Reducing waiting times in acute elderly care: a case study for the region of Twente

Author: C.G. van Loon

VU supervisor: Dr. R. Bekker

CWI supervisor: Prof. Dr. R.D. van der Mei

CWI co-supervisor: I. Aydemir

Second reader: Prof. Dr. B. Rink

Centrum Wiskunde & Informatica

Vrije Universiteit Amsterdam

July, 2024

Abstract

This thesis explores the reduction of waiting times and redundant hospital admissions at Intermediate Care services for elderly patients. This is done using a discrete event simulation (DES) model for the Twente region. The current practice of admission was explored after conversations and reading literature. Concluding three types of care arrive Low Complex, High Complex and Geriatric Rehabilitation care and are treated. From there, several interventions were implemented in the model. Key interventions evaluated include bed-sharing, centralization, prioritization, admission hours, and triage wards. Results indicate that a 40% partial bed-share between GR and High Complex care reduces the average waiting time by 3.8 hours. This is 7.8% of the total waiting time and 93.2% of the possible reduction by sharing all GR and High Complex beds. Centralizing services across multiple locations during high-traffic periods shows potential benefits, though further refinement of the allocation model is needed. Prioritizing patients from Emergency Departments or hospitals had minimal impact. Accelerating the admission process and extending admission hours to weekends significantly reduced waiting times. Opening Intermediate Care 24/7 for admission and neglecting all transfer times brings the average waiting time approximately to 0. Overall, targeted interventions can enhance Intermediate Care efficiency, especially since bed-sharing and flexible admission hours were proven effective. Further research is recommended to optimize these strategies.

Contents

1 Introduction

1.1 Problem description

1.1.1 Waiting times and Intermediate Care

Waiting times for acute elderly care are a severe problem in the Dutch healthcare system. In October 2023, 22.218 elderly people were waiting to be placed in a nursing home. Apart from that, 11.835 patients are waiting on a precautionary waiting list, despite there being a place available for patients in crises. However, it is becoming increasingly difficult to get a place in a nursing home of own choice. Also, a big problem within the healthcare system for elderly care is a staffing shortage. The combination of not getting preferred allocation and staffing shortage gives a big queue for healthcare for the elderly [1].

Prolonged hospital stays caused by delayed transfers of care lead to adverse events and high costs. Therefore Intermediate Care or ELV (Eerstelijns Verblijf in Dutch) is implemented by the government of the Netherlands. This is bed-based care for general health problems that do not require admission to the hospital but cannot be treated at home. This bed-based care aims to enable older adults to return home and live independently in the community [28].

After setting up the "wijkkliniek" (earlier version of Intermediate Care) successfully in Amsterdam for acute elderly care a rollout was planned for multiple regions following the Dolce Vita project, including the area of Twente. Multiple regions are considered in the Dolce Vita project, which all require a different modelling approach. So for example the model for the region of Twente will be different to the Intermediate Care model of Amsterdam. This thesis will focus on the region of Twente as a part of the Dolce Vita project, which is discussed in the next subsection, where after that the research question this thesis answers is posed.

1.1.2 Dolce Vita

This master thesis will focus on a part of the Dolce Vita Project. Dolce Vita stands for, “Data-Driven Optimization for a Vital Elderly Care System in the Netherlands”. Project Dolce Vita has been running since 2019. This project is run by the CWI, VU and Amsterdam UMC and aims to reduce waiting times in elderly care in the Netherlands [9]. The project’s objective is to map out elderly care as a chain, combining mathematical and medical knowledge. The chain is defined as a smooth operation where all components between the arrival process of the patients, and the service (treatment) they get are considered, including the transitions between working stations (where they get service). The goal of the chain is to map out the process of a patient from having an accident or illness, going to the hospital, until leaving the hospital as detailed as possible [28]. The Dolce Vita project considers different regions, all over the Netherlands, with varying structures of their Intermediate Care, which requires different modelling approaches.

1.2 Region Twente

As said before, this master’s thesis will focus on the Dutch region of Twente. Three different healthcare providers are considered, namely TMZ, ZA and Carint, shown in Figure 1. These are healthcare providers delivering Intermediate Care for elderly patients at multiple locations.

For the region Twente, two groups of care are present:

- High Complex, this is multi-disciplinary care. Designed for patients with significant medical, nursing, and rehabilitation needs. These patients often have multiple co-morbidities or complex conditions that require intensive and specialized care.
- Low Complex, this is mono-disciplinary care. It is intended for patients who need ongoing medical and rehabilitative support but at a lower intensity than High Complex care. These patients are generally more stable and have fewer medical complexities.

Other types of care are:

- Geriatric Rehabilitation (GR), focuses on patients who primarily need rehabilitation services to regain independence and functionality after an illness or injury. It is less intensive than High Complex care and more focused on rehab.
- Respite Care, provides temporary support to informal caregivers, helping to relieve their caregiving duties and replenish their ability to manage these responsibilities.

Examples of High Complex care are the treatment of patients who are recovering from a severe illness or major operation. These patients need a multi-disciplinary team and intensive monitoring. Low Complex care is for example patients recovering from a minor surgery or patients with stable chronic conditions needing some support. Lastly, GR care is for example for patients recovering from strokes or fractures or patients needing physical therapy after prolonged hospital stays. Respite care is a small amount of all arrivals at Intermediate Care and has the shortest length of stay.

Admission process In the Twente region, three types of care arrive at the Intermediate Care. Namely High Complex care, Geriatric Rehabilitation and Low Complex care. Due to practical importance, Respite care and Low Complex care are combined into Low Complex care. After visiting the General Practitioner or Emergency Department or having a hospital admission, patients can either go straight to Intermediate Care, wait from home or wait in the hospital. When patients wait in the hospital, this is known as a "redundant" hospital admission because these patients occupy a hospital bed, which can also be used for future arriving patients.

Currently, patients are assigned to Intermediate Care based on their caretype. So High Complex patients, Geriatric Rehabilitation patients and Low Complex patients have dedicated beds. Among these caretypes, no beds are shared. The advantages of this policy are limited training for healthcare employees and more specialistic treatment can be delivered. A significant disadvantage of this policy is that if one care type's bed group is fully occupied, patients needing that type of care cannot be admitted, even if other beds with adequate equipment are available.

Also, Intermediate Care has limited admission hours. According to regional experts, Intermediate Care is available between 8 A.m. and 5 P.m. on weekdays and closed during weekends. Also, the process of admitting a patient has some delays. These so-called admission days or transfer times are considered as extra waiting time between admitting a patient and giving the promised care.

Whenever a patient from the Emergency Department (EMD) arrives outside the admission hours, emergency beds are available at the hospital for those patients to be admitted.

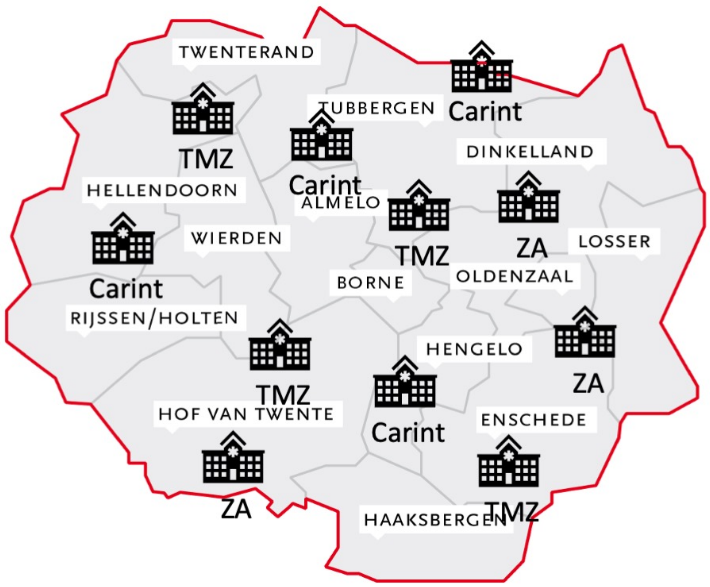

Figure 1: Overview of Intermediate Care locations in the region of Twente [17].

1.3 Research question

The research question is stated as follows:

"How can the organization of Intermediate Care of elderly patients, for the region of Twente, be adjusted such that waiting times and hospital admissions are reduced and the patient flow improves?"

The following subquestions answer this question:

1. How can the current admission policy and treatment at Intermediate Care be modelled?

2. How can changes in the model, called interventions, improve waiting times and patient flow at Intermediate Care?

The first subquestion involves defining the process of admitting acute elderly patients to Intermediate Care within a model, including the classification of care types, identification of outflow destinations, and specification of workstations where patients receive treatment. The second subquestion involves implementing various interventions in the model, such as bed-sharing among different care types, and analyzing how these changes affect the model's performance.

Plan of approach To answer the subquestions, which in turn answer the research question. The following steps will be taken:

1. First, consult experts in the field of Intermediate Care. Based on their insights, a visual model will be created to describe the possible pathways for elderly patients to Intermediate Care. Then, the current admission policy for Intermediate Care is lined out.
2. Model a DES (Discrete Event Simulation) model of the current admission policy. Using this policy and the DES model, simulations determine the waiting times, expected queue length and other metrics. Also, a sensitivity analysis is done to check what increasing or decreasing arrivals or service times do to the model.
3. Interventions. The current policy will be expanded with different changes in logistics, called interventions. The following interventions are considered.
    (a) Bed-sharing, in this intervention, it is checked whether sharing beds among caretypes influences the system. At this moment multiple sharing types are considered.
    (b) Centralization and allocation, as said in the region of Twente, the Intermediate Care locations are divided among the region Twente. This intervention looks at combining these locations or centralizing the Intermediate Care. Also, this intervention will look at the challenges of decentralized Intermediate Care. When multiple locations are considered, patients might have a preference for where they will be admitted. This intervention will also look at how the allocation of patients can be modelled.
    (c) Small interventions, small changes in the model. Three are present.
        i. The triage ward (TRW), previously known as triage beds or observation beds, was used for elderly patients whose care needs were uncertain, potentially requiring High Complex care, GR care, or possibly not Intermediate Care at all. This intervention examines the impact of reintroducing observation beds into the admission policy.
        ii. Priority, this intervention will look at prioritizing certain caretypes when a patient is admitted to Intermediate Care. For example, patients from the Emergency Department are in higher need of treatment so might be prioritized.
        iii. Admission hours, Intermediate Care is not open 24/7 for new admissions, which means patients might experience waiting time because of opening hours. This intervention looks at flexible opening hours and improvements in the admission process.
4. Verification and validation involve two key steps. First, the results obtained in the previous step are validated and verified against the literature, specifically using the Erlang-C model. Then, it is assessed whether the conclusions drawn from the research are valid, ensuring the model operates correctly and consistently handles all edge cases.
5. Conclusion and discussion, a conclusion and discussion will discuss the results and optional improvements needed.
6. Dashboard, to conclude the research a dashboard has been made to show outcomes for different sets of parameters. For further research in the region of Twente, this dashboard can be used.

1.4 Research guide

This research starts with some preliminary knowledge of queueing theory in section 2. After that, in section 3, literature on the subject is discussed and a hypothesis is stated. Next, section 4 describes the model and in section 5 methodology is described. At last, in section 6 results are shown, with a conclusion and discussion on further research in sections 7 and 8.

2 Preliminaries

In this section, some preliminary knowledge is discussed. This knowledge is about some basic queueing theory and its corresponding distributions. For this theory, the Poisson and exponential distribution are discussed. This section will show the mathematical notation which will flow into an explanation in words later in this research.

2.1 Poisson distribution in queueing models

This section will study the Poisson process and some properties thereof. It is shown how this random process naturally can be used to model simple queues. Start with a simple derivation of the Poisson process. The reason is two faults. First, it is from a mathematical point of view interesting to see how the Poisson process is derived. Second, and most importantly, it gives a better understanding of which assumptions are needed in this process and how these assumptions fit into the framework of this thesis. The approach is similar to [14].

Suppose the incoming patients in a hospital up to and including time t, with $t \in [0, \infty)$ is modelled. Let N_t represent the number of patients arriving in the time interval $[0, t]$. It is assumed that the random process $(N_t)_{t \in [0,\infty)}$ satisfies the following five conditions:

(i) N_t is a random variable taking values in $\{0, 1, 2, \ldots\}$,

(ii) $N_0 = 0$,

(iii) If $0 \leq s \leq t$, then $N_s \leq N_t$,

(iv) If $0 \leq s \leq t$, we assume that the number of patients which arrive during the time interval $(s, t]$ is independent of the number of arrivals before time s. This is called the independence property,

(v) We assume that there exists a number $\lambda > 0$ such that for small h we have:

$$\mathbb{P}(N_{t+h} = n + 1 | N_t = n) = \lambda h + o(h) \tag{1}$$

and

$$\mathbb{P}(N_{t+h} = n | N_t = n) = 1 - \lambda h + o(h). \tag{2}$$

Which are also implied by properties (iii) and (iv). The parameter λ is called the arrival rate.

The property (v) implies that the probability that a patient arrives in the hospital in a short interval $(t, t + h)$ is approximately a linear function of h. Also, (1) and (2) implies that

$$\mathbb{P}(N_{t+h} \geq n + 2 | N_t = n) = o(h).$$

From this, the probability that two patients enter the hospital during the interval $(t, t + h)$ is negligibly small. So in the model, two possible events are relevant; the event that no patients arrive in the interval $(t, t + h)$, or exactly one patient comes in this interval. This assumption can be questioned in the context of urgent hospital arrivals because, for instance, multiple victims may arrive simultaneously from an accident. This contradicts the assumption of exactly one arrival per time interval.

Now, set

$$T_i = \inf\{t \geq 0 : N_t = i\},$$

i.e. T_i is the first time in which i patients are arrived in the hospital. For this reason, call $T_1, T_2, \ldots$ the arrival times. If T_i is known, then N_t is given by

$$N_t = \max\{n \in \mathbb{N} : T_n \leq t\}.$$

A sketch of the graph of N_t together with four arrival times is displayed in Figure 2.

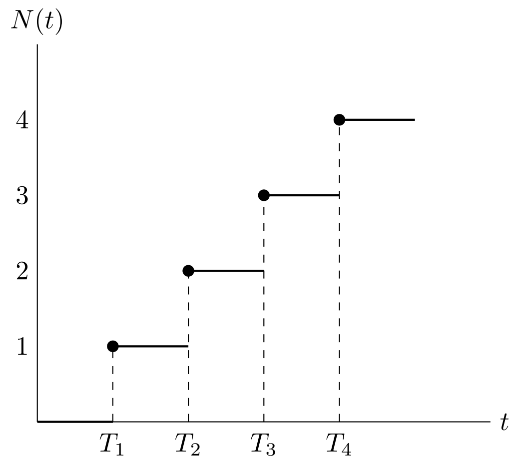

Figure 2: A sketch of the graph of N_t.

Below we indicate that the random process N_t has the Poisson distribution with parameter λt. By the "partition theorem" and using (1) and (2), after some algebra it can be seen that for $k \geq 1$,

$$\mathbb{P}(N_{t+h} = k) = \lambda h \mathbb{P}(N_t = k-1) + (1 - \lambda h)\mathbb{P}(N_t = k) + o(h).$$

Set $p_k(t) = \mathbb{P}(N_t = k)$, dividing both sides by h and letting $h \downarrow 0$ we get for $k = 0$

$$p_0'(t) = -\lambda p_0(t) \tag{3}$$

and for $k \geq 1$ we obtain

$$p_k'(t) = \lambda p_{k-1}(t) - \lambda p_k(t). \tag{4}$$

The boundary condition $N_0 = 0$ implies that $p_k(0)$ equals 1 if $k = 0$ and 0 if $k \neq 0$. An induction argument shows that

$$\mathbb{P}(N_t = k) = p_k(t) = \frac{1}{k!}(\lambda t)^k e^{-\lambda t}$$

satisfies (3) and (4). To conclude that N_t has the Poisson distribution with parameter λt.

Now consider the inter-arrival times X_1, $X_2 \ldots$, which are the times between two consecutive arrivals, i.e.

$$X_i = T_i - T_{i-1}, \quad i = 1, 2, \ldots.$$

A well-known property of the inter-arrival times $X_1, X_2, \ldots$ of a Poisson process with arrival rate λ, which follows an exponential distribution with parameter λ, see also [14]. The exponential distribution satisfies a so-called lack-of-memory property, i.e. for every $u, v \geq 0$

$$\mathbb{P}(X > u + v | X > u) = \mathbb{P}(X > v).$$

The memoryless property implies that the probability of an event occurring in the future is independent of the time that has already elapsed.

2.2 Exponential distribution in queueing models

In queueing theory, the exponential distribution is commonly used to model the service times in a queueing system. It describes the probability density of the time taken to complete a service for each entity in the queue. The exponential distribution is particularly suitable in single-server or multi-server queueing systems with memoryless service times.

The exponential distribution is characterized by its probability density function (PDF), given by:

$$f(y) = \begin{cases} \mu e^{-\mu y}, & \text{if } y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

μ Represents the service rate per unit of time.

Consider a scenario where entities arrive at a service facility and are served one at a time. Let Y be the random variable representing the service time required for each entity. Assume that service times follow an exponential distribution with parameter μ, representing the average service rate (or mean service rate) per unit of time. To derive the probability that an entity's service time is within a certain range, integrate the PDF over that range. Specifically, the probability that the service time Y is less than or equal to a given time t is given by [24]:

$$P(Y \leq t) = \int_0^t \mu e^{-\mu y}\, dy = 1 - e^{-\mu t}.$$

The probability of an entity's service time being exactly t is infinitesimally small. Therefore, the probability of observing exactly t service time is:

$$P(Y = t) = 0.$$

Literature by Shanthikumar et al. (2007) [25] states that when more servers are considered, the distribution of the service times will have a modest impact on the expected waiting time(s). When considering one, the variability in service times will have more impact on waiting time(s). When the only server is occupied for too long because of a high variation in service times, it holds up the system. When more servers are considered, the patients can get service at another server and skip the patient that blocks the busy server. So for this research, because multiple servers are considered, the distribution of service time(s) is expected to not impact the average waiting time too much. However, when the load per server is high and multiple servers are considered, the distribution of the service time(s) does matter. For now the service times are considered to have an exponential distribution.

2.3 Little's law

Little's Law is a fundamental principle in queuing theory that shows a connection between three key elements in a system: the average number of entities present (denoted by L), the arrival rate of those entities (λ), and the average time an entity spends in the system (W). Mathematically, it is expressed as:

$$L = \lambda \times W$$

It can be practically applied by:

- In service industries like customer support or retail, businesses can use Little's Law to optimize staffing levels and reduce wait times by understanding how customer arrival rates and service times affect the number of customers in the system.
- Little's Law is also valuable in computer science for analyzing system performance. It helps in tasks like optimizing server resources or improving network efficiency by providing a quantitative framework to evaluate system behaviour.
- In manufacturing and supply chain management, Little's Law aids in optimizing production processes and inventory management. By studying the flow of materials or products through a system, companies can identify bottlenecks and streamline operations to meet customer demands more effectively.

This law serves as a relationship between units in queuing theory, providing a framework for understanding and optimizing the flow of patients through the system. By analyzing this relationship, optimal strategies can be identified for resource allocation, balancing workload across various units, and minimising waiting times. This optimization involves adjusting staffing levels, bed allocation, and patient triage procedures to ensure that resources are used efficiently and patient care is delivered promptly [13].

2.4 Queuing models and Erlang C

2.4.1 Origin

Erlang C model and M/M/c queues have their roots in the work of A.K. Erlang, a Danish mathematician and engineer, who developed queueing theory in the early 20th century. Erlang's work laid the foundation for analyzing the performance of telecommunication systems, particularly in terms of call centre performance and telephone traffic engineering. The Erlang C model is a formula developed by A.K. Erlang to estimate the probability of a call being delayed or queued in a call centre system with a specific number of agents (c) and a given traffic intensity (ρ) [25]. The model assumes Poisson arrivals and exponentially distributed service times. It is widely used in call centre management to determine staffing levels and assess service quality. Following a Kendall notation, the Erlang C model can be described as an M/M/c queue. M/M/c queues are a fundamental type of queueing system characterized by having Poisson arrival processes, exponential service times, and multiple servers [15]. Numerous studies have explored the behaviour and properties of M/M/c queues, including their steady-state probabilities, mean waiting times, and throughput.

2.4.2 Queuing models

In queueing models, the exponential distribution and the Poisson distribution are used to model the service times and arrivals of entities in various systems, including, following the Kendall notation:

- M/M/1 Queues: Markovian arrival and service processes with one server.
- M/M/c Queues: Markovian arrival and service processes with multiple (c) servers.

The exponential distribution is a crucial assumption for defining the number of customers as a 1-dimensional Markov chain. This makes analyzing, such as average waiting time, system utilization, and queue length distributions possible [15].

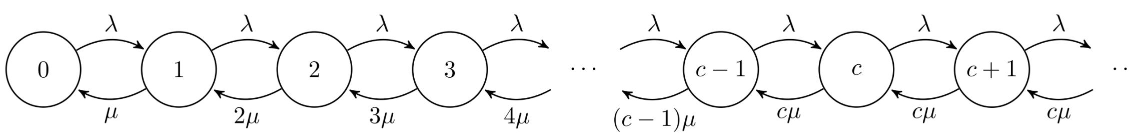

Figure 3: Transition diagram M/M/c queue

Consider an M/M/c queueing system with c servers. Figure 3, shows the transition diagram. It can be seen that patients arrive with rate λ. Because exponential service times are assumed, the framework as the Figure shows can be used for modelling the queue. Let L denote the number of entities in the system, and L_q denote the number of entities in the queue. Aim to derive the steady-state probabilities π_n of having n entities in the system. After solving the balance equations the following steady-state probabilities are derived.

$$\rho = \frac{a}{c} < 1$$

with

$$a = \frac{\lambda}{\mu}$$

$$\pi_j = \begin{cases} \frac{a^j}{j!}\pi_0 & \text{if } j \leq c \\ \frac{a^c}{c!}\left(\frac{a}{c}\right)^{j-c}\pi_0 & \text{otherwise} \end{cases}$$

with

$$\pi_0 = \left[\sum_{j=0}^{c-1}\left(\frac{a^j}{j!}\right) + \frac{a^c}{c!}\frac{1}{1-\rho}\right]$$

Where ρ represents the load per server. With arrival rate λ, service rate μ and c servers. The load per server cannot be greater than one for the system's stability. For an explanation of the assumption of identically distributed and independent LOS of patients, see [34] and [7], which states that this seems an appropriate assumption as long as the patient mix and medical practice do not change.

2.4.3 Application

In the M/M/c queueing model, Little's Law and the derived steady-state probabilities relate the average number of entities in the system (L) to the arrival rate (λ) and the average time entities spend in the system (W). Use Little's Law and the steady state probabilities to derive various M/M/c queue performance metrics.

- Average Queue Length (L_q) The average number of entities waiting in the queue (L_q) can be derived using Little's Law as follows:

  $$L_q = \lambda \cdot W_q$$

  where W_q is the average time spent by an entity waiting in the queue. This can be calculated using π_j. Using this, W_q can be calculated using Little's Law.

- Probability of delay

  The probability that all servers are busy, and an arriving entity has to wait or be lost. Call that $C(c, a)$, with c servers.

  $$C(c, a) = \sum_{j=c}^{\infty} \pi_j = \frac{1}{1-\rho} \frac{a^c}{c!} \pi_0$$

- Average Response Time

  The average response time represents the average time an entity spends in the system. It can be derived using Little's Law as follows:

  $$\text{Response Time} = \frac{L}{\lambda}$$

These metrics help in evaluating the performance of the M/M/c queueing system and are essential for system design, capacity planning, and optimization.

This comprehensive analysis provides a deeper understanding of the behaviour and efficiency of the M/M/c queue, enabling stakeholders to make informed decisions to improve system performance and customer satisfaction [25].

Findings Lots of research has been done for M/M/c queues. for this research, the most important finding is found by Vilaplana et al. (2013) [29] and Figure 4. This Figure shows when more servers are considered, more load per server is needed to make the waiting times increase. The Figure also explains why sharing can be effective because when more servers are considered, an extended load can be accepted per server. This is an important finding and something to reconsider when verifying the models used in this research.

