 | 5.2 |
| | 22:00-22:30 | 6.3 |
| | 22:30-23:00 | 4.2 |

**Table 7.2:** Variance-to-mean ratio for each time period

| Time period | Average # call requests | Variance | Variance-to-mean |
|---|---|---|---|
| 7:00-12:00 | 46.00 | 69.19 | 1.50 |
| 12:00-18:00 | 47.04 | 125.59 | 2.67 |
| 18:00-23:00 | 59.87 | 118.57 | 1.98 |

The observed overdispersion is confirmed using the following test statistic, which is also known as the Neyman-Scott test (see e.g., Lindsay, 1995):

$$T_{NS} = \sqrt{\frac{n-1}{2}} \left( \frac{S_n^2}{\bar{X}_n} - 1 \right)$$

This test rejects if $T_{NS} > \Phi^{-1}(1-\alpha)$ (see e.g., Brown and Zhao, 2002). For the morning, afternoon and evening the data shows values for $T_{NS}$ of respectively 3.25, 10.64 and 7.22. Using a significance level of 0.05, the null hypothesis is rejected for all three time periods. A substantial part of the observed dispersion is presumably attributable to a small number of frequent callers. From the data it can be seen that for only 22% of the clients the average number of call requests per day is more than 1. In addition, for most of these frequent callers, the total number of calls fluctuates considerably from day to day. Figure 7.1 shows the number of calls per day for four frequent callers. Especially the plots of clients 1 and 4 show some large spikes. One possible explanation for these spikes in the number of requests are temporarily poor health conditions.

The next section describes the simulation model used to assess the performance of the scenarios presented in Section 7.5.

## 7.4 Simulation model

### Demand for care and support

We consider a nursing home department consisting of 4 SSLFs. Each SSLF provides 24/7 care and support to six psychogeriatric clients in a small archetypical house-like facility. The nursing home department aims to deliver the necessary care and support as close as possible to the time preferences of the residents.

The data used as input for the simulation contains the following variables:

- *Client ID* – the ID of a specific client.

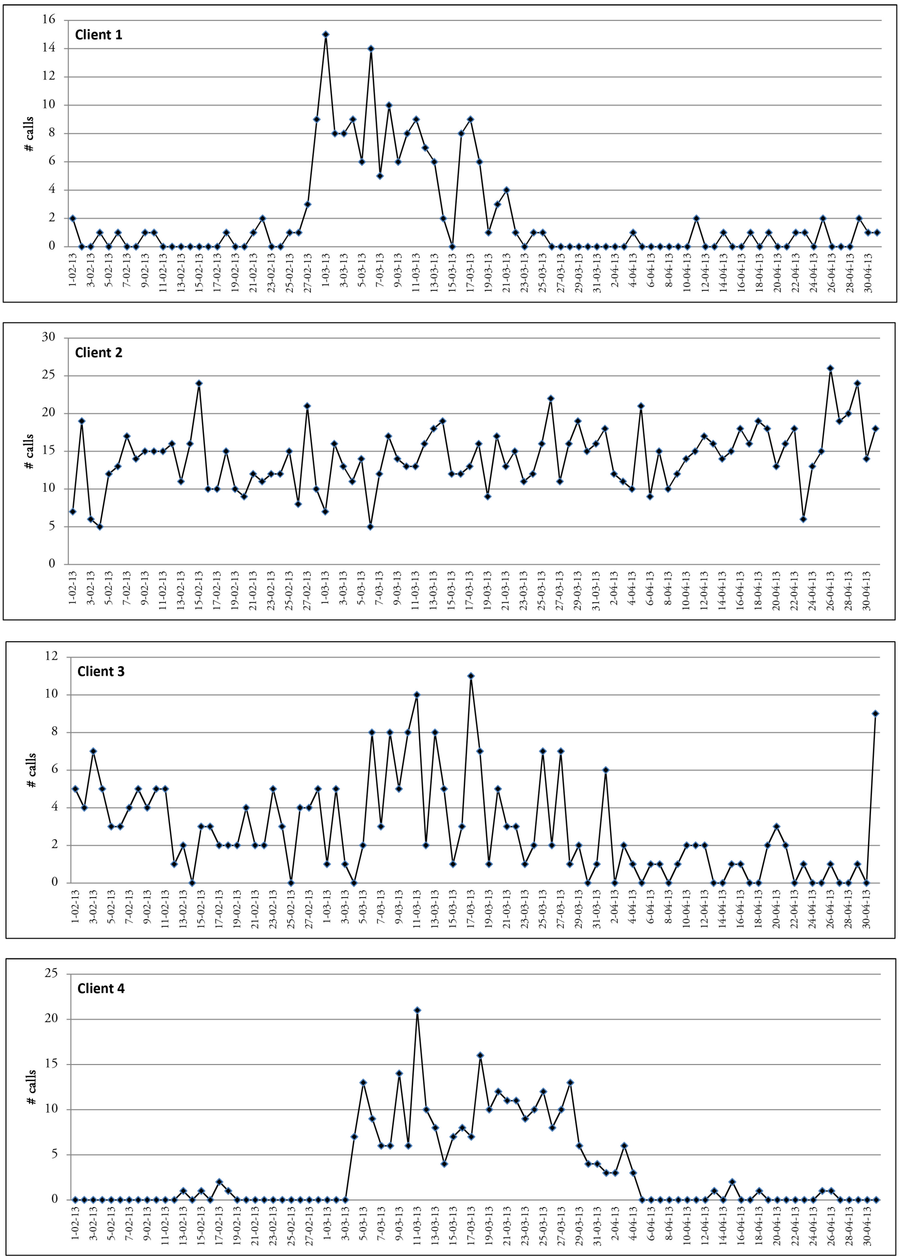


**Figure 7.1:** Number of calls during the day for client 1, 2, 3 and 4

- *Preferred Activity Time (PAT)* – the preferred starting time of the healthcare activity.
- *Date* – the date of the PAT.
- *Task* – either a scheduled or unscheduled healthcare task.
- *Service time* – duration of the care activity in minutes.

It should be noted that in this chapter an unscheduled care request is also referred to as a PAT.

**Scheduled care**

For the purpose of this simulation study, for each simulation round, the SSLFs are 'filled' with new clients. Together with their PATs for scheduled care and corresponding expected service times, each client is randomly chosen from the empirical datasets of departments A, B, C, D and E presented in Chapter 4. As such, we assume there are no differences between psychogeriatric and somatic clients in terms of PATs and the corresponding expected service times. We believe this assumption is plausible because scheduled care mainly consists of providing help and support with ADLs and there is no considerable difference (in terms of number and duration) between psychogeriatric and somatic clients with respect to these activities. The main characteristics of the datasets can be found in Table 4.1 and Table 4.2. Figure 7.2 shows how the 'expected' care delivery times (i.e. service times) for scheduled care are distributed. The average care duration for scheduled care activities is 12.64 minutes.

**Unscheduled care**

The test results in Section 7.3 show that the arrival rate of unscheduled care depends on the time of the day. In order to account for this time dependency, for each time period defined in Section 7.3 we calculate the ratio of the real average number of request during the time period to the average number of request when the total number of requests would have been evenly spread throughout the day (i.e., the expected number of arrivals). We refer to this ratio as the time period index (TPI). The TPIs for the morning, afternoon and evening are respectively 0.94, 0.96 and 1.10. We let $\lambda$ represent the arrival rate (request per hour). Using the TPIs, the average number of requests per hour for the three time periods would then be set as follows: $0.94\lambda$, $0.96\lambda$ and $1.10\lambda$.

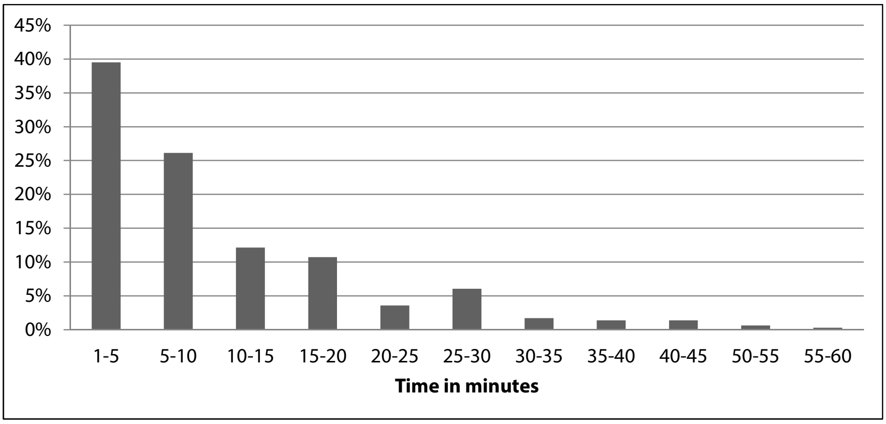


**Figure 7.2:** Distribution of care delivery times for scheduled care

**Table 7.3:** Time period index

| Time period | Average # arrivals | Expected # arrivals | TPI |
|---|---|---|---|
| 7:00-12:00 | 46.00 | 49.15 | 0.94 |
| 12:00-18:00 | 47.04 | 49.15 | 0.96 |
| 18:00-23:00 | 59.87 | 54.61 | 1.10 |

Based on observations in three different nursing homes, we estimate that in practice the average number of requests for unscheduled care is around 3 per hour per SSLF. However, as detailed insight into the number of request for the setting under study is lacking, the effect of different arrival rates will be examined (see Section 7.5).

