Master thesis Industrial engineering and management

Optimising CT scanner access times through data-driven appointment scheduling

Femke M.P. Geerts
University of Twente

May, 2025

Preface

Dear reader,

This research is part of my Master's graduation project for the Industrial Engineering and Management program. I conducted the study within the Integral Capacity Management team at UMC Utrecht. During my master's degree, I particularly enjoyed exploring various planning and logistics projects. I was especially drawn to mathematical and analytical business optimisation challenges that also had a social impact. Combined with my long-standing interest in healthcare, these were key reasons why I chose to conduct my thesis in a hospital setting.

Conducting research at UMC Utrecht gave me valuable insights into the significant added value that capacity management can bring to healthcare. At the same time, I discovered how challenging it can be to change an already well-functioning process. One of the greatest challenges I faced was identifying where to begin with optimisation and how to encourage openness to change among stakeholders. Through this experience, I gained practical knowledge in change management and had the opportunity to apply many of the concepts I had studied in a real-world context. Additionally, I developed both my independent working skills and my ability to collaborate as a team member and project owner within large project teams.

I would like to thank several people in particular. First of all, my company supervisor Bart van den Berg, for the many enjoyable and insightful feedback sessions we had. During these sessions, he taught me valuable management skills and gave me the trust and confidence to grow into a junior professional. Despite his busy schedule, he was always available for questions, feedback, or brainstorming sessions. I also want to thank all my colleagues who made time to answer questions, think along with me, and share a fun chat.

I'm also very grateful to my university supervisor, Daniela Guericke, for her helpful feedback and support. She taught me how to navigate the more difficult aspects of research, such as data gathering. I would also like to thank my second supervisor, Sebastian Rachuba, for his valuable contributions and feedback.

Finally, I would like to thank my family, my dispute, and my study friends, who have become incredibly important to me over the past few years. Personal circumstances made my master's journey challenging at times, and knowing they were always there for support and motivation meant a lot. A special thanks to Job, Aletta, and my parents for all the feedback and brainstorming sessions they provided during my thesis work.

I am proud of my work and hope you enjoy reading my master's thesis.

Femke Geerts

Utrecht, May 2025

Management Summary

Introduction and problem formulation

The CT radiology department of UMC Utrecht is under increasing pressure due to growing demand and limited resources. Despite efforts to manage capacity, the current average access time of 31 days for outpatient CT appointments still exceeds the target of 22.4 days set by Dutch health insurers. Therefore, the core issue is that the access time exceeds the agreed target set by health insurers. Interviews with hospital staff, observation studies and data review revealed that while demand is high, the current appointment schedule does not align with the high volume of walk-in and unplanned arrivals. This mismatch leads to inefficient scanner use, increased waiting times, and high access times. Based on this core problem, the main research question is:

How can we better align patient demand and CT appointment schedules to achieve target access times for scheduled appointments?

Problem solving approach

We developed a model based on the insights gained from the literature. We chose discrete-event simulation because it provides valuable performance insights, is user-friendly, and will help the hospital trust the model, thus increasing the likelihood of successful implementation. We chose discrete event simulation to evaluate the system, as it can incorporate variable arrival patterns, multiple patient types, and complex queuing behaviour with fewer simplifying assumptions than analytical methods. We developed a simulation-based optimisation framework combining Discrete Event Simulation (DES) with Simulated Annealing (SA). DES was used to evaluate how well different appointment schedules perform under realistic operating conditions, accounting for patient types, arrival rates, priority rules, scan durations, and scanner capacity. Simulated annealing was applied as a metaheuristic search method to explore the vast solution space and iteratively improve the appointment schedule. Figure 0.1 illustrates the relationship between DES and SA. SA proposes a new appointment schedule through controlled changes, and DES evaluates its performance using multiple replications to account for randomness. This loop continues until the stopping criteria are met.

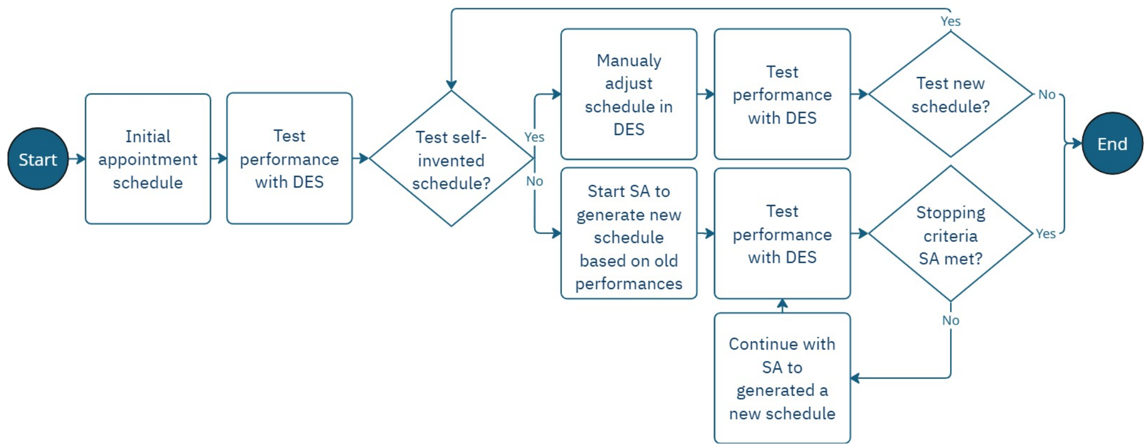

Figure 0.1: Relation between discrete-event simulation and simulated annealing in the model

A total of eight experiments were performed, including: evaluation of the current schedule; testing flexible versus hybrid scanner priority rules; testing allowing outpatient appointment slots on all scanners using simulated annealing; simulating additional outpatient slots using simulated annealing; exploring stakeholder-proposed schedules; and optimising combinations of these factors.

Key Findings:

- **Balanced objective function** ensured realistic trade-offs. The simulation did not just seek the lowest average waiting time, but also penalised excessive overtime and unmet service level targets. This multi-KPI approach led to better-performing schedules.
- **Flexible priority rules** lead to significantly better system performance. Allowing all CT scanners to dynamically select the next patient based on patient type (Emergency patient, inpatient, outpatient and walk-inpatient), rather than having dedicated walking scanners, resulted in lower average waiting times and higher service level performance.
- **Optimised schedules** increased utilisation without increasing overtime and capacity. Using simulated annealing, a schedule was generated that allowed an average of 5.8 additional appointments per day. This reduces the access time from 31 days to the target of 22.4 days. Compared to the initial schedule, the SL of emergency, walk-in and outpatients increased, and that of walking patients decreased but stayed within the target. The overtime and average waiting time also stay within the target. The comparison with the initial and proposed appointment schedule can be seen in Figure 0.2 and Figure 0.3. On all hours of the days extra appointments are added without violating the KPI targets.
- **Stakeholder involvement was key.** The simulation model is transparent and designed to be user-friendly, so planners and decision-makers can test new ideas and understand their effects. This approach led to more grounded experiments, better alignment with real-world constraints, and a greater chance of implementation.

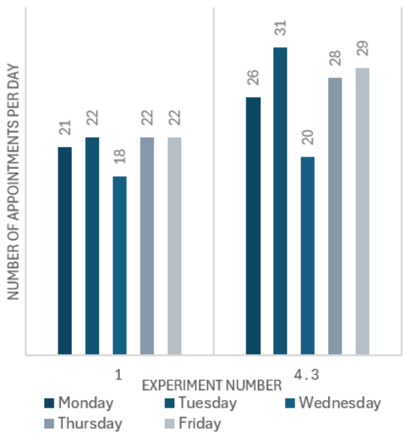

Figure 0.2: Number of appointment slots per day

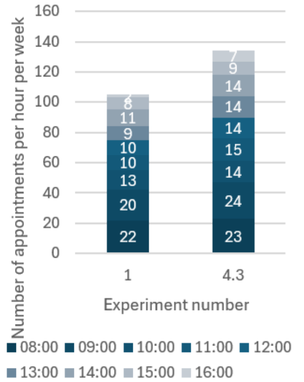

Figure 0.3: Number of appointment slots per hour

Conclusion:

This study demonstrates that significant improvements in access time and overall performance can be achieved through a combination of discrete-event simulation and metaheuristic optimisation (SA). This means that increasing capacity is not always needed to increase efficiency and access times. We proved that restructuring how and when appointments are scheduled and using flexible priority rules for all scanners, using DES and SA, delivers a scalable way to improve efficiency for UMC Utrecht. More case studies need to be performed to determine if flexible priority rules are best for all departments that deal with many unplanned arrivals.

The final model allows UMC Utrecht to evaluate existing schedules, explore alternative scenarios, and implement data-driven changes. While some simplifications were made, the model is robust and adaptable, making it a valuable tool for tactical decision-making.

Master thesis Industrial Engineering and Management

Optimising CT scanner access times through data-driven appointment scheduling

Author:
F.M.P. Geerts (Femke)

University of Twente
Drienerlolaan 5
7522 NB Enschede

UMC Utrecht
Heidelberglaan 100
3584 CX Utrecht

Supervisor University of Twente
Dr. D. Guericke (Daniela)
Dr. S. Rachuba (Sebastian)

Supervisors UMC Utrecht
Ir. B. Van den Berg (Bart)

Contents

Acronyms

Acronyms

MPSM: Managerial Problem-Solving Method

SL: Service Level

KPI: Key Performance Indicator

SA: Simulated Annealing

DE: Discrete Event Simulation

Symbols

O: Outpatient

I: Inpatient

W: Walk-in patient

C: clinical patient

1 Introduction

To complete the master's programme in Industrial Engineering and Management, research is conducted at UMC Utrecht. This chapter introduces the company in section 1.1. This is followed by describing the action problem and associated core problem in section 1.2. The research design is outlined in section 1.3.

1.1 Company introduction

This section provides an overview of the organisation and its key departments. Section 1.1.1 introduces UMC Utrecht and subsequently discusses the Integral Capacity Management Team in section 1.1.2. Finally, the Radiology Department is described in section 1.1.3.

1.1.1 UMC Utrecht

Healthcare systems worldwide are under increasing pressure due to the growing demands of patients and limited resources. Therefore, efficiently using these resources is essential to delivering timely and effective patient care. UMC Utrecht is one of the largest hospitals in the Netherlands, located in the Utrecht Science Park (Interchange, 2021). The hospital was founded in 1817 and has grown to more than 12,000 employees. The hospital has over 1,000 beds and treats more than 30,000 inpatients and 350,000 outpatients a year (UMC Utrecht, 2024).

Aside from patient care, another important role of the hospital is academic research and education. The hospital works closely with the Utrecht University Medical School and is a training centre for medical students, residents, and healthcare professionals. Their mission is to create patient care for tomorrow together, where the patient plays a central role. This means that in every part of their organisation, they aim for socially acceptable costs to keep healthcare accessible. For this, an agile organisation must apply innovations quickly and safely. Since it is an academic hospital, it often focuses on complex, high-quality healthcare. This means that the patient's care paths often include multiple specialists and types of diagnostics.

1.1.2 Integral capacity management team

The project owner of this thesis is the Integral Capacity Management team. They focus on projects that optimise utilisation and capacity to get the correct patients to the right places with the right resources and satisfy employees and patients. This department was founded in January 2021 during the COVID-19 pandemic. Their vision is to focus on continuous improvement instead of one superior improvement. For example, this means they prefer quick wins over one big multi-year project. Their strategy is to give specialists access to reliable data and manage the hospital based on its capacity limits. This will tackle the hospital-wide problem of the growing demands and limited resources.

1.1.3 Radiology department

Radiology is one of the hospital's core departments, with 24 radiologists. In addition to radiologists, this department consists of laboratory technicians. The laboratory technicians perform the scans. Senior laboratory technicians are responsible for staff planning and opening appointment time slots per category for each scan, where patient appointments can be planned. The radiology administration schedules the patients' appointments on the available time slots and registers the patients. Various imaging tests are performed in the radiology department, such as conventional X-ray diagnostics, ultrasound, angiography, mammography, nuclear examination, CT, and MRI. In addition to diagnostics, treatments are also performed using image-guided techniques. The CT department is special compared to others since it uses planned and walk-in appointments interchangeably.

The Integral Capacity Management department has not yet worked with the radiology department; therefore, there are many opportunities for capacity improvement problems. That is why we will focus on this department.

1.2 Problem identification

Section 1.2.1 describes how the action problem is chosen. Section 1.2.2 outlines the visualised problem cluster that helps to identify the core problem. Subsequently, section 1.2.3 explains how the core problem is chosen.

1.2.1 Action problem

The radiology department faces capacity problems. In the future, they need to deliver more healthcare with the same capacity. This is because the demand for scans will grow, and the resources will be the same or even less. The radiology department finds customer satisfaction very important. They believe that they can improve patient satisfaction by letting patients plan their appointments. On top of this, they think it will also help with the increased demand since it will save labour time of planners. The department has already started by letting patients plan their appointments by allowing walk-in appointments for CT scans. For now, the CT scan has the highest potential to let patients plan their appointments because for example, MRI scans or radiations require more preparations for both employees and patients and are therefore hard to plan by patients. To conclude, since this department wants to let patients schedule their appointments and they have already started with the CT, we will focus on the CT scan process.

To begin with, we conducted extensive interviews with different stakeholders to determine which part of the CT scan process requires the most improvement in capacity management. We asked these stakeholders about their most significant concerns and wishes for improvement.

Function	Wishes	Concerns
Business Office	- Supply & demand aligned - Better balance in onboarding new people	- Capacity cannot be adjusted to unknown demand - Know too late that employees should be trained
Medical managers radiology	- Patients plan their appointments - Better insight into performance	- Improve patient satisfaction - Too few laboratory technicians result in fewer scans - People on standby do not always have to work
Employee managers Radiology	- Better insight in demand to improve employee planning	- No insights in performance and demand
Senior Laboratory Technicians	- Know when walk-ins and emergencies arrive - Dynamic appointment slots	- Creating an agenda grid is complicated - Cannot adjust appointment duration
Laboratory Technicians	- Less variation in workload	- Variation in workload - Unhappy patients who waited long
Administrators Radiology	- Shorter waiting times for walk-ins	- Unhappy patients who must wait long
Heads of Clinic Departments	- Shorter access times	- Unclear what the access times are and if they are up to standard
CT Planners Clinic	- More performance insights - More appointment opportunities	- Long waiting lists

Figure 1.1: Stakeholders' wishes and concerns identified during the interviews.

Figure 1.1 shows an overview of the outcomes of the interviews. We see that the interests of the different stakeholders conflict. Some find the well-being of the patients most important, while others prioritise the employees, costs or insights into the performance. In a meeting with all the stakeholders, we agreed they first want to improve the long access times for scheduled appointments. They want to focus specifically on scheduled appointments since walk-ins and emergency patients do not have access times since they can arrive on the same day.

Every two weeks, each hospital has to calculate its access times. All hospitals in the Netherlands do this by counting the calendar days between the referral date and the third opportunity for an appointment (Zorgkaart Nederland, 2024). The third option is used because earlier appointment slots may become available due to cancellations, which do not accurately reflect typical waiting times. By considering the third available appointment, the measurement better represents the experience of the broader patient population. We calculated an average access time of 48 days for scheduled appointments over the last six months, with a maximum of 85 days using this method. Dutch health insurers negotiated with the hospitals that 80% of the patients should be able to get a CT scan within 21 days and the other 20% within 28 days (CZ, 2024). To simplify this, this means that the waiting time should have a maximum average of 22.4 days. At UMC Utrecht, currently 11% of the scans are within 28 days and 22% within 21 days. This means the waiting time for scheduled appointments is too long. If nothing changes, the access times will only get longer over time, as the demand for healthcare increases. While tackling the long access times, other wishes and problems will probably be solved, such as improved workload variation and better insight into supply, demand, and access times.

The conclusion is that the access times are too long, meaning they cannot even meet the current demand. Leading us to the following action problem: "The access time for scheduled CT scan appointments is longer than agreed with the health insurers." This research will thus focus on designing a method to improve the utilisation of CT scans and reduce the access times to the suggestions of the health insurance. To simplify this target, the average access time should be at least below 22.4 days.

1.2.2 Problem cluster

This section describes the process from the action problem to the core problem. The observed action problem is the too-long access times for fixed CT appointments. With these access times, UMC is not able to face future demand. The action problem captures a discrepancy between norm and reality (Heerkens and Van Winden, 2017). This research aims to resolve this gap by addressing its root causes. We made a problem cluster (Figure 1.2) to find the root causes, which are called the core problems. We gathered all the information with the help of interviews, observations and data research. The arrows run from effects to causes. There are four reasons for the long access times for CT scans.