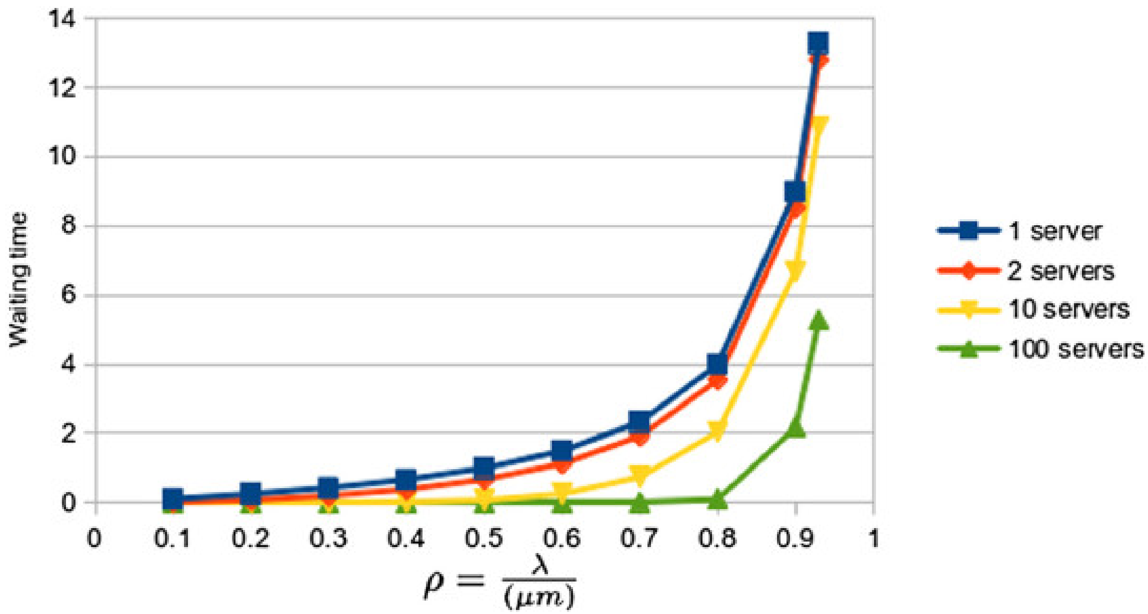

Figure 4: Result from Vilaplana et al. (2013) showing the load per server and waiting times

3 Literature review and hypothesis

This section provides essential background knowledge relevant to this research. It begins by summarizing previous research on Intermediate Care. Following this, it discusses literature related to studies on resource sharing and allocation, which is relevant to this research. Based on this collective knowledge, a hypothesis is formulated.

3.1 Intermediate Care

Melis et al. (2004) [22] describe multiple definitions of Intermediate Care globally. This paper states three definitions of Intermediate Care. The most relevant of them is from the British Geriatrics Society which states that Intermediate Care is defined as "The range of services designed to facilitate the transition from hospital to home, and from medical dependence to functional independence, where the objectives of care are not primarily medical, the patient's discharge destination is anticipated and a clinical outcome of recovery (or restoration of health) is desired." This definition is mostly the same as defined in the Netherlands. The paper of Melis, confirmed by other papers [27], states that the definition of Intermediate Care is difficult to standardize.

Steiner et al. (2001) [27] researched whether Intermediate Care is helpful. This paper does that by describing three types of Intermediate Care models. These are admission avoidance, post-acute care and either pre-acute or post-acute. admission avoidance wants to avoid unnecessary admissions in hospitals by for example assigning patients to GP nursing home beds. Post-acute care looks more at discharge schemes and supports targeting discharge to people who require medical or nursing intervention in emergencies. Research in this model concludes that Intermediate Care may be a good thing, but the question is what is needed to get the most out of Intermediate Care. This paper also references the high potential of Intermediate Care in hospitals.

Looking at the benefits of Intermediate Care, Vincent et al.(2015) [30] considered whether Intermediate Care can reduce costs. In this research, multiple economic researchers were considered to conclude that opening the Intermediate Care had a significant influence on the mortality of patients, mainly because of a decrease in the number of deaths in general medical wards. Also, other research in this paper reported that Intermediate Care helped prevent 'premature' discharges to the ward. The presence is also linked with a decrease in adjusted hospital mortality for adults admitted to Intermediate Care, notably, those admitted for full intensive care therapy rather than just monitoring. In this paper, it is also discussed whether Intermediate Care influences costs made, so more from an economic view than clinical opinions. One concluded that there is insufficient data available to support that Intermediate Care and general ward beds are cost-effective. Although some researchers state that Intermediate Care can be cost-efficient, it is neglected due to a lack of data available or practical issues such as a greater severity of illness in the patients admitted after the introduction of Intermediate Care and longer stays at the hospital not declared by the introduction of Intermediate Care itself. So concluding there is no evidence of the advantages of Intermediate Care. But more things need to be considered to conclude. Most important is that there is a need for careful patient triage and enough Intermediate Care beds need to be available for those who need them.

Intermediate Care is not only a facility in the Netherlands, but also in England Intermediate Care is available for the elderly who need care, but are good enough to stay out of the hospital. Young et al. (2009) [33] wrote about the developments of Intermediate Care in England. in England HaH (hospital at home) is the biggest IC model. Other models are day hospitals, Nursery-led units, Community hospitals and Short-term care. This research concludes that Intermediate Care has not filled its potential in England. There has been an increase of 7.8% in acute admissions since Intermediate Care is a health-service policy. Intermediate Care in England has also great challenges, namely changing staff members' skills and attitudes and engaging General Practitioners in the process. HaH is concluded to be the best random controlled trial supportive Intermediate Care service model. This research thus shows that Intermediate Care is not supposed to always be efficient.

3.2 Research and theory

3.2.1 Earlier research on Intermediate Care simulations

Arntzen et al. (2023) [5] showed a simulation model for Intermediate Care for the region of Amsterdam, which is similar research as this thesis. The research proposed different policies for reducing redundant and costly hospital admissions. This study concluded that the waiting times are not due to a lack of capacity or policy for the Amsterdam area, but due to admission days and admission hours. This research proposed sharing beds and changing admission hours. Also, observation beds are considered.

3.2.2 Sharing

Research by Palvannan et al. (2010) [23], Bekker et al. (2016) [7] and Adan et al. (2001) [2] showed that sharing can improve queues. In Bekker et al. (2016), several sharing techniques are used modelled by a theoretical queueing model. This paper also gives motivation for assuming a Poisson arrival process, which is widely accepted for urgent patients. This is confirmed by research performed by Worthington et al. (1987) [32] and Young et al. (1965) [34]. The paper also mentions the advantages and disadvantages of sharing types, which are also used for this thesis. Flexibility is the largest advantage for full sharing (All patients share all beds, in this thesis for High Complex care with GR care). Major disadvantages are e.g. a lack of specialization and the inability to prioritize patient types for example. In this paper, partial sharing (Earmarking) has the most advantages. This is supported by research by Smith and With et al. (1981) [26], they are the first to mention that full flexibility of resource sharing is not always beneficial. They state that it is not helpful to combine systems with different service time distributions. Mandelbaum et al. (1998) [20] showed that for light traffic, pooling is always useful, but for heavy traffic, pooling can have multiple effects, positive and negative.

Following Bekker's research, which assumes patients can be blocked from admission, the blocking probability can be reduced by a small amount of sharing. The same applies to other metrics. So the main message of this paper was that a little flexibility is generally sufficient. This is supported by research from Adan et al. and Palvannan et al. The literature also states that disadvantages of sharing in healthcare are also present. Namely, that sharing requires the medical staff to treat multiple caretypes, which may require costly additional training effort. Research by Huckman et al. (2008) [16] shows that focusing on a limited range of tasks improves efficiency so a little sharing can work best.

3.2.3 Square-Root-Staffing

Also in multiple papers, Bekker et al. (2016) [7] and Janssen et al. (2011) [18], the Square-Root-Staffing is discussed. This guidance shows how many beds should be allocated to each care unit. This allocation is a good starting point but also has some difficulties. Namely, it is dependent on some specific local conditions. Recalling ρ_i is the load for unit i, the capacity should roughly be.

$$s_i = b_i + \alpha\sqrt{b_i}$$

The first term b_i reflects the offered load. The second term $\alpha\sqrt{b_i}$ represents the safety capacity. α represents the safety factor, which is now set to 1. Literature states that when α is set to 2, the 95% confidence interval is met. [18] The Square-Root-Staffing principle originates from the high-traffic scenarios.

3.2.4 Cenralization and allocation

Centralization (1 location) In high-traffic scenarios, consolidating all servers into a single pooled resource may not always be advantageous. Mandelbaum et al. (1998) [20] highlight that while pooling resources can be beneficial under lighter traffic conditions, its effectiveness can diminish during periods of high demand. The complexity and variability in service demands during peak times can overwhelm a centralized system, potentially leading to inefficiencies and longer wait times. Therefore, maintaining separate servers or pools can often provide better performance and operational flexibility in such situations.

Research by Daigle et al. (2005) [10] supports this notion, emphasizing that maintaining separate locations during high-traffic periods allows for greater flexibility in resource allocation and management. This flexibility is crucial for dynamically adjusting resources to meet the diverse demands across different service types or patient categories. However, it's important to note that the distribution of beds among multiple locations plays a critical role in optimizing this process.

Allocation (multiple locations) When considering multiple locations, the allocation of patients becomes a challenge. For allocating elderly patients to nursing homes, an allocation model is published by Arntzen et al. (2022) [3]. This allocation model was based on preferences and utilities. This model serves as an alternative to regular waiting lines. The advantage of modelling the allocation is its patient-centred approach. In this case, patients can be served according to their individual preferences and waiting time (in)flexibility. As a result, the model performed approximately the same as a shared queue, so close to the minimum waiting time. Also, other research by Arntzen et al. (2022) [4] showed that allocation can be modelled optimally. In this case, the model was not modelled on bed capacity, but available hours. These models do not take the happiness and fairness of patients in the queues into account. So in this thesis for multiple locations of Intermediate Care.

With preference, fairness is an important metric. How fair is it to wait for service at your preferred location, instead of getting service earlier somewhere else? Research by Itzhak et al. (2008) [6] looked at the fairness of queues in different ways. The most effective one is the "absolute fairness difference" meaning the maximum extra waiting time compared to a FCFS policy. In "Stochastic fairness queueing" by McKenny et al. (1990) [21] a stochastic framework for fairness in queueing is given. Simulation results presented show that stochastic fairness queuing can achieve a performance that is within a factor of three of that of fairness queuing and is almost two orders of magnitude better than that of simple FCFS queuing. These results show that in queuing it is possible to take fairness into account and improve metrics at the same time.

3.2.5 Skill-Based-Routing

This thesis considers making a modelled allocation, based on an allocation model when Intermediate Care is decentralized. When looking at allocation globally, so not only in healthcare, the skill-based-routing principle, discussed in Legros et al. (2015) [19], Chen et al. [8] and Wallace et al. (2004) [31], is a nice framework. This is often used in call centres. Call centres usually handle several types of calls. To do so effectively, special training is considered for their agents. This is not cost-effective, because it is not efficient to have every agent be able to handle every type of call. Agents thus tend to have different skills and be able to treat different customers. Specific calls will then be assigned to different agents based on the customer's needs and the agent's skill. This is called Skill-Based-Routing (SBR).

Wallace et al. (2004) [31] look for simplification of SBR algorithms through resource pooling. This considers minimal cross-training (requiring agents to learn all skills, at most two) and a reasonable scheduling policy, trying to make the model perform almost as well as when all agents have all skills. After simulation studies, it is shown that with limited staff, knowing all skills, an SBR algorithm can be as functional as a maximal cross-training, making all agents learn every task.

This thesis will mostly refer to Chen et al. (2020) [8]. This paper provides a survey on the SBR framework and gives a comparison to healthcare. This paper gives a deep mathematical insight into the SBR algorithm. Various specialized queueing models have been made for healthcare systems, for example, to capture the fact that patients may return to service several times during their stay at the EMD. Other studies showed how to model the prioritization of new versus reentrant patients with deadline constraints. These examples lead to how this paper wants to describe how routing decisions may affect patient outcomes, and how that, in turn, may affect the routing decisions. This paper researched the effect of admission delay on patient outcomes, the effect of off-service placement on patient outcomes, readmissions and routing with predictive information. Chen et al only provided information on those topics but did not implement it in a simulation.

Mathematically this paper does allocation policies based on integer linear programs. To describe the model used, they defined at the start of each day t the queue length by variable $Q_i(t)$ and the number of busy servers in each pool/skill by $Z_j(t)$. Defining overflow costs $\phi_{i,j}$ and holding costs h_i per unit i. After observing $Q_i(t)$ and $Z_j(t)$, the admission decision $a_{ij}(t)$ is made, saying how many patients of class i are to be admitted in pool j. From there costs are defined by $\sum_{i,j} \phi_{i,j} a_{i,j}(t)$ and holding costs $\sum_i h_i(Q_i(t) - \sum_j a_{i,j}(t))$. This paper wants to minimize the long-run average costs and focus on the class of Markovian policies. The main differences between the approach in this paper and this research thesis are the objective function and the cost function. The methodology is a nice framework to use. Chen et al. define the objective function as a minimization of the average waiting time per location combined with the maximization of happiness and fairness. Also, a cost function is not considered in the allocation model, this is an interesting feature to add to further research.

This thesis uses this model and rebuilds it by seeing the language or skill as the preferred location of patients and the objective functions by fairness, happiness and waiting time, defining fairness by the literature of Itzhak et al. (2008) [6]. This means that fairness is the maximal difference between the model and the FCFS model. Happiness is defined by the number of times patients are treated at locations they do not prefer. This thesis will not optimize the metrics but will show how modelled allocation can come close to minimum waiting times while keeping happiness and fairness into account. A different strategy for solving the ILP of the SBR will be considered.

3.3 Hypothesis

From this literature review, it can be concluded that the interventions that were mentioned in section 1 can contribute to reducing waiting times. In terms of sharing, sharing all beds at Intermediate Care is expected to be the most effective in reducing waiting times. However, theory shows that partial sharing can achieve almost the same results with fewer changes in the system. Also for high traffic, which is the case, full flexibility cannot always be beneficial when service times of different caretypes have big differences. In terms of centralization, more locations are efficient during high-traffic periods. When decentralized, the allocation can be done by considering the fairness and happiness of patients based on previous research and a Skill-Based-Routing framework. The main hypothesis of this research is that sharing is expected to be the most efficient out of the interventions. Centralizing Intermediate Care can have its benefits, but during High-traffic periods decentralizing into multiple locations is expected to be more effective. When more locations are considered, the allocation can be done using an allocation model.

4 Model description

This section describes the DES model, explains what choices are made and how the different interventions are modelled. The input variable(s) are discussed and the stability condition is considered. At last, the evaluation shows how the performance of each intervention is measured.

4.1 DES model

Previous research by Arntzen et al. (2022) [4] implemented a DES model for Intermediate Care in Amsterdam. As said before, the different Intermediate Care regions the Dolce Vita project considers all have different structures of Intermediate Care. Therefore different DES models have to be considered. Figure 5 shows how patients arrive and what type of care is offered in Twente for Intermediate Care. This section compares the model of Twente to the one of previous research in Amsterdam.

The admission control for Intermediate Care locations in Amsterdam is regionally centralized. Therefore, one model at the aggregated level of the complete region sufficed. According to experts in the region of Amsterdam, the fraction of patients who need Low Complex care can be neglected. Therefore the Low Complex care is left out of the model for the region of Amsterdam shown in Figure 6. For the same reason, respite care is left out of the model in Twente shown in Figure 5. The biggest difference between the DES model of Amsterdam and Twente is that the different locations in Twente can offer different caretypes, according to experts within that region, also see Figure 1. As said before the different types of care Intermediate Care can offer are High Complex care, Geriatric Rehabilitation, Low Complex care and Respite care. Because of relevance, Respite care is left out of the model in Twente.

Looking at the Intermediate Care structure of Twente, shown in Figure 5, it can be seen that Low Complex patients arrive from the General Practitioner and are admitted to either Intermediate Care or waiting list 1 when there is no bed available immediately. In that case, the waiting time will be spent at home, so no hospital admission, due to the lack of urgency for Low Complex care. Patients who are considered High Complex are admitted from the same sources as in Amsterdam, namely the General Practitioner, Emergency Department and hospital. When patients cannot get an Intermediate Care bed immediately they either go to waiting lists 2, 3 or 4. When patients from the EMD are not immediately admitted to Intermediate Care, they wait in the hospital, which is a (redundant) hospital admission. Patients who are considered for Geriatric Rehabilitation care are only admitted from the hospital. When these patients cannot be admitted to Intermediate Care immediately, they are added to waiting list 4, which is in the hospital and is a prolonged stay in the hospital. For simplicity, this stay counts as a redundant admission, just as for patients from the EMD. It can be concluded that waiting on waiting lists 3 and 4 can cause major congestion in the hospital. Patients who do not need hospital admissions are admitted to the hospital blocking beds for patients who do need care in the hospital.

According to experts within the region of Twente, patients who are considered High Complex cannot move to Low Complex or vice versa, although this sometimes does happen in practice.

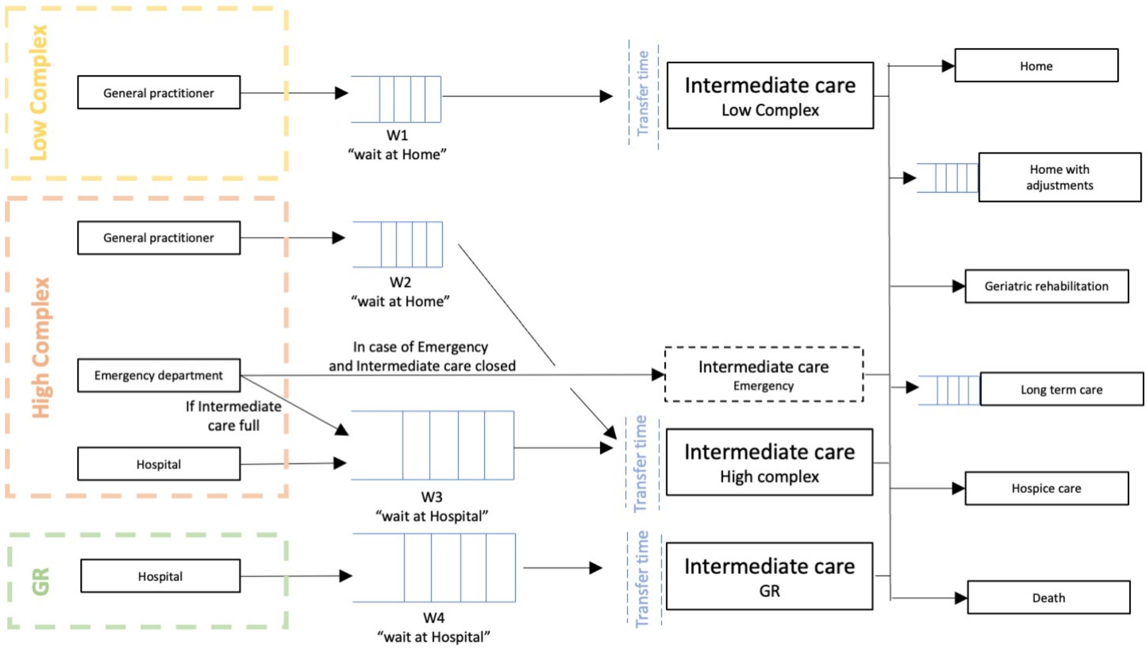

Figure 5: Intermediate Care structure of the region of Twente, No bed-sharing

4.2 Notation

The DES model in Twente has three different caretypes, which will be denoted as

$$i \in \{\text{High Complex, Low Complex, GR}\}.$$

In addition, there are also three different sources patients can be admitted from to Intermediate Care, denoted by

$$j \in \{\text{hospital, General Practitioner, Emergency Department}\}.$$

Then the outflow destinations (K) are denoted by

$$k \in \{\text{Home, Home with adjustments, Long-term care, Geriatric Rehabilitation, Hospice Care, Death}\}.$$

Denote the arrival rates as λ_{ijc}, where λ is the arrival rate of caretype i, inflow j and location c, with C number of locations. The average service is denoted as β_{ik} denoting the average service time for caretype i and outflow destination k, which does not depend on the admission source. Denote

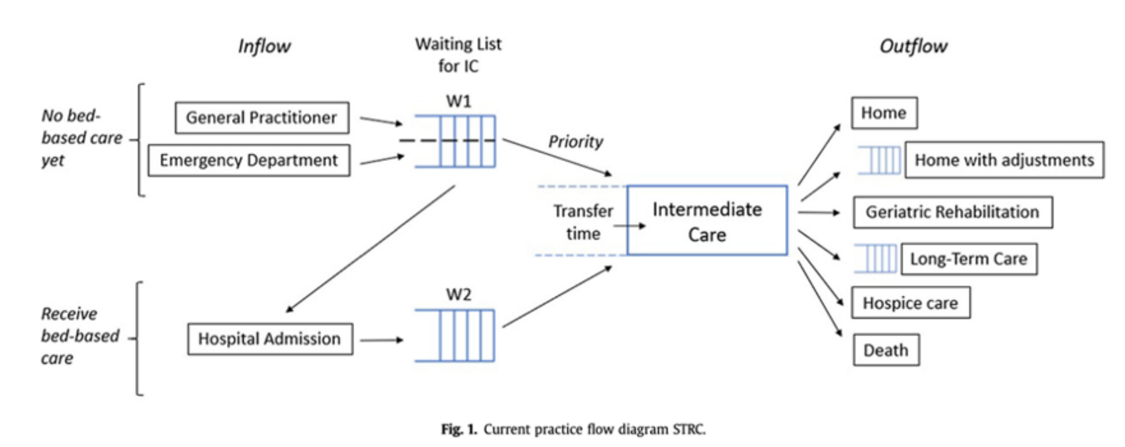

Figure 6: Intermediate Care structure of Amsterdam

the outflow probabilities as p_{ik} as the probability of caretype i going to outflow destination k. The number of beds is denoted by n_{ic} saying the number of beds for caretype i at location c, with N denoting the total number of beds over all locations. The distribution of beds among locations is unknown because no data is delivered on that. So the distribution of beds in different locations will be done randomly.

This thesis will use these variables to simulate different interventions within the DES model, for that denote

$$r \in \{\text{High Complex, GR, High Complex \& GR, Low Complex}\},$$

to give the difference between High Complex, GR and Low Complex care and combining High Complex and GR care. Denote $q_{r,c}$ the number of nurses available for model r at location c Lastly, denote s_r as the number of patients a healthcare employee can care for in model r.

For simplicity, the notation of the other variables used in the model will be in words instead of mathematical terms. So for example the sharing scenario will not have a mathematical term but will be expressed in words. The same goes for admission hours.

4.3 Input parameters

The parameters for the DES model are obtained from several surveys of the Intermediate Care locations in Twente and literature from previous researchers, Arntzen et al. (2023) [5] and Groot et al. [11]. These inputs have been combined into input parameters for one Intermediate Care location. The arrival rate for Geriatric Rehabilitation was taken from research conducted by de Groot et al. (2023) [11]. They looked at discharge rates from the hospital at Amsterdam UMC during the time period of 15th January to 15th May. They found out that 60 patients (12.2%) were referred to a Geriatric Rehabilitation facility. Note, that this number of older adults is not accurate, since this number is only for one location and this research considers multiple locations. The arrival rate for the High Complex patients was used from Arntzen et al. [5], which is for all of the Intermediate Care locations in Amsterdam. The input parameters of the surveys and the data used in the research of Arntzen et al. and Groot et al. are combined and shown in Table 2 for the outflow probabilities, Table 3 for the arrival rates and Table 4 for the service rates.

As shown, most arrivals come from the General Practitioner, especially for Low Complex care (including respite care). Most patients with Low Complex care go home, and few get GR or Hospice care. For High Complex care, patients arrive mostly from the General Practitioner, who is also most likely to go home after Intermediate Care. A full list of input parameters is given in Table 7 in the Appendix. Also, a default value is mentioned. Whenever a parameter is shifted for modelling an intervention, it is mentioned, if not the default value is used.

The mean arrival rates in Table 3 are assumed to have a Poisson distribution and be constant, saying a constant mean arrival rate per day. Previous research shows that it works fine assuming that [32], but in reality, the arrival of patients is random and not exactly constant. Also section 2 told that when inter-arrival times are exponentially distributed, the probability of 2 events happening at the same time is significantly small. This states that for example, the probability of two patients arriving at the same time is almost equal to zero. It can be asked if that's the case for Intermediate Care, since no arrival, nor service data is available. Also, all service times are independent and identically distributed. This property can be doubted in this research because no data is delivered to prove all service times are exponentially distributed with the rates mentioned in the table. For the average waiting time, it will not matter as explained in section 3.

Average Service time, $\beta_{i,k}$	High Complex	Geriatric Rehabilitation	Low Complex
Home	31.1	31.1	31.1
Home with adjustment	43.9	43.9	43.9
Long term care (WLZ)	47.8	47.8	47.8
Geriatric Rehabilitation	29.8	X	29.8
Hospice Care (WMO)	22.9	22.9	22.9
Death	22.9	22.9	22.9

Table 1: Average service time for the model, a combination of surveys in Twente. X-indicates not possible

Mean outflow probabilities, $p_{i,k}$	High Complex	Geriatric Rehabilitation	Low Complex
Home	0.578	0.6	0.7
Home with adjustment	0.107	0.107	0.14
Long term care (WLZ)	0.198	0.21	0.1
Geriatric Rehabilitation	0.034	X	0.02
Hospice Care (WMO)	0.023	0.023	0.02
Death	0.06	0.06	0.02

Table 2: Outflow probabilities for the model, a combination research in [5] and [11]. X-indicates not possible

Average arrivals per day, $\lambda_{i,j}$	High Complex	Geriatric Rehabilitation	Low Complex
General Practitioner	1.34	X	1.91
Emergency Department	0.83	X	X
hospital	0.94	0.54	X

Table 3: Arrivals per day for the model, a combination research in [5] and [11]. X-indicates not possible

Stability As said in the Literature review, this system is modelled in the spirit of an M/M/c queue. This means there is a minimum number of beds needed to take care of the patient's arrival. This is calculated by the stability condition. Saying that the load per bed, called ρ must not be greater than 1. This is calculated by the formula shown in the literature review in section 3. In Table 4 the average service times are given, calculated by the weighted average of the outflow probabilities and the respective service times, which gives.