To account for the observed additional variability in demand (i.e., over-dispersion) we model the average number of requests as a random variable using a Poisson-Gamma mixture model (see e.g., Jongbloed and Koole, 2001). Hence, we assume the demand for unscheduled care follows a Poisson process with random rate $\Lambda_t$, where $\Lambda_t \sim \text{Gamma}(\alpha_t, \beta_t)$. This mixture model can be considered as a negative binomial distribution with parameters $\alpha_t$ and $\frac{1}{1+\beta_t}$. The subscript $t$ emphasizes the dependence on the time period of the day (e.g., morning, afternoon and evening). See e.g., Cameron and Trivedi (2013) for a derivation of the negative binomial as a Poisson-Gamma mixture.

It remains to estimate parameters values $\alpha_t$ and $\beta_t$. In doing so, it suffices to solve the following set of linear equations:

Define

$A_t =$ estimated average number of arrivals during time period $t$

$I_t =$ observed rate of dispersion during time period $t$

$$\begin{cases} \frac{\alpha_t}{\beta_t} = A_t \\ \frac{1+\beta_t}{\beta_t} = I_t \end{cases} \Leftrightarrow \begin{cases} \alpha_t = \frac{A_t}{I_t - 1} \\ \beta_t = \frac{1}{I_t - 1} \end{cases} \tag{7.1}$$

where the first equation ensures that the average number of Poisson arrivals equals the number of arrivals observed, and, the second equation sets the appropriate rate of dispersion.

Next, the negative binomial distribution is fitted to the data. The values of the parameters are estimated using MLE. The unit of time is hours. Fig. 7.3, 7.4 and 7.5 visually show a good fit. Hence, despite the fact that for each of the time periods the fitted negative binomial distribution is rejected by the KS test ($\alpha = 0.05$), it seems practically useful to assume the number of arrivals are negative binomial distributed. The test results are presented in Table 7.4.

**Table 7.4:** Parameters and KS-statistic for the negative binomial distribution fitted with the number of arrivals per hour

| Period | Parameters | KS-value |
|---|---|---|
| Morning | $n = 26$, $p = 0.725$ | 0.1296 |
| Afternoon | $n = 15$, $p = 0.606$ | 0.1471 |
| Evening | $n = 40$, $p = 0.770$ | 0.1140 |

When it comes to the delivery times of unscheduled healthcare tasks, the analysis presented in Chapter 6 shows that the underlying distribution shows considerable variation. The statistical tests (see Table 6.2) and the plots presented in Figure 6.9 indicate that the hyperexponential distribution gives the best fit with the data.

Furthermore, in practice there is always a certain amount of time required for care workers to react to care requests and travel to the corresponding room. In Chapter 6 this is referred to as *travel time*. We model the travel time by incorporating a deterministic travel time between two successive tasks within the same SSLF and between two SSLFs.

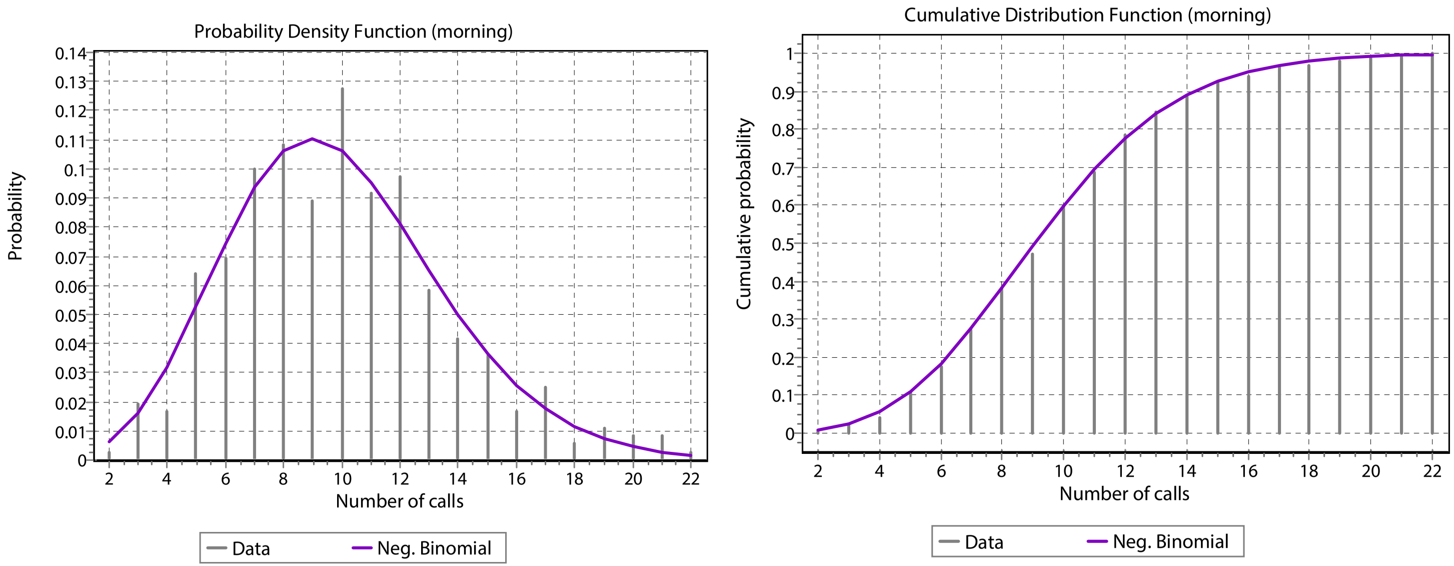


**Figure 7.3:** Probability density function (left) and cumulative distribution function (right) of the demand data for the morning fitted with a negative binomial distribution

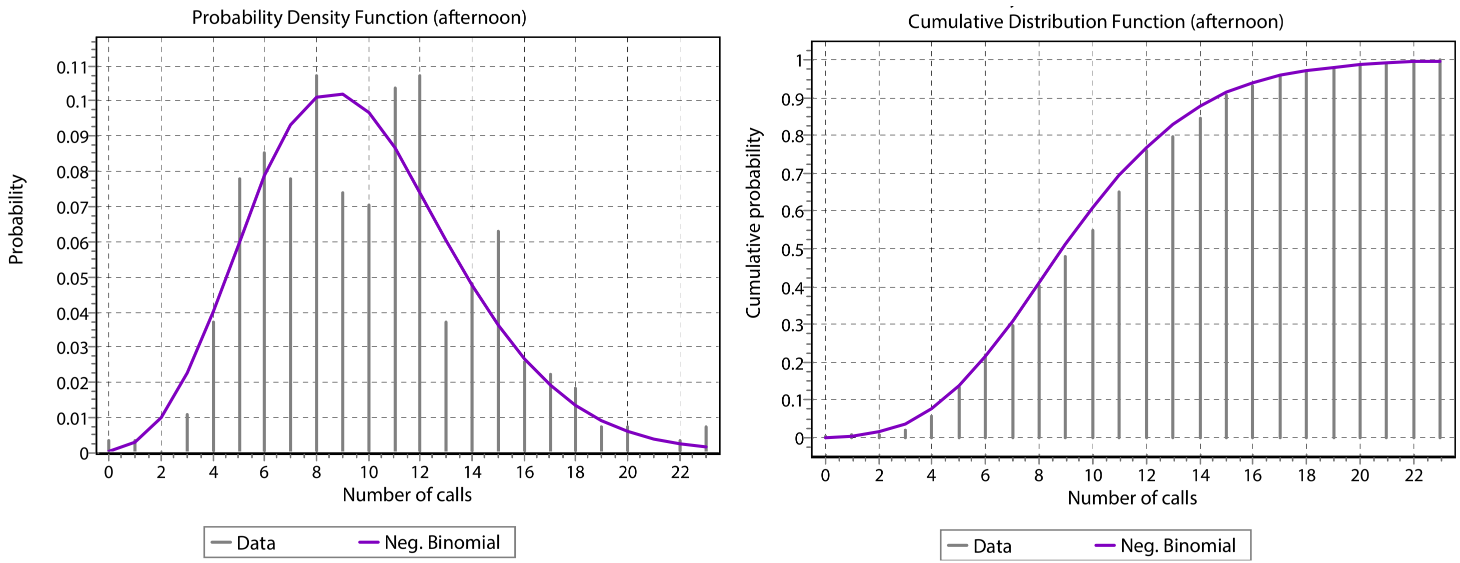


**Figure 7.4:** Probability density function (left) and cumulative distribution function (right) of the demand data for the afternoon fitted with a negative binomial distribution