1. The first reason (effect 1 in problem cluster) is an incorrect waiting list length. One explanation is that some people are on the list twice, while others no longer need a scan or voluntarily wait. On top of this, walk-in patients are on the list as well, and their access time is often their own choice. So, the calculated access time per speciality is not correct. It is, therefore, difficult to find the optimal number of appointments and to balance them between the specialities.
2. The second reason (2) is that there are not enough CT scans open, so they cannot face the demand for fixed appointments. The main reason CT scans are closed is due to the capacity of the laboratory technicians. The number of laboratory technicians available is not optimal because:
   (a) There are not enough technicians to always cover the demand (9).

(b) The technicians are scheduled according to supply (10). So, for example, they do not consider the demand when determining whether to approve requests for a day off.

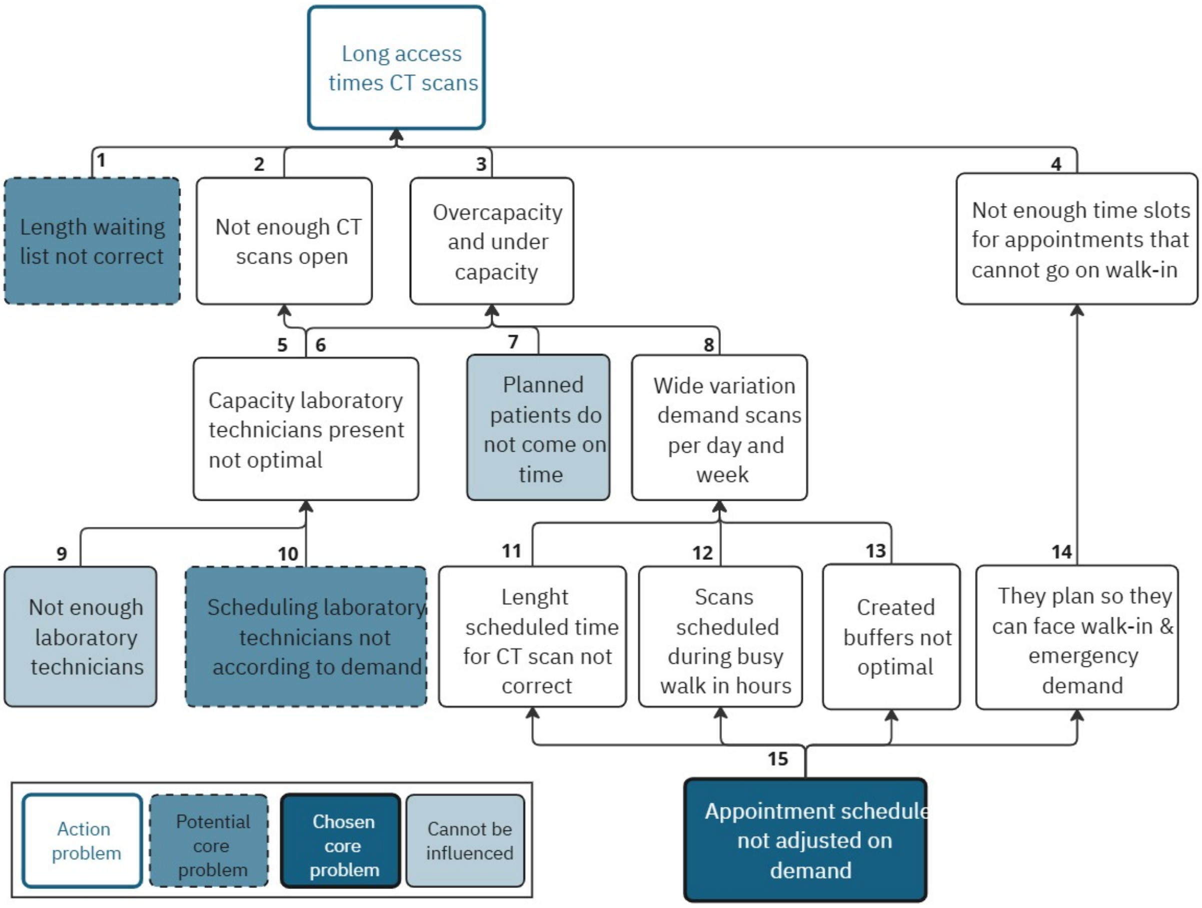

Figure 1.2: Problem cluster showing how we found the core problem.

3. The third reason is overcapacity and undercapacity (3). When there is an overcapacity of laboratory technicians, they cannot perform more scans because the demand cannot be increased immediately. The other way around is when too few technicians are present, appointments must be cancelled. This results in inefficient work, long waiting lists, and access times. There are three main causes for this.
   (a) The capacity of laboratory technicians present is not optimal (6).
   (b) The second motive is the arrival time of patients (7). If patients arrive late, staff may be unoccupied and, later on, overutilized.
   (c) Another cause is the wide variation in demand (8). For example, they are buried in work in the morning, and in the afternoon, it is the opposite. This has different causes.
      i. The first reason is the incorrectly scheduled appointment time. If appointments are scheduled for longer than the duration, the actual demand will be lower than expected.
      ii. Secondly, appointments are planned during busy walk-in hours. This results in undercapacity during the busy walk-in hours and overcapacity during quiet hours.
      iii. Lastly, buffers are created in many different ways. So, it is difficult to see how

much buffer time is planned. For example, too much buffer time will lead to idle working hours and low utilisation.

4. The fourth argument is that there are too few fixed appointments compared to a walk-in appointment (4).
   (a) The explanation for this is that the planners plan so they can face the walk-in demand and emergency demand (14). They do not know how many patients will arrive, so they save a lot of time for this to avoid overtime. This results in high access times for fixed appointments.

Reasons 11, 12, 13 and 14 all lead to the last potential core problem: the appointment schedule is not adjusted to demand. The appointment schedule can be adjusted by opening the optimal number of slots and planning the correct appointment length at the optimal time. This will decrease the variation in the number of scans daily and weekly. According to the Lean principle, reducing variation will make the work more efficient, so more patients can get a scan and the access time decreases (Theisens, 2021).

1.2.3 Core problem

In this section, we choose the core problem from the five candidates shown in Figure 1.2). Heerkens and Van Winden, 2017 states that the problems that cannot be addressed should be eliminated from the set of candidates.

It is, for now, impossible to hire more laboratory technicians because there is insufficient money and a shortage in the labour market (problem 9). Next, you cannot eliminate patients arriving late because you cannot influence the traffic or, for example, other emergencies that cause patients to arrive late (problem 7). Accordingly, the problems of "not having enough laboratory technicians" and "planned patients do not come on time" have been deleted from the list.

An administrative employee can remove duplicate patients or patients who do not need a scan anymore from the waiting list (problem 1). This is very time-consuming and very difficult. A better solution is to remind employees, with a notification in the scheduling system, to use an existing order instead of making a new order. This is a quick and practical fix to get a better insight into the waiting list. It will, however, not reduce the actual access time and, therefore, is not chosen as the core problem.

It is good to schedule employees according to demand to reduce over- and under-capacity. Right now, this is impossible since the exact demand is unknown (problem 10). Next to this, it is a cultural problem that they want to satisfy the time-off requests of their employees. However, we can make the most significant impact by adjusting the appointment schedule according to the demand (problem 15). An explanation for this is that four effects will be solved if the appointment schedule is adjusted to the demand. Therefore, the impact on access time of fixed appointments will be significant. On the contrary, scheduling laboratory technicians not according to demand will only solve one effect.

To conclude, the potential core problem: *"The appointment schedule for the CT scans are not adjusted on demand."* has been chosen as the core problem to solve. Because it is solvable in a relatively short time, it will have the most impact on the access times.

1.3 Research design

This section will describe the research scope, research questions and deliverables.

1.3.1 Scope

Certain boundaries are set within the topic of the core problem to ensure the research can be completed within the time constraints of one academic semester. Regarding the CT scans, we will only consider the demand for CT 1, CT 2, and CT 3. CT 4 is located at the Wilhelmina Child Hospital. There is little interaction with the other CT scans, so the demand for this patient group is left out of scope. The emergency patients are scanned at CT 5 (CT SEH), located at the emergency department, and at CT 1, CT 2, and CT 3. However, "normal patients" are not scanned on CT 5, so we will not adjust the appointment schedule for CT 5. However, we will consider the emergency patients' demand because they are also scanned on CT 1, 2 or 3.

This research aims to better align patient demand with appointment schedules so that the radiology department can work more efficiently. A demand forecast is needed as input, but it is not the primary focus of making a perfect demand forecast. In addition, walk-in patients can arrive any day, so technically, they do not have a waiting time until an appointment. So, the primary focus will be reducing access times for scheduled appointments.

1.3.2 Research questions

From the core problem, the main research question can be formulated as follows:

How can we better align patient demand and CT-appointment schedules to reach target access times for scheduled appointments?

This question helps to answer the action problem. Subsequently, sub-research questions are formulated to answer the main research questions. These questions form the structure of the report and are described below.

1. What is the current performance of the appointment schedule, and what was the historical demand pattern?

This research aims to reduce the access times for CT scans. It is essential to obtain detailed insight into the current situation, such as the current process for ordering a CT scan and the structure of the current appointment schedule. Next, we also focus on the Key Performance Indicator (KPI). To answer the following questions, we performed 23 stakeholder interviews, conducted observation studies, and gathered data from different sources. This is all described in Chapter 2.

1.1 What steps are involved in the current process for ordering a CT scan?

1.2 How is the scheduling of CT appointments currently managed?

1.3 What is the structure of the current appointment schedule?

1.4 Which KPIs (Key Performance Indicators) are currently in place and what are their performances?

2. According to the literature, what is the most effective method for creating an appointment schedule in a healthcare diagnostics setting that includes walk-in patients and access times?

The second research question has the purpose of gathering information for solution generation. With literature research, we found various methods to make appointment schedules that include unplanned arrivals and take access time into account. In the Scopus database, we selected relevant articles based on the search terms in the title, abstract, and keywords. With the help of the snowballing technique, more articles are reviewed. This literature review is presented in Chapter 3.

2.1 What level of detail must the appointment schedule have to incorporate walk-in patients and access time?

2.2 What are methods to incorporate walk-in appointments in an appointment schedule?

2.3 What are methods to prioritise different patient types in an appointment schedule?

2.4 What tactical decisions must be made to make an appointment schedule that incorporates walk-in patients?

2.5 Which methods can be used to develop an appointment schedule and test the performance?

3. What appointment schedule methods are most applicable, and how should they be designed?

Based on Chapter 3 and discussions with stakeholders, a method is developed that adjusts the current appointment schedule to meet the target access. For this, decisions about the policies, parameters, and model design had to be made. In Chapter 4 we describe the general solution approach and in Chapter 5 we described how the simulation should be used for this case study.

3.1 What method for creating an appointment scheduling is most suitable for the radiology department?

3.2 What should the input and output parameters of the method be?

3.3 How can the new appointment schedules be tested?

4. How can the performance of the new appointment schedules be measured?

In Chapter 6 the interventions are tested against the KPIs. The methods are compared with the current situation to test the performance.

4.1 What is the performance of the initial appointment schedule?

4.2 Which changes to the appointment schedule give the most improvement potential?

4.3 How robust is the method to relaxations of constraints and differences in input settings?

5. How can the proposed appointment schedule be effectively implemented?

Chapter 7 points out an implementation plan. This is important because the tool has no value without any guidance and proper implementation.

6. What recommendations can be drawn from the research conducted?

The last chapter, Chapter 8 we conclude the research and describe the limitations of this research and the practical and scientific contributions.

2 Current situation

This chapter describes the current situation of the CT department at UMC Utrecht to answer the research question: "What is the current performance of the appointment schedule, and what was the historical demand pattern?". Section 2.1 provides an overview of the CT radiology department. Section 2.2 explains the various order flows within the CT scan process. Section 2.3 details the current planning method, while Section 2.4 outlines the department's performance in its current state. We gathered all the data from existing data sheets and directly from systems. We organised the data with SQL. We also conducted numerous interviews and observational studies. To validate the data, we compared the output from the different sources and allowed stakeholders to review it.

2.1 Description of the CT Radiology department

The radiology department has multiple CT scanners. This section outlines their functionalities and the patient groups they serve, and explains the specific scan options available.

2.1.1 CT-scanner

As described in Section 1.3.1, there are five CT scanners (CT1, CT2, CT3, CT4, CT SEH) at UMC Utrecht. The CT scanner in the emergency department (CT SEH) is exclusively used for emergency patients. CT4 is located at the Wilhelmina Children's Hospital and is primarily used for children. When capacity is available, it is utilised by external hospitals or UMC Utrecht. CT1, CT2, and CT3 are situated next to each other at UMC Utrecht. CT1 and CT3 offer more advanced scanning functionalities, meaning some scans cannot be performed on CT2 and CT SEH. Emergency patients are therefore sometimes scanned on CT1. Transferring patients from CT1, CT2, or CT3 to the CT SEH scanner can be challenging because the CT SEH scanner is located in a different area. This, therefore, does not happen very often.

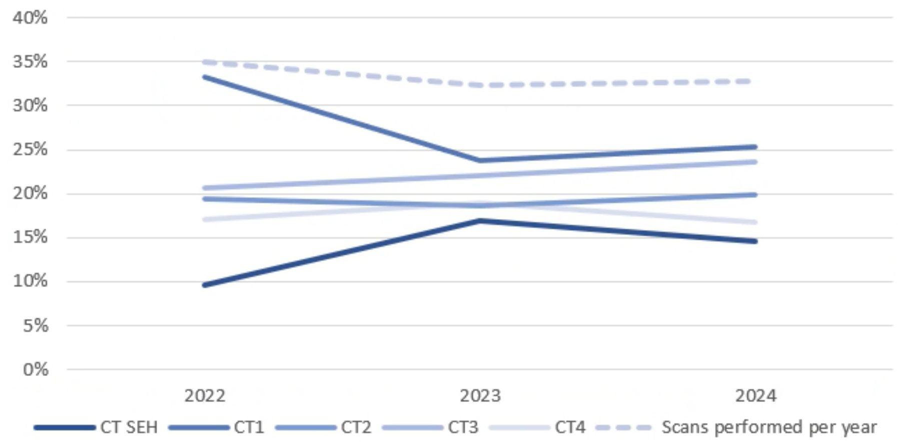

Figure 2.1: Percentage of all scans performed on each CT scanner and the production difference per year

Using production data of the number of scans performed in the previous years, we created a graph illustrating the percentage of scans performed on each CT scanner per year and the difference between the total number of scans performed annually (Figure 2.1). Due to confidentiality, the exact production numbers cannot be disclosed. However, the graph demonstrates that the total number of scans was higher in 2022 compared to 2023 and 2024. For example, in 2022, 10% of all scans were performed on the CT SEH. Most scans were performed on CT1, followed by CT3, while the CT SEH consistently had the lowest number of scans. The reasons for the discrepancy between CT1 and CT2 may be that CT2 was almost three times more closed during regular working hours. On top of this, it could be that on CT1, more scans are done outside regular working hours or during underutilization, since this scan has more functionalities. Section 2.4 will further explore the performance of the different CT scanners.

2.1.2 Overview of CT scan procedures

There are various ways to perform a scan. Some scans require patients to take oral contrast before the procedure, while others involve using an IV for contrast. Applying an IV is typically done in a separate room from the CT scanner and does not affect the CT scanner's production time. Additionally, some scans require patients to be sober. For specific rare procedures, patients need to be under anaesthesia. There are 171 types of scans, one might focus on the heart, while another targets the foot. These scans are requested by the ten different divisions of UMC Utrecht or by external parties. Over the past three years, 99% of all scans were scheduled for 20 minutes, while the remaining 1% were scheduled for 1 hour. These were often punctures.

2.2 System flow

This section begins by explaining the origin of an order. Next, it outlines how the order is handled within the system. Finally, it describes the production numbers.

2.2.1 Inflow

The order for a CT scan can originate from four different sources. Some of the orders come from the hospital's emergency department. These patients arrive at the hospital immediately after the order is made. Another source is the inpatient clinic, where hospitalised patients need a CT scan. Another source is the outpatient clinic, where doctors typically place an order after a patient has attended a consultation. Sometimes, it is clear that a patient requires a CT scan without a first consultation, and an order is placed immediately. Finally, external clinics such as private cardiology hospitals can directly place orders for CT scans, which are often outpatient appointments. These categories of orders have different priorities. Section 2.3 explains how these orders are scheduled.