High Complex care

$$\begin{aligned}\bar{\beta}_{High\ Complex} &= \sum_{k=1}^{6} \beta_{High\ Complex,k} \cdot p_{High\ Complex,k} \\ &= 31.1 \cdot 0.578 + 43.9 \cdot 0.107 + 47.8 \cdot 0.198 + 29.8 \cdot 0.034 + 22.9 \cdot 0.023 + 22.9 \cdot 0.06 \\ &= 35.05 \text{ days}\end{aligned}$$

Geriatric Rehbilitation

$$\begin{aligned}\bar{\beta}_{Geriatric\ Rehabilitation} &= \sum_{k=1}^{6} \beta_{Geriatric\ Rehabilitation,k} \cdot p_{Geriatric\ Rehabilitation,k} \\ &= 31.1 \cdot 0.6 + 43.9 \cdot 0.107 + 47.8 \cdot 0.21 + 22.9 \cdot 0.023 + 22.9 \cdot 0.06 \\ &= 35.30 \text{ days}\end{aligned}$$

Low Complex care

$$\begin{aligned}\bar{\beta}_{Low\ Complex} &= \sum_{k=1}^{6} \beta_{Low\ Complex,k} \cdot p_{Low\ Complex,k} \\ &= 31.1 \cdot 0.7 + 43.9 \cdot 0.14 + 47.8 \cdot 0.1 + 29.8 \cdot 0.02 + 22.9 \cdot 0.02 + 22.9 \cdot 0.02 \\ &= 34.21 \text{ days}\end{aligned}$$

For calculating the number of beds needed the load per bed is considered to be high, so close to one. Recalling the formula for the load per bed $\rho = \frac{\lambda}{c\mu}$. Rewrite this to $c = \lambda \cdot \mu$ gives for all caretypes.

	High Complex	Geriatric Rehabilitation	Low Complex
Average service time, β_i (d)	35.05	35.30	34.21

Table 4: Average service time

- For High Complex care, the total arrival is 3.11 patients per day. These patients have an average service time of 35.05 days. In this case, for this type of care, at least $3.11 \cdot 35.05 \approx 110$ beds need to be at least available to take care of the load.
- For Low Complex care, the total arrival is 1.91 patients (including respite care). These patients have an average service time of 34.12 days. This means at least $1.91 \cdot 34.12 \approx 66$ beds are needed to take care of the load.
- For GR care the total arrival is 0.54. These patients have an average service time of 35.30 days, meaning the minimum amount of beds needed equals $0.54 \cdot 35.30 \approx 20$ beds.

In the Twente case, they did not deliver any data on bed capacity. Because when the load per bed is equal to one, the system will explode, so one bed per caretype is added to make sure the load per bed is lower than one. So for the research, consider 111 beds for the High Complex, 21 beds for GR care and 67 for the Low Complex care to just be able to handle the load theoretically.

The Square-Root-Staffing principle is discussed in section 5.2.2. This principle adds a safety capacity, based on the square root, to the number of beds to not make the waiting times explode. Table 5 shows the beds needed for both approaches. For the rest of the research, the number of beds which is the outcome of Square root staffing is used to model the interventions. The reason for that can be found in the fact the model is more complex than a M/M/c queue, so more beds need to be considered than the number of beds used for stability. Considering transfer time and admission hours for example makes the system more complex so safety capacity is needed.

In Table 7, in the appendix, a full list of parameters for the model is given. Some of them include the mathematical notation.

Number of beds	High Complex	Geriatric Rehabilitation	Low Complex
Stability	111	21	67
Square-Root-Staffing, α =1	120	26	75

Table 5: Number of beds needed

4.4 Centralization and allocation

In terms of centralization, no data is delivered on the distribution of beds among the Intermediate Care locations. To model the influence of decentralizing Intermediate Care, the distribution of beds is done equally. One to five locations are considered, so the Intermediate Care is split into 5 equal locations. When multiple locations are considered the arrivals of patients at locations are distributed randomly from the total arrival rate. This is done because it is unknown how the arrivals among locations are distributed and no data on this topic is delivered. At last, when more than one location is considered, the allocation of patients becomes a challenge. How this challenge is modelled is discussed below.

Allocation model - overflow strategy Based on the literature, an ILP can be created based on the skill-based-routing algorithm [8] discussed in section 3. This algorithm assigns different skills to servers and allocates clients to those servers based on the required skills. This allocation is done using an integer linear program for the allocation policy, which delivers an optimized allocation to minimize the waiting time or expected queue length. This thesis tried to implement this algorithm in the allocation of patients to Intermediate Care in Twente. Because optimization is difficult to implement in the model, the ILP described above is modelled using an overflow strategy. This is a simple way of partly solving the ILP. This strategy does not assign patients to locations based on the total mean waiting time, which is minimized, but by looking at the patient itself. This is explained in algorithm 4. This so-called overflow strategy uses a threshold of how long a patient maximally wants to wait for its preferred location. This approach is chosen because in reality, preference and urgency of care change over time, which is different for every patient. When the urgency is high, a patient does not have a strong preference and the other way around. When the waiting time exceeds the threshold, the allocation goes to FCFS. The threshold is different for every patient and is assigned randomly between 0 and the variable max_waiting_time_model which can be shifted.

To show the results, no data is delivered on the capacities of different locations. By that, a designed situation is used considering five locations with different Intermediate Care capacities. This provides a discussion point on the distribution of beds over the various locations. The distribution of beds among locations is done by defining for example one location which delivers only High Complex care, one only Low Complex care and one only GR care, the other two locations deliver all types of care. Even though the model is approached using an overflow strategy, the waiting times, fairness, and happiness still are key metrics to evaluate the allocation approaches on.

4.5 Interventions

Based on previous research, and engagements with healthcare professionals and stakeholders in the Twente region, different interventions are considered. These interventions can influence the waiting times, patient transitions or blocking probabilities of Intermediate Care. The main interventions are bed-sharing, centralization and small interventions.

4.5.1 Intervention 1: bed-sharing

This intervention is about sharing beds among different caretypes. Sharing is a common practice which is proven to have a major influence on waiting times. Theoretically, all beds can be shared among all caretypes. Practically it is difficult to share or assign a Low Complex bed to a High Complex patient for example, due to a lack of equipment and intensity of care. Four variants of bed-sharing are considered:

- **Total bed-sharing**, this indicates sharing all beds in the Intermediate Care location. So Low Complex beds, GR beds and High Complex beds are all combined for all caretypes also the emergency beds are considered for regular admission. As said before this can be done theoretically, but delivers challenges practically. See Figure 27.
- **GR-High Complex bed-sharing**, this type of sharing considers two groups Low Complex and High Complex plus GR care. This is practically the maximal sharing possible in the Intermediate Care locations. Shown in Figure 28.

- **NO bed-sharing**, this sharing policy considers no sharing among Low Complex, High Complex and GR care. So three different bed groups are considered. see Figure 5. This is also the current practice.

- **Partial bed-sharing**, this type of sharing adds one group compared to the NO sharing policy, namely shared beds. These beds are used for sharing among High Complex and GR patients, whenever one of those capacities is full. See Figure 29.

As said in the literature review, it is assumed sharing will decrease waiting times. This is because flexibility in the system will make the system work better. The hypothesis from section 3.3 statesd sharing can deliver great results in terms of reducing waiting times.

Emergency beds Among all sharing types, two emergency beds are available per location. These beds are available for patients who come from the EMD and cannot go to Intermediate Care because it is closed. So when a patient from the EMD arrives after admission hours, they can go to an emergency Intermediate Care. This bed is then occupied and another free bed will be saved for emergencies. When a patient is discharged from an emergency bed, the bed will be free for new emergencies. The admission to these beds is modelled without transfer time because these beds are always available for admission and require quick admission because it is an emergency.

4.5.2 Intervention 2: centralization and allocation

As said before, the Intermediate Care admission in Twente is decentralized. Meaning there are multiple locations available to admit for Intermediate Care. Centralization of those locations means a greater capacity, but also a greater demand. Since the region of Twente is much bigger than the region of Amsterdam, see Figure 1, this intervention is combined with the preferences of patients. Assuming every patient has a location preference, for example preferring a location nearby so the family can visit easily. Considering three policies for allocating patients to different locations:

- **First Come First Served**, this policy indicates that whenever a bed is free at whatever location, the client who waited the longest is assigned first. This means that not every patient can get admitted to their preferred location. This is modelled as one big waiting list of patients for all locations.

- **Only preference**, this policy indicates that patients can only be admitted to their preferred location. This means that patients with shorter waiting times can be admitted earlier because patients are willing to wait for space at their preferred location. This is modelled as every location has its waiting list.

- **Modelled allocation**, this is a developed model in this thesis taking preference into account but also looks at fair allocation. This means not letting a patient wait for too long or skipping patients endlessly. Inspiration for this model is taken from previous research by Arntzen et al. (2022) [3] and the SBR framework discussed in section 3.2 and Chen et al. (2020) [8] and is approached using an overflow strategy discussed in section 4.4.

Centralization also gives advantages in terms of efficient capacity, meaning that nurses can be used more efficiently at a central location. This reduces the loss of capacity. Again defining s_r as the number of patients 1 nurse can take care of in model r, $r \in \{\text{High Complex, GR, Low Complex}\}$. Centralization is not always beneficial. For example, for high-traffic situations, it can be more efficient to divide the load over more locations. In its turn, delivers a challenge to not have a loss in capacity.

4.5.3 Intervention 3: small interventions

The last group of interventions are small policy changes which can help improve metrics. Three are considered.

- **Intervention 3a: priority**, for this intervention a GR-High Complex share is considered. Because for the current practice, a NO bed share, no priority can be given. Looking at Figure 28 for High Complex and GR care 2 waiting lines are considered namely 2 and 3. This intervention gives priority to patients from waiting list 3 over waiting list 2. This priority is because patients from the EMD, for example, have a high urgency to get care, also when not admitted immediately to the Intermediate Care, they are admitted to the hospital, which in turn delivers blocked beds. The same goes for patients who are already at the hospital.
- **Intervention 3b: triage ward**, previously the region of Twente worked with observation beds. These beds are used whether it is not certain if a patient needs Intermediate Care care or needs to go somewhere else. After a stay of approximately fourteen days maximum, the patient goes to Intermediate Care, or somewhere else, home for example.
- **Intervention 3c: admission hours**, previous studies showed that admission hours also affect waiting times [4]. Intermediate Care is during the week available between 8 A.m. and 5 P.m. and closed during weekends. Also, it takes time, called admission days, for hospitals to admit patients to Intermediate Care. For now, it is considered to be 1.5 days. This intervention checks whether changing the admission policy affects waiting times and queues. For patients that come from the EMD, emergency beds are available when the Intermediate Care is closed, they can also be admitted immediately.

The triage ward is not expected to have a big influence. This intervention has a financial advantage because observation beds are cheaper than Intermediate Care beds. Also, it has a practical advantage in making space for patients who are not sure to be considered Intermediate Care or something else. In Twente, they used observation beds previously, but nowadays they do not use them anymore, because the triage process is well-organized in the region.

5 Methodology

This section describes the methodology. First, the simulations and desired output are explained. Choices in modelling the DES model and interventions are highlighted. After that, the desired output variables and metrics are shown. Lastly, the methods for validation and verification are shown to verify the results.

5.1 DES model

Discrete Event Simulation (DES) models are computational models used to represent the operation of complex systems as a chronological sequence of events. Each event occurs at a specific point in time and causes a change in the state of the system. In DES models, time progresses by jumping from one event to the next, rather than advancing continuously. More on the DES model can be found in the literature [12].

As said before, the DES model of Twente considers three types of care, High Complex, Low Complex and GR care. The flow of the system is shown in Figure 5 and the input parameters are discussed in section 4.3. The model simulates the patient flow in the system. This means that a patient arrives at the Intermediate Care where they were admitted from either the EMD, GP or hospital. Consequently, their length of stay is determined by their caretype and their outflow destination.

The general idea of the simulation model, which is implemented in Python, is shown in algorithm 1. This algorithm takes as input the number of clients which are put through the system per sub-run, the number of sub-runs and the number of clients for the warm-up period. The initialization of the simulation takes part in the first sub-run where a dictionary of events is constructed. This event dictionary contains the arrivals of patients and then other types of events such as hospital admissions and discharges.

Because waiting times for empty systems differ from waiting times in a full system, else said steady state, the model first treats the patients for the warm-up period and then measures the metrics on the patients per sub-run after the warming period. The simulation runs until all of the patients per sub-run in all sub-runs have been discharged from the system (added to the clients for warming). In the simulation model, one second is treated as one day, this approach simplifies the determination of both the day of the week and the time of day within the simulation.

Preprocessing Before the model is run as in algorithm 1, the total effective beds are calculated using the input parameters from section 4.3. By defining $q_{r,c}$ as the number of nurses available for the care of model r at location c and s_r be the number of patients 1 nurse can take care of model r. This is calculated by;

$$\text{Number of effective beds for model } r \text{ at location } c = q_{r,c} \cdot s_r$$

If the number of effective beds is less than the total number of beds available, it means that not enough nurses are used leading to a loss of capacity. On the other side, if the number of effective beds is greater than the total number of available beds, it means that there are too many nurses for the capacity of beds available. This leads to a loss of duty.

In case there is a loss of capacity the number of beds used for the model changes to the effective number of beds. In case of a loss of duty, the number of beds will not change. So before the model starts it checks whether the input variables match each other.

Also, before running the algorithm the stability of the system is checked. As shown in section 4.3. In this part, the load ρ is calculated to check if it does not exceed 1. If so the system is not stable, initializing that the system cannot take the load with the provided capacity, which results in high waiting times. If there is an unstable system, the model will give an error but still give results, which then can be unreliable.

For the results of this thesis, it is assumed that the preprocessing works perfectly, so enough nurses are available and the system is stable (unless mentioned).

5.2 Arrivals

What happens to a patient when it arrives at the Intermediate Care depends on what policy or scenario is simulated. A flow diagram is shown in Figure 7 to give an insight into how the different interventions are simulated when a patient arrives. In the algorithm 2 the pseudocode for the arrival process is shown. These arrivals are assumed to have exponential interarrival times and thus follow a Poisson distribution. It depends on the scenario of sharing or allocation if the patient is admitted to the Intermediate Care directly or has to wait. In this pseudocode a centralized scenario, meaning one combined Intermediate Care location, is considered.

5.2.1 Triage ward

Figure 30 shows the flow diagram for this intervention. In the case of the triage ward intervention, observation beds are used for observing patients and nurses are not sure of what type of care is needed. When a patient arrives they can either go straight to Intermediate Care, to TRW (observation) or wait at the hospital or home. Research and opinions from experts have shown that approximately 10% of elderly patients need observation. This parameter can be shifted in the model. So when a patient arrives, it is determined whether it needs observation (with a probability of 10% in this case) or can go to Intermediate Care immediately. When a patient needs observation and the observation beds are occupied, it waits in the hospital (waiting list 3 or 4) or at home (waiting list 2).

When discharged from TRW they can either go home for example or to Intermediate Care and another patient from the waiting list can take the bed. A maximum of fourteen days is considered for observation. This variable can also be shifted in the model. The length of stay of patients that need observation is determined by the time the patient spent in observation, the waiting time for observation and the time it spends in Intermediate Care.

On the other hand, when a patient does not need observation, it follows the original arrival scheme from algorithm 2. According to experts, the advantage of the triage ward intervention is that observation beds are cheaper than Intermediate Care beds. Also, the triage ward scenario is considered when an Intermediate Care location is very large.

5.2.2 Sharing

As seen in Figure 7, the type of bed-sharing plays a major role in whether a patient can be treated immediately or has to wait from home or in the hospital. When a total bed-share is considered, the caretype does not play a role in allocating beds. When a GR-High Complex, partial- or NO-bed-share is considered the type of care of the patient determines whether the patient can be assigned directly or needs to wait. Partial bed-sharing can be effective for high-traffic queues. For all bed-sharing types, emergency beds are available for admission of EMD patients after admission hours.

Again a total bed share means that Low Complex care, High Complex care and GR Care are combined and share all beds also the emergency beds are used for admission. This is theoretically the most sharing possible, but practically difficult to implement. A GR-High Complex share means that GR and High Complex care share beds, and Low Complex care have their beds. This is almost the most sharing possible. A partial sharing is considering more groups of beds. Namely Low Complex beds, High Complex beds, GR beds and shared beds. Each caretype has committed beds, but shared beds are available for admission when their committed space is full for High Complex and GR patients. NO sharing means that three-bed groups are available: Low Complex, High Complex and GR care. As the name says, no sharing among caretypes is possible.

5.2.3 Centralization and allocation

In algorithm 2, only one combined Intermediate Care location is considered and an FCFS (First Come First Served) policy is applied. This means when patients with the right caretype are waiting when a patient of that same caretype is discharged, the one with the longest waiting time gets the bed first, without looking at what location is preferred. Which indicates that patients get admitted to locations which they do not like or prefer, which lowers the so-called "happiness" of the patient. In reality, patients have a preference for locations that are nearby or more easy to reach. In the Twente case preference plays a bigger role than for example in Amsterdam, due to the difference in the size of the two regions. Twente is a much bigger region, so patients add more value to being placed closer to home.

When patients are allocated in line with their preferences, the happiness of the patients is maximal. This causes the longest waiting patient not to get treatment first all the time, which is called "skipping" or the "fairness" of the queue. An allocation model is made to find the balance between the fairness and happiness of the queues but also tries to minimize the waiting time. This model is described in algorithm 4. When a patient arrives at the Intermediate Care, the allocation is done by preference. So only the preferred location is considered. When an FCFS model is considered, at arrival, when Intermediate Care is full, other locations can be considered to check if the patient can get treatment there, otherwise, the patient is assigned to the waiting list. When a patient is assigned to a waiting list multiple metrics are tracked, namely.

- Arrival time in queue, the time the client is assigned to the queue.
- Waiting time, the time the patient is spending in the queue.
- Utility, a value that increases with waiting time linearly, telling the amount of utility for keeping your preference. this variable tries to represent urgency.
- Possible alternative locations, representing the possible other locations a patient can go, based on its caretype.
- Skipped, a True or False value which tells if a patient is skipped once or more before.
- Timestamp skipped, this is a timestamp telling at which point the patient got skipped the first time.
- Extra waiting time, the current time minus the timestamp skipped. Saying how much time the patient has to wait longer. This value is used for the fairness definition.

Using these metrics, the overflow strategy discussed in section 4.4 is applied for modelling allocation of patients. This tries to balance fairness and happiness while reducing the waiting time(s) and considering multiple locations. For modelling this intervention regarding allocation, 5 different locations are modelled with different numbers of beds. Because no access to data from other locations is available, this designed situation is used.

Happiness is defined by the percentage of patients during simulation which are placed in other locations than they preferred before. Fairness is defined by literature and considered to be the maximal difference in waiting time between an FCFS and the model used.

5.2.4 Admission hours

Since not every Intermediate Care location is open 24/7, this intervention checks what would happen if opening weekend hours were available. Also, it is checked whether speeding up the transfer time influences waiting times. Transfer time refers to the time it takes for the patient to be admitted to the Intermediate Care from when the patient was referred to the Intermediate Care, also shown in Figure 5. Therefore it is checked what day it is at the time of the arrival. First, it is checked if the day of the arrival is a weekday, and subsequently, whether the Intermediate Care is still open for admission or not. If the Intermediate Care is still open then the patient is admitted to the Intermediate Care at the time of the transfer time. Whereas if the Intermediate Care is closed for admission then the patient is admitted the next day plus at the time of transfer time.

Alternatively, if the arrival time is during the weekend and the Intermediate Care is closed then if the day is Saturday, the patient is admitted two days later plus the transfer time. However, if the day is a Sunday and the Intermediate Care is closed, then the patient's arrival time is changed to the next day plus the transfer time. When a patient comes from EMD, emergency beds are available for admission outside opening hours for urgency. On the other hand, if the Intermediate Care isn't closed for admission during the weekend the patient is admitted at the time of transfer time. Additionally, if the Intermediate Care is open on the weekends but is still closed for admission then the admission to the Intermediate Care is changed to the next day plus the transfer time.

Evening opening hours are not considered to be changed, because General Practitioners are certainly closed for admission in the evening and in the evening it is not expected to have many elderly patients coming for Intermediate Care admission. If they do, they can be considered emergency issues and can be admitted to an emergency bed.

5.3 Discharge

When a patient is discharged from Intermediate Care, a patient from the waiting list can replace that patient. Which patient will be assigned to the bed is based on, again, the policy. Insights are shown in the flow diagram in Figure 8. In the pseudocode in algorithm 3 it can be seen that discharge can happen for observation beds, emergency beds, hospital admission and Intermediate Care beds. When a patient is discharged, the arrival times, discharge times and destination are added to the output list. This is so that the waiting times can be determined. When a patient leaves the system another one might be on the waiting list. This is where the priority comes into play. If priority is for the waiting lists 3 and 4 then it is first checked if there is a patient on those waiting lists. If not, a patient from the waiting list 2 is admitted.

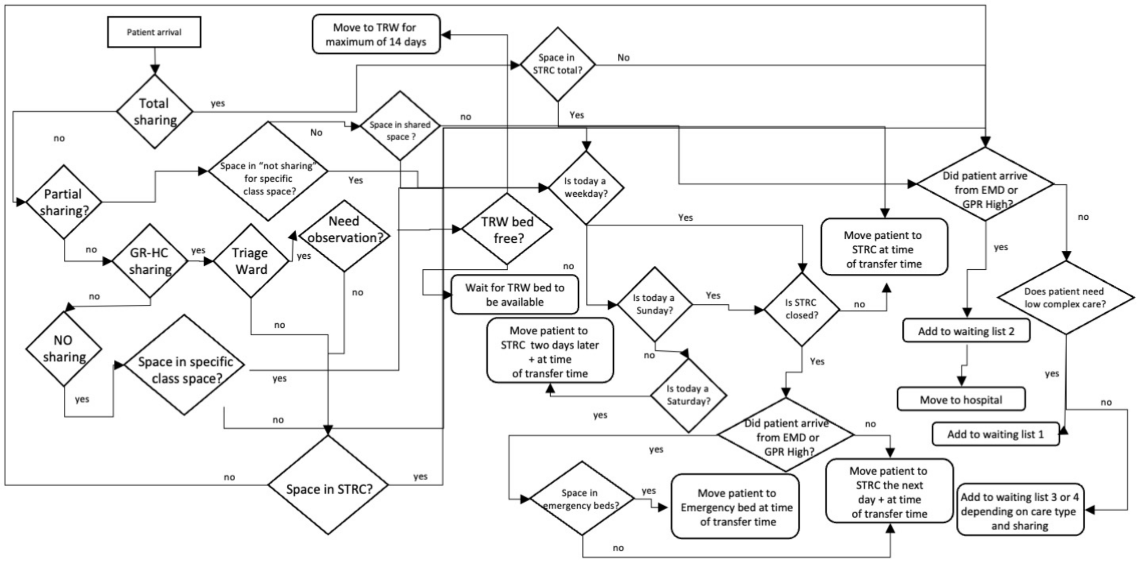

Figure 7: Flow diagram of arrivals

5.3.1 Sharing

When doing a total sharing scenario there is no need to look at the classes of patients, but when looking at partial- GR and High Complex or NO-sharing scenarios the model takes the classes into account. For a partial sharing scenario, it is checked if the discharged patient was in the "shared" bed space. If so, then any type of class (within High Complex and GR care) can be admitted to Intermediate Care. However, if the patient was not in the shared bed space, then a patient of that same class is searched. Conversely, if there is no priority, then it is checked who came first to the waiting line and admitted that particular patient. The same as above is done for the sharing scenarios, where the class is also dependent on which patient is admitted directly and who is not.

5.3.2 Centralization and allocation

For implementing the Centralization intervention, a designed situation is used for modelling different locations of Intermediate Care because no data on the distribution of beds among locations is delivered. In this example, a NO bed-sharing scenario for all the different locations is considered. As the algorithm 4 shows, when the waiting time is bigger than the utility of the patient waiting, the allocation will be FCFS. Utility represents the maximum amount of days a patient wants to wait for care. This is the overflow strategy from section 4.4. The utility function can also change to exponential or lognormal. Previous research by Arntzen et al. (2022) [3] showed that this way of allocating patients can work effectively.