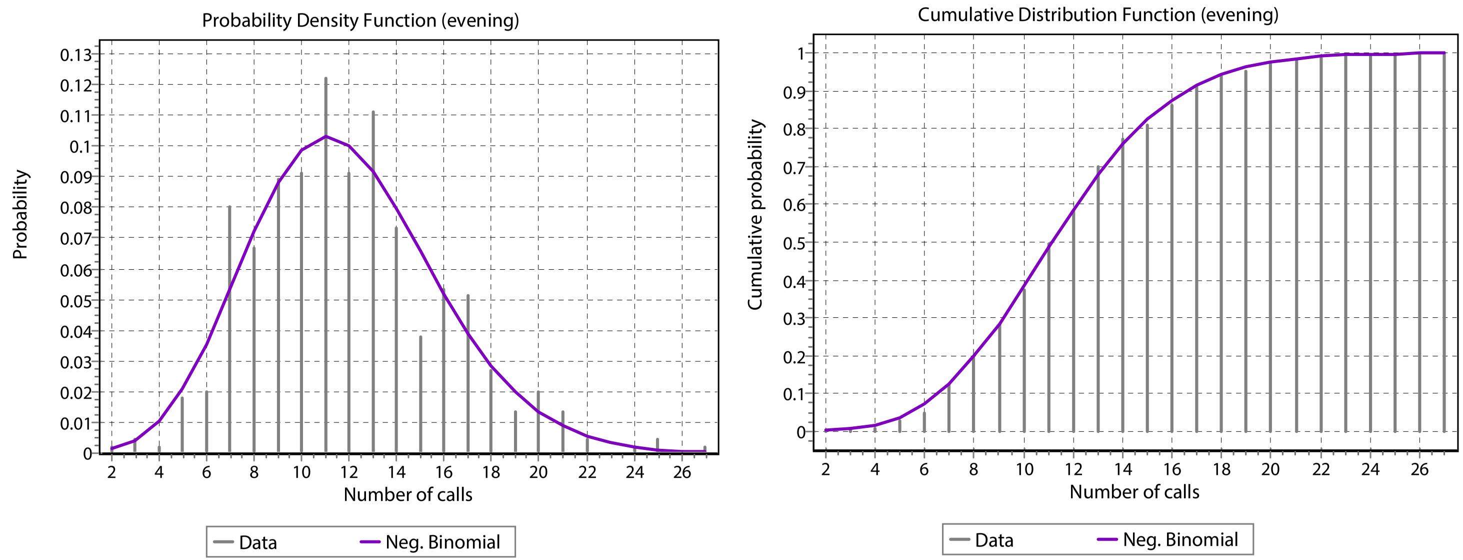


**Figure 7.5:** Probability density function (left) and cumulative distribution function (right) of the demand data for the evening fitted with a negative binomial distribution

### Staffing and task assignment

In line with the common characteristics of SSLFs mentioned in Section 7.2, we assume each care worker has the required qualifications to perform all tasks. Furthermore, a care worker can be assigned to either a single or to multiple SSLFs. With respect to the assignment of tasks the following strategies are distinguished:

*(1) First-come first-serve strategy* – The provision of care and support is on first-come first-serve (FCFS) basis. In this fully reactive approach, no distinction is made between scheduled and unscheduled care. A Python script was developed to simulate the FCFS approach. The FCFS procedure can be described as follows:

- Each day is divided into time buckets of 1 minute ($t = 1, 2, ..., 960$).
- Starting at 7:00 hours ($t = 1$), for every time step, healthcare tasks based on PATs are assigned to the first available care workers.
- When a client has a care request during the handling of an earlier PAT (i.e., an additional PAT), this additional request will be adhered to by the same care worker.
- When all care workers are busy, a virtual queue is filled with the remaining activities.
- PATs in the queue have priority over a 'new' PAT.

*(2) Task scheduling* – In this approach scheduled and unscheduled care activities are organized separately. Using the OERplanner algorithm (for more detailed info see Chapter 5), scheduled healthcare tasks are planned (if possible) within 15 minutes from the time preference of the client. In addition, for unscheduled healthcare care tasks (i.e, unscheduled care), the FCFS approach is used.

### Performance measures

To assess the performance of the scenarios presented in the next section, for each simulation run, we calculate the following performance measures:

- *Average waiting time at time $t$* – The average time a task has spent in the queue (for all tasks in the queue) at time $t$.

- *Total average waiting time* – The average waiting time over all $t$.
- *Queue length at time $t$* – The number of tasks waiting in the queue at time $t$.
- *15 minute service-level in time-interval $\Delta t$* – The % of PATs which originated in time-interval $\Delta t$ for which a care worker was present at the client within 15 minutes.
- *Overall service-level* – The average service-level over all $\Delta t$.
- *Occupancy rate at time $t$* – The number of care workers handling care care requests at time $t$ divided by the total number of care workers present at time $t$.
- *Total average occupancy rate* – The average occupancy rate over all $t$.

Subsequently, the performance measures are averaged over all simulation runs. As stated in the previous section, when a client has a new care request during the handling of a earlier PAT (i.e., an additional PAT), this additional request will be adhered to by the same care worker. We assume the waiting time for this additional PAT to be 0.

### Required number of simulation runs

Because of the random characteristics of the simulation model a certain number of runs is needed in order to obtain sufficient accurate results. To determine the required number of simulation runs, for each minute between 7:00 and 23:00 hours, we examine the attained confidence intervals progressively throughout the simulation. We continue to iterate until the predefined confidence interval has converged to an acceptable level of accuracy. This approach is referred to by Robinson (2014) as the graphical method. Using a 99% confidence interval, and a maximum acceptable confidence interval width of 2 minutes, it was found that a minimum number of 2000 simulation runs is required. Figure 7.7 shows the average waiting time during the course of the day for the base case over 2000 runs (the black line) and the corresponding confidence intervals (red dotted lines).

In the following section, we first present the specifics of the base case. Next, using the base case as reference, we examine the performance of alternative scenarios.

## 7.5 Scenarios

### Base case

The base case, which resembles a real-life SSLF-setting, can be described as follows. We consider a Dutch nursing home department consisting of 4 SSLFs. Each of the four SSLFs provides 24/7 care and support to six psychogeriatric clients in a small archetypical house-like facility. In the Netherlands the type and amount of care needed by a client is based upon a Care Intensity Package score (ZZP score). Each package is assigned a maximum tariff set by the Dutch Health Care Authority. On average, the clients in the department under study have an indication for ZZP 7, which can be described as a need for 'sheltered living with very intensive caring due to specific disorders with an emphasis upon guidance'.

During daytime (7:00-23:00 hours) each SSLF has one fixed care worker. In addition, from 7:30-11:30 hours two more care workers are available. These care workers are not dedicated to a single SSLF, but provide support to two SSLFs (i.e., flexible care workers). From 16:00-20:00 hours one extra flexible care worker is available who provides support to all four SSLFs. Details on the shifts are given in Table 7.5. See Figure 7.5 for a visualization of the allocation of the care workers. The provision of care and support services in each of the SSLFs is on FCFS basis. Hence, in the base case setting no distinction is made between scheduled and unscheduled care. When it comes to unscheduled care, we assume the arrivals to be negative bionomial distributed and the average number of requests to be around 3 per hour per SSLF. Using the TPIs of Table 7.3, the average number of requests per hour for the three time periods is set as follows: 2.8 (morning), 2.9 (afternoon) and 3.3 (evening).

Based on the findings presented in Chapter 6 we assume the care delivery times to be hyperexponentially distributed with $\hat{p}_1 = 0.10$, $\hat{p}_2 = 0.90$, $\hat{\mu}_1 = 0.11$, $\hat{\mu}_2 = 0.56$ (see also Table 6.2). These parameters can be interpreted as follows: with probability $\hat{p}_2 = 0.90$, the client has a minor request taken on average $1/\hat{\mu}_2 = 1.79$ minutes, whereas with probability $\hat{p}_1 = 0.10$ the client has a large request taking $1/\hat{\mu}_1 = 9.28$ minutes on average. In the base case we assume the travel time to be 1 minute between two successive tasks within the same SSLF and 2 minutes between two separate SSLFs.