2.2.2 Process flow

We developed a flowchart that illustrates the high-level process flow of a CT scan, including the percentage of patients per category during weekdays (Figure 2.2). For emergency patients and walk-in patients, no appointment is planned. After placing the order, the patient can proceed to the waiting room for their CT scan. Outpatients may include both Walk-in patients and outpatients. The flowchart distinguishes them as "outpatient appointment order created" and "walk-in order created". The walk-in orders make up the most significant part, with 43% of all orders. *Scheduling time* in the flowchart refers to the period between the order creation and the appointment scheduling. *Access time* measures the interval from when the order is created to when the actual appointment occurs. *Waiting time* is the period spent in the waiting room, while *processing time* is the duration of the scan itself. *Throughput time* is the total time spent in the hospital, from entering the waiting room to leaving the CT scan.

2.2.3 Production flow

Figure 2.3 illustrates the difference in the number of scans performed per month for each year. For example, in 2022 10% of all scans took place in January. In some months, there is a consistent trend across all years where production increases or decreases. For instance, in February the production was always lower than January. In other months, such as August, there is no trend.

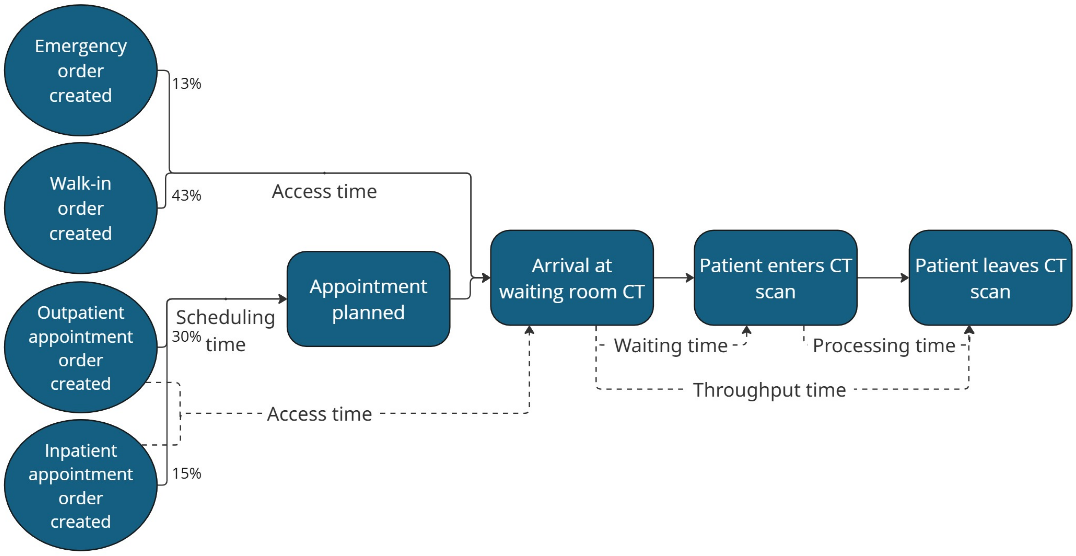

Figure 2.2: High-level process flow with waiting times

These differences can be caused by fluctuating demand or inconsistent opening hours. We only have data from the opening hours from March to December of 2024. The data show that closing hours were significantly higher in May, August, and December compared to other months. As illustrated in Figure 2.3, production in 2024 was also lower during these months. In contrast, June, July, and October had relatively low closing hours, with July and October showing higher production levels. However, due to limited data, it is not possible to determine whether these trends reflect seasonal variation in demand or whether production is directly influenced by changes in opening hours.

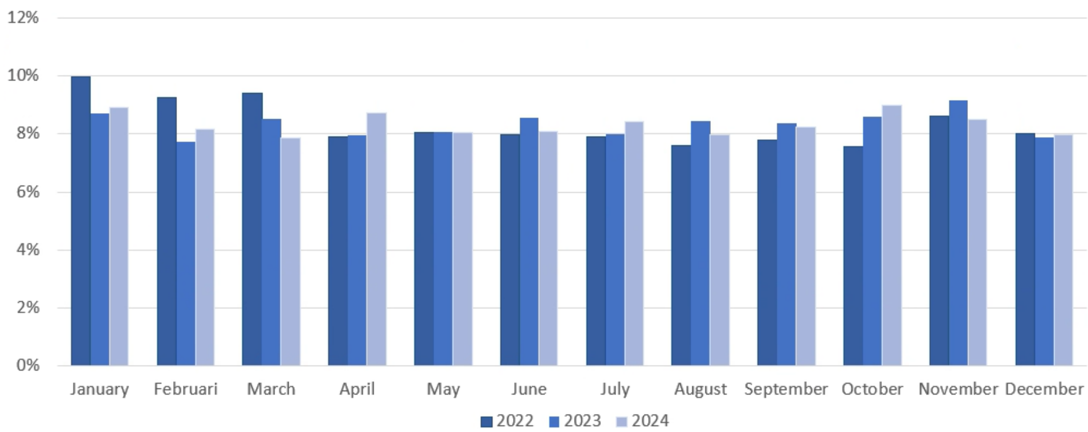

Figure 2.3: Difference in the number of scans performed per month for each year

Figure 2.4 shows the daily production across different years. The production during the weekend is significantly lower than on weekdays. This is logical since CT1 and CT SEH are open for emergency patients, and the others are closed. Very occasionally, CT1 is open for appointments to work overtime when the waiting lists are too long and personnel are available.

2.3 The current planning method

This section provides a more detailed explanation of the planning process for CT scans. In the hospital, there are different employees responsible for the strategic, tactical and operational planning. It is therefore important to know what takes place at each planning level. Moreover, understanding the current situation will help us in finding and solving the problem. We achieved this with the help of the healthcare planning and control framework outlined by Hans et al. (2012).

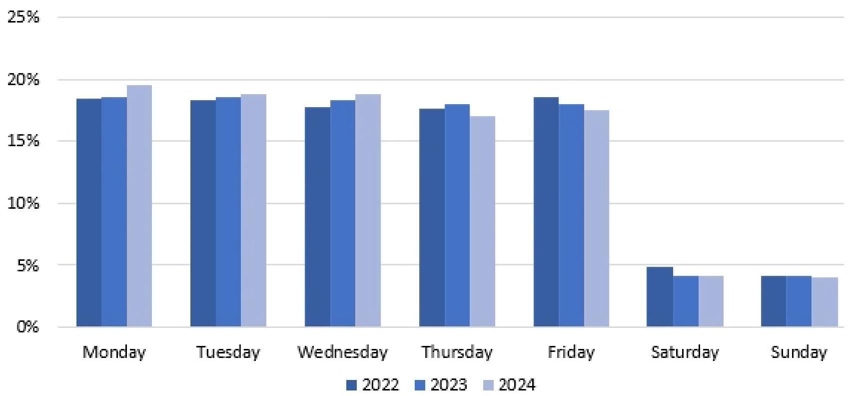

Figure 2.4: Percentage of number of scans per day of the week per year

2.3.1 Health care planning and control framework

Hans et al. (2012) states that decision-making varies at different points as information becomes available gradually. This concept forms the basis of their framework, which consists of three hierarchical levels: strategic, tactical, and operational levels of control. As shown in Table 2.1, we adapted this framework for the CT scan department. Medical planning involves researching and developing new CT scan software and protocols. Resource capacity planning regards decisions about flexible or dedicated CT scanners and appointment scheduling. Material planning involves designing the supply chain for consumable resources. Financial planning addresses budget allocation for CT machine maintenance and staff salaries. Improving resource capacity planning aligns closely with the focus of this research (better aligning the appointment schedule with the demand to meet target access times). The following sections will provide further details on resource capacity planning for appointments and demand at each hierarchical level.

Area / Level	Medical planning	Resource capacity planning	Materials planning	Financial planning
Strategic	Research of software and CT protocols	Long-term laboratory technicians planning, capacity CT scans, flexible or dedicated CTs	Supply chain design	Investments in CT scans or room upgrades, contracting insurance companies
Tactical	Case mix protocol planning	Appointment scheduling, staffing, admission planning	Supplier selection and contracting.	Budget allocation for supplies, CT maintenance, salaries
Offline operational	Individual protocol selection	Appointment scheduling, workforce shift planning	Medical supplies purchasing, choosing order sizes	Cost tracking
Online operational	Immediate protocol adjustments	Emergency and walk-in coordination, adjustments to daily staff schedules	Rush ordering, inventory replenishing	Billing complications

Scope colour

Table 2.1: CT scan healthcare planning and control framework (Hans et al., 2012)

2.3.2 Strategic planning

Strategic-level planning involves long-term decision-making to define an organisation's mission based on high-level forecasts (Hans et al., 2012). Due to cost constraints, UMC Utrecht cannot purchase new CT scanners and extra staff. On top of this, hiring additional staff is impossible because of labour market shortages. As a result, the CT department has a fixed capacity. However, decisions about flexible or dedicated CT scanners are also strategic (Zonderland et al., 2021) and could influence the access times.

2.3.3 Tactical planning

In tactical planning, decisions are made over a longer horizon than operational planning but shorter than strategic planning. The CT department schedules its laboratory technicians 4 weeks in advance. The number of technicians scheduled depends on availability rather than predicted demand. When there is an insufficient number of technicians available, one or more CT scanners may be closed. The scanners can be opened during the evening or weekend if necessary.

Appointment schedules allocate specific time slots for particular patient groups, specialities, or treatments. Some slots are reserved for a heart scan, intensive care patients, or walk-in patients. These blocks are scheduled based on experience. This means that a specific category will get less space in the appointment schedule when it is consistently underutilised. There is no overview of the actual waiting list per category. However, if a department requests more time and the CT department observes that its appointment slots are quickly filled, they may try to allocate more time. To limit overtime, they buffer spare time by planning more time than needed, leaving gaps in the schedule, and blocking time for unexpected situations (Figure 2.5). This leads to underutilization and a wide variation in access times for different patient groups.

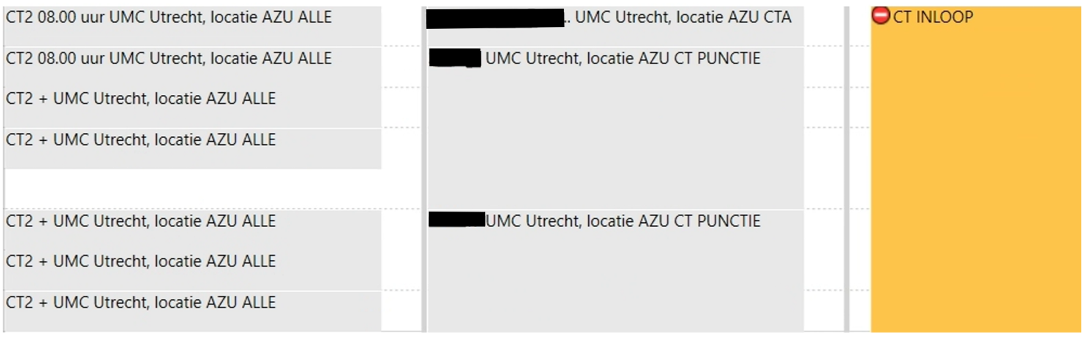

Figure 2.5: Example of appointment schedule of the morning of three CT scanners

2.3.4 Offline operational planning

Operational planning involves short-term decision-making (Hans et al., 2012). We developed a flow chart of the current appointment scheduling process through stakeholder interviews, shown in Figure 2.6. This appointment scheduling takes place both centrally and decentrally. The outpatient clinic administration is responsible for scheduling appointments. They can choose a walk-in date, in this case, the patient is expected to go to the radiology department on a specific date. They can also select a walk-in period, where a patient is asked to come before a specific date. Or the patient gets a particular appointment date and time. They contact the central radiology administration if they cannot find a suitable time slot. The central administration often has reserved time slots for emergency and clinical patients that can be used. The central radiology administration schedules appointments for external patients. Generally, the patients are planned on a first-come, first-served basis.

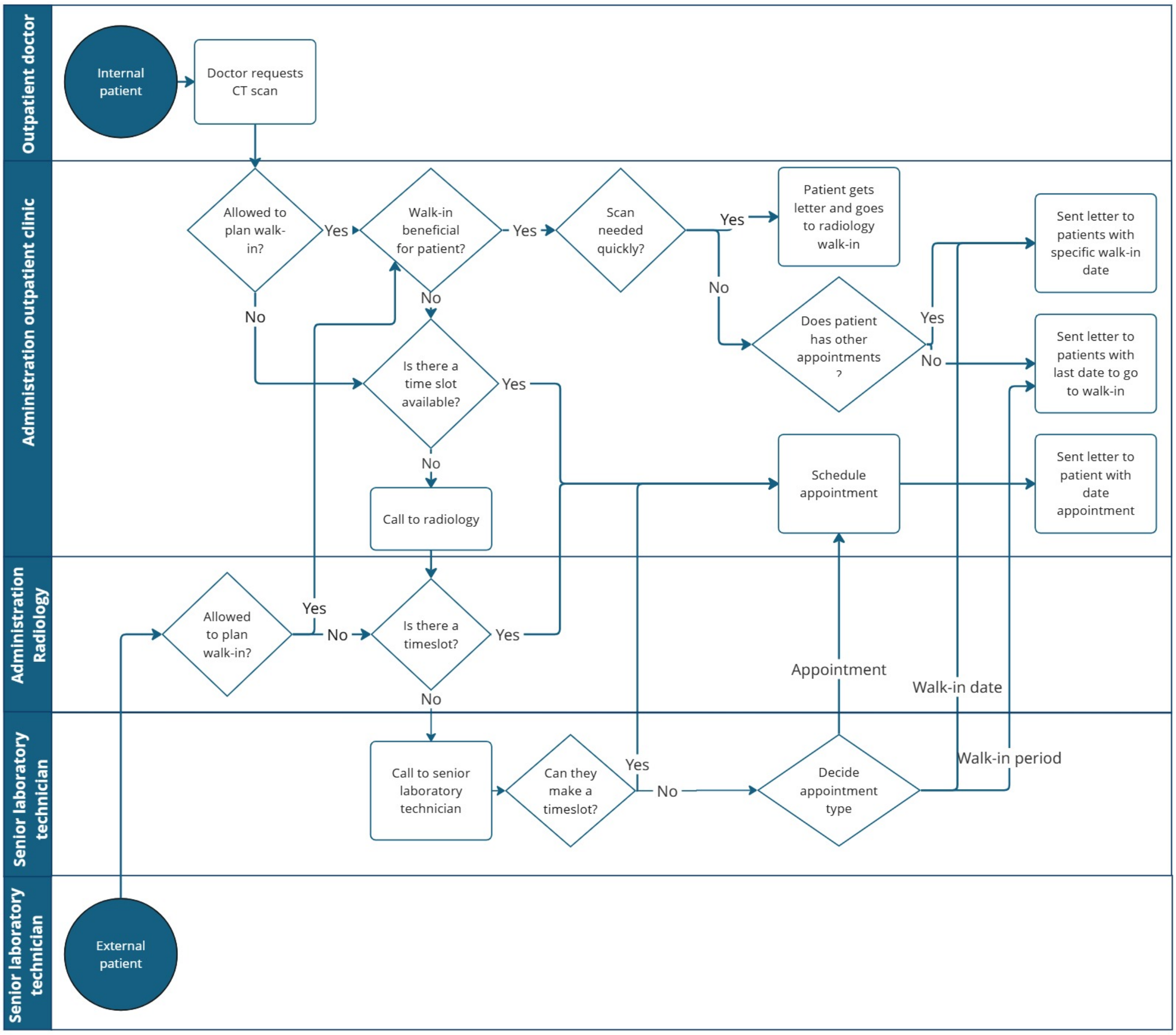

Figure 2.6: Flow chart appointment planning process of the CT scans

2.3.5 Online operational planning

Online operational planning concerns reactive decision-making. In this case, it refers to the add-on scheduling of emergency patients and determining the order in which patients in the waiting room are scanned.

2.3.6 Scope planning level

As outlined in Section 2.3.1 the primary focus of this research is on resource capacity planning. As Zonderland et al. (2021) suggests, we will start the capacity management at the strategic level and work it through till the operational level. We have to decide on flexible or dedicated CT scanners for the strategic level, not the capacity, since this is fixed. At the tactical level, suggestions will be made to improve the block scheduling of appointments to meet the target access times. Furthermore, operational planning significantly affects the waiting times of different patient groups, which makes decisions in prioritising different patient groups important.

2.4 Performance

Together with our stakeholders and the literature, we devised multiple categories to measure our system's performance (Bhattacharjee and Ray, 2016, Kortbeek et al., 2014). We gathered the data for this analysis through real-life observational studies, reenactments, and multiple meetings to identify connections between different datasets and explore them. Most of the data pertains only to the data of 2024, as analysing earlier years was impossible due to a transition to a new EPD system (Electronic patient file).