The FCFS policy is modelled by setting all utilities to -1, so it is always open to switching locations. The only preference is modelled by setting all utilities equal to infinity, saying patients

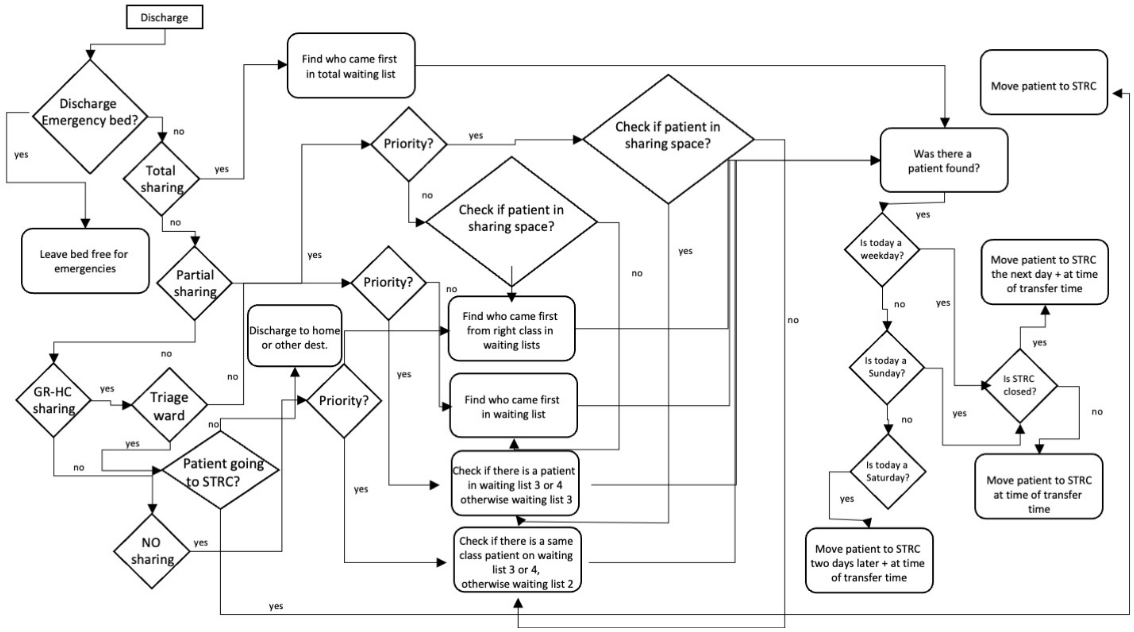

Figure 8: Flow diagram of discharge

will never be open to switching locations. The allocation model assigns random variables to the utility of patients on the waiting lists. The fairness of the waiting list is defined based on literature by Itzhak et al. (2008) [6]. Mentioning fairness is the maximal waiting time difference between the FCFS policy and the model used. In this research, happiness is chosen as the percentage of patients placed in other locations than they preferred before.

5.4 Output

In Table 6 the output measures are shown. These output measures are chosen in consultation with healthcare experts from the region of Twente. For every intervention, the performance measures are checked. The main output of this research is a dashboard that Intermediate Care organizations can use to calculate their expected waiting times. With this, the model can perform with different parameters.

5.4.1 Confidence interval

When the metrics are calculated a confidence interval for all the sub-runs is calculated. The confidence interval provides an estimate of the uncertainty around the sample estimate. In particular, given a confidence interval of 95%, the confidence interval is a random interval that will contain the true metric 95% of the time. A 95% confidence interval is calculated by.

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

Where $\bar{x}$ is the sample mean of all sub-runs, σ is the standard deviation of that sample, n is the size of the sample, so the number of subruns and z is the z-score corresponding to the confidence interval. For the 95% case, the z-score equals 1.96. The smaller the confidence interval, the more reliable the result of the simulation is. For this research, the confidence interval needs to be less than 0.1 day. So the width of the confidence interval needs to be.

$$W = 2 \cdot (z \cdot \frac{\sigma}{\sqrt{n}}) < 0.1d$$

To achieve this result it is decided to have 100 subruns, 1000 patients per subrun and 500 patients for warming. Which is equal to the number of simulations done in previous research in [5] and provides a significant confidence interval. Other metrics are then considered also to have a significant confidence interval. In total 100.000 patients are treated for the results of the simulation(s). It is checked whether the simulation ensures the system reaches its so-called steady state and converges to an average waiting time. This is done by calculating the cumulative mean of the waiting time at a subrun. Assume a sequence of n patients $x_1, x_2, \ldots, x_n$, the cumulative mean of the waiting time $\bar{W}_i$ for each index i (where i ranges from 1 to n) is defined as:

$$\bar{W}_i = \frac{1}{i} \sum_{j=1}^{i} x_j$$

Where:

- $\bar{W}_i$ is the cumulative mean up to index i.
- $\sum_{j=1}^{i} x_j$ denotes the sum of elements from the first element up to the i-th element. In this research, it is the mean of subrun j.
- i is the current index for which the cumulative mean is calculated.

In a subrun 1000 patients will be treated for measuring the performance, approximately $\frac{1.34+0.83+0.94+0.46}{1.34+0.83+0.94+0.46+1.91} \cdot 1000 \approx 652$ GR and High Complex patients are treated per subrun for example. It is checked whether, within this number of patients treated per subrun, the mean waiting time will converge.

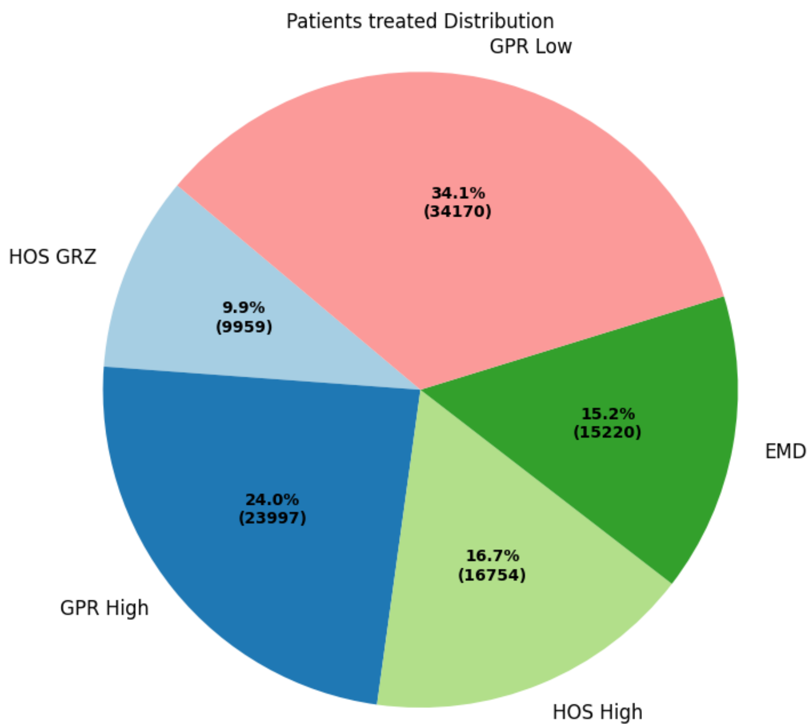

Figure 9: Pie chart of distribution patients treated in simualtion, NO-bed share (base case)

5.4.2 Dashboard

The Dashboard is programmed in Python with the Streamlit package. This package makes it possible to convert the interface to HTML so everybody can use it. The interface is shown in Figure 10. This dashboard has hidden input spots for the model input, divided by kind. Also, the locations can be named to make the interface more attractive. When the simulation is running, the effective beds are checked and the stability of the whole system is shown in dropdown menus.

The results are displayed under hidden blocks and divided into four sections. Namely, Waiting times (Wachttijden), hospital admissions (Ziekenhuisopnames), Length of stays (Verblijven) and Number of beds (Aantal bedden). The system is run by clicking on the 'Start Simulatie' button. The default values from Table 7 are also used as the default for the dashboard.

Output Measure	Explanation
Wait time Intermediate Care High	Waiting time of a patient who needs High Complex or GR treatment.
Wait time from hospital GR	Waiting time for patients arriving from hospital with GR care.
Wait time from hospital High	Waiting time for patients arriving from hospitals with High Complex care.
Wait time from GP High	Waiting time for patients arriving from GP High Complex.
Wait time from EMD	Waiting time for patients arriving from the Emergency Medical Department (EMD).
Wait time from Intermediate Care Low	Waiting time of a patient who needs Low Complex treatment.
Wait time to TRW	Waiting time of a patient, which needs observation to be observed. If necessary.
Occupancy rate High	Occupancy rate for High Complex Intermediate Care.
Occupancy rate Low	Occupancy rate for Low Complex Intermediate Care .
Number of patient replacements	The average number of replacements or transitions of a patient during the process.
Los High and Low Complex	Average length of stay for patients Intermediate Care High or Low Complex.
Percentage with hospital admission from EMD	Percentage of patients from EMD who get a (redundant) hospital admission.
Percentage with hospital admission from hospital	Percentage of patients from EMD who get a (redundant) hospital admission.
Number of patients with hospital admission from EMD	Number of patients from EMD who get a (redundant) hospital admission.
Number of patients with hospital admission from the hospital	Number of patients from the hospital who get a (redundant) hospital admission.
Service level	Percentage of patients that have been treated within 3 days.

Table 6: Explanation of Output Measures

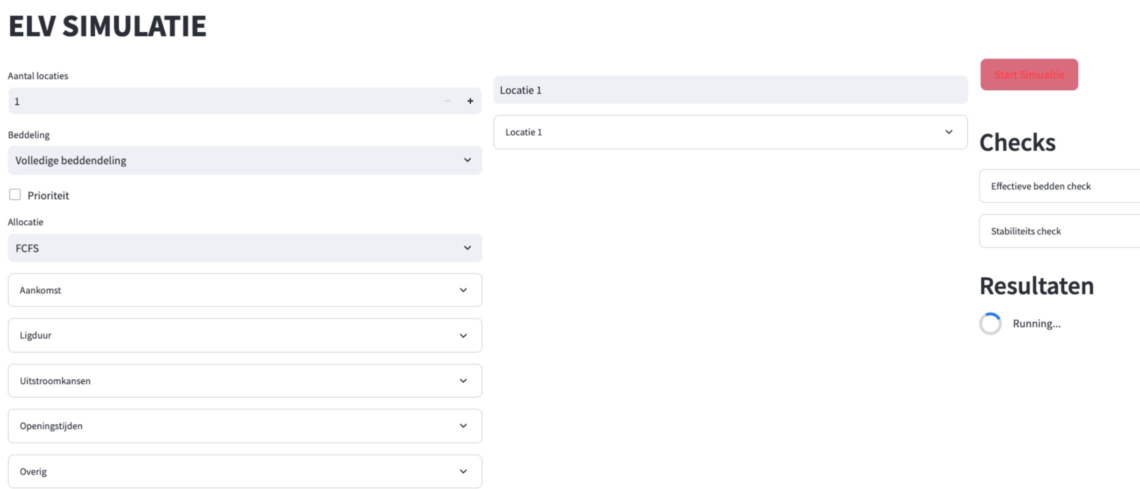

Figure 10: Interface Dashboard (in Dutch)

6 Results

In this section, the results are shown. First, the results of the current policy are explained. A sensitivity analysis shows how the model reacts to increments and decrements of arrivals and services. Then the results of the interventions are shown and evaluated. At last, the validation and verification are done. The Figures in this section contain a double y-axis. The Tables rounded the results to 3 decimals and are considered to have a significant confidence interval (unless mentioned) of 0.1 days for the waiting times as discussed in section 5.4.1. The other metrics are then also considered to have a significant 95% confidence interval.

6.1 Current practice and convergence

The current practice of admission at Intermediate Care in the Twente is a NO-bed share. This means no beds are shared among caretypes, so Low Complex High Complex and GR care have their beds. The number of beds was determined by stability and the Square root staffing principle. For modelling the Square root staffing principle was used because safety capacity was expected to be needed in this model. Table 5 in section 4.3 shows the distribution of beds for both types of calculations. The results of the simulation are shown in Table 13. The table shows that the waiting time for GR care is not stable when 21 GR beds are used. It is extremely high and was shown to have a large confidence interval. The Square-Root-Staffing method showed significant confidence intervals and is therefore used to model the interventions.

The table also shows that the average waiting time for GR and High Complex care is 2.02 days and Low Complex patients wait 1.80 days on average. The percentage of patients with a hospital admission is 2.4% from the EMD and 1.6% from the hospital. So from the hospital 1.6% gets an extension of stay in the hospital. The occupancy rate is 84% at Intermediate Care.

Convergence Looking at Figure 11, it can be seen that for all caretypes, the mean waiting time will converge to the mean at the red-dotted line. For GR care it is less convincing than for the other caretypes. However, considering a 0.1-day interval, the waiting times have converged. Figure 9 shows the distribution of patients treated in the simulation. In the pie chart, it can be seen that from GR care the least patients were treated. This can explain the less convincing convergence. In Figure 26 a density plot of occupied beds is shown after the simulation. Here it can be seen that on average of the 223 beds (including emergency beds), 185 are occupied, which will give the approximate 84% occupancy rate.

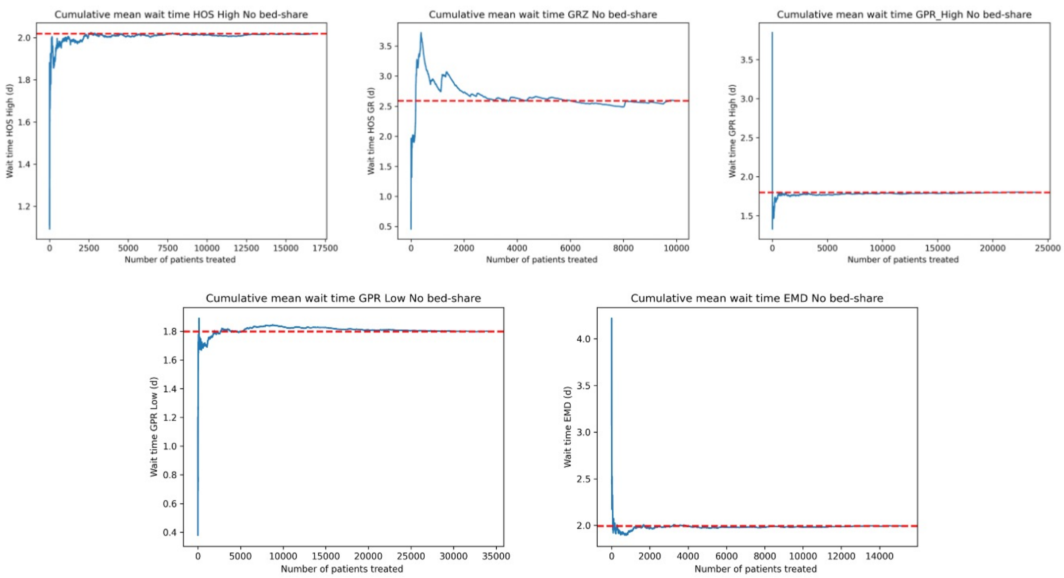

Figure 11: Cumulative mean of the waiting times No bed share.
(UL: HOS High, M: HOS GR, UR: GPR High, DL: GPR Low, DR: EMD)

6.2 Sensitivity analysis on arrivals and service times

To check the model's sensitivity, several scenarios are run to check how the model reacts in case of changes in arrivals and service times. The base is set to a NO bed share scenario with 75 Low Complex beds, 120 High Complex beds and 26 GR beds for stability. This was following the queuing theory discussed in section 3 the capacity needed following the square-root staffing principle. All in one combined Intermediate Care location.

- Increase of service and arrival of 20% and 10%. This simulates a sudden increase in arrival (for example a COVID-19 crisis) or an increase in average service time (for example when fewer nurses are available).
- Decrease of service and arrival of 20% and 10%. This simulates a sudden decrease in arrival or service.

The results are shown in Table 10, and 15 and Figure 12 and 13. Those tables and figures show that when the arrivals and service times increase by 10% the waiting times for Low, GR and High Complex care increase, but are still considered stable because of the 0.1 days confidence interval. However, the system shows signs of instability when a 20% increase in arrivals or service times occurs, especially in GR care. Waiting times increase heavily and the confidence interval gets wider. This indicates that when this is expected in the future, the current capacity may not be enough to handle the load and sharing or other interventions may be necessary.

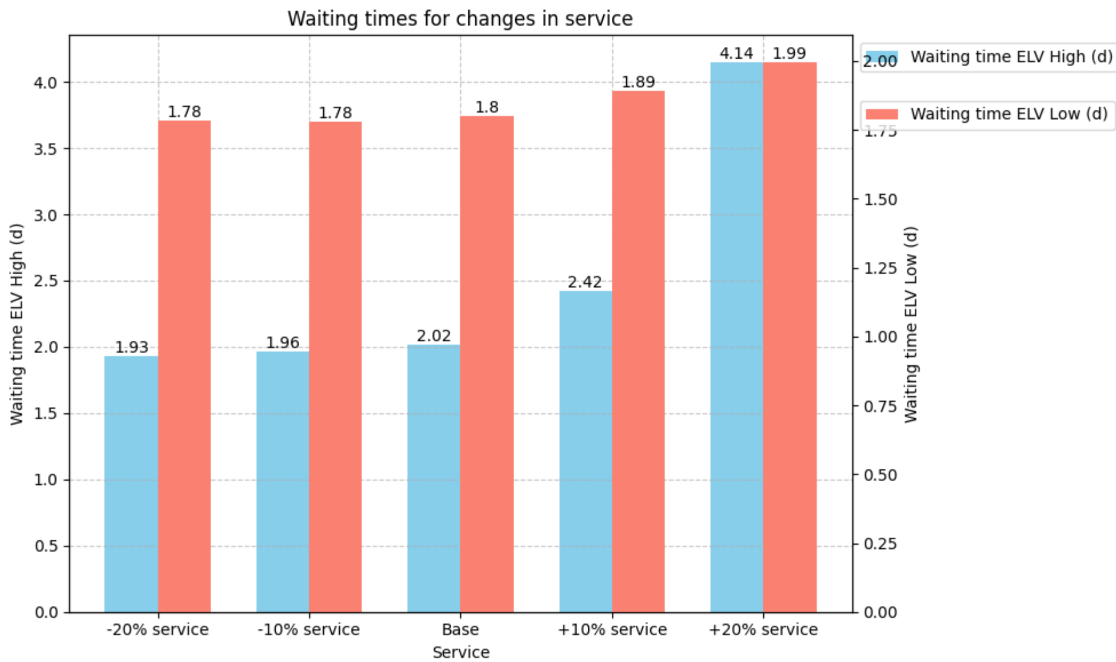

Figure 12: Waiting times for changes in service

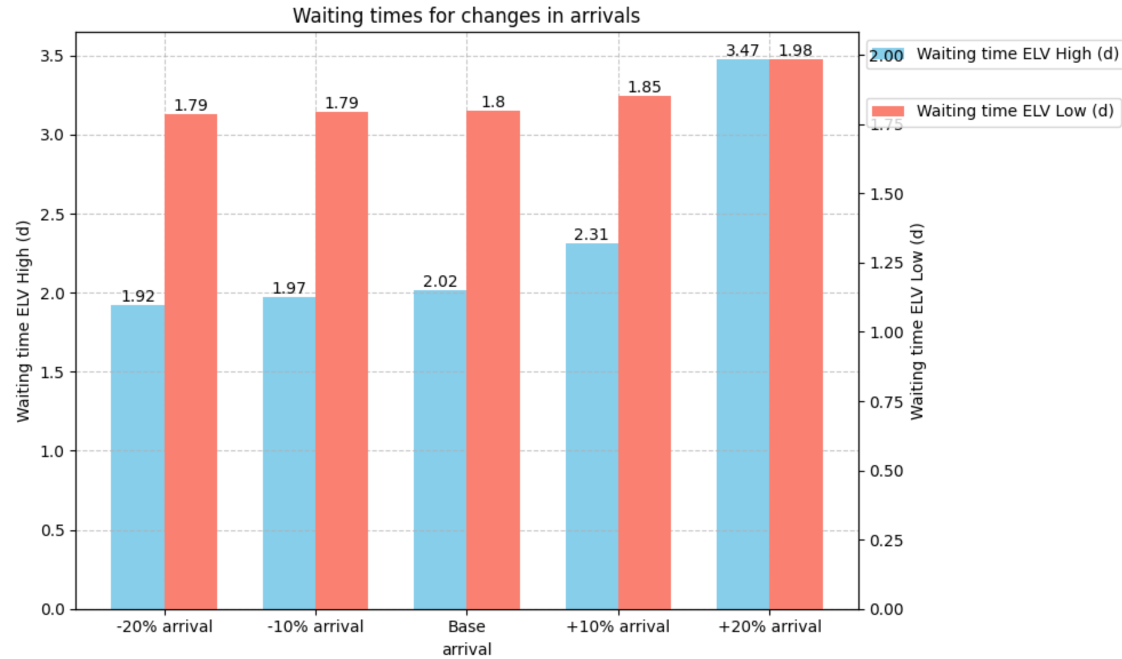

Figure 13: Waiting times for changes in arrivals

Conlcusion This subsection concludes that when an increase of 20% in arrivals or service times occurs, the current capacity could not be enough to handle the load offered. In this case sharing, centralization or another intervention in the system should be necessary to handle the bottleneck with the current capacity.

6.3 Effectiveness of interventions

In this subsection, the results of the interventions are discussed. Starting with the influence of bed-sharing on waiting times, followed by the centralization of Intermediate Care. At last, the results of the small interventions will be shown and discussed.

6.3.1 Intervention 1: bed-sharing

As said, four different types of bed-sharing are considered.

- **Total bed-sharing** indicates sharing all beds in the Intermediate Care location. So Low Complex beds, GR beds and High Complex beds are combined, and the emergency beds are also used for regular admission. As said before this can be done theoretically, but delivers challenges practically. This scenario considered 223 beds in total at the Intermediate Care.
- **GR-High Complex-bed-sharing**, this type of sharing considers two groups Low Complex and High Complex plus GR care. This is practically the maximal sharing possible in the Intermediate Care locations. This scenario considered 146 GR and High Complex beds and 75 Low Complex beds.
- **NO bed-sharing**, this sharing policy considers no sharing among Low Complex, High Complex and GR care. So three different bed groups are considered. In this scenario, 120 High Complex beds, 26 GR beds and 75 Low Complex beds are considered.
- **Partial bed-sharing**, this type of sharing adds one group compared to the NO sharing policy, namely shared beds. These beds are used to share among High Complex and GR patients. For the results 40% sharing among GR and High Complex is considered. Motivation is given below.

These are all combined in one combined Intermediate Care location, so a centralized scenario is considered.

NO bed-sharing Table 14 and Figure 14 consider GR and High Complex bed distributions. As can be seen, the waiting times for GR patients can increase quickly when not enough beds are available, less than 23. The table shows that the waiting times explode when more GR beds are taken away also the confidence intervals are wider. This indicates the instability of the system and when more simulations are considered the waiting times for GR increase even more. It supports the fact that safety capacity is needed compared to the theoretical capacity needed.

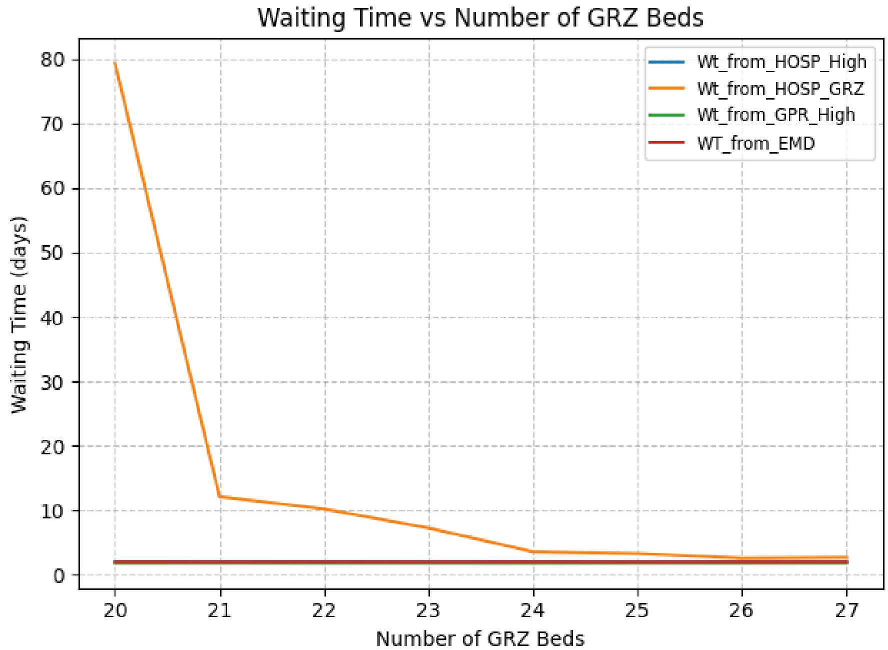

Figure 14: Waiting time for different numbers of GR beds

Low Complex For the waiting times of Low Complex patients, Table 18 shows that the waiting times increase when a Total bed-share scenario is considered. This sharing scenario was the only one affecting the Low Complex patients and is proven to have a negative influence on the waiting time(s). The Table also shows that the average waiting time for Low Complex patients is around two days for the other sharing interventions.