The simulation results of the base case are presented in Figure 7.7 and Table 7.6. The results show that the total average waiting time per task is 5.73

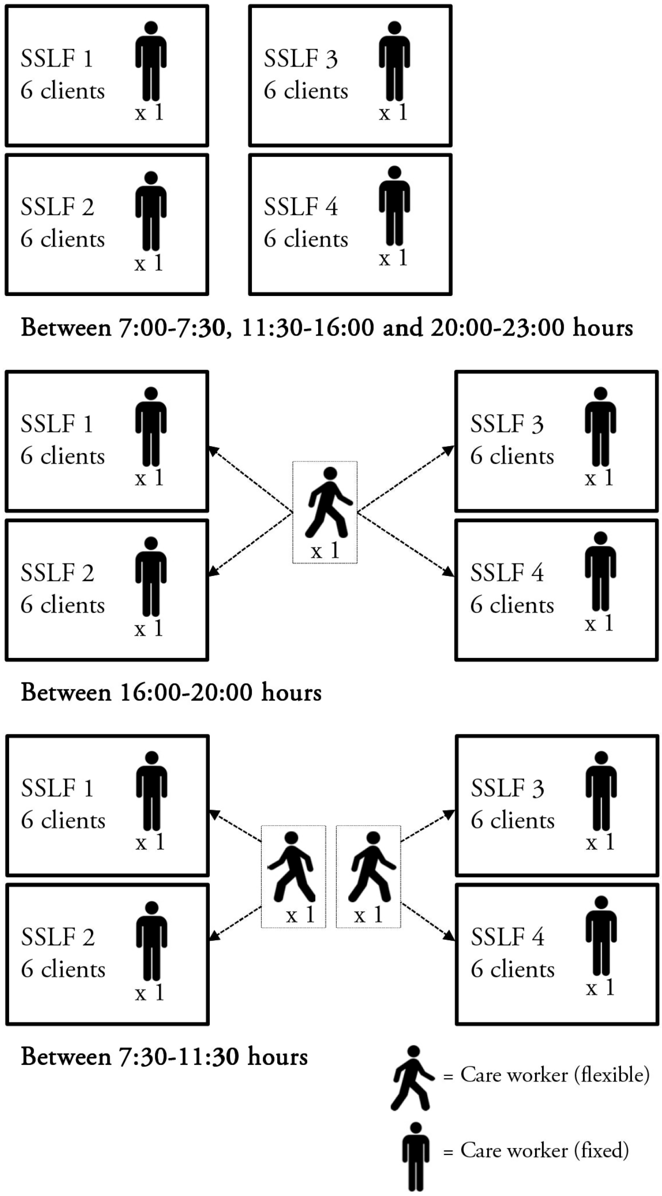


**Figure 7.6:** Allocation of care workers base case

**Table 7.5:** Shifts base case

| Shift | Time | Number of hours | Number of shifts | Total per 24 hours |
|---|---|---|---|---|
| Day | 7:00-15:00 | 8 | 4 | 32 |
| Morning | 7:30-11:30 | 4 | 2 | 8 |
| Afternoon/evening | 15:00-23:00 | 8 | 4 | 32 |
| Evening assistance | 16:00-20:00 | 4 | 1 | 4 |
| Total | | | | 76 |
| Hours/client | | | | 3.17 |

minutes. Furthermore, 88.35% of the healthcare tasks are provided within 15 minutes of the preferred delivery time. The total share of unscheduled tasks is 55% (i.e., the total share of scheduled task is 45%). And although the total average waiting time is less than 6 minutes, the average waiting time fluctuates considerably over the course of a day with average waiting times up to 20 minutes around 8:30 and 22:00 hours (see Figure 7.7 top). Most of the observed spikes in average waiting time can be explained by major activities in daily living (see also Chapter 4). Hence, during the morning, most residents need assistance with getting out of bed and/or dressing. Around lunch and dinner time, several residents need help with feeding. Finally, some of the residents need assistance with getting to bed in the evening. Furthermore, Figure 7.7(middle) shows striking peaks (i.e., long queues) around 12:00, 17:00 and 21:00 hours. These peaks are caused by clients who need their medicine. Despite the fact that two additional care workers are available during the morning, the occupancy rate (i.e., the workload) remains high during this period of the day (see Figure 7.7 bottom). This notion is supported by the findings presented in Chapter 5. The total average occupancy rate is 52.81%.

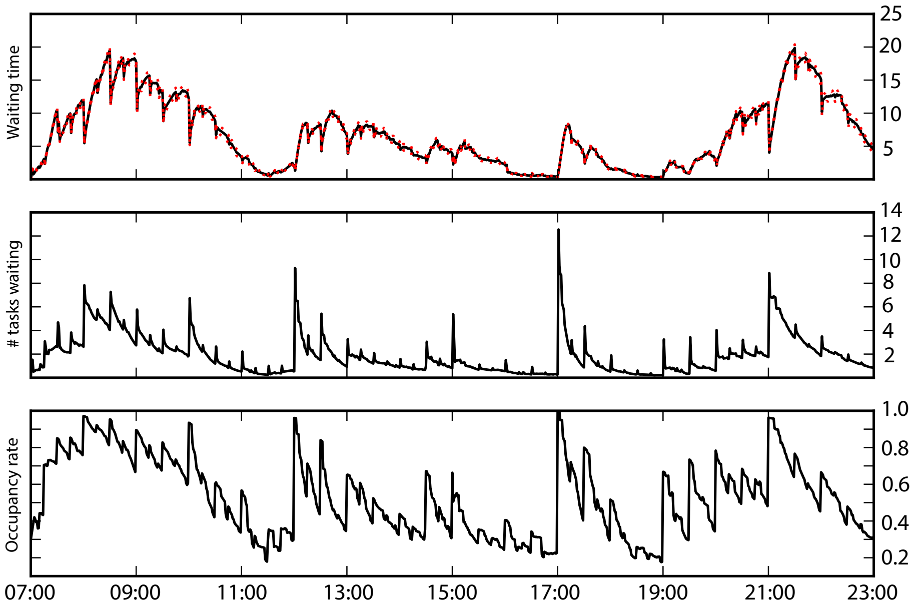


**Figure 7.7:** Simulation results base case. Average waiting time (top), average queue length (middle), average occupancy rate (bottom)

**Table 7.6:** Results base case: overall performance

| 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|
| 88.35% | 5.73 minutes | 52.81% |

### Alternative scenarios

With the base case scenario as point of departure, we explore the impact of a change in:

1. Demand for unscheduled care
2. Duration of care delivery for unscheduled care
3. Travel time
4. Allocation flexibility
5. Shifts
6. Staffing level
7. Number of clients
8. Allocation policy

#### Scenario 1: Demand for unscheduled care

In this scenario we examine the effects of changes in the assumptions on the arrival process of unscheduled care.

First, we look at the effect of changing the parameters under time-dependent negative binomial assumptions (i.e., the base case assumption). The top part of Table 7.7 provides an overview of the configurations. From the results (see Figure 7.8 and Table 7.8) it can be seen that an increase in the arrival rate of unscheduled care has only a moderate impact on the overall service-level. Hence, an increase in the average arrival rate from 3/hr to 6/hr results in a decrease in service-level of only 3.87%. An explanation for this observation would be that, compared to scheduled care, the demand for unscheduled care is more evenly spread throughout the day. Therefore, an increase in demand for unscheduled care puts relatively little additional pressure on the busy periods during the day. Furthermore, the average service time of unscheduled healthcare tasks is considerably less then that of scheduled healthcare tasks. This explanation seems plausible as Scenario 2 (see Figure 7.10) shows that an increase in the average care duration has substantial impact on the overall service-level. Furthermore, compared to the base case, Scenario 1c shows a 1.55 minute longer total average waiting time. Lastly, the results presented in Table 7.8 show that

**Table 7.7:** Configurations Scenario 1

| Time-dependent negative binomial arrivals | | | |
|---|---|---|---|
| | Rate | | |
| Scenario | Morning | Afternoon | Evening |
| Base case | 2.8/hr | 2.9/hr | 3.3/hr |
| 1a | 1.9/hr | 1.9/hr | 2.2/hr |
| 1b | 3.7/hr | 3.8/hr | 4.4/hr |
| 1c | 5.6/hr | 5.7/hr | 6.6/hr |
| **Homogeneous Poisson arrivals** | | | |
| Scenario | Rate | | |
| 1d | 2/hr | | |
| 1e | 3/hr | | |
| 1f | 4/hr | | |
| 1g | 6/hr | | |

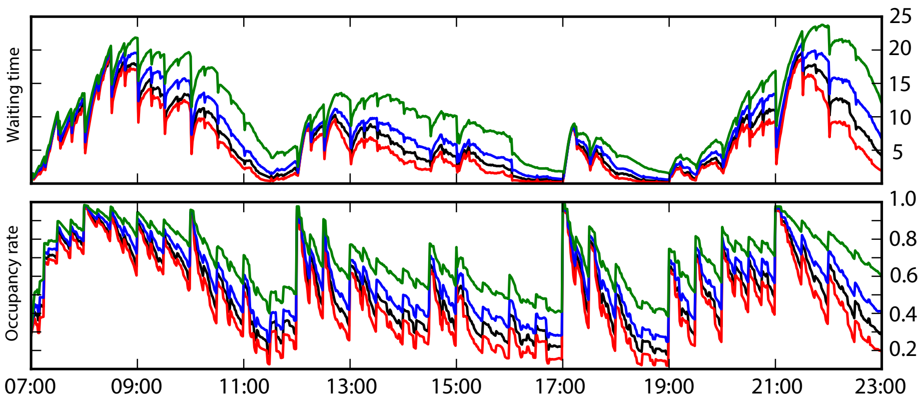


**Figure 7.8:** Simulation results Scenario 1. The plots show the average waiting time (top) and average occupancy rate (bottom) during the day. Base case (black), Scenario 1a (red), 1b (blue) and 1c (green)

an increase in the arrival rate of unscheduled care has a substantial impact on the total average occupancy rate.

Next, we examine the performance under the assumption of time-homogeneous Poisson arrivals. The bottom part of Table 7.7 shows the configurations. As can be seen from the results presented in Figure 7.9 and Table 7.9, the performance differs only slightly from the performance under negative

**Table 7.8:** Overall performance Scenario 1a, b and c

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.38% | 5.72 minutes | 52.42% |
| 1a | 88.99% | 5.47 minutes | 46.60% |
| 1b | 87.31% | 6.20 minutes | 58.53% |
| 1c | 84.96% | 7.27 minutes | 69.37% |

binomial assumptions.