2.4.1 Appointment duration

Appointment duration is a key factor in system performance and appointment scheduling (Ma et al., 2016). In this radiology department, appointment end times are often poorly registered, making durations in the data appear longer than in reality. To correct this, we adjusted the end times to the start of the next appointment. Start times are generally accurate, as confirmed by our observations and laboratory technicians. This is because the recorded start time corresponds to the moment a technician opens the scan protocol and initiates the scan. However, idle time is still included in the total appointment duration. Since the recorded end time is based on the start of the next appointment, we cannot accurately determine the duration of any idle time between appointments. As a result, it is not possible to filter out this idle time from the data. Yet, the average appointment time remains shorter than planned (Table 2.2). We concluded with the laboratory technicians that the maximum appointment duration is still exceptionally high and not representative of real life. With the help of professionals and the existing data, we concluded that the maximum scan duration of appointments planned for 60 minutes is 90 minutes and appointments of 20 minutes are a maximum of 35 or 25 minutes, depending on whether they are more complex scans. Therefore, an extra column is included with the average appointment duration without outliers. As you can see in the table, the actual average appointment duration differs significantly from the raw data. Distinguishing scan durations between more categories, such as patient types, is challenging. Moreover, observations showed shorter durations during busy periods due to increased staff efficiency. We calculated how often the scan duration in the data was longer than the maximum time stated above. With these calculations, we can conclude that at least 58% of the appointment durations in the data deviate from real life.

	Planned appointment duration	Average appointment duration in data	Maximum duration per scanner in data	Average appointment duration without outliers
CT 1	1:00:00	1:02:19	5:37:00	58:00 max duration 1:30:00
CT 1	20:00	16:56	5:37:00	13:51 max duration 35:00
CT 2	20:00	15:06	3:45:00	12:12 max duration 25:00
CT 3	20:00	14:24	3:59:00	12:04 max duration 25:00
CT Average	20:00	14:28	3:58:00	13:14

Table 2.2: Average appointment duration

2.4.2 Capacity

A key factor in calculating available capacity is the opening hours. Most departments display capacity using appointment slots in the block schedule, marking unavailable times with orange blocks. However, this department also uses blocks for intended purposes, such as staffing shortages and maintenance, and to reserve time for walk-in or emergency patients. Although this is the actual capacity, it is not shown in the system (see Figure 2.7). As a result, we could not rely on the system data to determine capacity. Instead, we manually reviewed the CT scanner agendas and data for 218 weekdays. Table 2.3 shows the number of closed days. A typical working day runs from 08:00 to 16:30, with one scanner operating until 17:30 to reduce overtime. For our calculations, we assumed this extended schedule always applies to CT scanner 1.

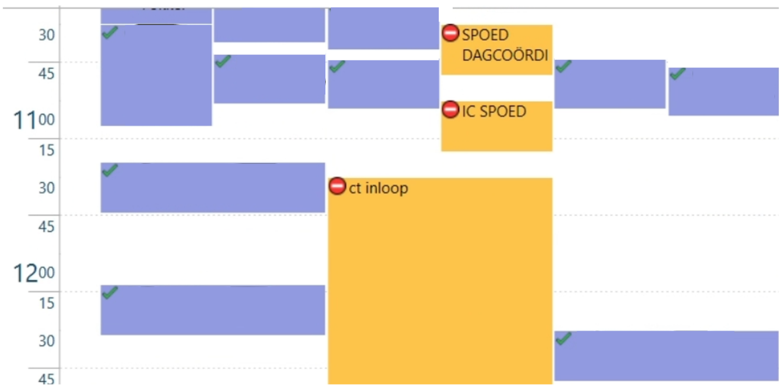

Figure 2.7: Agenda CT scan with blockings

2.4.3 Utilization

Utilisation is essential for both management and patients. Higher utilisation enables more scans to be performed, potentially reducing access times, but it can also increase workload and lead to overtime. Therefore, maintaining a balanced level of utilisation is essential. We distinguish two types: staff utilisation and CT scanner utilisation. Currently, the department lacks insight into either. Therefore, we calculated it by dividing the appointment durations by the available capacity, as shown in the formulas below. Table 2.3 presents the utilisation per scanner. A notable observation is the large discrepancy between planned and actual utilisation, mainly due to inaccurate planned appointment durations. This indicates that staff and CT scanners are often underutilised, highlighting room for improvement.

	CT1	CT 2	CT 3
Total days	218	218	218
Days not open	11.6	31.1	16.1
Planned utilization staff	86%	96%	100%
Realized utilization staff	60%	64%	67%
Planned utilization CT scanners	81%	82%	93%
Realized utilization CT scanners	56%	55%	62%

Table 2.3: Utilisation of the CT scanners of 2024

$$\text{Planned utilization staff} = \frac{\text{Total number of scans} \times \text{Average planned appointment duration}}{(\text{Full days open} - \text{Days not open}) \times \text{Workinghours per day} \times 60} \times 100$$

$$\text{Realized utilization staff} = \frac{\text{Total number of scans} \times \text{Average realized appointment duration}}{(\text{Full days open} - \text{Days not open}) \times \text{Workinghours per day} \times 60} \times 100$$

$$\text{Planned utilization scanners} = \frac{\text{Total number of scans} \times \text{Average planned appointment duration}}{\text{Number of days} \times \text{Workinghours per day} \times 60} \times 100$$

$$\text{Realized utilization scanners} = \frac{\text{Total number of scans} \times \text{Average realized appointment duration}}{\text{Number of days} \times \text{Workinghours per day} \times 60} \times 100$$

2.4.4 Access time

As mentioned in Section 1.2.1, only outpatient appointments are associated with access times. UMC Utrecht defines future access time based on the third available appointment. Over the past six months, this averaged 48 days, with a maximum of 85 days. We also calculated access time using historical data by measuring the time between appointment creation and the appointment date. For outpatients, this resulted in an average access time of 31 days. We discussed with the administration employees that 91 days was the maximum access time in the past. We excluded access times over 91 days to avoid misleading results, as these likely reflect voluntary waiting or a half-year checkup. Figure 2.8 presents a box plot of the access times. You see that the average waiting time is 31 days. Moreover, 25% of the patients have an access time below 4 days, 50% below 22 days and 75% below 55 days.

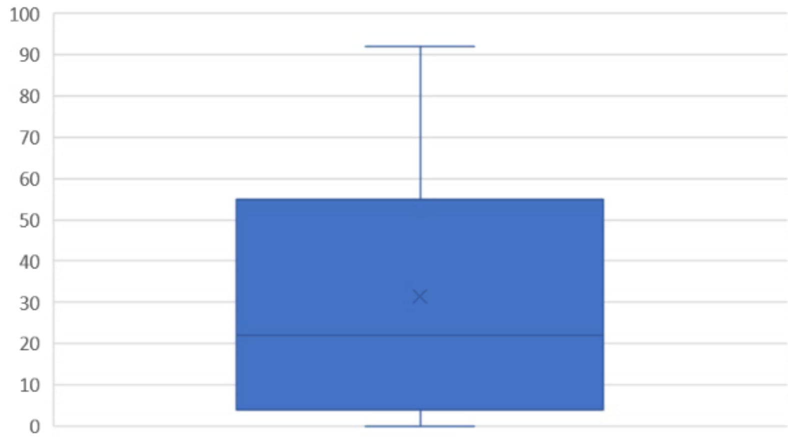

Figure 2.8: Boxplot of the access times until 100 days

2.4.5 Variation number of scans

Section 2.1.1 showed significant variation in the number of scans performed per CT scanner. Section 2.2.3 indicated some fluctuation in the number of scans per month and per day. Figure 2.9 displays the average number of walk-in patients arriving per hour, per day. This varies substantially, suggesting a significant difference in the arrival rate across both weekdays and hours of the day. Figure 2.7 illustrates an example of daily variation. Many appointments are scheduled in the first hour, while almost none are planned in the second. To explore this further, we created heatmaps for all three CT scanners, showing the average number of appointments scheduled per hour. Table 2.4 shows the heatmap of one CT scanners. The results were striking: the average ranged from 0.3 to 4.2 scans per scanner per hour. This variation is not reflected in the block schedule, potentially affecting patient waiting times, staff workload, and overall utilisation.

2.4.6 Waiting time

We calculated waiting time as the difference between the scheduled appointment time and the actual start time. If a patient arrives early, the scheduled time is used, as this reflects voluntary waiting. On average, emergency patients wait 7.11 minutes, inpatients 6.26 minutes, outpatients 5.14 minutes, and walk-ins 19.10 minutes. We observed that most inpatient and outpatient appointments start on or before the scheduled time. Walk-in patients get an appointment when they arrive in the hospital, and experience the longest average waiting times. Therefore, we focused on this group in Figure 2.10, which shows the percentage of walk-in patients who wait longer than a given duration. For instance, 55% wait more than 15 minutes, and 20% wait more than 35 minutes.

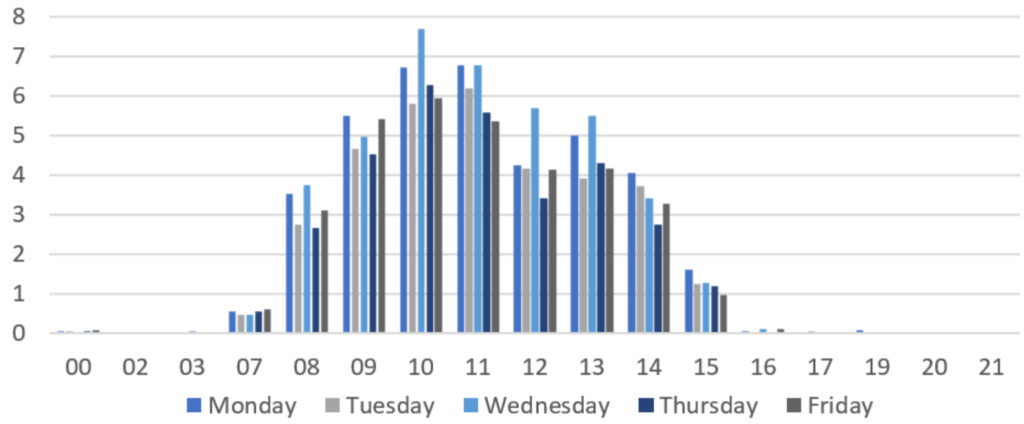

Figure 2.9: Average number of walk-ins per hour per day

	Monday	Tuesday	Wednesday	Thursday	Friday
08:00:00	2,5	2,7	2,1	3,2	3,2
09:00:00	4,4	3,5	3,0	3,8	4,4
10:00:00	4,3	3,6	4,2	3,9	4,4
11:00:00	4,0	3,3	3,3	3,5	3,3
12:00:00	2,5	2,3	3,1	1,9	2,2
13:00:00	2,6	1,9	3,4	2,2	2,7
14:00:00	3,9	3,5	2,9	3,1	3,4
15:00:00	1,9	2,0	2,8	1,5	1,9
16:00:00	1,4	1,0	0,3	0,9	0,7

Table 2.4: Heatmap of one CT scanner (number of scans planned per hour)

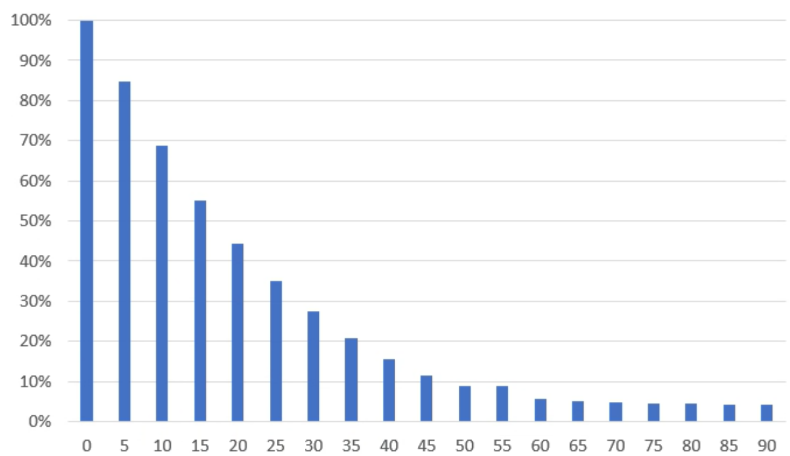

Figure 2.10: Percentage of the walk-in patients that have to wait longer then ..minutes

2.4.7 Overtime

There is both planned and unplanned overtime. One laboratory technician is always available to scan patients during evenings and weekends. This technician can perform additional scans if patients are still in the waiting room after regular hours. When a scan is already in progress at the end of a shift, it is completed, though this overtime is minimal due to the short scan durations. If multiple patients remain in the waiting room, overtime may be required. However, this has not occurred in the past year. Planned overtime, such as evening or weekend shifts, is occasionally scheduled to meet demand.

2.4.8 Number of no-shows

Unfortunately, no data on the number of no-shows is available. However, since the introduction of the walk-in hour, no-shows appear to have become less frequent. Inpatient appointments are usually scheduled on short notice, making no-shows unlikely. Outpatient appointments account for approximately 30% of all appointments, and no-shows still occur in this group, although the exact number is unknown.

2.5 Chapter conclusion

This chapter provided a comprehensive overview of the current CT scheduling process. The focus of this research will be on resource capacity planning. Key decisions include whether to have dedicated or flexible CT scanners, how to improve block scheduling, and how to prioritise different patient groups. We identified issues such as inconsistent data registration and the use of CT agenda blocks that do not accurately reflect actual capacity. Utilisation is currently low, and there is significant variation in patient arrivals by hour and day. This variability is not effectively incorporated into the block schedule, which likely contributes to longer waiting times, increased staff workload, low utilisation, and high access times. These findings emphasise the need for a more adaptive appointment scheduling approach that balances access time, workload, and flexibility for walk-in patients. This provides the foundation for the literature review in Chapter 3, where we explore methods for improving healthcare appointment scheduling under uncertainty and fluctuating demand.

3 Literature review

This chapter presents a literature review to address the research question: "According to the literature, what is the most effective method for creating an appointment schedule in a healthcare setting that includes walk-in patients?". Appointment schedules can be classified at various levels of detail, as outlined in Section 3.1. Section 3.2 explores methods for prioritising different patient types. The process of developing an appointment schedule involves several tactical decisions, which are summarised in Section 3.3. Finally, Section 3.4 reviews methods from the literature used to evaluate the performance of appointment schedules.

3.1 The level of detail of appointment scheduling methods

This research aims to improve appointment scheduling methods in a healthcare setting, accommodating walk-in patients. The primary objective is to reduce patient access times by aligning demand with available appointment slots. Zomer (2022) introduced a framework for comparing different appointment scheduling methods. This framework has three dimensions: the various goals, the level of detail, and the characteristics of appointment scheduling methods. Using this framework as a foundation, we developed a customised framework tailored to this research (Table 3.1).

For this research, the primary goal of a scheduling method must be to minimise access time. Consequently, all scheduling methods included in our framework address access time. According to Zomer (2022), appointment scheduling methods can be categorised into five types. Among these, the categories percentage of patient types, block scheduling, and slots filled with service types include scheduling methods that account for access time. Therefore, these three categories are included in the first dimension of our framework (Table 3.1). The primary distinction between the categories lies in the level of detail in the appointment schedules. These categories will be further discussed in the following sections. The second dimension of our framework incorporates the characteristics of the appointment scheduling methods, as it is essential to determine whether they accommodate the needs of walk-in patients. The planning process of walk-in patients can be seen as emergency arrivals since they arrive at the clinic without an appointment. However, the methods must take into account the different priority rules of emergency patients and walk-in patients.

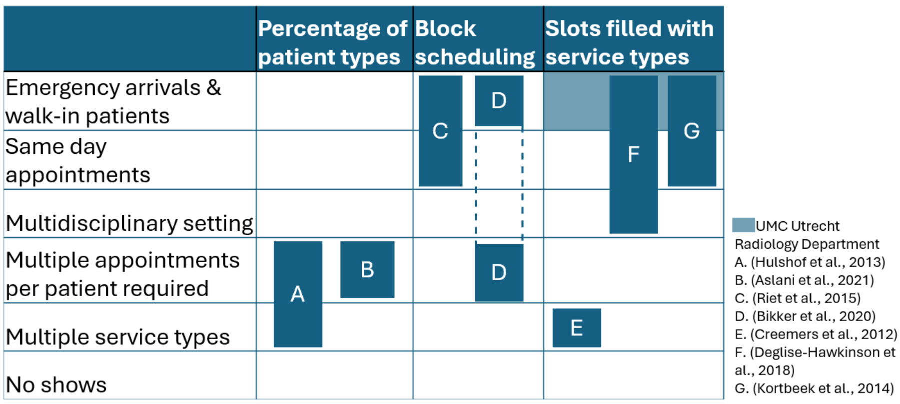

Table 3.1: Scheduling methods that keep in mind the access time (Zomer, 2022)

3.1.1 Percentages for patient types

This appointment scheduling method operates at the lowest level of detail. Specifically, it only specifies the percentage of capacity allocated to different patient types (Zomer, 2022). It does not consider emergency arrivals, walk-in patients, or same-day appointments.