Total sharing Table 18 show that a Total sharing scenario is inefficient. When all caretypes share their beds, the waiting times will increase. In contradiction, the percentage of hospital admissions dropped. But following the results it can be concluded that sharing all beds is not efficient. Literature supports this finding, for example, Smith and With et al. (1981) [26] stated that full flexibility of resource sharing is not always beneficial. And they state that it can be explained by different service time distributions, which is now the case. Not all caretypes have the same service time. This also explains why the confidence interval is a bit wider than the other sharing types. However because the lower bound of the interval is higher than the upper bound of the other sharing types, the conclusion stays the same. For further research, the simulation can be expanded for total sharing.

Partial bed-sharing In Figure 15 and Table 17 the results are shown for various distributions of the partial beds. This is done by increasing the amount of partial beds with 10% every step until fully partially shared. As can be seen, with little sharing big results can be achieved. With 40% of sharing almost all reduction of waiting time advantages can be realised. The reduction with

40% sharing is 3.8 hours on average. Resulting in 7.8% of the total waiting time and 93.2% of the possible advantages of sharing all GR and High Complex beds. The table also shows that the percentage of redundant hospital admissions from the EMD and hospital decreased using partial sharing. This type of sharing is thus proven to add flexibility to the system.

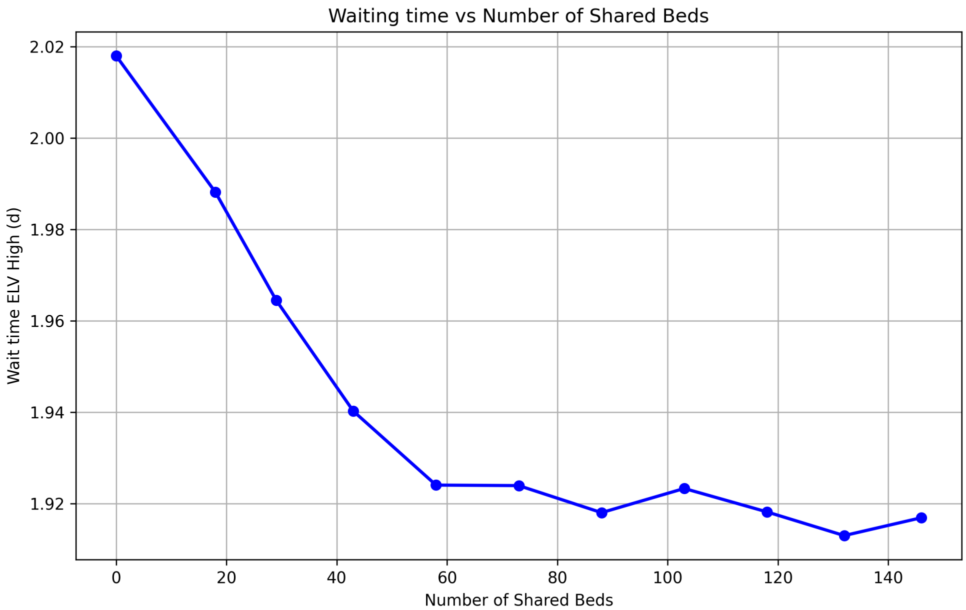

Figure 15: Waiting times for bed-sharing interventions for multiple numbers of shared beds

Findings The literature stated that flexibility will reduce waiting times. So more sharing can be effective in terms of waiting time(s). Also, little sharing can already give significant improvements. The results are shown in Figure 16 and Table 18. GR and High Complex sharing is in terms of waiting times and percentage of hospital admissions from the hospital, more effective than the other types of sharing. This type of sharing thus reduces the percentage of patients who are admitted to the hospital to wait for an Intermediate Care bed by 1.6% to 0%.

As said before, total sharing is practically difficult and is shown to not have positive results. Regarding waiting times and practical implementations, combining GR and High Complex care give the best results. In terms of waiting times, this result can almost be achieved by 40% partial sharing.

Conclusion This intervention concludes that a little sharing can already be effective in terms of waiting times. In terms of redundant hospital admissions combining the High Complex and GR departments can reduce the percentage from the hospital from 1.6% to 0%. Overall it proves that sharing is a nice way to reduce waiting times with the same capacity.

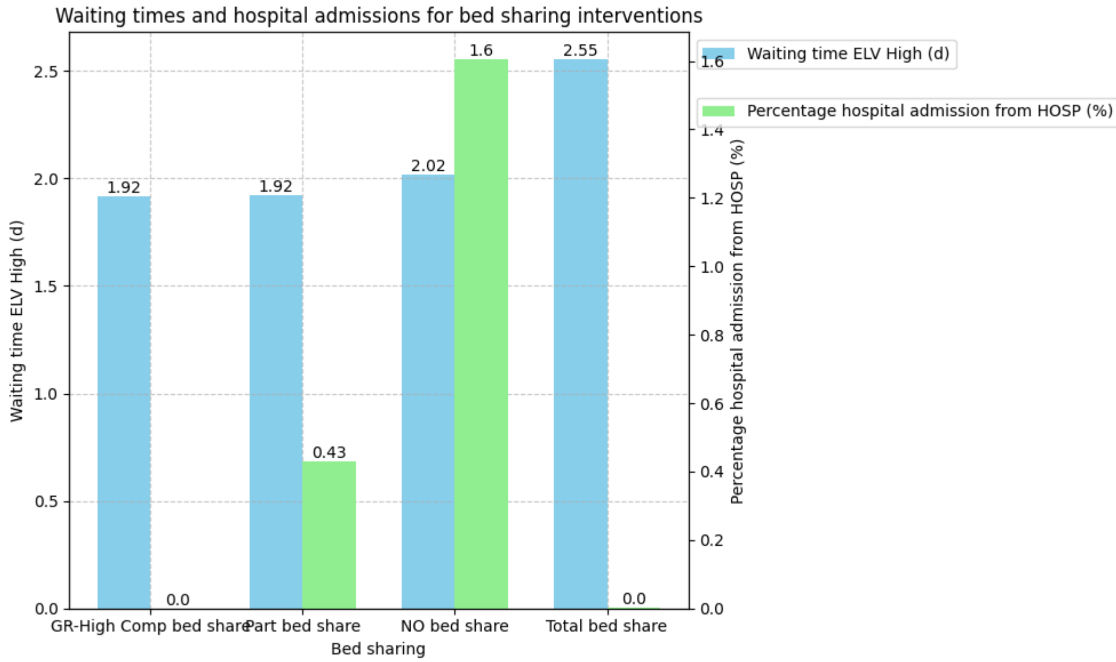

Figure 16: Waiting times and hospital admissions for bed-sharing interventions

6.3.2 Intervention 2: centralization and allocation

This intervention first looks at whether combining all locations of Intermediate Care is beneficial. On the other hand, this intervention looks at the allocation policy of patients.

Number of locations The results of extending locations of Intermediate Care into equal locations are shown in Figure 17 and Table 9. These patients are allocated FCFS. In the Table and Figure it can be seen that the waiting times drop a little with extending the number of locations but the hospital admissions increase. An explanation can be found by the distribution of the beds among the locations, which was now equal, and the fact that for heavy traffic considering multiple locations can be beneficial according to the literature [20]. Also, the arrivals were distributed randomly across the different locations. In reality, bigger locations get more arrivals and the number of arrivals per location is also dependent on the geographical location.

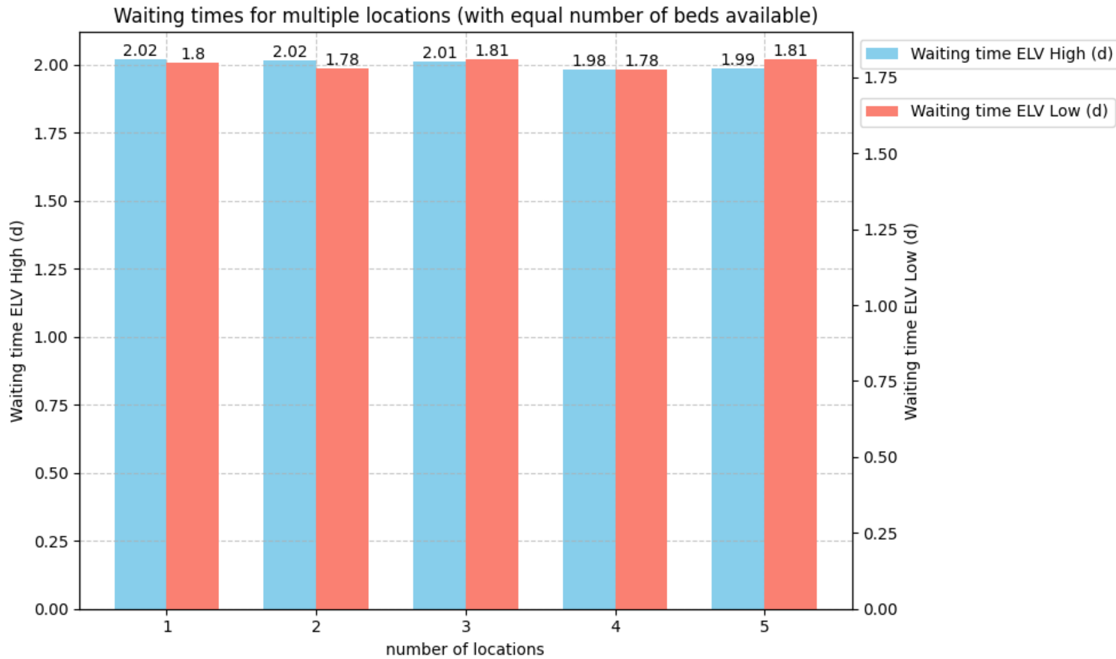

Figure 17: Results of expanding the number of locations. (All have an equal number of beds)

Allocation In terms of allocation, three different allocation scenarios were considered. Namely

- **First Come First Served**, this policy indicates that whenever a bed is free at whatever location, the client who waited the longest is assigned first. This means that not every patient can get admitted to their preferred location.
- **Only preference**, this policy indicates that patient can only be admitted to their preferred location. Patients with shorter waiting times can be admitted earlier because they are waiting for space at their preferred location.
- **Modelled allocation (overflow strategy)**, this is a developed model, shown in algorithm 4 in this thesis taking preference into account but also looking at fair allocation. This means not letting a patient wait too long or skip patients endlessly. This overflow strategy is discussed in section 4.4

The results are shown in Figure (s) 18, 19 and Table 16. The allocation model developed approaches the fairest allocation possible, but FCFS is shown to be the most efficient in this case in terms of waiting time. Literature proves that more research on the allocation model can make the allocation model approach or even beat the FCFS policy. The only preference policy shows signs of instability looking at the high waiting times and wide confidence interval, so cannot be assumed to be significant.

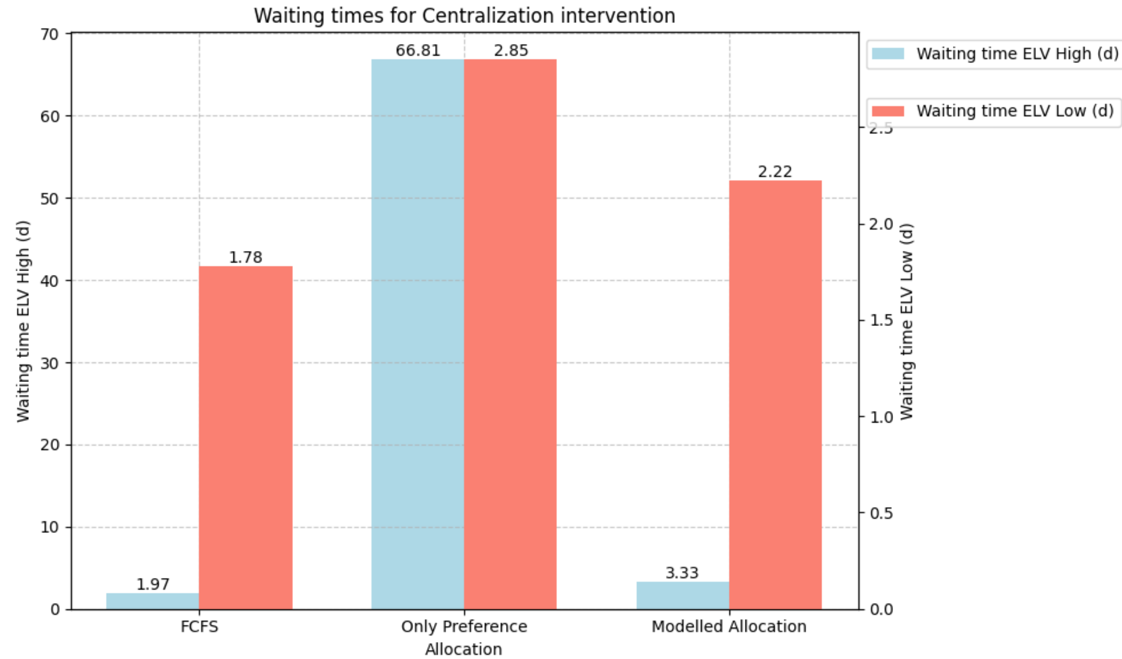

Figure 18: Waiting times for allocation research

When looking at the fairness and happiness in Figure 19 it can be seen that the happiness is the worst for the FCFS policy. This percentage indicates that on average 33% is allocated in a location the patient did not prefer before. When looking at the only preference policy, it again implies instability. The fairness of this policy is very high, indicating patients have to maximally wait infinitely long for treatment when they only want to be treated at their preferred location. From here it can be concluded that the metric of fairness is practically not the right measure and needs refinement in further research.

Again because no data was available on the distribution of beds, a designed situation is used for defining different locations. What's interesting is that the allocation model still generates a high waiting time, but reduces the average percentage of patients who are treated at locations they did not prefer by 25%. Despite that, the results indicate the overflow strategy needs refinement.

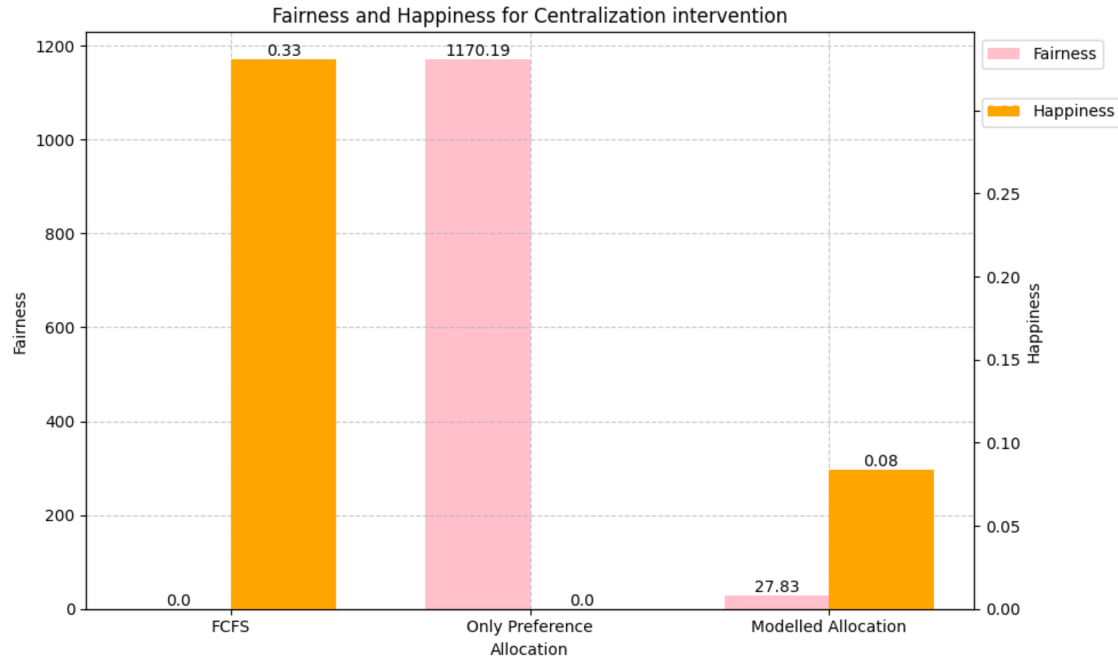

Figure 19: Fairness and happiness for allocation research

Conclusion This subsection concludes that decentralizing Intermediate Care can reduce the waiting time (on a low scale). This is supported by literature, which states that dividing the load over multiple locations can positively affect waiting times for high-traffic scenarios (which is now the case) [20]. In this case, a designed situation (saying equal size of locations) was used because no data was delivered on the distribution of beds among the locations. When looking at decentralized Intermediate Care, an allocation model shows the potential to find the balance between the fairness and happiness of the allocation of Intermediate Care for multiple locations but needs more research. Because, again, no data was delivered about the distribution of beds among locations also a designed situation (different number of beds per location) was used. This makes the results less significant and gives potential for further research.

6.3.3 Intervention 3a: Triage ward

For this intervention, multiple values of several observation beds are considered. Figure 20 and Table 8 show the results. The average waiting time increases when the number of observation beds increases. This can be explained by the random number of patients needing observation. Now 10% are deemed to have observation, but when more observation beds are considered, the number of beds available for Intermediate Care decreases. This will explain a sudden increase in waiting time(s). Also when considering a split in observation or not the system gets more complicated and delivers a new optimization problem. This intervention did not influence Low Complex patients.

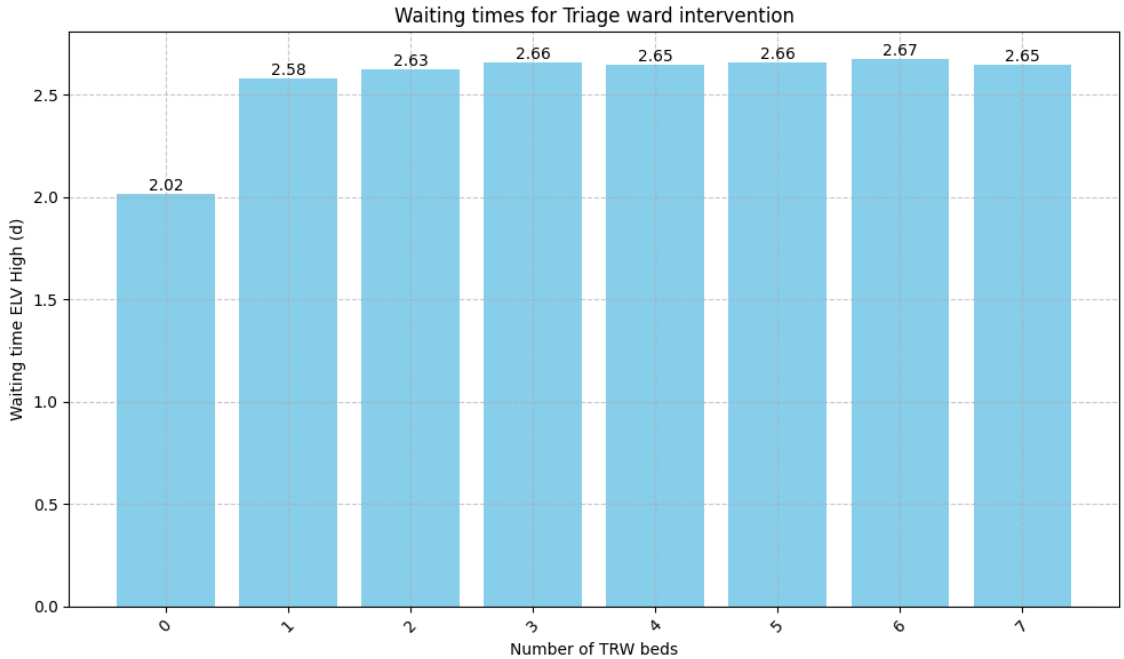

Figure 20: Waiting times for triage ward

Conlcusion To determine whether observation beds are beneficial for reducing waiting times in Intermediate Care, it is essential to establish the percentage of patients requiring observation beds and the number of such beds available. Additionally, it is crucial to consider that increasing the number of observation beds will reduce the availability of High Complex beds.

6.3.4 Intervention 3b: priority queues

GR-High Complex share This intervention prioritises patients from a hospital or EMD over patients from the General Practitioner. In Table 20 and Figures 21 and 22 the results are shown. The waiting time is minimal because a GR-High Complex share scenario is considered. So in terms of waiting time(s) not much changes, for hospital admissions it seems prioritizing only delivers bottlenecks for admissions from the EMD. A decrease in waiting times for highly complex patients in hospitals and an increase in waiting times for patients admitted by the GP can be seen, but no conclusions can be made based on that.

Partial sharing The results of using a 20% or 10% partial share and priority are shown in Table 20. It can be seen that for a 20% partial sharing scenario, not much changes. The same conclusion as for a GR-High Complex share can be taken. But for a 10% share, it can be concluded that the waiting times increase and the percentages of hospital admissions also increase. So prioritizing only delivers bottlenecks in this case.

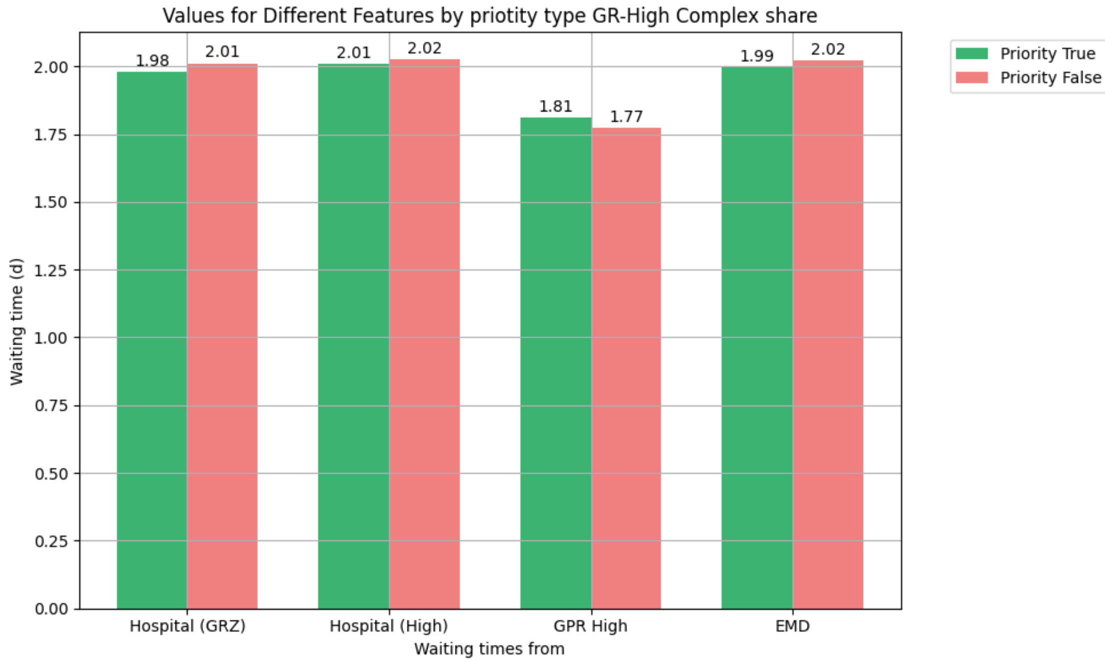

Figure 21: Barplot Priority GR-High Complex share

Conlcusion For GR-High Complex and 20% partial share, no clear influence can be concluded on the influence of prioritizing patients from the EMD and hospital. But for a 10% partial share between GR and High Complex care, it can be concluded that hospital admissions and waiting times increase. But as a side note, these waiting times are already close to the minimum, which implies that the chance of priority having a positive influence was quite small already.

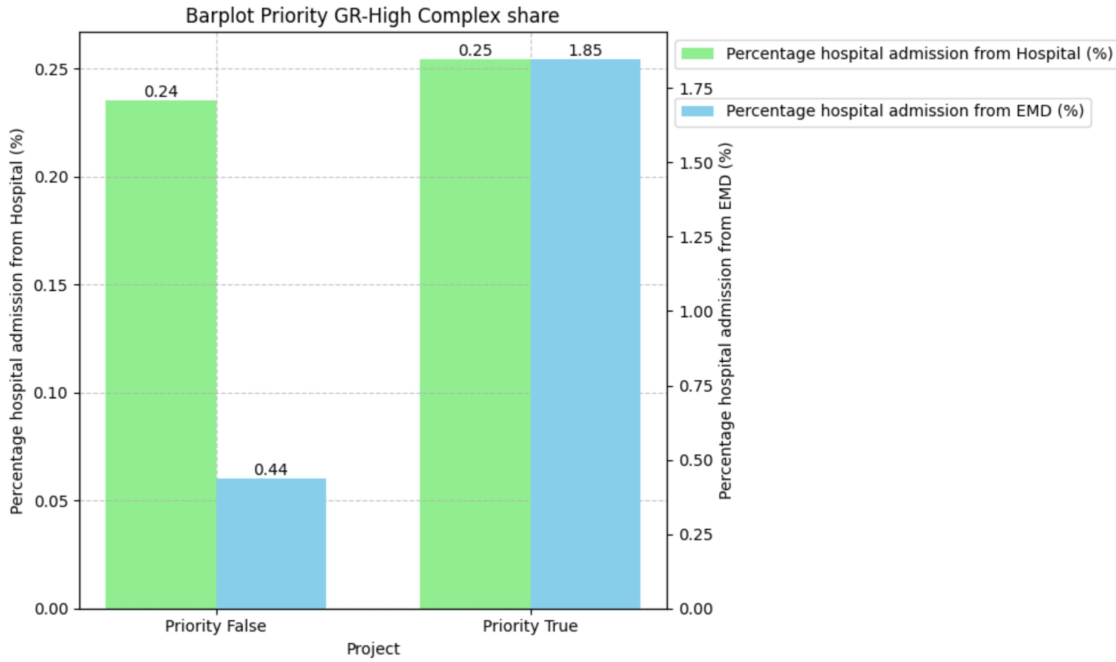

Figure 22: Barplot Priority and waiting times

6.3.5 Intervention 3c: admission hours

First, how much waiting time is expected because of opening hours is determined to determine the effect of admission hours. For this situation (open 8-17h and closed during weekends) the flow of waiting through admission hours is shown in Figure 23. The expected waiting time is given by the area under the curve which is,

$$\mathbb{E}(\textit{Waiting time due to admission hours}) = \frac{\frac{1}{2} \cdot 63 \cdot 63 + \frac{1}{2} \cdot 15 \cdot 15 \cdot 4}{7 \cdot 24} = 14.49 \text{ hours} \approx 0.60 \text{ days}$$

Add the admission days (1.5d) to this, giving an average expected waiting time of approximately 2.1 days. Figure 16 shows the waiting times can be reduced to roughly that number. This intervention will look at lowering the admission hours or expanding the admission hours in the weekend, the waiting times will drop even more.