Hence, when it comes to the arrival process, the overall service-level and total average waiting time are relatively insensitive to the assumptions explored.

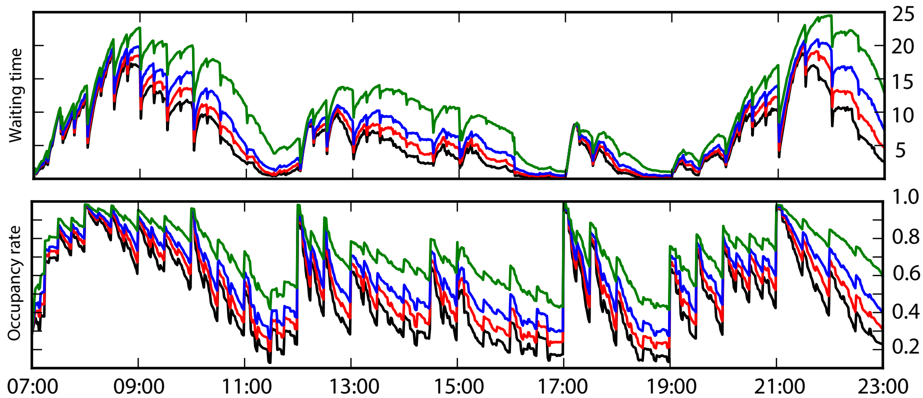


**Figure 7.9:** Simulation results Scenario 1.The plots show the average waiting time (top) and average occupancy rate (bottom) during the day. Scenario 1d (black), 1e (red), 1f (blue) and 1g (green)

**Table 7.9:** Overall performance Scenario 1d, e and f

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| 1d | 89.16% | 5.36 minutes | 48.23% |
| 1e | 88.74% | 5.56 minutes | 54.65% |
| 1f | 87.95% | 5.90 minutes | 60.60% |
| 1g | 85.67% | 6.92 minutes | 71.95% |

### Scenario 2: Duration of care delivery for unscheduled care

In this scenario we assess the effect of less fluctuation in the care delivery times for unscheduled care by assuming exponentially distributed care delivery times. In addition, we examine the impact of an increase in the care duration for unscheduled care activities. The different configurations are presented in Figure 7.10.

**Table 7.10:** Configurations Scenario 2

| | | |
|---|---|---|
| Base case | Hyperexponential | $\hat{p}_1 = 0.10$, $\hat{p}_2 = 0.90$, $\hat{\mu}_1 = 0.11$, $\hat{\mu}_2 = 0.56$ |

| Scenario | Average care duration in minutes |
|---|---|
| 2a | 3 |
| 2b | 4 |
| 2c | 6 |
| 2d | 8 |

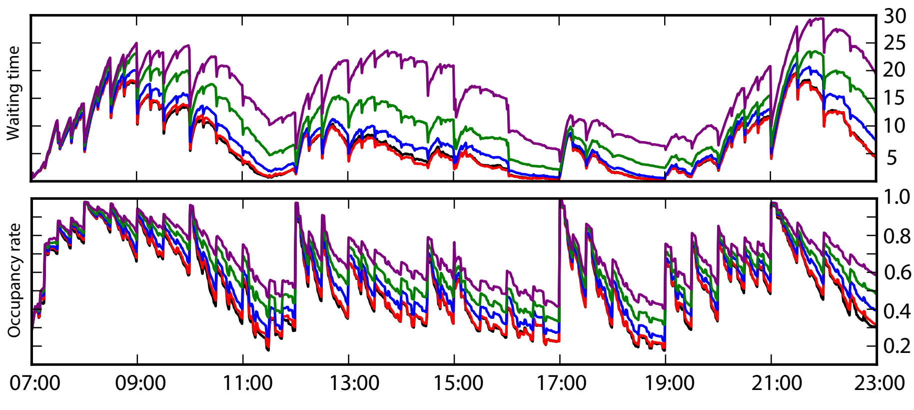


**Figure 7.10:** Simulation results Scenario 2. The plots show the average waiting time (top) and average occupancy rate (bottom) for base case (black), 2a (red), 2b (blue) and 2c (green), 2d (purple)

The results show that the performance of Scenario 2a closely resembles the performance obtained for the base case. Furthermore, a change in the average care duration for unscheduled care from 3 to 8 minutes results in an increase of the total average waiting time of almost 6 minutes. The 15 minute service-

**Table 7.11:** Overall performance Scenario 2

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.30% | 5.74 minutes | 52.63% |
| 2a | 88.04% | 5.87 minutes | 53.60% |
| 2b | 85.84% | 6.76 minutes | 57.78% |
| 2c | 81.85% | 8.68 minutes | 63.09% |
| 2d | 77.69% | 11.54 minutes | 69.26% |

level drops from 87.90% to 77.64%. Hence, it can be concluded that the overall performance is relatively sensitive to changes in the average care duration of unscheduled care.

**Scenario 3: Travel time**

To assess the impact of travel time on the simulation results, multiple configurations are examined (see figures 7.11 and 7.12). We vary both the travel time within and between SSLFs. The travel times are kept deterministic.

**Table 7.12:** Configurations Scenario 3

| Scenario | Travel time in minutes within a SSLF |
|---|---|
| Base case | 1 |
| 3a | 0 |
| 3b | 2 |
| 3c | 4 |
| **Scenario** | **Travel time in minutes between SSLFs** |
| Base case | 2 |
| 3d | 0 |
| 3e | 4 |
| 3f | 6 |

Figures 7.11 and 7.12 show that the performance is more sensitive to a change in travel time within SSLFs than between SSLFs. This can be explained by the fact that most of the care workers are allocated to a specific SSLF (i.e., fixed care workers). Consequently, a change in travel time between two suc-

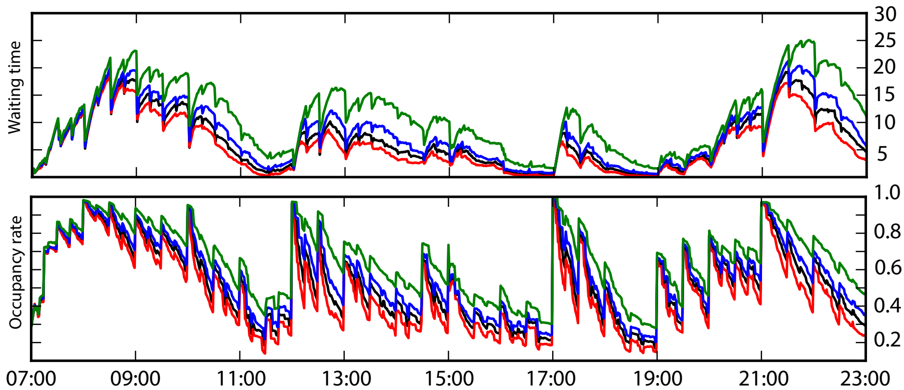


**Figure 7.11:** Simulation results Scenario 3. The plots show the average waiting time (top) and average occupancy rate (bottom) for base case (black), 3a (red), 3b (blue) and 3c (green)

**Table 7.13:** Overall performance Scenario 3a, b and c

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.36% | 5.72 minutes | 52.57% |
| 3a | 90.56% | 4.27 minutes | 47.53% |
| 3b | 85.68% | 7.17 minutes | 56.84% |
| 3c | 77.67% | 10.31 minutes | 65.06% |

**Table 7.14:** Overall performance Scenario 3d, e and f

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.40% | 5.68 minutes | 52.25% |
| 3d | 89.20% | 5.33 minutes | 51.29% |
| 3e | 87.53% | 6.08 minutes | 53.36% |
| 3f | 86.60% | 6.51 minutes | 54.61% |

cessive tasks within the same SSLF has a relative large impact on the overall performance.

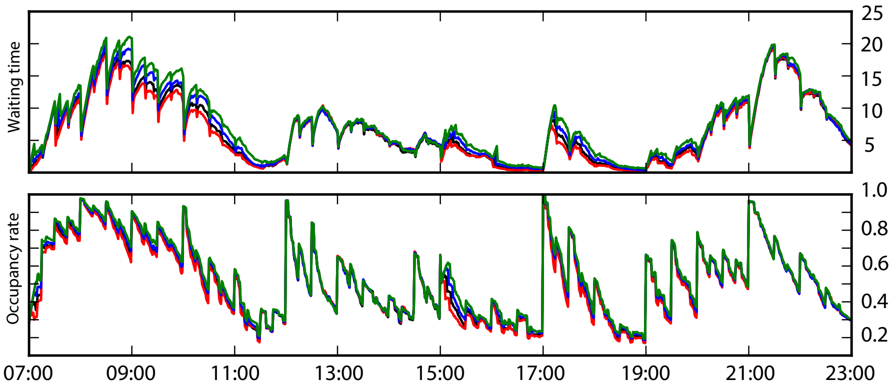


**Figure 7.12:** Simulation results Scenario 3. The plots show the average waiting time (top) and average occupancy rate (bottom) for base case (black), 3d (red), 3e (blue) and 3f (green)

### Scenario 4: Allocation flexibility

In this scenario we examine the impact of an increase in the allocation flexibility of care workers. Figure 7.13 shows the examined configurations and the corresponding results. It can be observed that especially an increase in allocation flexibility of fixed care workers has a substantial positive effect on the overall performance. Making fixed care workers fully flexible, results in a total average waiting time of 28.07% below the base case and a 6.67% improvement in service-level. In addition, the total average occupancy rate increases with 8.40%. This increase in the overall performance due to more flexibility is in line with the findings presented in Chapter 5.