3.1.2 Block scheduling

The methods in this category allocate capacity among different patient types and assign specific lengths to each block (Zomer, 2022). This category includes the method developed by Van Riet and Demeulemeester (2015), which considers emergency arrivals and same-day appointments. Additionally, it emphasises the access times. Appointment scheduling methods that reserve appointment blocks for specific patient types are commonly used in surgery planning (Van Riet and Demeulemeester, 2015). Most of these methods use block-wise scheduling (Van Riet and Demeulemeester, 2015). Van Riet and Demeulemeester (2015) outlines three strategies for incorporating emergency arrivals into surgery appointment scheduling: the dedicated policy, the flexible policy, and the hybrid policy. These strategies will be further explained in Section 3.3.3.

3.1.3 Slots filled with patient types

This scheduling method has fixed slots with predetermined start and end times allocated to specific patient types (Zomer, 2022). This category includes the methods proposed by Kortbeek et al. (2014) and Deglise-Hawkinson et al. (2018), which incorporate emergency arrivals and same-day appointments while focusing on access times. The method developed by Kortbeek et al. (2014) specifies the number of appointments to schedule each day and their optimal timing. It also determines the availability of slots for walk-in patients, aiming to balance unscheduled waiting times with scheduled access times. Access time is evaluated using a discrete-time cyclic queuing model, while the daily-level process is assessed with a Markov reward process (Kortbeek et al., 2014).

Deglise-Hawkinson et al. (2018) presents an appointment scheduling approach that allocates slots to patient groups based on urgency while minimising access delays for all appointments. Access time, also called indirect delay, is a KPI often overlooked in the literature. Gupta and Denton (2008) were among the first to incorporate access time into planning and scheduling methods. This problem is formulated as a queueing network optimisation and approximated using deterministic linear optimisation. The approach seeks to balance workloads and meet targeted access times, ensuring efficient scheduling for urgent and non-urgent patient groups.

3.2 How to prioritise different patient types

This section explores methods for prioritising different patient types in appointment scheduling, focusing on balancing access and waiting times for various patient groups. Prioritising one patient group can significantly affect resource utilisation and patient satisfaction. The literature highlights multiple approaches to prioritise patients, each with unique benefits and trade-offs. By analysing these strategies, this chapter aims to identify how prioritisation affects appointment scheduling methods and how the preferred patient needs can be met.

3.2.1 Priority access times for urgent patients

The method described by Deglise-Hawkinson et al. (2018) is an example of an appointment scheduling approach that allocates a specific number of slots to patient groups based on urgency. This method reduces access times for urgent patient groups, but it may increase the average access time across all patients. It balances the trade-offs between throughput, overtime, and access delay, making it particularly useful for a system with multiple patient classes, specialists, and competing performance metrics. To implement this method, the patient groups must have different priorities regarding access times.

3.2.2 Priority access times scheduled patients

Sequential appointment scheduling prioritises scheduled patients over walk-in patients. This approach allocates unfilled time slots not assigned to fixed appointments to walk-in patients (Yan et al., 2014). This method primarily focuses on cost efficiency and considers access times and high resource utilisation factors. It can, however, also cause more overtime. Suppose the department handles a large number of walk-in patients. In that case, this method may be less suitable, as it can lead to significant fluctuations in workload, overtime, and patient waiting times.

3.2.3 Priority access times for scheduled patients and freedom for walk-in patients

This method, called delay scheduling, allows walk-in patients to decide whether to accept the waiting time or return later (Reilly et al., 1978). This approach focuses on giving walk-in patients more control over their scheduling decisions, as they can decide whether to wait or not. The workload of the patients who choose not to stay in the queue is shifted to a different time, which can help improve overall waiting times, reduce workload variation, and minimise overtime. However, predicting when the appointment will occur is more challenging as patients have more freedom. Additionally, access times may increase since no specific slots are reserved for walk-in patients, potentially resulting in longer waiting times or patients returning later.

3.2.4 Priority access times scheduled patients and waiting time walk-in patients

The method developed by Kortbeek et al. (2014) considers both the access times for scheduled patients and the waiting time for walk-in patients. This method relies on the fact that demand for both scheduled and walk-in patients often follows a cyclic pattern. This allows appointment schedules to be designed cyclically as well. The cycle length can range from days to weeks or even months. The method determines the number of appointments to schedule each day and the optimal timing for these appointments using queue length probabilities. It uses fixed appointment slot lengths, and the service always takes one time slot. Patients who cannot be treated within their desired access times will be rejected, and no overtime is allowed.

3.2.5 Priority waiting time walk-in patients

One method that prioritises the waiting time of walk-in patients is off-peak scheduling developed by Gupta and Denton (2008). This method predicts the arrival patterns of walk-in and emergency patients. The remaining time slots can be used for scheduled patients in a way that complements the arrival of walk-in patients. As a result, patient arrivals become more evenly distributed, leading to a more homogeneous workload and reduced overtime (Zonderland et al., 2009). For this method, it is crucial that the arrival times of walk-in patients are predictable and that the remaining appointments can be planned flexibly. However, since the primary focus is reducing the waiting time for walk-in patients, the access time for scheduled appointments may be longer.

3.2.6 Priority on appointment characteristics

In addition to prioritising patient types, appointment scheduling can also prioritise appointment characteristics. The literature identifies various sequencing rules for scheduling, such as first-come, first-served, shortest-case-first, and longest-case-first (Hulshof et al., 2012). For this method to be effective, appointment characteristics must be more important than patient types. Most importantly, the characteristics of appointments must differ significantly from one another.

3.2.7 Summary

This section has explored various methods for prioritising patient groups in scheduling systems. Table 3.2 summarises these approaches by categorising them by their priority group, focus, requirements, and limitations. Depending on the patient group's needs, each method balances trade-offs such as access times, waiting times, and resource utilisation. By analysing these methods, healthcare providers can choose the most suitable approach based on their priorities.

Priority group	Methods	Focus	Requirements	Limitations
Access urgent patients	Integrated care, access management	Complex care	Different priority groups	Increase mean access time
Access scheduled patients	Sequential appointment scheduling	Costs, utilization	Small group of walk-ins	Variation: overtime, waiting time, workload
Access scheduled patients, freedom walk-ins	Delay scheduling	Freedom of patients, workload	Possibility to leave, flexible system	Uncertainty, longer accesstimes walk-ins
Access scheduled patients, waiting time walk-ins	Cyclic appointment schedules	Access-, waiting-, over-time	Cyclic demand, fixed appointment lengths	Patients can be rejected
Waiting time walk-ins	Off-peak scheduling	Overtime, workload, waiting time	Prediction of walk-in arrivals, flexible remaining slots	Long access times scheduled patients
Appointment characteristics	Sequencing rules	Utilization, efficiency	Different appointment characteristics	May overlook patient type priorities

Table 3.2: Summary of options to prioritise patient groups with the methods

3.3 Tactical decisions for appointment schedules

This chapter explores tactical decisions, queueing dynamics, and policies for managing elective and emergency patient flows, which are necessary for designing efficient scheduling systems.

3.3.1 Tactical decisions

Zonderland et al. (2021) identifies eight key tactical decisions to address when designing an effective appointment schedule. These decisions include capacity allocation, the number of patients per consultation session, patient overbooking, the length of the appointment slot, the number of patients per appointment slot, the sequence of appointments, queue discipline in the waiting room, and anticipation for unscheduled patients. Each decision influences access times, waiting times, and overtime. For instance, the number of appointment slots per consultation session directly impacts patient access time and waiting times (Hulshof et al., 2012). Increasing the number of patients scheduled in a session can reduce access times, but often also leads to longer waiting times for patients and increased staff overtime. Similarly, the decision to allow overbooking can improve utilisation rates and reduce access times by compensating for the effects of no-shows. However, overbooking may also lead to extended patient waiting times and higher staff overtime (Hulshof et al., 2012). These trade-offs highlight the importance of carefully balancing priorities when designing appointment schedules.

3.3.2 Queueing theory in a healthcare setting

To optimise the current appointment schedule, it is essential to understand the existing system flow and the underlying queueing mechanisms. This understanding is crucial, for example, when simulating the current system or evaluating potential improvements. Healthcare systems can be described as complex queueing networks, as highlighted by Creemers and Lambrecht (2008). The patient flow through the healthcare system significantly impacts the outcomes of an appointment schedule (Srinivas and Ravindran, 2020). Queueing theory provides a valuable framework for understanding and modelling this flow. In systems with parallel servers, patients need to utilise only one server to complete their service (Winston, 2003), which we see in pictures a and c of Figure 3.1. In contrast, serial treatments require a sequence of different services, needing patients to go through all stages sequentially to complete the process, as we see in pictures b and d of Figure 3.1.

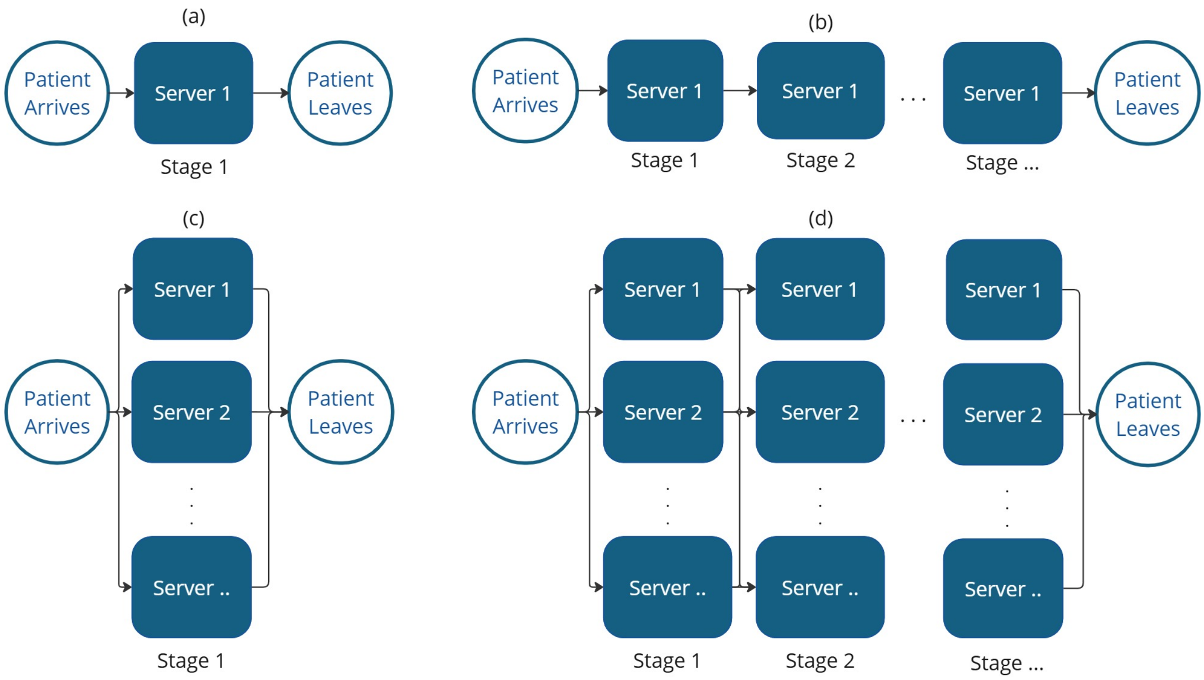

Figure 3.1: Queueing systems (Srinivas and Ravindran, 2020)

The total waiting time experienced in single-stage and multi-stage systems can differ significantly (Srinivas and Ravindran, 2020). In multi-stage systems, overall waiting times may vary considerably due to uneven waiting times among servers, where some servers may become bottlenecks. With a single server, the waiting time only depends on the queue length for that specific server. This difference highlights the effects of server dynamics and designs when evaluating or optimising appointment schedules.

3.3.3 Dedicated policy versus flexible and hybrid policy

As outlined in section 3.1.2, there are three primary methods for managing emergency patients. However, the literature does not yet agree on which of these methods best fits a system that includes walk-in patients. These methods are described in detail below. The dedicated policy assigns separate operating rooms for elective and emergency patients. This approach reduces the waiting time for non-elective patients but increases the waiting time for elective patients (Van Riet and Demeulemeester, 2015). Additionally, it leads to higher utilisation of regular operating rooms while lowering the utilisation of dedicated emergency rooms (Van Riet and Demeulemeester, 2015). This method aims to improve access to care for both patient groups (Hans and Vanberkel, 2011).

Under the flexible policy, all operating rooms are available for elective and emergency cases. According to a survey by Cardoen et al. (2010), 85% of hospitals implement this strategy.

Elective surgeries are scheduled in advance, while emergency surgeries are accommodated using pre-scheduled buffers (break-in moments) or reserving slack capacity. Moreover, scheduling appointments in complement to walk-in arrivals can lead to a more consistent patient arrival pattern (Zonderland et al., 2009). Ferrand et al. (2010) argue that elective and emergency patients benefit from greater flexibility in highly variable systems. However, many studies indicate that the increased flexibility often disrupts the elective schedule by incoming non-electives and emergency patients, leading to longer waiting times for elective patients (Van Riet and Demeulemeester, 2015). On the other hand, waiting times for non-elective patients and emergency patients are expected to decrease. Furthermore, this policy will result in an overall improved utilisation and increased overtime (Van Riet and Demeulemeester, 2015). The objective of this policy is to minimise waiting time for emergency patients (Hans and Vanberkel, 2011).

The hybrid policy combines elements of both the dedicated and flexible policies, aiming to achieve a better balance between flexibility and access time. Tancrez et al. (2009) proposes a model in which there is one dedicated operating room for non-elective patients, two rooms in which non-electives can still enter with priority if the dedicated room is occupied, and the other rooms are just for elective patients. Despite its potential advantages, research on the hybrid policy remains limited, and its benefits are not yet fully validated. However, Ferrand et al. (2014) suggest that the hybrid approach outperforms the dedicated and flexible policies by improving waiting times for elective and non-elective patients while reducing overtime. These findings indicate that the hybrid policy could provide promising solutions for systems seeking to optimise access and waiting times across diverse patient groups.

3.4 Solution approaches

Several methodologies exist to optimise appointment schedules. Some can evaluate the performance of a single schedule rule, compare various alternative appointment schedules, or search for near-optimal schedules (Bhattacharjee and Ray, 2016). According to Bhattacharjee and Ray (2016), all these approaches require an accurate representation of the system's inputs, processes, and outputs to evaluate an appointment schedule's performance effectively. Two primary techniques are commonly used to model patient flows within a system: analytical queueing-theoretic models and discrete-event simulation. Among these, discrete-event simulation is the most widely used due to its flexibility and detailed representation capabilities (Cayirli and Veral, 2003). This section explores the advantages and limitations of these two methodologies used for appointment scheduling.

Analytical theoretic models

Mercer was the first to conduct a queueing theoretic analysis of appointment systems (Bhattacharjee and Ray, 2016). In his model, patients were scheduled to arrive at equal time intervals. The patient could come at the start of the scheduled interval with a general lateness distribution or may not show up at all. Doi, Chen, and Osawa (Bhattacharjee and Ray, 2016 analysed an appointment system where scheduled inter-arrival times were assigned randomly and were modelled as independent and identically distributed random variables. In this model, patients who cannot enter the system within their assigned interval leave. Hassin and Mendel (2008) developed an analytical queueing model that considered the impact of no-shows. However, integrating all the unique features of an appointment system, such as scheduled and unscheduled arrivals, non-punctuality, and no-shows, into a single queueing model is complex and often intractable. Additionally, issues related to sequencing and including multiple patient classes have seldom been addressed in the literature (Kolisch and Sickinger, 2007). To conclude, this method can be helpful when one of the characteristics of an appointment system needs to be tested and improved (Huh et al., 2013).

Discrete-event simulation

Analytical queueing models encounter several limitations when applied to appointment systems that involve no-shows, walk-ins, multiple patient types, sequencing rules, and non-exponential inter-arrival and service times (Bhattacharjee and Ray, 2016). In contrast, discrete-event simulation is widely used for studying complex systems and accounting for uncertainty (Van Riet and Demeulemeester, 2015). This method can model performance within a stochastic environment, making it highly suitable for unpredictable patient flow scenarios (Bikker et al., 2015). Another unique feature of simulation is that scheduling approaches can be easily tested in various circumstances (Joustra et al., 2012). Law (2013) classifies simulation models based on three key characteristics. First, models can be either static or dynamic. A static model represents a system at a single point in time, whereas a dynamic model accounts for changes over time (Law, 2013). Second, models are categorised as deterministic or stochastic. Deterministic models have predictable, non-random components, while stochastic models include random variables that introduce variability (Law, 2013). Finally, simulation models can be discrete or continuous. In discrete models, changes occur at specific points or events, whereas continuous models progress and change continuously over time (Law, 2013).