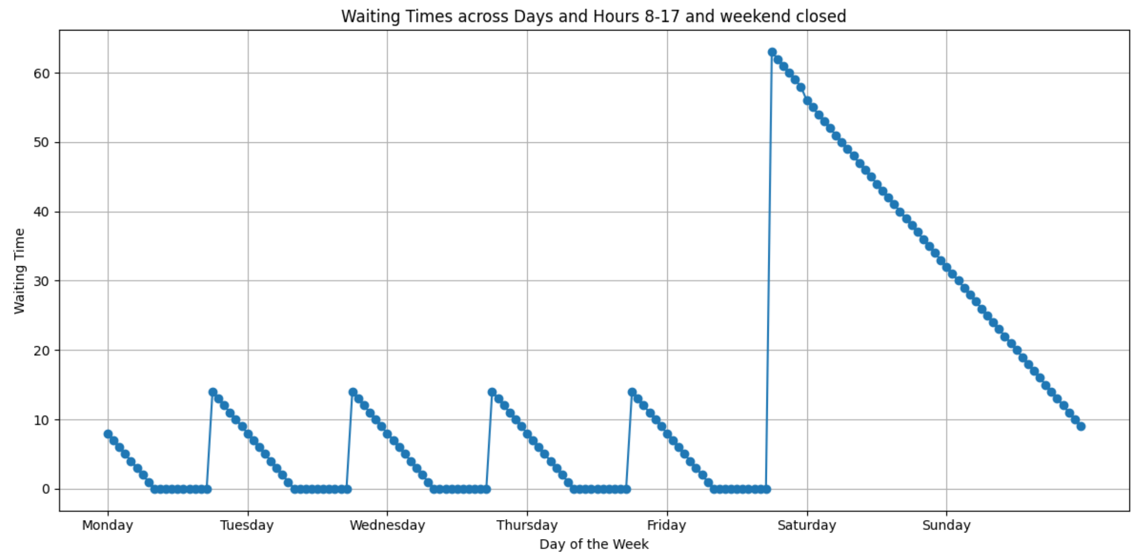

Figure 23: Waiting time because of admission hours

The results of the simulations are shown in Table 11 and Figure 24. As can be seen, most waiting times can be reduced by speeding up the admission process. Opening hours on weekend days, also affect the average waiting time. And when Intermediate Care is open 24/7 and has no transfer time, the waiting time is almost zero.

Conlcusion This intervention concludes that speeding up the admission process and extending admission hours in weekends can have a big contribution to reducing waiting times.

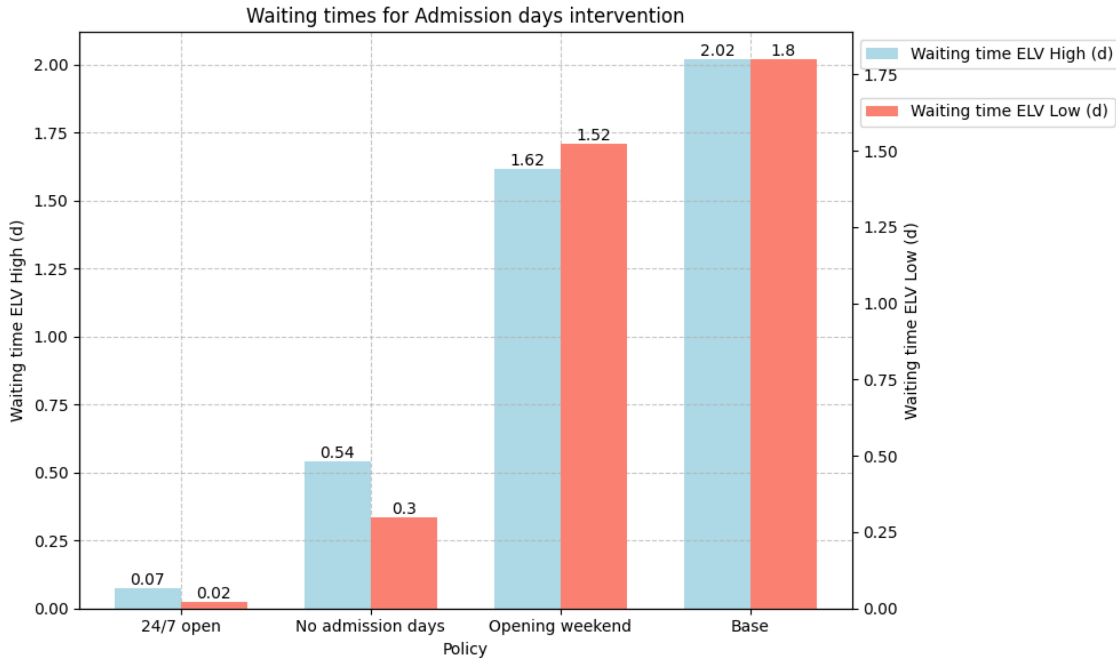

Figure 24: Waiting times for admission hours intervention

6.4 Validation and verfication

The validation is done in 2 parts, first, the model is compared to an Erlang C model, discussed in section 3. This is done by replicating the plot in Figure 4 checking an M/M/c scenario and comparing values. Secondly, the model is discussed by four different checks:

- A triage ward scenario with no observation beds, should be the same as a GR-High Complex share scenario.
- A partial bed-share scenario with no partial shared beds, should give the same result as a No bed-share scenario.
- A partial bed share scenario with all shared beds should give the same result as a GR-High Complex share scenario.
- Whenever 1 location is considered, the three different allocation types should give the same result. Because no other locations can be considered, the preferences of the patients cannot change.

The model cannot be verified properly, because not enough data is delivered. So the comparison with real data is not possible. To verify the model healthcare experts gave monthly input on the modelling and results.

6.4.1 Erlang C

When running the model posed with no admission days, only looking at High Complex care and 24/7 opening hours during the week. The model should return an M/M/c queue. To show that the Figure from the literature review is replicated.

Figure 25 shows this model can reproduce the plot from previous research of Vilaplana et al. [29] and Figure 4. This is done by decreasing the arrival rate and thus the load per bed with steps of 10%, and also by simulating more patients than for the interventions, for different numbers of beds. The maximum load is set to 95% because a load of 1 means an overload of the system. This Figure shows that waiting time increases later when more servers are considered.

Also, this Figure shows that when more beds are considered, more load per bed can be accepted. Moreover, the Figure shows that when considering more servers, the load per bed can be closer to one than 0.95 before the waiting times increase.

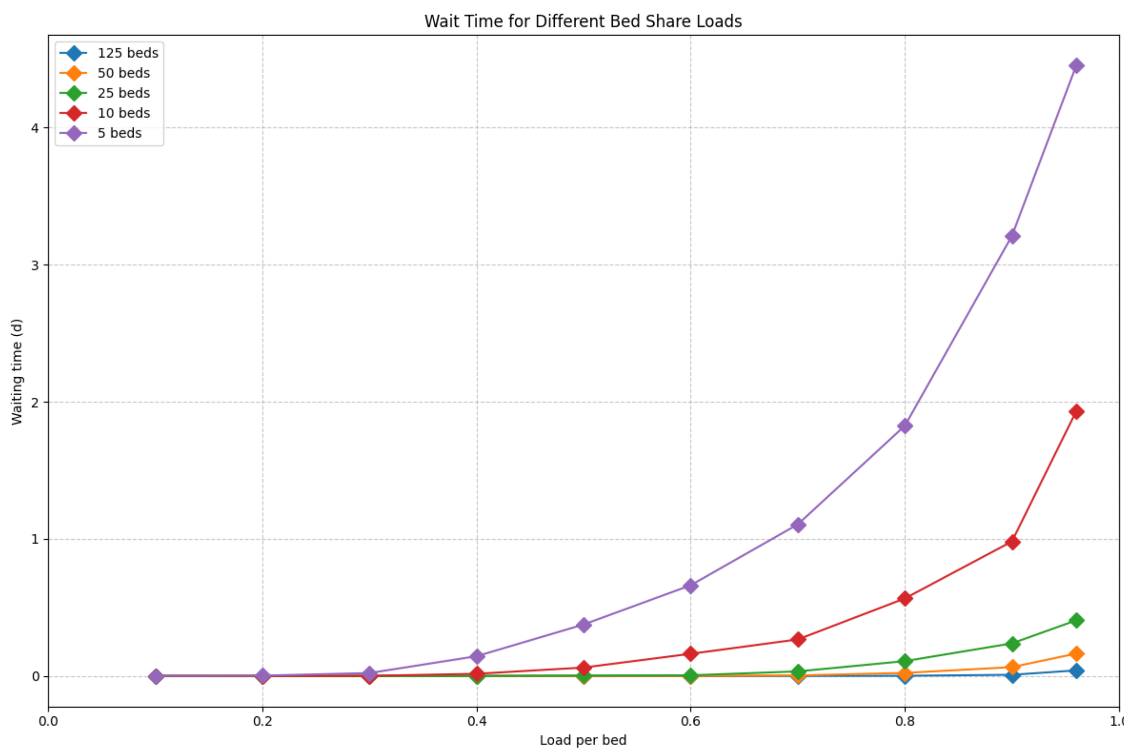

Figure 25: Waiting times for different bed loads in the model

Replicating the Figure shows roughly that the Erlang C model applies to the model proposed in this thesis. This verifies that the values, coming from the model are realistic compared to the theory. Even if the values in both graphs on the y-axis do not match, clearly the same patterns can be seen. The values on the y-axis do not match, because in this model a different service time is considered compared to the research of Figure 4.

6.4.2 Model comparrison

Tables 19, 21and 20 show the results of the comparisons in the model. The numbers are a match. This verifies the model. This also verifies the correct implementation of the model, which in turn implies that the results can be valid. To try a validation of the model, Figure 26 shows the density of occupied beds in a histogram. When this is compared to the occupied bed distribution of research in Amsterdam by Arntzen et al. (2023) [5] some similarities can be seen, but no conclusion can be made on that because no real data has been delivered.

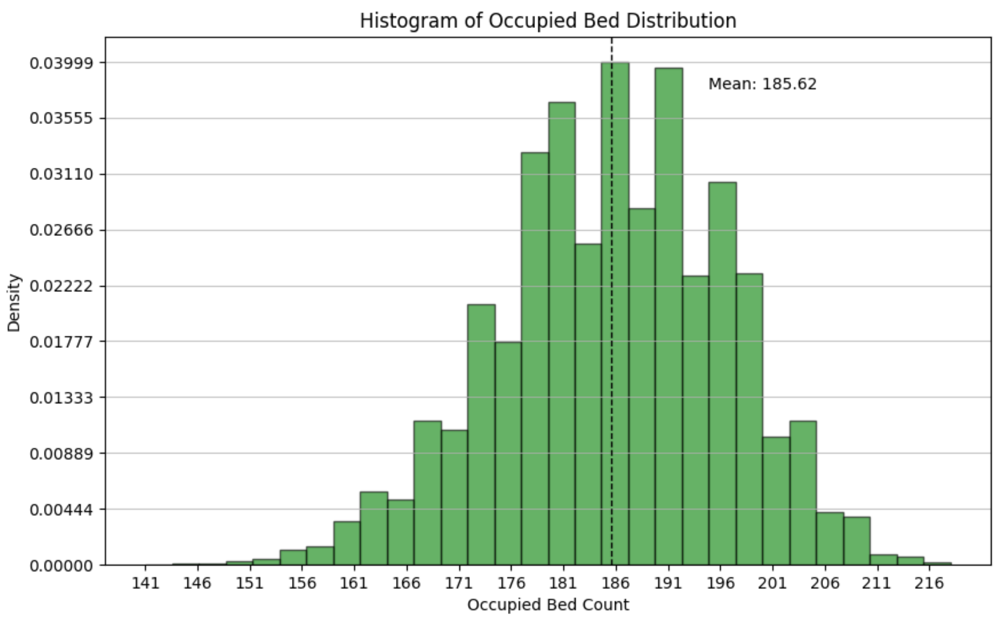

Figure 26: Histogram density of occupied beds

7 Conclusion

To conclude this thesis, the waiting times, redundant hospital admissions and multiple more metrics for Intermediate Care in the region of Twente can be reduced significantly by applying interventions in the system. In the case of bed-sharing partial sharing, with 40% of the capacity to be exact, can be considered to be the best way to maximize the flow in the system while minimizing the waiting times. Also, this way of sharing has more practical advantages than total sharing or GR-High Complex sharing. The 40% partial sharing of GR and High Complex care delivers 93.2% of the waiting time reduction fully sharing both departments delivers and reduces the average waiting time by 7.8%. Sharing all departments was proven to provide more bottlenecks than advantages. Other interventions, like priority or triage wards, have their advantages but are concluded to not have a major influence on the system. Dividing the Intermediate Care into more locations delivers an allocation problem, which can be solved using an allocation model. This finds a balance between fairness and happiness. However, changing the admission policy was shown to have the biggest influence on reducing the waiting times and maximizing the flow of the elderly through the system.

8 Discussion

This section discusses multiple possibilities for improving the model and achieving more insights.

- The main point of criticism is the number of patients simulated. For now, 100.000 patients are treated in total and the warming-up period is set to 500 patients. This is considered to achieve a significant confidence interval of 0.1 days maximally. The number of beds per caretype is chosen by looking at the stability of the system and adding and adding safety capacity following the Square-Root-Staffing principle. It is shown that this number needs to be chosen accurately because the system (especially for GR care) can explode quickly if not enough beds are available. For further research, more simulations can be considered to check whether the unstable results still are unstable after more simulations or converge.

- When the research is performed with fewer beds available at Intermediate Care, the same results can be obtained, but the impact of interventions is greater. But those results contain parts of instability because the number of beds available must be chosen accurately. Concluding that interventions can have major impact on waiting times even though not enough beds are available.

- The 95% confidence interval is chosen to be 0.1 days. This is a relatively large interval when compared to the two days of waiting time, which is the minimum. Further research should consider more simulations to make the interval tighter and more significant. This can also give different results and conclusions indicating the insights of the previous discussion point.

- When considering a 0.1-day confidence interval, this is compared to an average waiting time of 2 days 5% of that waiting time. This can be considered as a large part of the waiting time(s). So again for further research, this interval can be narrowed by increasing the number of simulations.

- This model uses designed situations for the centralization and allocation interventions, this example assumes arrivals and services which may not be realistic. Also, beds are distributed among different locations based on a designed situation. This makes it difficult to make recommendations on the policy for Intermediate Care. For further research declaration data of patients, available in the CBS (Central Bank of Statistics), can be used. For the multiple locations, the dashboard finds a nice solution. Here the supervisors from the region of Twente can fill in their bed distribution among their locations.

- This model assumes no abandonment of patients in queues. So when a patient needs care it gets care even though it might wait long. In reality, patients might abandon the system because of death or long waiting times. Further research might take abandonment rates into account for allocation or modelling interventions.

- During the allocation of patients, the Skill-Based-Routing principle is used for inspiration. This algorithm determines an optimal waiting time with limited staff knowing all skills. This research made an allocation model based on the SBR but did not minimize waiting time or maximize happiness or fairness. This algorithm made use of an overflow strategy, which means making a threshold of how long a patient wanted to wait before considering other locations. This was a nice approximation and a way to easily generate a solution to the problem, but optimization can be considered for further research.

- This research looks at the fairness of an allocation. This was defined by the maximum difference between an FCFS and the model considered. The advantage of this approach is the fact that when this value is minimized, the allocation is fair for all patients. However, the big disadvantage is that the maximum value is infinity. So it isn't easy to define when an allocation is fair and if the value considered is good. It is expected that the more patients are simulated, the higher the fairness value will be. For further research, it is suggested that a different definition of fairness be defined. Literature by Itzhak et al. (2008) [6] defined stochastic fairness, which can be interesting for further research.

- In section 6 an sensitivity analysis is performed. When looking at a 20% increase in arrivals and service times, it can be seen that the waiting times explode, especially for the GR department. This indicates that the results are not steady-state results and cannot be trusted. When more simulations are considered those waiting times will increase even more. The only conclusion that can be made in this subsection is the fact that when arrivals or service times increase, the load per bed will be too high.

- The fairness of allocation is modelled by setting a timestamp of when a patient on the waiting list is skipped. this is then considered to be the FCFS time of allocation and the fairness uses this for comparison. However, when looking at this modelling, the allocation of earlier patients determines the timestamp of other patients. This means that the timestamp is not exactly the time of allocation when FCFS is considered. Which delivers discussion for further research.

- This research assumes Poisson arrivals and exponential service times. This is assumed based on the literature and surveys. At CBS (Centraal bureau statistiek) declaration data is available for analyzing arrivals and service times within Intermediate Care. For further research, this dataset can be used for analyzing the distributions of the service times and arrival rates.

- In the allocation model, an overflow strategy is considered. This strategy is considered when a patient waits for too long, it will be handled FCFS. This so-called Utility depends on the waiting time linearly. For further research, it can be considered to look at different functions of Utility, for example, exponential or Lognormal. Exponential can be considered because the urgency for treatment gets bigger in more an exponential way than a linear way.

- For now the arrival rates are assumed to be constant. However, the arrivals at Intermediate Care are expected to be time-dependent. So a constant arrival rate per day can be doubted and is something to consider when doing further research.

- Figure 11 shows the convergence of the mean waiting time for the bed-share intervention. Convergence for GR care can be seen but not clear enough. Increasing the number of patients per subrun can make this mean converge better and tighten the confidence interval. A major disadvantage is that the running time will then increase which was for now not possible to do.

References

[1] ActiZ. Hoe staat het met: de wachtlijsten voor verpleeghuizen? Hoestaathetmet: dewachtlijstenvoorverpleeghuizen?|ACTIZ.(z.d.).Âăhttps://www.actiz.nl/hoe-staat-het-met-de-wachtlijsten-voor-verpleeghuizen#:~:text=Hoe%20lang%20zijn%20de%20wachtlijsten,nog%20eens%2011.835%20uit%20voorzorg. [Accessed 15-01-2024].

[2] Ivo Adan and Jacques Resing. *Queueing theory.* 1 2001.

[3] Rebekka Arntzen, René Bekker, and Rob van der Mei. Preference-based allocation of patients to nursing homes. *Available at SSRN 4670165.*

[4] Rebekka J Arntzen, René Bekker, Oscar S Smeekes, Bianca M Buurman, Hanna C Willems, Sandjai Bhulai, and Rob D van der Mei. Reduced waiting times by preference-based allocation of patients to nursing homes. *Journal of the American Medical Directors Association*, 23(12):2010–2014, 2022.

[5] Rebekka J Arntzen, Judith H van den Besselaar, René Bekker, Bianca M Buurman, and Rob D van der Mei. Avoiding hospital admissions and delayed transfers of care by improved access to intermediate care: A simulation study. *Journal of the American Medical Directors Association*, 2023.

[6] Benjamin Avi-Itzhak, Hanoch Levy, and David Raz. QUANTIFYING FAIRNESS IN QUEUING SYSTEMS. *Probability in the engineering and informational sciences (Print)*, 22(4):495–517, 9 2008.

[7] René Bekker, Ger Koole, and D. Roubos. Flexible bed allocations for hospital wards. *Health Care Management Science*, 20(4):453–466, 4 2016.

[8] Jinsheng Chen, Jing Dong, and Pengyi Shi. A survey on skill-based routing with applications to service operations management. *Queueing systems*, 96(1-2):53–82, 10 2020.

[9] CWI. Halvering wachtlijsten in de ouderenzorg mogelijk door nieuw wiskundig model — cwi.nl. https://www.cwi.nl/nl/samenwerkingen/showcases/cwi-sigra-and-amsterdam-umc-join-forces-to-improve-elderly-care/. [Accessed 15-01-2024].

[10] John N. Daigle. *Queueing Theory with Applications to Packet Telecommunication.* 1 2005.

[11] Aafke J de Groot, Elisabeth M Wattel, Romke van Balen, Cees MPM Hertogh, and Johannes C van der Wouden. Discharge to rehabilitation oriented care after acute hospital stay; association with vulnerability screening on hospital admission. *Annals of Geriatric Medicine and Research*, 2023.

[12] George S. Fishman. *Discrete-Event Simulation.* 1 2001.

[13] Gerald J. Lieberman Frederick S. Hillier. Introduction to operations research. http://www.maths.lse.ac.uk/Personal/stengel/HillierLieberman9thEdition.pdf. [Accessed 15-01-2024].

[14] Geoffrey Grimmett and Dominic Welsh. *Probability An introduction, second edition.* Oxford, 2017.

[15] Donald Gross. *Fundamentals of queueing theory.* John wiley & sons, 2008.

[16] Robert S Huckman and Darren E Zinner. Does focus improve operational performance? lessons from the management of clinical trials. *Strategic Management Journal*, 29(2):173–193, 2008.

[17] IKT. Ons netwerk - IKT.

[18] AJEM Janssen, Johan SH Van Leeuwaarden, and Bert Zwart. Refining square-root safety staffing by expanding erlang c. *Operations Research*, 59(6):1512–1522, 2011.

[19] Benjamin Legros, Oualid Jouini, and Yves Dallery. A flexible architecture for call centers with skill-based routing. *International journal of production economics*, 159:192–207, 1 2015.

[20] Avishai Mandelbaum and Martin I Reiman. On pooling in queueing networks. *Management Science*, 44(7):971–981, 1998.

[21] Paul E McKenney. Stochastic fairness queueing. In *IEEE INFOCOM'90*, pages 733–734. IEEE Computer Society, 1990.

[22] René J. F. Melis, Marcel G M Olde Rikkert, Stuart G. Parker, and Monique I. J. Van Eijken. What is intermediate care? *BMJ. British medical journal*, 329(7462):360–361, 8 2004.

[23] R. Kannapiran Palvannan and Kiok Liang Teow. Queueing for healthcare. *Journal of Medical Systems*, 36(2):541–547, May 2010.

[24] Hossein Pishro-Nik. Exponential Distribution | Definition | Memoryless Random Variable — probabilitycourse.com. `https://www.probabilitycourse.com/chapter4/4_2_2_exponential.php`. [Accessed 15-01-2024].

[25] J George Shanthikumar, Shengwei Ding, and Mike Tao Zhang. Queueing theory for semiconductor manufacturing systems: A survey and open problems. *IEEE Transactions on Automation Science and Engineering*, 4(4):513–522, 2007.

[26] David R Smith and Ward Whitt. Resource sharing for efficiency in traffic systems. *Bell System Technical Journal*, 60(1):39–55, 1981.

[27] Andrea Steiner. Intermediate care–a good thing? *Age and ageing*, 30(suppl 3):33–39, 8 2001.

[28] R Thijssen. Dolce vita: datagedreven optimalisatie voor de ouderenzorg. actie leer netwerk. `Juist.(2023,15mei).ÂăDolceVita:datagedrevenoptimalisatievoordeouderenzorg.ActieLeerNetwerk.Âăhttps://www.actieleernetwerk.nl/dolce-vita-datagedreven-optimalisatie-ouderenzorg/`. [Accessed 15-01-2024].

[29] Jordi Vilaplana, Francesc Solsona, Francesc Abella, Rosa Filgueira, and Josep Rius. The cloud paradigm applied to e-health. *BMC medical informatics and decision making*, 13:1–10, 2013.

[30] Jean-Louis Vincent and Gordon D. Rubenfeld. Does intermediate care improve patient outcomes or reduce costs? *Critical care*, 19(1), 12 2015.

[31] Rodney Wallace and Ward Whitt. Resource pooling and staffing in call centers with skill-based routing. 04 2004.

[32] DJ Worthington. Queueing models for hospital waiting lists. *Journal of the Operational Research Society*, 38(5):413–422, 1987.

[33] John Young. The development of intermediate care services in England. *Archives of gerontology and geriatrics*, 49:S21–S25, 12 2009.

[34] John P Young. Stabilization of inpatient bed occupancy through control of admissions. *Hospitals*, 39(19):41–48, 1965.