**Table 7.15:** Configurations Scenario 4

| Scenario | Allocation |
|---|---|
| 4a | Fixed care workers provide support to two SSLFs. |
| 4b | Fixed care workers provide support to all four SSLFs. |
| 4c | All flexible care workers provide support to all four SSLFs. |

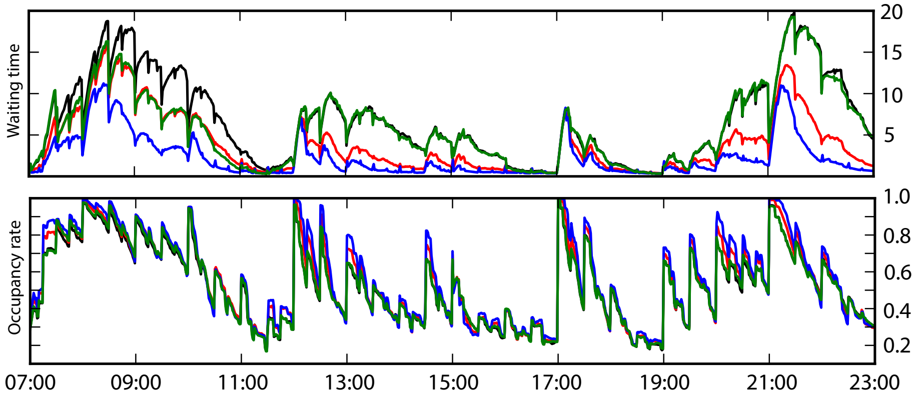


**Figure 7.13:** Simulation results Scenario 4. The plots show the average waiting time (top) and average occupancy rate (bottom) for base case (black), 4a (red), 4b (blue) and 4c (green)

**Table 7.16:** Overall performance Scenario 4

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.39% | 5.70 minutes | 52.38% |
| 4a | 91.57% | 4.83 minutes | 55.02% |
| 4b | 94.29% | 4.10 minutes | 56.78% |
| 4c | 88.67% | 5.52 minutes | 52.78% |

### Scenario 5: Shifts

To examine the effect of a change in the shifts of care workers we look at the following two scenarios: (5a) creating an additional morning shift from 7:30-11:30 hours at the expense of the current evening assistance, (5b) changing the evening assistance from 16:00-20:00 to 18:00-22:00 hours. These scenarios are chosen because, in the base case, the workload is relatively high between 7:30-11:30 and 18:00-22:00 hours. For details on the scenarios, see Table 7.17.

The impact of a change in shifts on the overall performance is very small with a maximum reduction of the total average waiting time of less than 15 seconds and an increase in the overall service-level of only 0.45% (see Table 7.18).

**Table 7.17:** Configurations Scenario 5

| Scenario 5a | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Morning assistance | 7:30-11:30 | 2 | 2 | 8 |
| Morning assistance | 7:30-11:30 | 4 | 1 | 4 |
| Total | | | | 76 |
| Hours/client | | | | 3.17 |

| Scenario 5b | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Morning assistance | 7:30-11:30 | 2 | 2 | 8 |
| Evening assistance | 18:00-22:00 | 4 | 1 | 4 |
| Total | | | | 76 |
| Hours/client | | | | 3.17 |

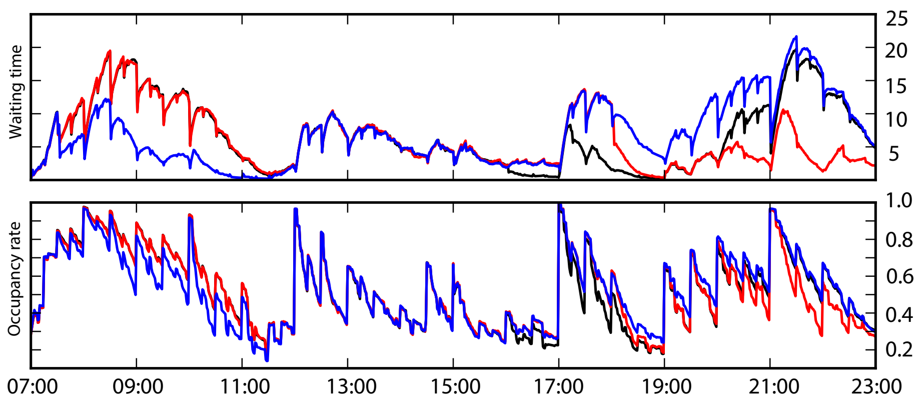


**Figure 7.14:** Simulation results Scenario 5. The plots show the average waiting time (top) and average occupancy rate (bottom) for base case (black), 5a (blue) , 5b (red)

**Table 7.18:** Overall performance Scenario 5

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.30% | 5.75 minutes | 52.59% |
| 5a | 88.70% | 5.52 minutes | 52.54% |
| 5b | 88.41% | 5.69 minutes | 52.68% |

### Scenario 6: Staffing level

In this scenario we look at the effect of an increase in the number of care workers. Table 7.19 shows the configuration of the two scenarios we examined. From the results presented in Table 7.20 it appears that an increase in the number of care workers has a considerable impact on the overall performance. Under both scenarios, the overall service-level largely exceeds 95%, with a total average waiting time of around 3 minutes. Compared to the base case, the maximum average waiting time during the day drops from 20 to around 12 minutes. However, the occupancy rate declines strongly to 46.09% for Scenario 6a and 40.71% for Scenario 6b.

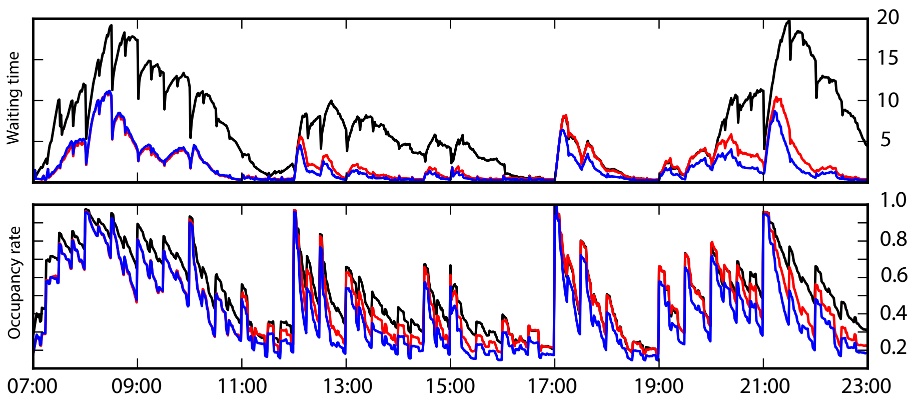


**Figure 7.15:** Simulation results Scenario 6. The plots show the average waiting time (top) and average occupancy rate (bottom) for base case (black), 6a (red) and 6b (blue)

**Table 7.19:** Configurations Scenario 6

| Scenario 6a | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Morning assistance | 7:00-11:00 | 2 | 2 | 8 |
| Morning assistance | 7:00-11:00 | 4 | 1 | 4 |
| Afternoon assistance | 11:00-19:00 | 4 | 1 | 8 |
| Evening Assistance | 19:00-23:00 | 4 | 1 | 4 |
| Total | | | | 88 |
| Hours/client | | | | 3.67 |

| Scenario 6b | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Morning assisance | 7:00-11:00 | 2 | 2 | 8 |
| Morning assistance | 7:00-11:00 | 4 | 1 | 4 |
| Afternoon assistance | 11:00-19:00 | 2 | 2 | 16 |
| Evening assistance | 19:00-23:00 | 2 | 2 | 8 |
| Total | | | | 100 |
| Hours/client | | | | 4.17 |

**Table 7.20:** Overall performance Scenario 6

| Scenario | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|
| Base case | 88.40% | 5.74 minutes | 52.59% |
| 6a | 95.72% | 3.15 minutes | 46.09% |
| 6b | 96.94% | 2.62 minutes | 40.71% |

#### Scenario 7: Number of clients

To assess the impact of more clients per SSLF under constant average costs, three alternative scenarios are examined. Table 7.22 shows the shifts for each scenario. The results (see Figure 7.16 and Table 7.21) show that a change in the number of clients per SSLF has a substantial impact on the performance. Reducing the number of clients per SSLF from 6 to 5 leads to a 49.13% increase in the total average waiting time. In addition, the top plot in Figure 7.16 shows that reducing the number of clients from 6 to 5 results in average waiting times up to 40 minutes around 9:30 and 11:00 hours. With 8 clients per SSLF the total average waiting drops below 5 minutes with a 15 minute service-level of almost 92%. The maximum average waiting time drops to around 13 minutes.