The literature indicates that simulation is the most widely used approach for addressing challenges related to patient flow and capacity allocation (Vieira, 2020). For instance, Joustra et al. (2012) employed discrete-event simulation to model the radiotherapy process and evaluated strategies to minimise patient waiting times. Similarly (Ma et al., 2016), utilised simulation to improve patients' access times by determining the optimal number of appointment slots and testing various scheduling policies. Another benefit of simulation models is that they are more visualised and are, therefore, more accessible for healthcare management to understand. The downside of using a simulation model is that it can be very time-consuming to make and run. Overall, discrete-event simulation is an effective tool for incorporating stochastic elements and testing complex scenarios to improve appointment schedules (Van Riet and Demeulemeester, 2015).

Combining Analytical models and Discrete-event simulation

As mentioned earlier, some processes are too complex to be accurately captured by analytical queueing models alone. Conversely, identifying optimal appointment schedules using simulation models can also be challenging. Simulation is more commonly used to test different configurations and visualise system performance, rather than determining optimal solutions. To effectively represent the complexity of healthcare processes, which often involve numerous interdependent variables, a combination of analytical queueing models and discrete-event simulation can be advantageous. This hybrid approach leverages the strengths of both techniques. Analytical queueing models are well-suited for addressing scheduling problems, but they are limited in handling dynamic optimisation tasks (Bikker et al., 2015). In contrast, discrete-event simulation is capable of modelling dynamic and complex systems in detail, but it is not ideally suited for optimisation (Bikker et al., 2015).

A practical application of this hybrid approach involves using an analytical metaheuristic to search for good (often near-optimal) solutions within a complex search space, while the simulation serves as the environment for evaluating those solutions. Simulated annealing is a commonly used analytical optimisation technique in this context, as it helps avoid convergence to local optima (Zheng et al., 2025). It does so by swapping or moving appointments within a schedule based on various parameters (Hans and Vanberkel, 2011).

3.5 Chapter conclusion

There have been valuable proposals in the literature for classifying the level of detail in appointment schedules. We created a structured overview of appointment scheduling methods that incorporate walk-in patients and access time. The proposed method, slots filled with patient types, will be used to develop a tailored appointment schedule. This method uses fixed slots with predetermined start and end times allocated to specific patient groups. This approach is especially helpful for decentralised administrations, which often lack the necessary knowledge to optimally schedule appointments. It will therefore positively impact patient inflow variability, access times, and waiting times.

In addition, we provided an overview of methods for prioritising different patient types, each of which affects access and waiting times in different ways. In our case, there are no clear priority distinctions within patient types, nor are patient characteristics known in advance. Moreover, it is impractical to focus solely on either scheduled patients or walk-in patients. Therefore, cyclic appointment schedules, a technique that balances the access times of scheduled patients with the waiting times of walk-in patients, is the most suitable, as it aligns best with the focus of this research.

Many tactical decisions are required when creating a new appointment schedule. To the best of our knowledge, the literature has not yet addressed whether it is better to use a flexible or hybrid policy when considering the high demand of unplanned patients, along with access and waiting time constraints. We will experiment with both settings and evaluate the outcomes, forming the theoretical contribution of this research.

To improve the current CT appointment scheduling system, we explored several methods for designing appointment schedules under uncertainty and with walk-in demand. We chose discrete-event simulation to evaluate the system, as it can incorporate variable arrival patterns, multiple patient types, and complex queuing behaviour with fewer simplifying assumptions than analytical methods. Analytical queueing models lack the scalability and flexibility needed for settings involving diverse KPIs and operational constraints. To optimise the appointment schedule, we will use simulated annealing. This metaheuristic was preferred over exact optimisation techniques due to the vast solution space and the inability to analytically express the system's performance. Simulated annealing enables effective exploration while reducing the risk of getting stuck in local optima. It also integrates well with simulation models. This combined approach allows us to evaluate and improve appointment schedules realistically and transparently, making the results more accessible and trustworthy for stakeholders.

4 Solution approach

Based on insights from Chapter 3, it remains unclear whether flexible or hybrid rooms are more effective with many walk-in patients in reducing access time. This chapter provides an overview of the solution approach. It presents a simulation model to evaluate appointment schedules and priority rules, focusing on how patient prioritisation and room pooling affect key performance indicators (KPIs). Additionally, the level of detail in the conceptual model is explained through its assumptions and simplifications. The inputs, outputs, model setup, simulation settings, and validation and verification processes are described. The chapter ends with an explanation of the experiments.

4.1 Overview solution approach

This section describes the solution approach, based on the process outlined by Robinson (2014). As discussed in Section 3.4, various methods exist to model appointment systems, including analytical queueing models and simulation. We selected a combination of discrete-event simulation and a metaheuristic optimisation method, simulated annealing, to improve the appointment schedule. Figure 4.1 provides a high-level overview of the modelling process. The following sections describe each step in more detail.

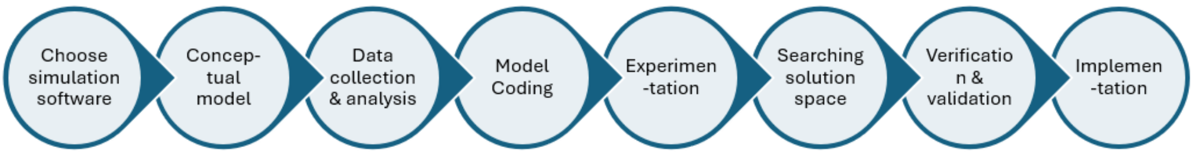

Figure 4.1: The simulation steps

Simulation allows us to evaluate different configurations without disrupting actual operations. It provides a controlled, repeatable environment to test scenarios and assess performance. In contrast, analytical queueing models, rely on simplifying assumptions, such as exponential intertarrival times and service times, and are difficult to extend to real-world complexities like walk-ins, sequencing rules, and multiple patient types (Kolisch and Sickinger, 2007; Bhattacharjee and Ray, 2016; Hassin and Mendel, 2008). These models are better suited to studying isolated aspects of appointment systems rather than integrated scheduling policies.

Discrete-event simulation overcomes these limitations by capturing system variability and operational complexity. It models stochastic arrivals, dynamic resource constraints, and priority-based sequencing with minimal simplification. This makes it especially suitable for healthcare diagnostic settings, where patient flow is highly variable. Moreover, the visual nature of simulation improves accessibility for stakeholders, increasing trust in model outcomes.

Manually testing all possible schedules with the simulation would be computationally infeasible. Therefore, we use simulated annealing, designed for large, complex search spaces, to guide the optimisation. It iteratively modifies the current schedule and evaluates performance using the simulation model. This means that discrete-event simulation evaluates candidate schedules, while simulated annealing navigates the search space. The simulation model can also be used independently to analyse fixed schedules, such as the current planning or stakeholder-proposed changes. Figure 4.2 illustrates this interaction.

4.2 Model objective

This research aims to develop a new appointment schedule for CT scans at UMC Utrecht that better aligns with patient demand while ensuring that target access times are met and no KPIs are negatively affected. To achieve this, the simulation model has the following objectives:

- Provide a reliable representation of patient flow and appointment scheduling procedures.
- Offer insights into the performance of the current schedule, stakeholder-proposed schedules, and schedules generated through simulation-based optimisation, based on KPIs, access time, utilisation, waiting time, and overtime.
- Ensure user-friendlyness by maintaining only the necessary level of detail.
- Present simulation outcomes understandable through visualisations and summary tables.
- Ensure that the recommended appointment schedule is flexible and easy to understand and implement.

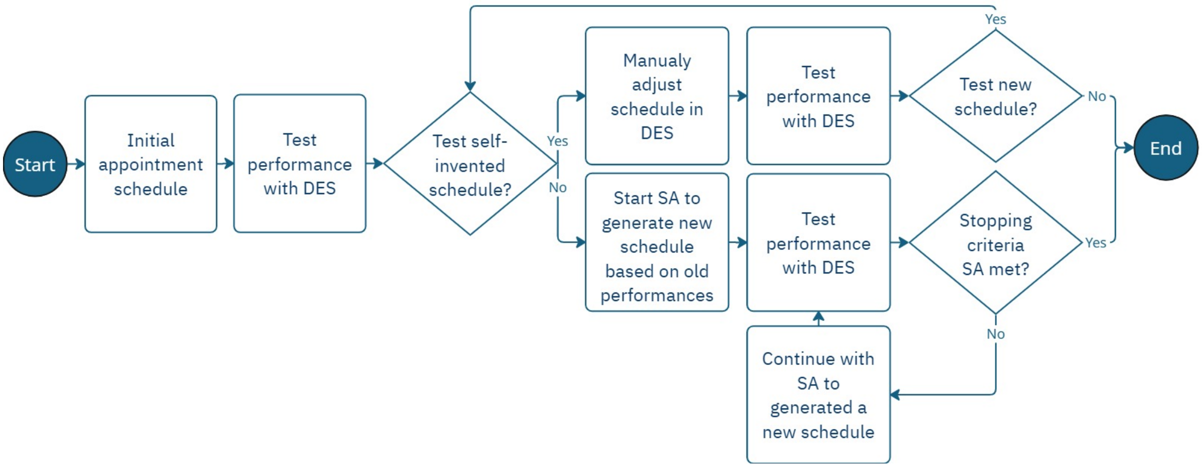

Figure 4.2: Relation between discrete-event simulation and simulated annealing in the model

4.3 Input

A well-designed conceptual and computer model is essential, but even the best model will produce inaccurate results if the input data is poor (Robinson, 2014). This section outlines what data is needed as input for the model.

Capacity

Capacity is a key factor in determining the feasibility and efficiency of appointment schedules, as it directly influences how many patients can be served and under what operational constraints. Capacity consists of several components.

1. **Staff:** The model considers only laboratory technicians, excluding radiologists, planners, and administrative staff, as they do not directly impact the number of scans that can be performed or their speed.
2. **Scanners:** Number of scanners available and how many hours they are available for patient scans.
3. **Appointment scheduling:** The appointment schedule specifies the capacity for different patient types and scan procedures.

Patient attributes

Patient attributes are essential for accurately modelling system behaviour, as they determine how and when patients enter the system, how long they require resources, and how their needs vary.

1. **Patient classification:** Patients are typically classified as inpatients, outpatients, or emergency. Outpatients may have scheduled or walk-in appointments, which can be booked well in advance or last minute. To ensure the system can accommodate non-elective arrivals, we simulated how many appointments can be planned ahead. Since not all scan types qualify for walk-ins, it's important to estimate how many require scheduling and reserve enough time slots to handle them promptly.

2. **Arrival rates:** The outpatient clinic schedule could predict non-elective patient arrivals by using referral rates per appointment type. For example, the number of times a cardiologist refers a patient for a CT scan during an initial consultation. In some cases, even the scan type may be predicted. This requires consistent workflows and accurate data registration.

   Arrival rates can also be derived from historical data, but it's essential to first test for statistical differences across weeks, days, and hours to determine if separate rates are needed. Input rates can be based on traces, empirical, or statistical distributions. Traces use raw historical data, while empirical distributions estimate the distribution from data and generate random values accordingly (Robinson, 2014). Sufficient data points are needed to capture variability and support generalisation (Robinson, 2014). Statistical distributions rely on theoretical models, offering less transparency but greater flexibility for parameter tuning and sensitivity analysis, enhancing model robustness.

   The steps to derive a statistical distribution that best fits the data using empirical data are: creating a histogram to visualise the data's shape, estimating parameters, testing the goodness of fit through, do a chi-square test, and making a Q-Q plot (Robinson, 2014). With the chi-squared test, you calculate the expected frequency error with the distribution and the actual frequency. When the total relative error is smaller than the critical value, it can be said that with 95% confidence, the data fit the distribution. When no clear distribution emerges it is important to use scientific knowledge to match the data with the right distribution.

3. **Service times:** The service time distributions can be developed in the same way as the arrival rates.

Sequencing and priority rules

It is important to simulate the priority rules from the real systems as precisely as possible. In Section 3.2 we described that when focusing on access time, waiting time, and overtime, using cyclic appointment schedules is the most effective approach, particularly when fixed appointment lengths are considered. Therefore, it is good to use the method of Kortbeek et al. (2014) and implement a discrete-time cyclic queuing model in a discrete-event simulation to assess the access and waiting times for different patient types. This will help balance the access time for scheduled patients and the waiting time for unscheduled patients.

4.4 Output

Outputs are crucial for validating the simulation, assessing whether the objectives have been met, and identifying the reasons for any failure. In collaboration with stakeholders, we identified relevant KPIs.

Calculating access times

Forecasting demand based on historical production is unreliable, as production is constrained by available capacity. Furthermore, access time is not always reflected accurately in waiting lists. Zonderland et al. (2021) notes that waiting lists often contain patients who no longer require care, and that 10% of waiting time may be voluntary. Zonderland et al. (2021) states that the number of appointment slots directly impacts access times. Hulshof et al. (2012) even suggests an exponential relationship between appointment volume and access time. With this relationship, we formulated the formula below to estimate the number of slots required and the corresponding access time.

$$\text{New access time} = \text{Old access time} \times \left(1 - \frac{\text{New appointments} - \text{Old appointments}}{\text{Old appointments}}\right)$$

Waiting time per patient type and service level

According to Kortbeek et al. (2014), waiting time directly influences access time, making it essential to balance both. Minimising waiting time improves not only the patient experience but also workflow efficiency for example, due to happier patients. Since perceived waiting time differs across patient types, we will measure it separately for each group. To evaluate performance, we will track both the average waiting time and the percentage of patients seen within the target, referred to as the service level. This is important because a low average may still hide long waits for many patients. The service level provides a more complete assessment.

Utilization

Many researchers have confirmed the effect of utilisation on access time (Hulshof et al., 2012, Zonderland et al., 2021), and generally agree on the relationship between open slots, utilisation, and waiting time (Srinivas and Ravindran, 2020). However, these studies provide limited guidance on effectively balancing these factors. To address this gap, we will experiment with different configurations in our simulation model and assess their impact on utilisation. One potential strategy to improve utilisation is to temporarily adjust capacity (Vermeulen et al., 2009). The simulation results will also be used to help validate the model by comparing utilisation outputs to realistic expectations.

Overtime

To ensure that improvements in access time do not come at the expense of staff well-being and extra costs, we include overtime as a key performance indicator. Measuring overtime allows us to evaluate whether demand can be met within regular working hours. It also helps identify when scheduling changes lead to excessive workload.

Patients rejected

To evaluate the model's ability to meet demand, we include patient rejection as a key performance indicator. This KPI reflects whether the system can accommodate all patient types within the available capacity.

Outputs to verify the model

Some outputs are not essential for measuring the objective but are crucial for verifying the model. They can help checking that the simulation behaves as intended and aligns with the conceptual model.

1. **Average number of arrivals per patient type per hour of the day:** This can be used to verify whether the simulation correctly produces the expected number of arrivals per hour according to the defined inputs and does not delete or add extra patients.
2. **Variance of the number of arrivals per patient type per hour of the day:** Ensuring that the model replicates the intended variance is important, as it can influence downstream behaviours such as queue lengths and staff workload.

3. **Average processing time:** This output enables a check that the simulation is correctly implementing the defined processing logic, without unintended changes to patient flow or timing.

4.5 Level of detail

This section outlines the level of detail in the simulation model. Simplifications and assumptions are needed to ensure feasibility and keep the simulation manageable and effective while maintaining a reasonable level of realism. They are made based on the knowledge of UMC Utrecht however, we expect that they are similar for other hospitals.

4.5.1 Assumptions

We made several assumptions due to data availability and uncertainties surrounding the actual processes.

1. **All available time slots are assumed to be filled:** There is a long waiting list, along with last-minute scan requests, which help fill any gaps in the schedule.
2. **Patients are not prioritised based on expenses:** While hospitals receive different payments for various patient types, there is no prioritisation based on these expenses.
3. **No no-shows:** The number of no-shows has drastically decreased since the implementation of walk-in hours. The exact number of no-shows is unknown.
4. **No delays due to equipment shortages:** All necessary medical supplies are assumed to be always available.
5. **No changeover time:** The time between scans is considered negligible for this model.
6. **No failed scans:** Technicians report that almost no scans require a repeat procedure.
7. **Uniform skill levels among technicians:** There has been no observed correlation between technician skill levels and efficiency across procedures.
8. **Consistent staff availability:** Scheduled technicians or their substitutes have consistently been present during their shifts, according to the data.
9. **Scanners are only accessible for patient scans:** In the past, research was conducted outside regular working hours.
10. **The arrival rate of walk-ins does not vary per week:** Although the number of scan types available for walk-ins is growing, the lack of sufficient data makes it challenging to compare walk-in demand week by week.
11. **Only CT1 works overtime:** Overtime use has historically been limited to just one scanner working after regular hours.
12. **Patients are sent home at 23:00:** Any patients still in the system at 23:00 are assumed to be sent home. This has never happened in the past.