9 Appendix

Algorithm 1: DES-model pseudocode

```
Procedure simulateELV
Check effective beds;
if Effective beds < beds available then
    beds available = effective beds;
    else
        beds available = beds available
    end
end
Check system is stable;
if System is not stable then
    print "Error, system is not stable"
end
Initialize global simulation parameters;
for each simulation sub-run do
    Initialize specific sub-run variables;
    Set counter and trackers for the sub-run;
    while sub-run condition is not met do
        Identify the next event to occur;
        Process the event;
        Update the state of involved patients;
        Modify relevant data structures;
        Schedule future events if necessary;
        Update simulation metrics;
        Calculate and record metrics;
    end
    Aggregate sub-run statistics;
end
Compile overall simulation results;
Return compiled results;
```

Algorithm 2: Arrivals-DES Model pseudocode

```
Procedure ProcessArrival
Create new client record;
Determine the source of arrival (e.g., hospital, EMD, GP High Complex or GP Low
  Complex);
if client arrives from EMD then
    Determine if hospital admission is needed or direct placement at Intermediate Care is
      possible;
    if hospital admission is needed then
        if Intermediate Care is closed and the Emergency bed is free then
            Assign Patient to Emergency bed;
            Schedule future events for discharge from Intermediate Care;
        end
        else
            Schedule a hospital admission event;
            Add patient to waiting list 3;
        end
    end
    else
        Check availability for Intermediate Care bed;
    end
    if bed is available then
        Place the patient in Intermediate Care;
        Schedule future events for discharge from Intermediate Care;
    end
end
else if Patient arrives from hospital (GR or High Complex) then
    Determine if direct Intermediate Care admission is possible;
    if bed is available then
        Place the patient in Intermediate Care;
        Schedule future events for discharge from Intermediate Care;
    end
    else
        The patient stays in the hospital and is assigned to the waiting list 3 or 4;
    end
end
else if Patient arrives from GP (Low Complex or High Complex) then
    Check the availability of beds at Intermediate Care;
    if bed is available then
        Assign the patient to the Intermediate Care bed;
        Schedule future events for discharge;
    end
    else
        Assign patient to waiting list 1 or 2, wait from home;
    end
end
```

Algorithm 3: Process Discharge Procedure

```
Procedure ProcessDischarge
Identify the client to be discharged for either observation bed, emergency bed, hospital or
  Intermediate Care bed;
Determine the client's next destination after discharge;
if next destination is a care facility (e.g., Intermediate Care) then
    Check for available beds in the care facility;
    if bed is available then
        Place the client in the care facility;
        Schedule future events for further treatment or final discharge;
    end
    else
        Add client to the care facility waiting list;
    end
end
else if next destination is home or another non-medical facility then
    Update client record to indicate discharge to home;
    Finalize the client's discharge from the healthcare system;
end
Update overall system metrics and records;
if patient is discharged then
    if patient is discharged from Emergency bed then
        Leave bed free for emergencies.
    end
    if patient is discharged from Intermediate Care bed then
        Check if someone from the waiting list can replace the discharged patient.
    end
    if patient is discharged from observation bed then
        Check if someone from the waiting list or hospital needs observation and can replace
          the discharged patient
    end
end
```

Algorithm 4: Allocation model for multiple locations, overflow strategy

```
Procedure ProcessAllocation
The arrival of patient;
Assign a preferred location to the patient where the type of treatment needed is offered;
if Arrival at location i ∈N then
    if Intermediate Care bed is available immediately at preferred location then
        Assign the patient to Intermediate Care at that location;
    end
    else if Patient arrives from EMD then
        Assign patient to waiting list 3 plus hospital admission;
    end
    else
        Assign patient to their respective waiting list and wait for the Intermediate Care bed
          to be available;
    end
    if Patient assigned to waiting list then
        Set utility for the patient;
        Determine possible alternative locations;
        Set Alternative location to False;
    end
end
if Discharge at location i ∈N then
    Check who is the first patient on the waiting list;
    for All patients on waiting list (in order of waiting time) do
        if First waiting patient's preferred location is i or waiting time in queue > utility
          (e.g Alternative location is true) then
            Assign bed at location i to that patient;
        end
        if Bed is assigned because Alternative location is True then
            Add "Transfer" to the patient's journey;
            Change the patient's preferred location;
        end
        else
            Check for all clients on the waiting list if possible to assign a bed to them based
              on preferred location or utility;
        end
        if Bed is assigned to patient on waiting list then
            Check if the patient was the first patient on the list;
            if Patient was not first on waiting list then
                Patients were skipped because of location;
                Start extra waiting time for patients who were skipped, because they did not
                  want to be assigned to location i;
            end
        end
    end
end
```

Variable Name	Description	Default Value
Scen_shared_beds_Full	Full sharing scenario for beds	FALSE
Scen_NO_Sharing	No sharing scenario for beds	TRUE
Scen_Part_bed_share	Partial sharing scenario for beds	FALSE
Scen_Triage_ward	triage ward scenario	FALSE
Scen_Total_Sharing	Total sharing scenario for beds	FALSE
Priority	Priority of scheduling	False
Preference	Preference setting	FCFS
n_loc (C)	Number of locations	1
$n_{i,.}$	**Number of beds caretype i at location .**	
n_beds_ELV (N)	Number of ELV beds [[[GR,H.C],L.C]]	[[[21,111],67]]
n_beds_TRW	Number of TRW beds	[0]
n_beds_Emergency	Number of emergency beds	[2]
n_nurses ($q_{r,c}$)	Number of nurses [[[GR,H.C],L.C]]	[[[40,30],10]]
n_subruns	Number of subruns	100
n_clients_per_subrun	Number of clients per subrun	1000
n_clients_for_warming	Clients for warming	500
$\lambda_{i,j,.}$	**Arrival rate of caretype i inflow j location .**	
arr_HOS_High	Arrival rate of HOS (High)	0.94
arr_GPR_High	Arrival rate of GPR (High)	1.34
arr_EMD	Arrival rate of EMD	0.83
arr_HOS_GR	Arrival rate of HOS (GR)	0.54
arr_GPR_Low	Arrival rate of GPR (Low)	1.91
$p_{i,k}$	**Outflow probabilities caretype i to k**	
out_p_Home_High	Outflow probability to home rate (High) (%)	0.578
out_p_Home_GR	Outflow probability to home rate (GR) (%)	0.6
out_p_Dead_High	Outflow probability to dead rate (High)(%)	0.06
out_p_Dead_GR	Outflow probability to dead rate (GR) (%)	0.06
out_p_WMO_High	Outflow probability to WMO rate (High) (%)	0.023
out_p_WMO_GR	Outflow probability to WMO rate (GR) (%)	0.023
out_p_WLZ_High	Outflow probability to WLZ rate (High) (%)	0.198
out_p_WLZ_GR	Outflow probability to WLZ rate (GR) (%)	0.198
out_p_GRV_High	Outflow probability to GRV rate (High) (%)	0.034
out_p_GRV_GR	Outflow probability to GRV rate (GR) (%)	0
out_p_Pall_High	Outflow probability to palliative rate (High) (%)	0.107
out_p_Pall_GR	Outflow probability to palliative rate (GR) (%)	0.107
out_p_Home_Low	Outflow probability to home rate (Low) (%)	0.7
out_p_Dead_Low	Outflow probability to dead rate (Low) (%)	0.02
out_p_WMO_Low (%)	Outflow probability to WMO rate (Low) (%)	0.02
out_p_WLZ_Low	Outflow probability to WLZ rate (Low) (%)	0.1
out_p_GRV_Low	Outflow probability to GRV rate (Low) (%)	0.02
out_p_Pall_Low	Outflow probability to palliative rate (Low) (%)	0.14
$\beta_{i,k}$	**Service time of caretype i to k**	
serv_HOS_High	Service time of HOS (High)(days)	30

Variable Name	Description	Default Value
serv_Home_High	Service time of home care (High)(days)	31.1
serv_Home_GR	Service time of home care (GR)(days)	31.1
serv_Dead_High	Service time of dead (High)(days)	22.9
serv_Dead_GR	Service time of dead (GR)(days)	22.9
serv_GRV_High	Service time of GRV (High)(days)	29.8
serv_GRV_GR	Service time of GRV (GR)(days)	0
serv_Pall_High	Service time of palliative (High)(days)	43.9
serv_Pall_GR	Service time of palliative (GR)(days)	43.9
serv_WMO_High	Service time of WMO (High)(days)	22.9
serv_WLZ_High	Service time of WLZ (High)(days)	47.8
serv_WMO_GR	Service time of WMO (GR)(days)	22.9
serv_WLZ_GR	Service time of WLZ (GR)(days)	47.8
serv_Home_Low	Service time of home care (Low)(days)	31.1
serv_Dead_Low	Service time of dead (Low)(days)	22.9
serv_GRV_Low	Service time of GRV (Low)(days)	29.8
serv_Pall_Low	Service time of palliative (Low)(days)	43.9
serv_WMO_Low	Service time of WMO (Low)(days)	22.9
serv_WLZ_Low	Service time of WLZ (Low)(days)	47.8
time_Max_Opn_EMD	Max opening time of EMD	24 (Corresponding to 00:00)
Opening_weekday	Opening of ELV on weekdays weekday	8 (Corresponding to 8:00)
EMD_start_time	EMD start time	0 (Corresponding to 00:00)
EMD_end_time	EMD end time	24 (Corresponding to 24:00)
GPR_start_time	GPR start time	0 (Corresponding to 00:00)
GPR_end_time	GPR end time	17 (Corresponding to 17:00)
p_opn_weekend	Probability of opening on weekend	0 (Always closed)
Adm_days	Admission days	1.5
Max_days_TRW	Max days in TRW	14
n_patients_per_nurse (s_r)	Number of patients a nurse can take care of	3
time_max_opn_ELV	Max opening time in ELV	17 (Corresponding to 17:00)
observation_prob	Observation probability (%)	10
max_wait_time_model	Max waiting time before changing preference	30
Print_modus	Print modus	FALSE

Table 7: All input parameters for the model with default values and short descriptions

Scenario Triage Ward								
Number of TRW beds	[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Wait_time_ELV_High	2.018	2.582	2.626	2.657	2.646	2.659	2.674	2.645
Wait_time_ELV_Low	1.799	1.823	1.808	1.831	1.811	1.814	1.819	1.826
Wt_from_HOSP_GRZ	2.557	3.156	3.189	3.551	3.427	3.451	3.578	3.380
Wt_from_HOSP_High	2.018	2.901	2.976	2.980	2.948	2.989	2.987	2.951
Wt_from_GPR_High	1.798	1.795	1.806	1.789	1.797	1.800	1.790	1.814
Wt_from_GPR_Low	1.799	1.823	1.808	1.831	1.811	1.814	1.819	1.826
Wt_to_TRW		0.011	0.013	0.015	0.020	0.022	0.021	0.034
WT_from_EMD	1.993	3.007	3.048	2.967	3.021	3.020	2.974	3.014
Perc_with_HOSP_adm	0.024	0.011	0.015	0.019	0.020	0.020	0.026	0.032
Number with hosp adm EMD	3.661	1.692	2.395	2.860	3.127	3.095	4.031	4.943
Perc_with_HOSP_adm_HOSP	0.016	0.008	0.008	0.010	0.011	0.010	0.009	0.013
Number with hosp adm HOSP	16.055	7.771	8.055	9.699	11.017	10.029	8.841	13.297
nr_pat_repl	2.022	2.067	2.072	2.073	2.074	2.076	2.076	2.075
los_ELV_High	36.965	36.256	36.473	36.388	36.691	36.636	36.672	36.595
los_ELV_Low	36.170	36.395	36.267	35.990	35.844	36.129	35.989	36.002
bez_gr_total	0.832	0.813	0.820	0.816	0.824	0.827	0.829	0.832
bez_gr_High	0.842	0.812	0.818	0.820	0.830	0.833	0.836	0.841
bez_gr_Low	0.836	0.838	0.844	0.831	0.834	0.838	0.839	0.836
serv_level	0.804	0.788	0.788	0.788	0.786	0.787	0.786	0.786
Priority	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS
Number of transfers between locations	0	0	0	0	0	0	0	0
Number of Locations ELV	1	1	1	1	1	1	1	1
Number of beds ELV_High	[146]	[145]	[144]	[143]	[142]	[141]	[140]	[139]
Number of beds ELV_Low	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]
Number of beds GRZ	[26]	[26]	[26]	[26]	[26]	[26]	[26]	[26]
Number of beds High Complex	[120]	[119]	[118]	[117]	[116]	[115]	[114]	[113]
Number of shared beds	[]	[]	[]	[]	[]	[]	[]	[]
Fairness	0	0	0	0	0	0	0	0

Table 8: Results TRW beds

Number_of_Locations_ELV	1	2	3	4	5
Wait_time_ELV_High	2.018	2.017	2.013	1.981	1.985
Wait_time_ELV_Low	1.799	1.780	1.808	1.777	1.808
Wt_from_HOSP_GRZ	2.557	2.562	2.569	2.462	2.492
Wt_from_HOSP_High	2.018	2.007	2.019	2.001	2.029
Wt_from_GPR_High	1.798	1.808	1.804	1.791	1.775
Wt_from_GPR_Low	1.799	1.780	1.808	1.777	1.808
Wt_to_TRW					
WT_from_EMD	1.993	1.967	1.933	1.917	1.892
Perc_with_HOSP_adm	0.024	0.029	0.024	0.026	0.023
Number with hosp adm EMD	3.661	4.506	3.845	4.073	3.580
Perc_with_HOSP_adm_HOSP	0.016	0.026	0.027	0.028	0.026
Number with hosp adm HOSP	16.055	26.077	27.364	28.191	26.019
nr_pat_repl	2.022	2.078	2.085	2.121	2.101
los_ELV_High	36.965	37.198	37.046	36.994	37.009
los_ELV_Low	36.170	35.835	35.919	36.048	36.119
bez_gr_total	0.832	0.606	0.523	0.481	0.450
bez_gr_High	0.842	0.832	0.816	0.799	0.799
bez_gr_Low	0.836	0.807	0.787	0.795	0.754
serv_level	0.804	0.802	0.802	0.806	0.803
Priority	FALSE	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS	FCFS
Number of transfers between locations	0	0.02904	0.02733	0.06649	0.04069
Number of beds ELV_High	[146]	[73, 73]	[48, 49, 49]	[36, 36, 37, 37]	[29, 29, 29, 29, 30]
Number of beds ELV_Low	[75]	[37, 38]	[25, 25, 25]	[18, 18, 18, 19]	[15, 15, 15, 15, 15]
Number of beds GRZ	[26]	[13, 13]	[8, 9, 9]	[6, 6, 7, 7]	[5, 5, 5, 5, 6]
Number of beds High Complex	[120]	[60, 60]	[40, 40, 40]	[30, 30, 30, 30]	[24, 24, 24, 24, 24]
Number of shared beds	[]	[]	[]	[]	[]
Number of TRW beds	[0]	[0]	[0]	[0]	[0]
Fairness	0	0	0	0	0

Table 9: Results for multiple ELV locations

Project	Base_case_arr_10%	Base_case_arr_20%	Base_case_arr_-10%	Base_case_arr_-20%
Wait_time_ELV_High	2.314	3.47 (2.97,3.97)	1.972	1.924
Wait_time_ELV_Low	1.851	1.98 (1.94,2.03)	1.794	1.785
Wt_from_HOSP_GRZ	4.260	10.47 (7.65,13.28)	2.297	2.024
Wt_from_HOSP_High	2.043	2.09 (2.06,2.12)	2.011	2.018
Wt_from_GPR_High	1.804	1.89 (1.86,1.92)	1.785	1.777
Wt_from_GPR_Low	1.851	1.98 (1.94,2.03)	1.794	1.785
Wt_to_TRW				
WT_from_EMD	1.974	2.02 (1.99,2.05)	1.996	1.994
Perc_with_HOSP_adm	0.068	0.14 (0.13,0.14)	0.008	0.000
Number with hosp adm EMD	10.827	21.542	1.242	0.011
Perc_with_HOSP_adm_HOSP	0.040	0.08 (0.08,0.08)	0.006	0.001
Number with hosp adm HOSP	40.181	79.433	6.253	1.326
nr_pat_repl	2.056	2.11 (2.1,2.12)	2.007	2.001
los_ELV_High	37.263	37.87 (35.04,40.7)	36.987	36.962
los_ELV_Low	35.986	33.98 (30.72,37.23)	35.855	36.139
bez_gr_total	0.877	0.91 (0.91,0.91)	0.772	0.696
bez_gr_High	0.888	0.92 (0.92,0.92)	0.783	0.701
bez_gr_Low	0.880	0.9 (0.9,0.9)	0.773	0.705
serv_level	0.786	0.74 (0.74,0.75)	0.809	0.814
Priority	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS
Number of transfers between locations	0	0	0	0
Number of Locations ELV	1	1	1	1
Number of beds ELV_High	[146]	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]	[75]
Number of beds GRZ	[26]	[26]	[26]	[26]
Number of beds High Complex	[120]	[120]	[120]	[120]
Number of shared beds	[]	[]	[]	[]
Number of TRW beds	[0]	[0]	[0]	[0]
Fairness	0	0	0	0

Table 10: Results sensitivity arrivals

Project	24/7_open + No admission days	No_admission_days	Opening_weekend (8 A.m.-5 P.m.)
Wait_time_ELV_High	0.072	0.542	1.618
Wait_time_ELV_Low	0.020	0.300	1.523
Wt_from_HOSP_GRZ	0.389	1.212	2.138
Wt_from_HOSP_High	0.007	0.509	1.522
Wt_from_GPR_High	0.009	0.286	1.514
Wt_from_GPR_Low	0.020	0.300	1.523
Wt_to_TRW			
WT_from_EMD	0.014	0.516	1.509
Perc_with_HOSP_adm	0.015	0.017	0.026
Number_with_hosp_adm_EMD	2.213	2.665	4.026
Perc_with_HOSP_adm_HOSP	0.010	0.012	0.015
Number_with_hosp_adm_HOSP	9.916	12.138	15.024
nr_pat_repl	2.013	2.018	2.022
los_ELV_High	35.056	35.525	36.832
los_ELV_Low	34.431	34.662	35.897
bez_gr_total	0.853	0.843	0.841
bez_gr_High	0.866	0.852	0.851
bez_gr_Low	0.851	0.847	0.844
serv_level	0.995	0.991	0.855
Priority	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS
Number_of_transfers_between_locations	0	0	0
Number_of_Locations_ELV	1	1	1
Number_of_beds_ELV_High	[146]	[146]	[146]
Number_of_beds_ELV_Low	[75]	[75]	[75]
Number_of_beds_GRZ	[26]	[26]	[26]
Number_of_beds_High_Complex	[120]	[120]	[120]
Number_of_shared_beds	[]	[]	[]
Number_of_TRW_beds	[0]	[0]	[0]
Fairness	0	0	0

Table 11: Results for different opening scenarios

Project	Priority False GR-High complex	Priority False part 10%	Priority False part 20%	Priority True part 20%	Priority True part 10%	Priority True GR-High complex
Wait_time_ELV_High	1.917	1.982	1.965	1.962	2.109	1.928
Wait_time_ELV_Low	1.808	1.795	1.818	1.814	1.818	1.806
Wt_from_HOSP_GRZ	2.033	2.101	2.037	1.993	2.060	1.978
Wt_from_HOSP_High	1.992	2.164	2.053	2.069	2.187	2.008
Wt_from_GPR_High	1.771	2.002	1.838	1.861	2.057	1.811
Wt_from_GPR_Low	1.808	1.795	1.818	1.814	1.818	1.806
Wt_to_TRW						
WT_from_EMD	1.989	2.097	2.025	1.985	2.130	1.995
Perc_with_HOSP_adm	0.001	0.077	0.050	0.068	0.110	0.018
Number with hosp adm EMD	0.205	11.741	7.627	10.303	16.359	2.849
Perc_with_HOSP_adm_HOSP	0.000	0.020	0.011	0.013	0.024	0.003
Number with hosp adm HOSP	0.000	20.367	10.983	13.464	24.335	2.545
nr_pat_repl	2.000	2.052	2.027	2.030	2.062	2.006
los_ELV_High	37.128	37.039	36.870	37.091	37.386	37.151
los_ELV_Low	35.488	36.004	35.953	36.212	36.007	35.741
bez_gr_total	0.840	0.842	0.843	0.842	0.847	0.843
bez_gr_High	0.856	0.856	0.856	0.855	0.862	0.864
bez_gr_Low	0.830	0.836	0.840	0.840	0.839	0.826
serv_level	0.811	0.791	0.805	0.804	0.785	0.812
Priority	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
Preference	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS
Number of transfers	0	0	0	0	0	0
Number of Locations ELV	1	1	1	1	1	1
Number of beds ELV_High	[146]	[146]	[146]	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]	[75]	[75]	[75]
Number of beds GRZ	[0]	[23]	[21]	[21]	[23]	[0]
Number of beds High Complex	[0]	[108]	[96]	[96]	[108]	[0]
Number of shared beds	[146]	[15]	[29]	[29]	[15]	[0]
Number of TRW beds	[0]	[0]	[0]	[0]	[0]	[0]
Fairness	0	0	0	0	0	0

Table 12: Comparison of Project Metrics for Different Scenarios

Project	Scen_no_bed_Sharing_**SQRT**	Scen_no_bed_Sharing_**stability**
Wait_time_ELV_High	2.018	4.65 (4.08,5.23)
Wait_time_ELV_Low	1.799	1.89 (1.86,1.91)
Wt_from_HOSP_GRZ	2.557	18.63 (15.13,22.14)
Wt_from_HOSP_High	2.018	2.04 (2.02,2.07)
Wt_from_GPR_High	1.798	1.84 (1.82,1.87)
Wt_from_GPR_Low	1.799	1.89 (1.86,1.91)
Wt_to_TRW		
WT_from_EMD	1.993	1.98 (1.96,2.01)
Perc_with_HOSP_adm	0.024	0.07 (0.06,0.07)
Number with hosp adm EMD	3.661	10.222
Perc_with_HOSP_adm_HOSP	0.016	0.09 (0.08,0.09)
Number with hosp adm HOSP	16.055	85.150
nr_pat_repl	2.022	2.1 (2.09,2.11)
los_ELV_High	36.965	39.72 (37.06,42.38)
los_ELV_Low	36.170	34.32 (30.27,38.38)
bez_gr_total	0.832	0.89 (0.89,0.89)
bez_gr_High	0.842	0.9 (0.9,0.9)
bez_gr_Low	0.836	0.88 (0.88,0.88)
serv_level	0.804	0.74 (0.74,0.75)
Priority	FALSE	FALSE
Preference	FCFS	FCFS
Number of transfers between locations	0	0
Number of Locations ELV	1	1
Number of beds ELV_High	[146]	[132]
Number of beds ELV_Low	[75]	[67]
Number of beds GRZ	[26]	[21]
Number of beds High Complex	[120]	[111]
Number of shared beds	[]	[]
Number of TRW beds	[0]	[0]
Fairness	0	0

Table 13: Results for no bed-sharing and stability scenarios

Number of beds GRZ	[27]	[26]	[25]	[24]	[23]	[22]	[21]	[20]	[19]
Wait_time_ELV_High	2.040	2.018	2.128	2.174	2.766	3.239	3.458	13.533	55.071
Wait_time_ELV_Low	1.814	1.799	1.809	1.820	1.807	1.811	1.822	1.812	1.810
Wt_from_HOSP_GRZ	2.657	2.557	3.274	3.553	7.233	10.192	12.084	79.331	364.079
Wt_from_HOSP_High	1.997	2.018	1.989	2.012	2.007	2.027	2.008	1.997	2.014
Wt_from_GPR_High	1.807	1.798	1.800	1.780	1.785	1.786	1.804	1.793	1.764
Wt_from_GPR_Low	1.814	1.799	1.809	1.820	1.807	1.811	1.822	1.812	1.810
Wt_to_TRW									
WT_from_EMD	1.989	1.993	1.985	1.997	1.984	1.991	1.986	2.003	1.988
Perc_with_HOSP_adm	0.026	0.024	0.027	0.017	0.021	0.019	0.020	0.015	0.013
Number with hosp adm EMD	4.047	3.661	4.131	2.673	3.353	2.780	2.991	2.254	1.985
Perc_with_HOSP_adm_HOSP	0.011	0.016	0.021	0.024	0.039	0.066	0.063	0.066	0.094
Number with hosp adm HOSP	11.245	16.055	21.315	24.499	39.064	65.978	62.715	65.770	94.556
nr_pat_repl	2.022	2.022	2.027	2.033	2.049	2.060	2.067	2.088	2.101
los_ELV_High	37.081	36.965	37.312	37.108	37.954	38.335	38.595	48.507	90.318
los_ELV_Low	36.336	36.170	36.185	36.305	35.799	36.008	36.125	36.190	36.139
bez_gr_total	0.835	0.832	0.840	0.838	0.844	0.840	0.844	0.845	0.847
bez_gr_High	0.845	0.842	0.852	0.847	0.858	0.855	0.857	0.860	0.862
bez_gr_Low	0.837	0.836	0.838	0.843	0.840	0.834	0.842	0.838	0.840
serv_level	0.802	0.804	0.799	0.793	0.783	0.772	0.767	0.747	0.736
Priority	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS
Number of transfers between locations	0	0	0	0	0	0	0	0	0
Number of Locations ELV	1	1	1	1	1	1	1	1	1
Number of beds ELV_High	[146]	[146]	[146]	[146]	[146]	[146]	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]
Number of beds GRZ	[27]	[26]	[25]	[24]	[23]	[22]	[21]	[20]	[19]
Number of beds High Complex	[119]	[120]	[121]	[122]	[123]	[124]	[125]	[126]	[127]
Number of shared beds	[]	[]	[]	[]	[]	[]	[]	[]	[]
Number of TRW beds	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Fairness	0	0	0	0	0	0	0	0	0