**Table 7.21:** Overall performance Scenario 7

| Scenario | Number of clients | 15 minute service-level | Total average waiting time | Total average occupancy rate |
|---|---|---|---|---|
| Base case | 6 | 88.35% | 5.73 minutes | 52.81% |
| 7a | 5 | 84.28% | 8.62 minutes | 55.41% |
| 7b | 7 | 89.98% | 5.19 minutes | 51.57% |
| 7c | 8 | 92.88% | 4.15 minutes | 49.29% |

**Table 7.22:** Configurations Scenario 7

| Scenario 7a<br>5 clients | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total hr per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Total | | | | 64 |
| Hours/client | | | | 3.20 |

| Scenario 7b<br>7 clients | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Morning asssistance | 7:30-11:30 | 2 | 4 | 16 |
| Evening assistance | 16:00-20:00 | 4 | 2 | 8 |
| Total | | | | 88 |
| Hours/client | | | | 3.14 |

| Scenario 7c<br>8 clients | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 1 | 4 | 32 |
| Evening | 15:00-23:00 | 1 | 4 | 32 |
| Morning assistance | 7:30-11:30 | 2 | 4 | 16 |
| Morning assistance | 7:30-11:30 | 4 | 1 | 4 |
| Evening assistance | 16:00-20:00 | 4 | 2 | 8 |
| Evening assistance | 19:00-23:00 | 4 | 2 | 8 |
| Total | | | | 100 |
| Hours/client | | | | 3.13 |

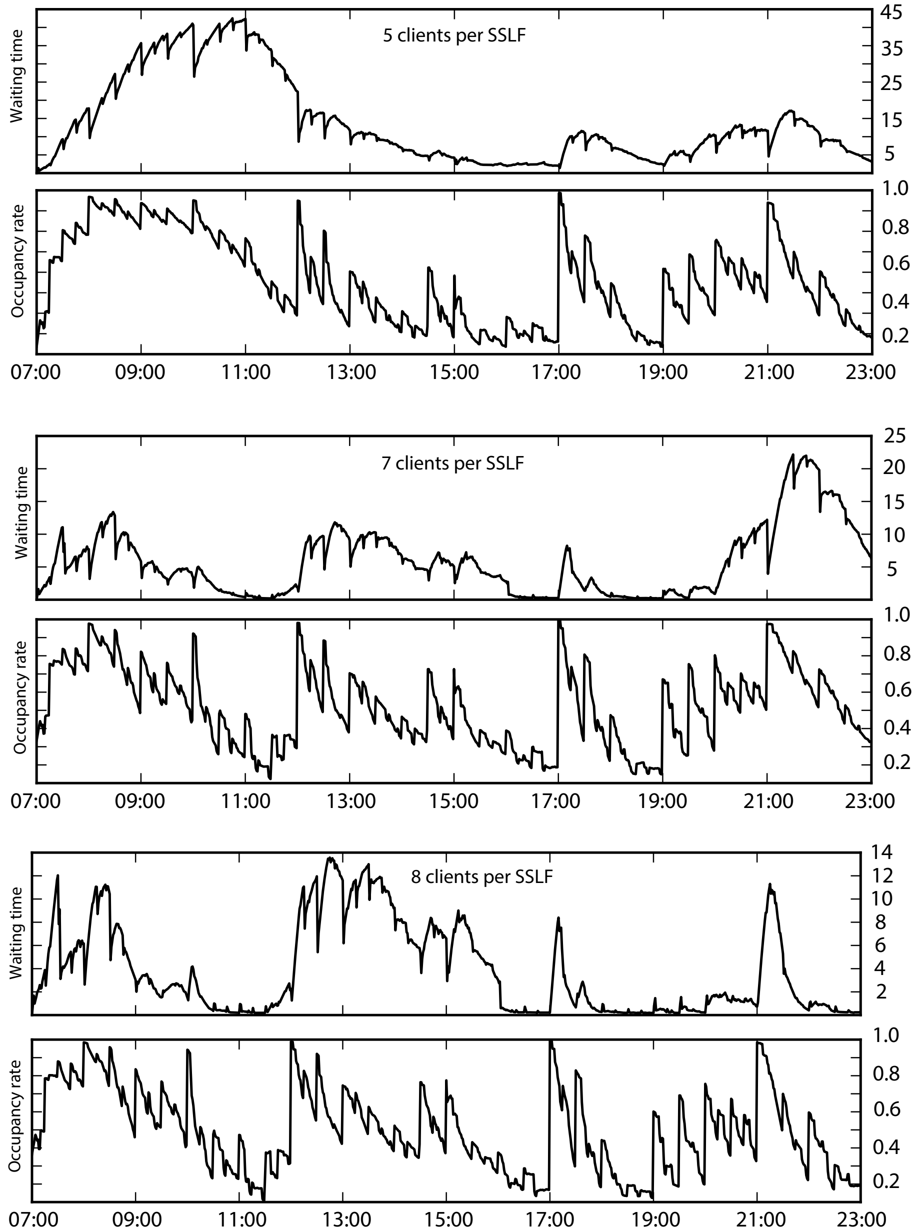


**Figure 7.16:** Simulation results Scenario 7: The plots show the average waiting time and average occupancy rate during the day

#### Scenario 8: Allocation policy

In Scenario 8a we examine the effect of organizing scheduled and unscheduled care separately. This approach allows to increase the demand-side flexibility of scheduled care by using time windows during which the activities are supposed to be carried out. More specifically, using the OERplanner algorithm, scheduled healthcare tasks are planned (if possible) within 15 minutes from the time preference of the client. In addition, for unscheduled healthcare care tasks, the FCFS approach is used. See Chapter 5 for more detailed information on the OERplanner algorithm. In this scenario, the supply of care and support is organized on a larger scale, i.e., care workers are assigned to multiple SSLFs. During day-time (7:00-23:00 hours), four care workers are assigned to handle scheduled healthcare tasks according to a predetermined schedule. These predetermined schedules are also referred to as 'care routes'. In addition, between 7:00-23:00 hours, two care workers are available for the provision of unscheduled care. These care workers are fully flexible as they can handle tasks for all of the four SSLFs.

Scenario 8a is compared with a FCFS approach in which, between 7:00-23:00 hours, there are four fixed care workers. These care workers have the flexibility to handle tasks of two SSLFs. Additionally, there are two care workers available between 7:00-23:00 hours. These care workers can handle tasks of all four SSLFs. This scenario is referred to as Scenario 8b. See Table 7.23 for the configurations of both scenarios.

See Table 7.24 for the 15 minute service-level of the two scenarios under study. The results show that Scenario 8b outperforms Scenario 8a in terms of service-level. This can be explained by the fact that the allocation flexibility is reduced when scheduled and unscheduled tasks are organized separately. However, the advantage of working with a predetermined schedule is that it provides a more clear structure and coordination in the provision of care and support.

Table 7.23: Configurations Scenario 8

| Scenario 8a | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Sched. care day | 7:00-15:00 | care routes | 4 | 32 |
| Sched. care evening | 15:00-23:00 | care routes | 4 | 32 |
| Unsched. care morning | 7:00-11:00 | 4 | 2 | 8 |
| Unsched. care afternoon | 11:00-19:00 | 4 | 2 | 16 |
| Unsched. care evening | 19:00-23:00 | 4 | 2 | 8 |
| Total | | | | 96 |
| Hours/client | | | | 4.00 |

| Scenario 8b | | | | |
|---|---|---|---|---|
| Shift | Time hours | Allocation to # of SSLFs | Number of shifts | Total per 24 hrs |
| Day | 7:00-15:00 | 2 | 4 | 32 |
| Evening | 15:00-23:00 | 2 | 4 | 32 |
| Morning assisance | 7:00-11:00 | 4 | 2 | 8 |
| Afternoon assistance | 11:00-19:00 | 4 | 2 | 16 |
| Evening assistance | 19:00-23:00 | 4 | 2 | 8 |
| Total | | | | 96 |
| Hours/client | | | | 4.00 |

Table 7.24: Results Scenario 8

| Scenario | 15 minute service-level | | |
|---|---|---|---|
| | Overall | Scheduled care | Unscheduled care |
| Base case | 88.36% | - | - |
| 8a | 90.79% | 89.59% | 91.75% |
| 8b | 96.48 % | - | - |

## 7.6 Service-level-based staffing

In other service industries such as call-centers, staffing decisions are often based on minimizing staffing costs while maintaining an acceptable level of service, where level of service refers to the responsiveness to the customers' demand (e.g., Koole and Mandelbaum, 2002). When it comes to call-center performance management, service-level is often defined as the percentage of calls ($X\%$) answered within $Y$ seconds. In this section we use the following context specific definition: *the % of PATs in time-interval* $\Delta t$ *for which a care worker was present at the client within 15 minutes.* Although, in the base case the overall service-level is 88.35%, the average service-level fluctuates considerably during the course of the day (see Figure 7.17). Around 7:30-8:00 and 20:30-21:00 hours a serious drop in the average service-level can be observed. This example shows that basing staffing decisions on averages without taking variability into account can be dangerous, something which is also referred to as "the flaw of averages" (Savage, 2002). To maintain an acceptable level of service throughout the day, call-centers often base their measurements and staffing decisions on half-hour intervals.