4.5.2 Simplifications

We made simplifications, to facilitate the rapid development and maintain transparency in the simulation model.

1. **Uninterrupted functioning of CT scanners:** CT scanners are assumed to operate without breakdowns or planned maintenance. In reality, they are out of service 2.1% of the time, but this is ignored to simplify the simulation.

2. **Infinite waiting room capacity:** There are no constraints on the physical capacity of the waiting room.
3. **Fixed one-week block schedules:** Although scheduling adjustments can occur and the schedule may vary each week, the simulation uses a fixed weekly schedule. This approach is ideal for employees, as they often desire structure.
4. **Two types of consultation duration distributions:** The data indicates a significant difference in appointment duration between regular consultations and puncture procedures. Due to the limited data available, no further distinctions are made between other types of consultations.
5. **Consistent staff availability:** Staff holidays are not considered in the simulation to simplify the model and reduce run time.
6. **Arrivals are simulated between 8:00 and 17:00:** The simulation is designed to test the block schedule within regular working hours, from 8:00 to 17:00.
7. **Patients arrive on time:** Scheduled patients are assumed to arrive on time, as there is no data on late arrivals.

4.6 Model setup

As described, the simulation model will be developed using discrete event simulation to replicate the scheduling process. The design must be as user-friendly as possible, so that users without specialised modelling knowledge can operate and interpret it. During the simulation, entities (patients) progress through the process, and data is collected in real time. Each arriving patient is assigned a unique ID, arrival time, patient type, processing time, and scan type. When a patient arrives or when a scanner becomes available, the model determines which patient will be served next based on the priority rules. The model also checks if the scanner is compatible with the patient's scan type. Once a patient completes the scan, all relevant attributes are saved to assess the performance and validity of the appointment schedule.

4.7 Simulation settings

Before experimentation, it is crucial to ensure that the results of the model's performance are accurate (Robinson, 2014). To achieve this, it is essential to determine the type of simulation being used, assess whether a warm-up period is necessary, and ensure that the simulation has an appropriate run length and a sufficient number of replications.

Type of simulation

Understanding the nature of a simulation model and its output is essential to obtain accurate results. A simulation can be classified as a terminating or a non-terminating model (Robinson, 2014). A terminating simulation has a natural endpoint, whereas a non-terminating simulation does not. For accurate output, it is also crucial to determine whether we are dealing with transient or steady-state production. A transient output refers to a distribution that changes over time.

Run length and replications

Often many simulation inputs are stochastic, which makes the outputs naturally variable as well (Robinson, 2014). So, a single run does not reliably reflect the behaviour of the system. Instead, the model must run for a sufficient duration or be repeated across multiple replications. Each replication uses a fixed stream of random numbers, ensuring reproducibility and facilitating comparison across experiments (Robinson, 2014).

To determine the required number of replications, we apply the confidence interval method described by Robinson (2014), using a 95% confidence level ($\alpha = 0.05$). We monitor the relative half-width of the interval and aim for a maximum deviation of 5% from the mean. We need as many replications until this criterion is consistently met. The confidence interval is calculated using the formula below. where: $\bar{x}$ = Sample mean, $t_{\alpha/2}$ = Critical value from the t-distribution, s = Sample standard deviation, and n = Number of replications.

$$CI = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

4.8 Verification & validation

Verification and validation are critical components of the simulation modelling process. Verification ensures that the conceptual model has been accurately translated into a functioning computer model (Robinson, 2014). Validation, on the other hand, evaluates whether the model's outputs are sufficiently accurate for the user's intended purpose Robinson, 2014.

4.9 Experiment design

This section describes the experiments needed to find the appointment schedule that meets the objective and the experimental setup.

4.9.1 Experiments description

As shown in Figure 4.2, we evaluate the performance of appointment schedules using discrete event simulation (DES) and generate new schedules using simulated annealing (SA). The goal is to find appointment schedules that improve access time and patient flow without increasing overtime or reducing service levels. To achieve this, we test various experimental configurations. One key question is whether it is more effective to apply flexible priority rules to all scanners or to use hybrid scanners. Additionally, we explore whether fixed appointments should be scheduled on all scanners or only on a subset. To improve access time, we also experiment with increasing the number of appointment slots. Based on insights from previous chapters, employees can have ideas for new schedules. Furthermore, different combinations of these factors can yield varying outcomes. Based on this information, and in collaboration with scheduling staff, we have designed eight distinct experiments and outlined their potential effects on the different stakeholders in Table 4.1.

1. Experiment 1 evaluates the performance of the current appointment schedule and is used for validation, verification and comparison with the other experiments.
2. Experiment 2 uses the current schedule but applies flexible priority rules to all scanners. As the literature states, it is unclear what the effect is on waiting time, overtime and work pressure. On the other KPIs, it will have no effect.
3. Experiment 3 builds on Experiment 2 by using SA to generate alternative appointment schedules with fixed appointments on all scanners, which are then tested using DES. Having appointments are more optimal times will improve access time, work pressure and overtime.
4. Experiment 4 also starts from the settings in Experiment 2, but SA is used to design schedules with an increased number of fixed appointments. More appointments will positively affect the access time and production, but negatively waiting time and overtime.
5. Experiment 5 builds on Experiment 3, where the input schedule already includes fixed appointments on all scanners. With SA, schedules can be designed with an increased number of fixed appointments. It is unclear what the effect is of all flexible scanners.

	Patient 1) Waiting time, 2) Access time	Employee 1) Overtime, 2) Work pressure	Management 1) Production, 2) Utilisation
1) Baseline schedule	-	-	-
2) Baseline schedule, Flexible scanners	1) ?, 2)	1) ?, 2) ?	1) , 2)
3) Baseline schedule, Flexible scanners & Appointments on all scanners (SA)	1) +, 2)	1) +, 2) +	1) , 2)
4) Baseline schedule, Flexible scanners, More appointments with same capacity (SA)	1) -, 2) +	1) -, 2) -	1) +, 2) +
5) Input exp 3, Flexible scanners, More appointments with same capacity (SA)	1) ?, 2) +	1) ?, 2) ?	1) +, 2) +
6) Schedule staff	1) ?, 2) +	1) ?, 2) ?	1) +, 2) +
7) Schedule staff, Flexible scanners	1) ?, 2) +	1) ?, 2) ?	1) +, 2) +
8) Schedule staff, Flexible scanners, Appointments on all scanners (SA)	1) +, 2) +	1) +, 2) +	1) +, 2) +

Table 4.1: Expected effect of experiments

6. Experiments 6, 7, and 8 use appointment schedules proposed by employees as the basis for further experimentation. We will test whether the staff or SA can generate better-performing appointment schedules.

4.9.2 Simulated Annealing

Simulated annealing works by generating new solutions through small changes to a current solution. Worse solutions can be accepted to avoid local optima. Over time, the acceptance probability of worse solutions decreases with the temperature. Zheng et al. (2025) and van Essen et al. (2014) have demonstrated that simulated annealing is effective for schedule optimisation in healthcare environments. As shown in Figure 4.3, the algorithm follows a structured process of generating neighbourhood solutions, evaluating them, and adjusting the temperature to control acceptance.

4.9.3 Performance Function

The DES evaluates the performance function for each proposed appointment schedule. The performance function consists of the average waiting time with penalty terms for service level violations and overtime. This allows the model to balance performance across multiple KPIs. Penalties are added when a patient category's service level falls below its target or when total overtime exceeds the limit.

We use penalties to ensure that unacceptable schedules are discouraged. This enables the simulated annealing algorithm to consider trade-offs, such as a slight increase in waiting time in exchange for a better service level or reduced overtime. Penalties must be high enough to discourage solutions that severely violate service level targets or exceed acceptable overtime, but not so high that the algorithm rejects all slightly suboptimal solutions. Therefore, it is important to calibrate the penalty weights (Thomas Schneider et al., 2020). This ensures a smooth and meaningful neighbourhood search during simulated annealing.

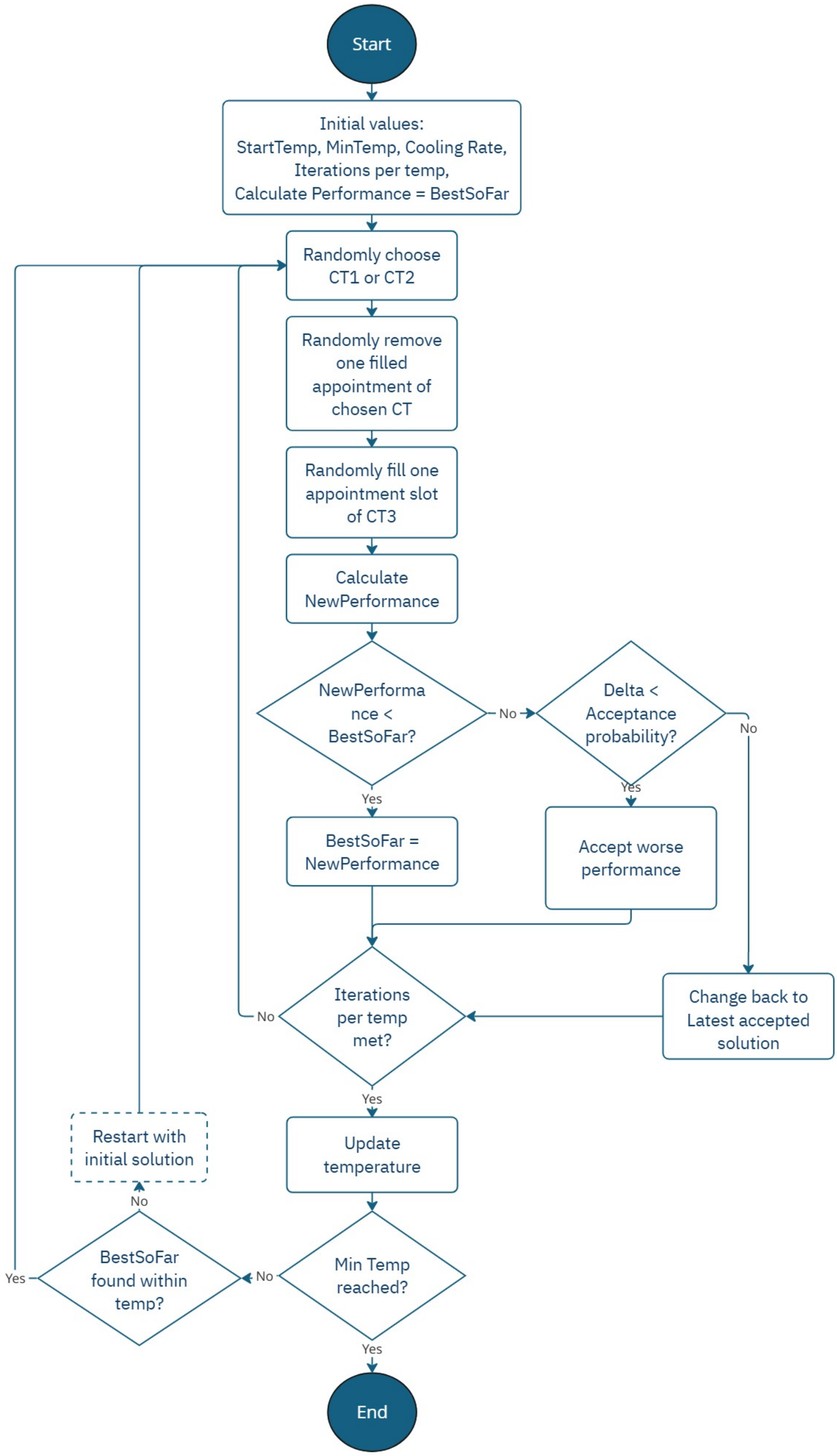

Figure 4.3: Simulated annealing approach

4.9.4 Parameter calibration of simulated annealing parameters

Parameter calibration improves the algorithm's effectiveness. This includes selecting appropriate values for the start temperature, cooling rate, number of iterations per temperature, and stopping conditions. The goal is to achieve high-quality solutions within a reasonable computation time, while avoiding early convergence to local optima. The calibration ensures that solutions close to the best are accepted early, while poor solutions are less often accepted as the temperature decreases. Zheng et al. (2025) suggests adding a restart mechanism to further reduce the risk of getting stuck in a local optimum.

4.10 Chapter conclusion

In this chapter, we presented the complete solution approach used to develop and evaluate new appointment schedules that meet target access times and other KPIs. We selected discrete-event simulation to model patient flow, capacity, and priority rules in a realistic, transparent, and flexible way. DES helps us to measure the performance of the appointment schedules. SA will be used to navigate the large and complex solution spaces without exhaustive search. These new appointment schedules will again be tested with the DES.

We described how the model's performance will be measured and the scope, which ensures a feasible and focused model. The model aims to evaluate and improve appointment schedules based on the KPIs: access time, waiting time, utilisation, and overtime. A performance function balances the KPIs using penalties, calibrated to navigate the search space efficiently. This solution approach forms the foundation for the case study discussed in Chapter 5.

5 Case study

In this chapter, we apply the conceptual model and solution approach of Chapter 4 to the CT department of UMC Utrecht. In Section 5.1 and 5.2 we describe the specific input en outputs. Section 5.3 and Section 5.4 explains the simulation model and its settings for UMC Utrecht. This is followed by a verification and validation study in Section 5.5. Section 5.6 describes the experimental setup. The chapter ends with a conclusion of this case study in Section 5.7.

5.1 Input

This section outlines the data used as input for the simulation model. We collected this data with observational studies, interviews with stakeholders and by using the available data.

5.1.1 Capacity

Staff

Typically, a CT scanner operates when two laboratory technicians are available. Regular working hours are from 08:00 to 16:30, Monday to Friday. During the lunch break (12:00–14:00), technicians alternate, and an additional staff member assists to prevent delays. From 16:30 to 17:30, one technician is present to handle appointment overruns, ensuring one scanner remains operational. On top of this, outside regular hours, one technician is always available. As previously noted, we assume technician capacity remains constant each week. Although in reality, staff availability decreases during school holidays, along with scan demand. Our focus is limited to the appointment schedule for regular weeks.

CT Scanners

As described, CT1, CT2, and CT3 operate when staff are available. In case of an employee shortage, CT2 is the first to close, followed by CT3. We assume no scanner breakdowns occur, so the staff capacity matches the capacity of the CT scanners.

Appointment scheduling

Currently, two appointment schedules are used, one for even weeks and one for odd weeks, but the differences are minimal. Therefore, we use a single schedule as the initial input for our simulation model. The current schedule reserves slots for outpatient appointments while also accommodating walk-in, emergency, and inpatient cases. An example of the initial appointment schedule for CT1 is shown in Figure A.1 in Appendix A.1.

Currently, specific appointments are reserved for certain scan types, such as heart scans. However, due to the lack of data on waiting lists for different scan types, we cannot recommend the exact number of blocks that should be allocated to each type. Furthermore, reserving dedicated time slots for specific scans could negatively impact overall access time, as it increases the likelihood of unfilled slots. As discussed in Section 3.1, the appointment schedules should include fixed slots with predetermined start and end times allocated to specific patient types, in this case, outpatient clinic appointments.

5.1.2 Patient attributes

Patient classification

The following scans cannot be scheduled as walk-ins and must be scheduled in advance: CTA-Arterial, Drainage, Puncture, Heart-Coronary, Heart-Standard/Vascular Ring, Heart-Valve, Anaesthesia, and scans from the Princess Máxima Centrum, due to their complexity.

Arrival rates

In our case, predicting non-elective patient arrivals based on outpatient clinic schedules is not feasible. Appointment registration and reporting vary across departments, making consistent analysis difficult. Additionally, clinics often deviate from their schedules, and we lack accurate, usable data to identify reliable patterns. Further, some departments only allow walk-ins immediately after consultation, while others permit walk-ins on later days.

Given these limitations, we estimate patient arrival rates using historical data. Since we focus on regular weekly appointment schedules, we assume arrival rates remain stable week to week. However, Section 2.4.5 showed variation in walk-in arrivals per hour and day. With statistical tests we confirmed that 8 out of 10 time slots had p-values below 0.05, indicating significant differences in arrival rates by hour and day (see Table A.2 in Appendix A.1), and must be calculated per hour and day. Our arrival time data spans March to December 2024, with only about 40 data points per hour and day. Relying solely on traces or empirical distributions would limit the model's ability to generalise. To balance flexibility and accuracy, we use a combination of empirical and statistical distributions where possible.