Table 14: Results for different numbers of GRZ beds

Project	Base_case_serv_10%	Base_case_serv_20%	Base_case_serv_-10%	Base_case_serv_-20%
Wait_time_ELV_High	2.425	4.14 (3.53,4.76)	1.965	1.935
Wait_time_ELV_Low	1.892	1.99 (1.95,2.04)	1.782	1.783
Wt_from_HOSP_GRZ	4.814	14.34 (10.91,17.78)	2.212	2.043
Wt_from_HOSP_High	2.050	2.1 (2.07,2.13)	2.027	2.012
Wt_from_GPR_High	1.837	1.91 (1.89,1.93)	1.790	1.788
Wt_from_GPR_Low	1.892	1.99 (1.95,2.04)	1.782	1.783
Wt_to_TRW				
WT_from_EMD	1.968	2.01 (1.98,2.03)	2.005	2.015
Perc_with_HOSP_adm	0.064	0.12 (0.12,0.12)	0.008	0.000
Number with hosp adm EMD	9.902	19.216	1.300	0.019
Perc_with_HOSP_adm_HOSP	0.038	0.08 (0.07,0.08)	0.009	0.001
Number with hosp adm HOSP	38.483	75.957	8.733	0.910
nr_pat_repl	2.057	2.11 (2.1,2.11)	2.009	2.002
los_ELV_High	40.897	45.6 (42.29,48.91)	33.581	30.028
los_ELV_Low	39.451	44.79 (39.63,49.96)	32.620	29.128
bez_gr_total	0.881	0.91 (0.91,0.91)	0.773	0.695
bez_gr_High	0.895	0.93 (0.93,0.93)	0.783	0.701
bez_gr_Low	0.877	0.91 (0.91,0.91)	0.774	0.700
serv_level	0.782	0.74 (0.74,0.75)	0.809	0.812
Priority	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS
Number of transfers between locations	0	0	0	0
Number of Locations ELV	1	1	1	1
Number of beds ELV_High	[146]	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]	[75]
Number of beds GRZ	[26]	[26]	[26]	[26]
Number of beds High Complex	[120]	[120]	[120]	[120]
Number of shared beds	[]	[]	[]	[]
Number of TRW beds	[0]	[0]	[0]	[0]
Fairness	0	0	0	0

Table 15: Results for different service levels

Project	First Come First Served	Preference based	Model based (overflow strategy)
Wait_time_ELV_High	1.967	66.81 (59.54,74.08)	3.327
Wait_time_ELV_Low	1.776	2.85 (2.72,2.99)	2.223
Wt_from_HOSP_GRZ	2.317	348.62 (310.14,387.1)	4.308
Wt_from_HOSP_High	2.007	2.27 (2.23,2.31)	3.360
Wt_from_GPR_High	1.802	2.06 (2.02,2.1)	3.123
Wt_from_GPR_Low	1.776	2.85 (2.72,2.99)	2.223
Wt_to_TRW			
WT_from_EMD	1.926	1.97 (1.94,2.01)	2.876
Perc_with_HOSP_adm	0.027	0.05 (0.05,0.05)	0.120
Number with hosp adm EMD	4.123	7.295	18.863
Perc_with_HOSP_adm_HOSP	0.013	0.06 (0.06,0.06)	0.068
Number with hosp adm HOSP	13.461	58.118	67.693
nr_pat_repl	2.365	2.08 (2.07,2.08)	2.332
los_ELV_High	36.957	189.41 (146.76,232.05)	38.172
los_ELV_Low	35.967	33.53 (30.21,36.85)	36.466
bez_gr_total	0.512	0.41 (0.41,0.41)	0.450
bez_gr_High	0.942	0.81 (0.81,0.81)	0.895
bez_gr_Low	0.961	0.8 (0.8,0.81)	0.800
serv_level	0.807	0.73 (0.73,0.74)	0.708
Priority	FALSE	FALSE	FALSE
Preference	FCFS	NO	Pref
Number of transfers between locations	0.33054	0	0.08365
Number of Locations ELV	5	5	5
Number of beds ELV_High	[24, 10, 24, 34, 30]	[24, 10, 24, 34, 30]	[24, 10, 24, 34, 30]
Number of beds ELV_Low	[15, 15, 0, 30, 15]	[15, 15, 0, 30, 15]	[15, 15, 0, 30, 15]
Number of beds GRZ	[0, 10, 0, 10, 6]	[0, 10, 0, 10, 6]	[0, 10, 0, 10, 6]
Number of beds High Complex	[24, 0, 24, 24, 24]	[24, 0, 24, 24, 24]	[24, 0, 24, 24, 24]
Number of shared beds	[]	[]	[]
Number of TRW beds	[0]	[0]	[0]
Fairness	0	1170.191039	27.82543112

Table 16: Results for Different Project Models

Percentage shared	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Wait_time_ELV_High	2.018	1.988	1.965	1.940	1.924	1.924	1.918	1.923	1.918	1.913	1.917
Wait_time_ELV_Low	1.799	1.811	1.818	1.819	1.799	1.815	1.797	1.812	1.804	1.824	1.808
Wt_from_HOSP_GRZ	2.557	2.033	2.037	2.023	2.009	2.010	1.989	2.016	1.988	2.006	2.033
Wt_from_HOSP_High	2.018	2.070	2.053	2.042	2.036	1.992	2.002	2.025	2.010	1.989	1.992
Wt_from_GPR_High	1.798	1.899	1.838	1.811	1.789	1.812	1.800	1.781	1.771	1.782	1.771
Wt_from_GPR_Low	1.799	1.811	1.818	1.819	1.799	1.815	1.797	1.812	1.804	1.824	1.808
Wt_to_TRW											
WT_from_EMD	1.993	2.008	2.025	1.975	1.964	1.973	1.969	1.976	2.004	1.978	1.989
Perc_with_HOSP_adm	0.024	0.088	0.050	0.029	0.017	0.014	0.011	0.006	0.004	0.003	0.001
Number with hosp adm EMD	3.661	13.646	7.627	4.477	2.545	2.132	1.689	0.931	0.563	0.423	0.205
Perc_with_HOSP_adm_HOSP	0.016	0.020	0.011	0.008	0.004	0.004	0.003	0.001	0.001	0.000	0.000
Number with hosp adm HOSP	16.055	19.631	10.983	7.513	4.309	4.126	3.043	1.271	0.810	0.465	0.000
nr_pat_repl	2.022	2.042	2.027	2.015	2.009	2.007	2.006	2.003	2.002	2.001	2.000
los_ELV_High	36.965	37.083	36.870	36.969	36.788	37.046	37.068	36.812	37.011	36.934	37.128
los_ELV_Low	36.170	35.891	35.953	35.993	35.788	35.970	35.990	36.125	36.095	36.224	35.488
bez_gr_total	0.832	0.843	0.843	0.841	0.839	0.843	0.841	0.839	0.841	0.844	0.840
bez_gr_High	0.842	0.859	0.856	0.855	0.854	0.861	0.858	0.850	0.858	0.856	0.856
bez_gr_Low	0.836	0.834	0.840	0.835	0.832	0.831	0.831	0.840	0.830	0.843	0.830
serv_level	0.804	0.801	0.805	0.809	0.813	0.811	0.813	0.812	0.811	0.813	0.811
Priority	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS	FCFS
Number of transfers	0	0	0	0	0	0	0	0	0	0	0
Number of Locations ELV	1	1	1	1	1	1	1	1	1	1	1
Number of beds ELV_High	[146]	[146]	[146]	[146]	[146]	[146]	[146]	[146]	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]
Number of beds GRZ	[26]	[22]	[21]	[19]	[16]	[13]	[10]	[7]	[4]	[2]	[0]
Number of beds High Complex	[120]	[106]	[96]	[84]	[72]	[60]	[48]	[36]	[24]	[12]	[0]
Number of shared beds	[0]	[18]	[29]	[43]	[58]	[73]	[88]	[103]	[118]	[132]	[146]
Number of TRW beds	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Fairness	0	0	0	0	0	0	0	0	0	0	0

Table 17: Percentage sharing scenarios

Project	GR-High-Complex_share	Total_bed_share	partial_bed_Sharing (40%)	no_bed_Sharing_SQRT
Wait_time_ELV_High	1.917	2.55 (2.21,2.89)	1.940	2.018
Wait_time_ELV_Low	1.808	2.81 (2.19,3.43)	1.819	1.799
Wt_from_HOSP_GRZ	2.033	2.59 (1.67,3.51)	2.023	2.557
Wt_from_HOSP_High	1.992	2.49 (1.93,3.04)	2.042	2.018
Wt_from_GPR_High	1.771	2.99 (2.19,3.8)	1.811	1.798
Wt_from_GPR_Low	1.808	2.81 (2.19,3.43)	1.819	1.799
Wt_to_TRW				
WT_from_EMD	1.989	2.14 (1.88,2.4)	1.975	1.993
Perc_with_HOSP_adm	0.001	0.0 (0.0,0.0)	0.029	0.024
Number with hosp adm EMD	0.205	0.019	4.477	3.661
Perc_with_HOSP_adm_HOSP	0.000	0.0 (0.0,0.0)	0.008	0.016
Number with hosp adm HOSP	0.000	0.082	7.513	16.055
nr_pat_repl	2.000	2.68 (2.42,2.95)	2.015	2.022
los_ELV_High	37.128	33.83 (31.29,36.38)	36.969	36.965
los_ELV_Low	35.488	36.78 (32.94,40.63)	35.993	36.170
bez_gr_total	0.840	0.78 (0.78,0.78)	0.841	0.832
bez_gr_High	0.856	nan (nan,nan)	0.855	0.842
bez_gr_Low	0.830	nan (nan,nan)	0.835	0.836
serv_level	0.811	0.81 (0.81,0.81)	0.809	0.804
Priority	FALSE	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS	FCFS
Number of transfers	0	0	0	0
Number of Locations ELV	1	1	1	1
Number of beds ELV_High	[146]	[]	[146]	[146]
Number of beds ELV_Low	[75]	[]	[75]	[75]
Number of beds GRZ	[0]	[]	[19]	[26]
Number of beds High Complex	[0]	[]	[84]	[120]
Number of shared beds	[146]	[]	[43]	[]
Number of TRW beds	[0]	[0]	[0]	[0]
Fairness	0	0	0	0

Table 18: Comparison of Project Metrics for Different Scenarios

Project	PREFERENCE FCFS CHECK	PREFERENCE CHECK	PREFERENCE model CHECK
Wait_time_ELV_High	2.018	2.018	2.018
Wait_time_ELV_Low	1.799	1.799	1.799
Wt_from_HOSP_GRZ	2.557	2.557	2.557
Wt_from_HOSP_High	2.018	2.018	2.018
Wt_from_GPR_High	1.798	1.798	1.798
Wt_from_GPR_Low	1.799	1.799	1.799
Wt_to_TRW			
WT_from_EMD	1.993	1.993	1.993
Perc_with_HOSP_adm	0.024	0.024	0.024
Number with hosp adm EMD	3.661	3.661	3.661
Perc_with_HOSP_adm_HOSP	0.016	0.016	0.016
Number with hosp adm HOSP	16.055	16.055	16.055
nr_pat_repl	2.022	2.022	2.022
los_ELV_High	36.965	36.965	36.965
los_ELV_Low	36.170	36.170	36.170
bez_gr_total	0.832	0.832	0.832
bez_gr_High	0.842	0.842	0.842
bez_gr_Low	0.836	0.836	0.836
serv_level	0.804	0.804	0.804
Priority	FALSE	FALSE	FALSE
Preference	FCFS	Pref	Model
Number of transfers between locations	0	0	0
Number of Locations ELV	1	1	1
Number of beds ELV_High	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]
Number of beds GRZ	[26]	[26]	[26]
Number of beds High Complex	[120]	[120]	[120]
Number of shared beds	[]	[]	[]
Number of TRW beds	[0]	[0]	[0]
Fairness	0	0	0

Table 19: Results for Different Preference Checks

Project	Scen_GR-High-Complex_share	Part sharing 100%
Wait_time_ELV_High	1.917	1.917
Wait_time_ELV_Low	1.808	1.808
Wt_from_HOSP_GRZ	2.033	2.033
Wt_from_HOSP_High	1.992	1.992
Wt_from_GPR_High	1.771	1.771
Wt_from_GPR_Low	1.808	1.808
Wt_to_TRW		
WT_from_EMD	1.989	1.989
Perc_with_HOSP_adm	0.001	0.001
Number with hosp adm EMD	0.205	0.205
Perc_with_HOSP_adm_HOSP	0.000	0.000
Number with hosp adm HOSP	0.000	0.000
nr_pat_repl	2.000	2.000
los_ELV_High	37.128	37.128
los_ELV_Low	35.488	35.488
bez_gr_total	0.840	0.840
bez_gr_High	0.856	0.856
bez_gr_Low	0.830	0.830
serv_level	0.811	0.811
Priority	FALSE	FALSE
Preference	FCFS	FCFS
Number of transfers between locations	0	0
Number of Locations ELV	1	1
Number of beds ELV_High	[146]	[146]
Number of beds ELV_Low	[75]	[75]
Number of beds GRZ	[0]	[0]
Number of beds High Complex	[0]	[0]
Number of shared beds	[146]	[146]
Number of TRW beds	[0]	[0]
Fairness	0	0

Table 20: Comparison of Metrics for Different Scenarios

Project	0 TRW beds	Scen_no_bed_Sharing_SQRT	0% partial share
Wait_time_ELV_High	2.018	2.018	2.018
Wait_time_ELV_Low	1.799	1.799	1.799
Wt_from_HOSP_GRZ	2.557	2.557	2.557
Wt_from_HOSP_High	2.018	2.018	2.018
Wt_from_GPR_High	1.798	1.798	1.798
Wt_from_GPR_Low	1.799	1.799	1.799
Wt_to_TRW			
WT_from_EMD	1.993	1.993	1.993
Perc_with_HOSP_adm	0.024	0.024	0.024
Number with hosp adm EMD	3.661	3.661	3.661
Perc_with_HOSP_adm_HOSP	0.016	0.016	0.016
Number with hosp adm HOSP	16.055	16.055	16.055
nr_pat_repl	2.022	2.022	2.022
los_ELV_High	36.965	36.965	36.965
los_ELV_Low	36.170	36.170	36.170
bez_gr_total	0.832	0.832	0.832
bez_gr_High	0.842	0.842	0.842
bez_gr_Low	0.836	0.836	0.836
serv_level	0.804	0.804	0.804
Priority	FALSE	FALSE	FALSE
Preference	FCFS	FCFS	FCFS
Number of transfers between locations	0	0	0
Number of Locations ELV	1	1	1
Number of beds ELV_High	[146]	[146]	[146]
Number of beds ELV_Low	[75]	[75]	[75]
Number of beds GRZ	[26]	[26]	[26]
Number of beds High Complex	[120]	[120]	[120]
Number of shared beds	[]	[]	[0]
Number of TRW beds	0	[0]	[0]
Fairness		0	0

Table 21: Comparison of Project Metrics for Different Scenarios

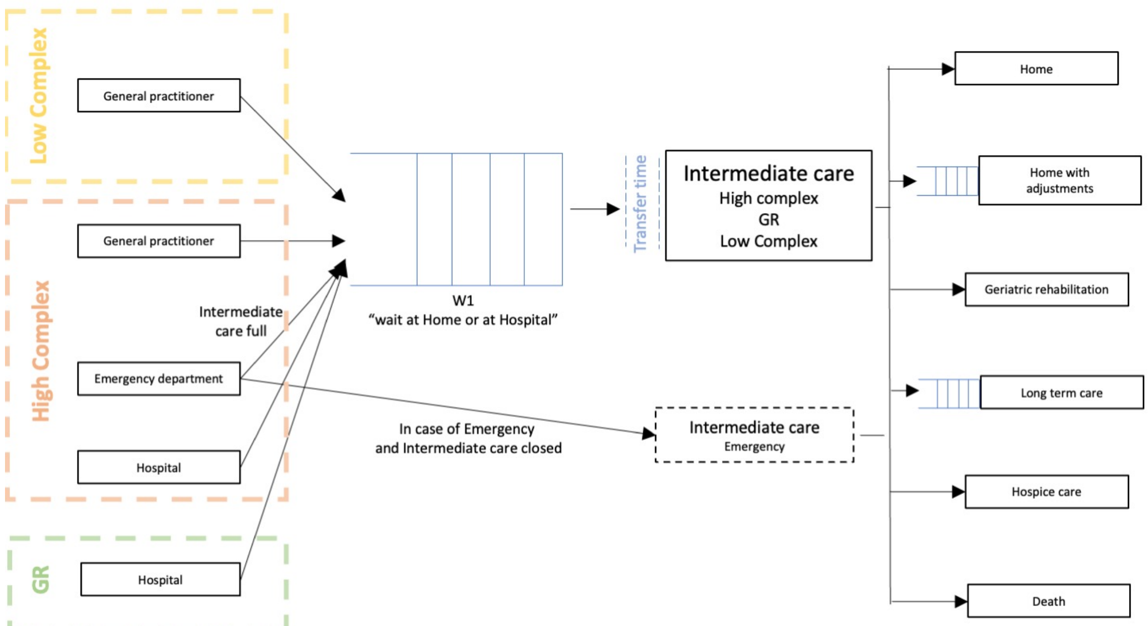

Figure 27: Total share of Intermediate Care

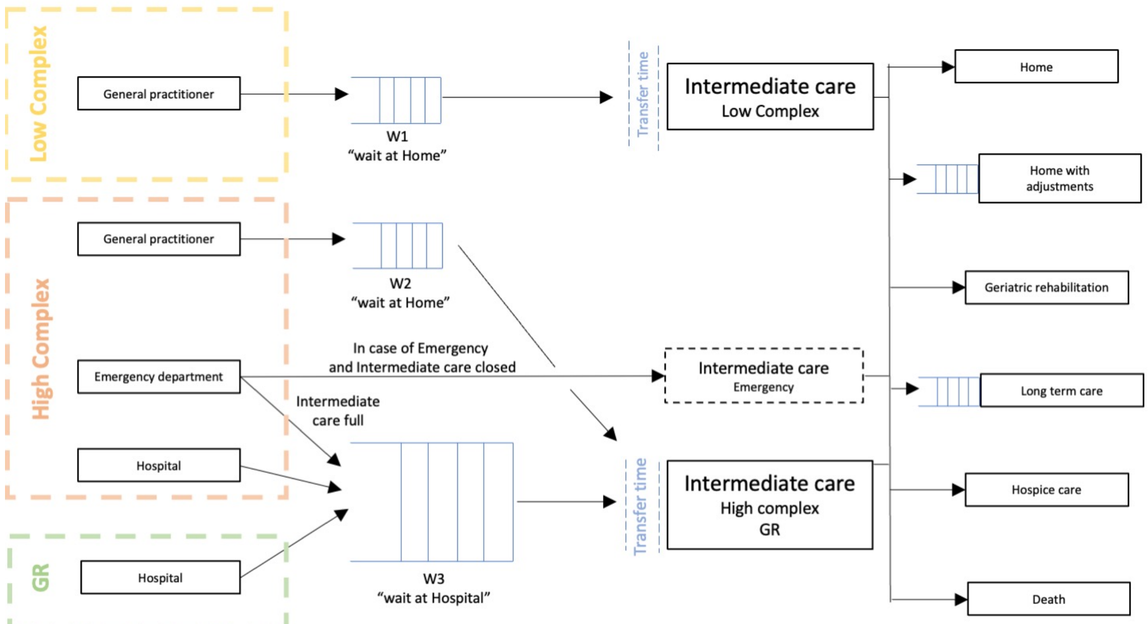

Figure 28: Full share of Intermediate Care

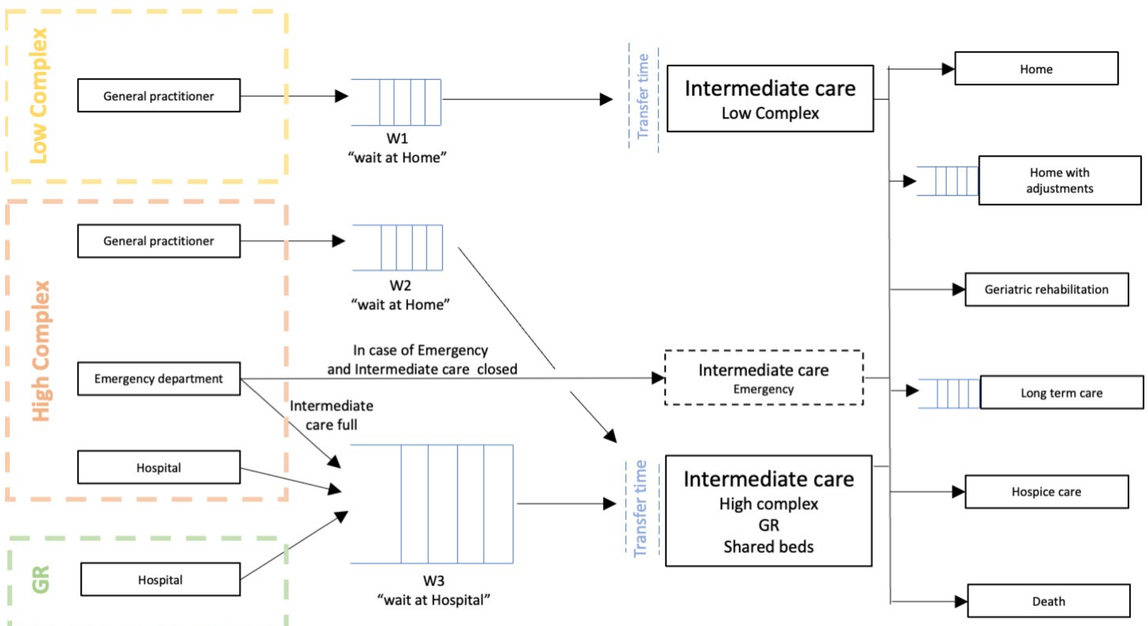

Figure 29: Partial share of Intermediate Care

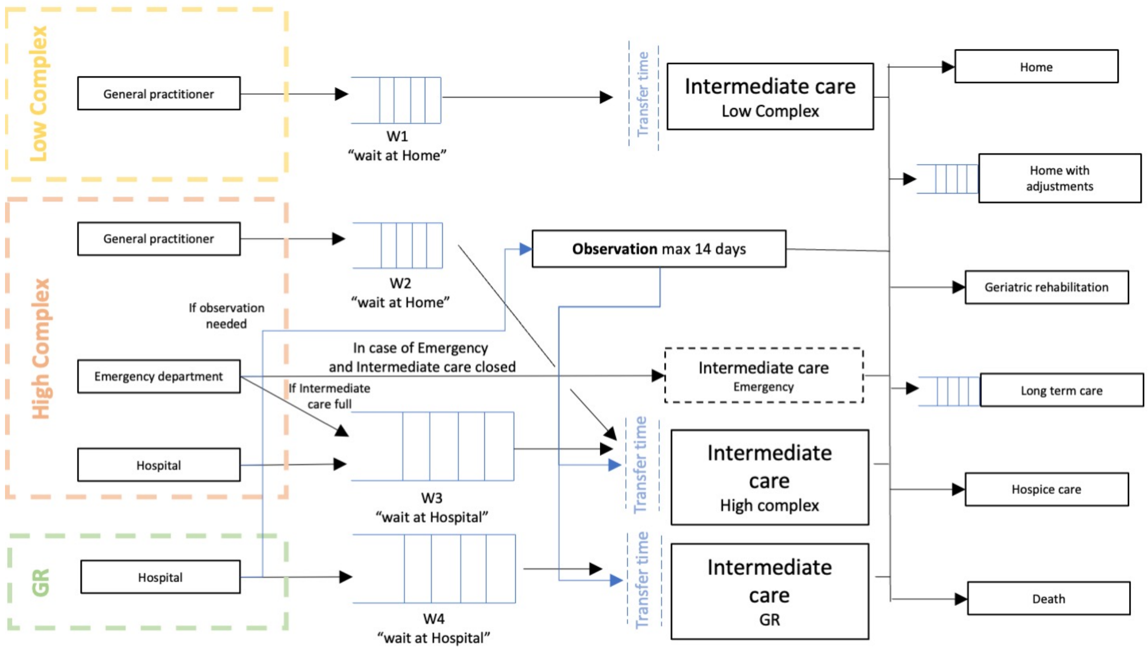

Figure 30: Flow diagram for the triage ward scenario