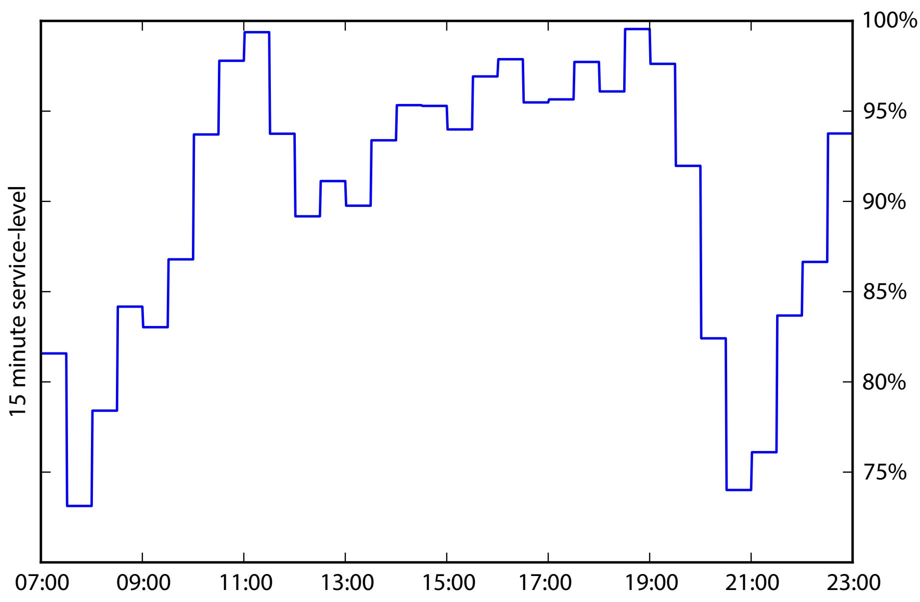


**Figure 7.17:** Base case: 15 minute service-level during the day

In line with this reasoning, we consider the problem of minimizing the total number of care workers while maintaining an acceptable level of service throughout the day. The average service-level goal is set at 95/15. Hence, for 95% of the PATs in each half-hour time-interval a care worker should be at the client within 15 minutes. Furthermore, we assume each of the care workers to be fully flexible. That is, they can handle tasks for all of the four SSLFs. We use the simulation tool to determine how many care workers are needed in each time-interval to obtain the required service-level. Starting at the first time-interval (7:00-7:30 hours) we determine the required number of care workers and step by step include the next following time interval.

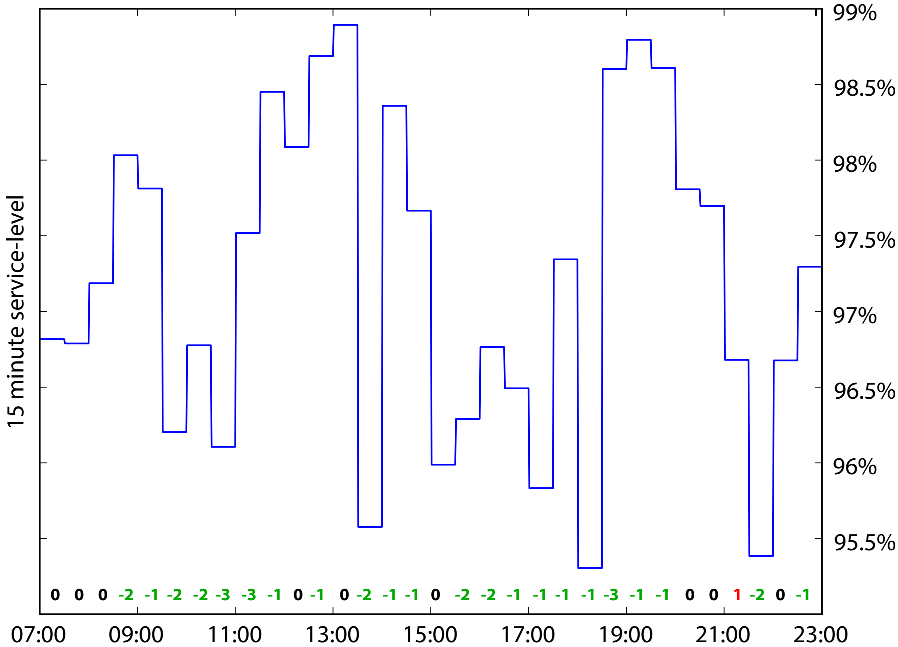


**Figure 7.18:** Service-level based staffing policy: 15 minute SL during the day

Figure 7.18 shows the average service-level during the day when applying service-level based staffing. The overall average service-level and total average occupancy rate under this staffing policy are respectively 96.38 and 61.18%. On the top side of the x-axis, for each half-hour interval, the required number of care workers is compared with the base case. A total of 59 working hours is needed to achieve a 95/15 service-level in each half-hour interval, which is 22% less than in the base case. The reasons for this are (1) the increased allocation flexibility of the care workers (see also Scenario 4) and (2) a reduction of overstaffing during the less busy periods of the day.

## 7.7 Conclusions and discussion

In this chapter we analyzed the performance of small-scale living facilities (SSLFs), focusing on the problem of meeting the time preferences of residents regarding the delivery of care and support. The case of SSLFs is interesting as small-scale care is under increasing financial pressure due to budget cuts. A simulation model is developed which resembles the current care delivery process of a Dutch nursing home department with four SSLFs. The model is used to examine the performance under various scenarios.

With a total average waiting time of 5.73 minutes and a 15 minute service-level of 88.35%, the overall performance under the base case is beyond our expectations. Nonetheless, the average waiting time fluctuates considerably over the course of a day, with average waiting times up to 20 minutes around 8:30 and 22:00 hours. When it comes to the arrival process, the results show that the overall performance is relatively insensitive to an increase in the average arrival rate of unscheduled care. In addition, the performance under time-homogeneous Poisson arrivals differs only slightly from the performance under negative binomial assumptions. With regard to the average care duration, the results show that, the overall performance is relative sensitive to changes. This is an interesting finding as under current Dutch healthcare reforms, the care needs of nursing home clients are likely to become more severe (see Chapter 1). The results also indicate, that an increase in allocation flexibility of fixed care workers has a substantial positive effect on the overall performance, which is in line with the findings presented in Chapter 5. As such, when meeting the time preferences of clients is considered to be an important QoL factor, the SSLF principle of working with a small fixed team of care workers is open to question. With regard to a change in shifts, the results show that the impact on the overall performance is very small with a maximum reduction of the total average waiting time of less than 15 seconds and an increase in service-level of only 0.45%. In contrary, an increase in the number of care workers has a large impact on the average waiting times. An additional shift during the morning results in a drop of the maximum average waiting time during the day from 20 to around 12 minutes. Furthermore, the results also show that a change in the number of clients per SSLF under constant average costs also has a substantial impact on the performance. A reduction in the number of clients per SSLF from 6 to 5 leads to a 49.13% increase in the total average waiting time. With 8 clients per SSLF the total average waiting drops below 5 minutes with a 15 minute service-level of almost 92%. In the last scenario we examined the effect of organizing scheduled and unscheduled care separately, whereas dedicated workers are assigned to handle scheduled healthcare tasks according to a predetermined schedule. The results of this scenario show that it does not lead to better performance in terms of service-level compared to a fully FCFS policy, in which no distinction is made between scheduled and unscheduled healthcare tasks. The main reason for this is that the allocation flexibility is reduced when scheduled and unscheduled tasks are organized separately. However, in practice it would not be realistic to make all care workers fully flexible as it would most likely lead to a lack of overview and stressful situations. Hence, an important advantage of working with a predetermined schedule for unscheduled care is that it provides a more clear structure and coordination in the provision of care and support.

In Section 7.6 we considered the problem of minimizing the total number of care workers, while maintaining a 95/15 service-level throughout the day. For each half-hour interval we determined how many care workers are needed to obtain the required service-level. We did this under the assumption that care workers can handle tasks for all of the four SSLFs. The results show that allocation flexibility together with numerical flexibility has a substantial positive effect on the performance. In practice however, a service-level-based staffing approach is not fully applicable. For example, the collective labor agreement restricts the way in which personnel can be deployed. Nevertheless, we do feel there is considerable scope for improvement when it comes to increasing flexibility. In this respect, nursing homes could learn from other service industries.

In conclusion, we can say that the shifts presented in the base case provide a good point of departure. In addition, to further improve the performance in the short run the focus should lie on increasing (1) the allocation flexibility of care workers and (2) the number of clients per SSLF.

This study also shows that simulation is useful for assessing and improving daily nursing home operations. The presented simulation model is an important step to better understand the real-life performance of SSLFs and the underlying relationships involved. It provides a basis for building a decision support tool for nursing home managers. An important next step is building an interface to allow nursing home managers to use the simulation tool in daily practice.