Due to limited data, no clear distribution emerged from histograms, goodness-of-fit tests, or Q-Q plots. Since patient arrivals are discrete events per hour, we considered standard discrete distributions such as the binomial and Poisson distributions. According to Robinson (2014), random arrivals per time interval often follow a Poisson distribution. We therefore did a correlation test between the number of arrivals on one day and hour and the following hour to test dependencies. The correlation coefficients can be seen in Table A.1 of Appendix A.1). All correlations are close to zero, which indicates minimal correlation. We also tested correlation with a hypothesis test. When the absolute value of the t-statistic is equal to or less than the critical value (2.023), we fail to reject the null hypothesis, meaning the data do not provide evidence of a significant correlation. With all our t-statistics smaller than the critical value, we can confirm that arrivals occur independently, supporting our choice to model patient arrivals using the Poisson distribution.

The Poisson distribution is defined by a single parameter, the mean (Robinson, 2014). Since our focus is on regular weeks, we used historical data from these periods to calculate the mean arrival rate for each hour of the week across the three patient types. The arrival rates were calculated using the following formula, where X_1 is the number of arrivals observed over T_1 periods (Robinson, 2014):

$$\lambda_1 = \frac{X_1}{T_1}$$

As previously noted, we assume all available time slots are filled due to high demand, so the arrival rate of outpatient clinic scheduled appointments corresponds to the number of available time slots.

Service times

As described in Section 2.4.1, limited data on appointment durations makes it difficult to distinguish between service times of CT scanners, patient types, or scan types. According to stakeholders, scan type is the main factor influencing appointment duration. For example, CT puncture scans typically take significantly longer, while heart scans average about five minutes more than other types. However, over 60% of scans are unplanned and their type is unknown in advance, limiting the impact of this knowledge on the simulation. Currently, CT puncture scans are allocated 60 minutes, and all other planned scans 20 minutes.

Although we initially expected inpatients to take longer due to mobility constraints, they often arrive prepared and with support, resulting in durations similar to outpatients. Walk-in appointments tend to be shorter, likely because puncture scans are not performed on a walk-in basis. Given the limited data and stakeholder input, we categorise appointment durations into two groups: CT puncture scans and all other scans. For the non-puncture group, we assume average durations will balance out.

As with arrival rates, we aimed to fit a statistical distribution to the observed appointment durations. For the 20-minute group, the gamma distribution provided the best fit based on the histograms (Figure 5.1). However, the chi-squared test statistic (515) exceeded the critical value (26), likely due to the data being rounded to whole minutes, while actual appointment durations are recorded in seconds. The gamma distribution remains the best fit, with parameters:

$$\alpha = \frac{\text{mean}^2}{\text{variance}} = \frac{14.10^2}{47.44} = 4.19, \quad \beta = \frac{\text{variance}}{\text{mean}} = \frac{47.44}{14.10} = 3.36$$

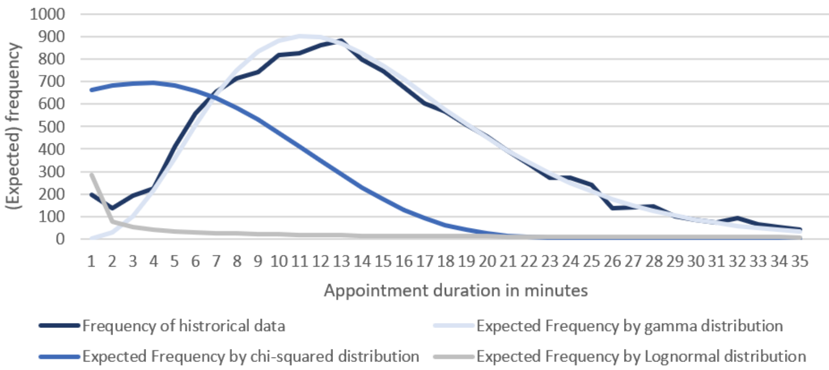

Figure 5.1: Histogram: 20-minute appointment duration

For CT puncture scans, both gamma and normal distributions seemed fitting (Appendix A.2, Figure A.3). The chi-squared test results showed a total relative error of 19.8 for the gamma distribution and 20.4 for the normal, with a critical value of 19.7. We therefore selected the gamma distribution, with parameters: $\alpha = 8.57$ and $\beta = 6.87$. Appendix A.2 provides a full overview of all chi-squared tests performed.

5.1.3 Sequencing and priority rule

Figure 5.2 illustrates the priority rules in the waiting room for both the current system and the simulation. In the current system, CT 3 prioritises walk-in patients, while CT 1 and CT 2 offer more flexibility in handling different patient types. In the simulation, the priority rules have been simplified because we have no data on how many clinical patients arrive with a nurse, nor the number of early arrivals. On top of this, laboratory technicians cannot see in the system if a patient arrives with a nurse. They only know this when they see it in the waiting room, it is then a human choice to let this patient go first or not.

5.2 Output

The department previously lacked established KPIs. We identified relevant KPIs, set targets and observed the current performance. These KPIs are outlined below.

Number of outpatient clinic appointments per week for calculating access times

In the current system at UMC Utrecht, only outpatients wait for appointment slots. Stakeholders confirmed that the waiting list often includes outdated referrals and patients postponing scans voluntarily, making it an unreliable indicator of access time. We therefore used the exponential relationship between the number of slots and the access time as described in Section

4.4. Currently, the average access time is 31 days with 21 outpatient slots per day. Applying the exponential access time model, using the formula below, would require an average of 26.8 outpatient appointment slots per day to achieve the target access time of 22.4.

$$\text{Access time} = 31 \times \left(1 - \frac{\text{Number of outpatient appointments} - 21}{21}\right)$$

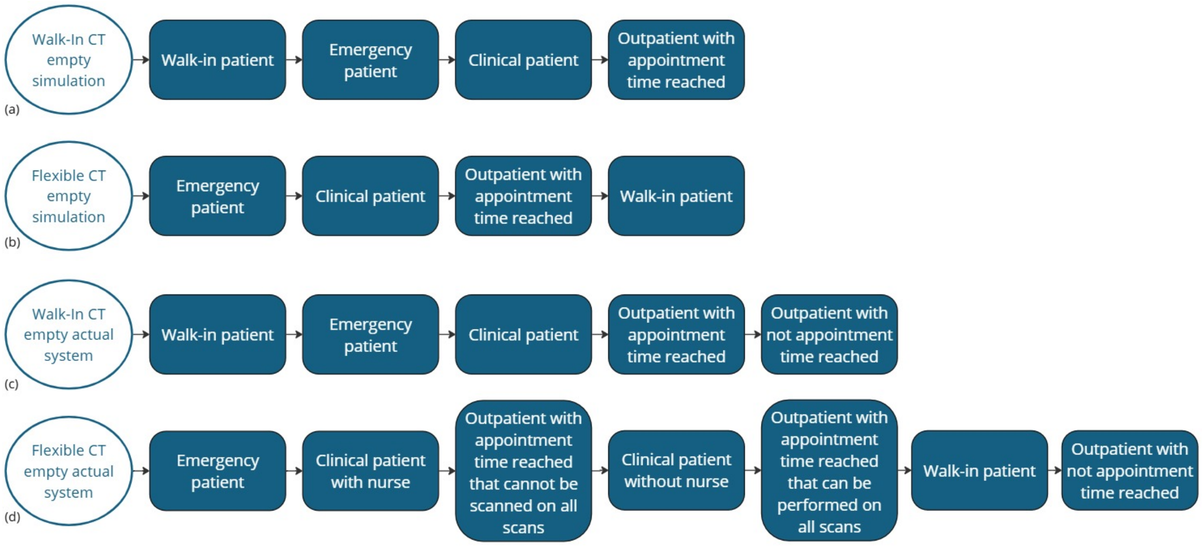

Figure 5.2: Priority rules actual system and simulation

Waiting time per patient type and service level

Through interviews, we discovered that both administration employees and laboratory technicians benefit from shorter waiting queues. Administration employees handle fewer complaints, while laboratory technicians work with less stressed patients, which can positively impact processing times. In collaboration with the staff, we defined target waiting times for each patient category: 5 minutes for emergency patients, 10 minutes for inpatients, 15 minutes for outpatients, and 30 minutes for walk-in patients. The target service level is to meet 95% of these waiting times.

Utilization

The hospital aims to maximise CT scanner utilisation while keeping waiting times within target limits. High utilisation is important, as CT laboratory technicians are in high demand and can also support MRI or X-ray departments when not needed for CT. Therefore, maintaining high utilisation not only improves efficiency but also optimises staff efficiency across departments.

Overtime

The hospital prioritises employee satisfaction and aims to minimise overtime. Therefore, any improvements in access time must not result in additional overtime for staff.

Patients rejected

In the current system at UMC Utrecht, no patient are rejected. The department places great importance on ensuring that every patient receives care. Therefore, the target for this KPI is to maintain zero patient rejections.

5.3 The simulation model

Using all the information gathered for UMC Utrecht, we programmed the simulation model in Plant Simulation. Figure 5.3 displays the control panel of the simulation model. In the input quadrant, users can adjust all variables relevant to the case study, such as target waiting times, closing times, and arrival rates.

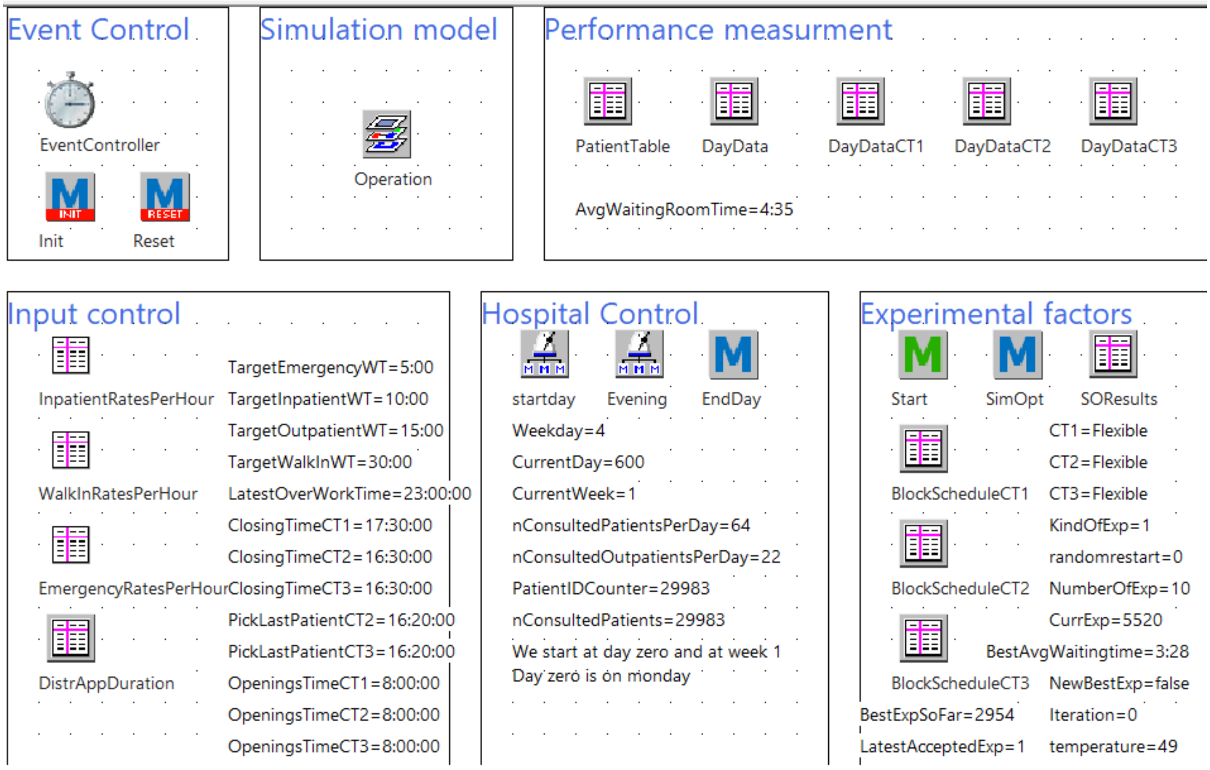

Figure 5.3: Control panel discrete event simulation CT radiology department

Figure 5.4 illustrates the operation of the model. The simulation follows the defined logic using real-world inputs, allowing users to test and see the performance of appointment schedules.

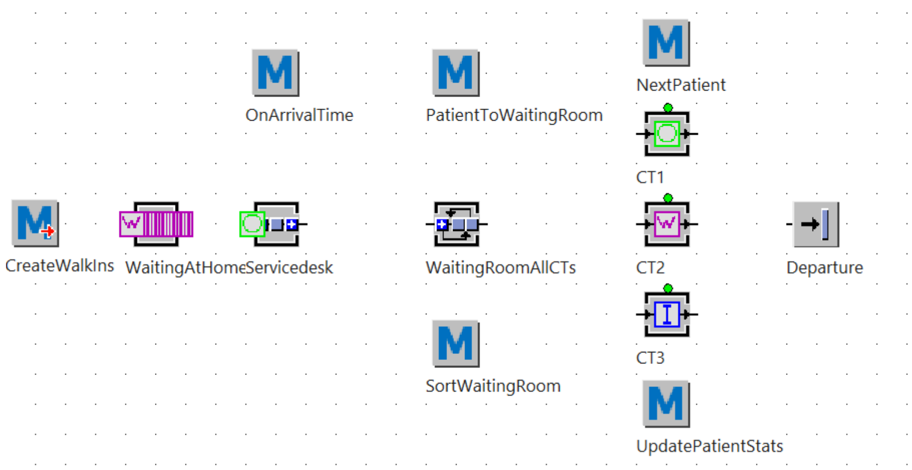

Figure 5.4: Operation discrete event simulation CT radiology department

5.4 Simulation settings

This simulation is classified as a terminating model, as it has a defined endpoint. The system is empyt at the end of each working day when staff and still waiting patients are sent home and starts empty the next morning. While each day terminates individually, the full run length spans one week to reflect varying appointment schedules and arrival patterns across weekdays. The system starts and ends empty each day, and we excluded the waiting lists due to the earlier described data quality and model simplifications, therefore no warm-up period is required (Robinson, 2014). This setup ensures consistent starting conditions and captures typical weekly operations.

To determine the required number of replications, we calculated confidence intervals for five key performance indicators (KPIs). To optimise performance, we selected inputs with the highest variance or lambda values, since they demand more replications. Figure 5.5 shows that with 42 replications, all relative errors fall below the 0.05 threshold. To ensure accuracy and stability, we simulate 48 one-week replications, totalling a run length of 336 days.

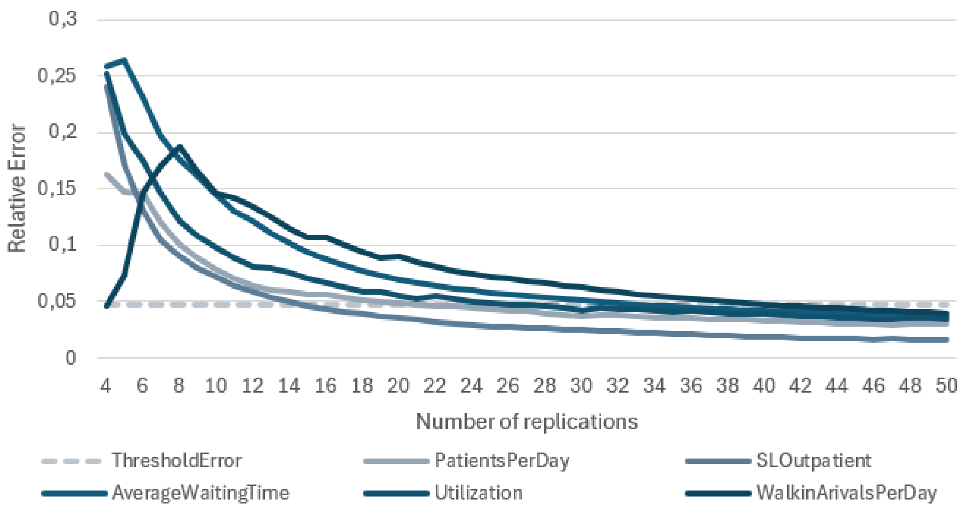

Figure 5.5: Number of replications of one week needed

5.5 Verification & validation

This section outlines the verification and validation process of our simulation model.

5.5.1 Verification

Code inspection and debugging: Throughout the simulation process, we consistently checked whether the methods and logic aligned with the conceptual model. To support this, we explained the code to external experts for review and feedback. For each technique, we created pseudo-code before implementation to ensure clarity and consistency with the conceptual design. We then carefully stepped through the simulation, method by method, to confirm that each functioned as intended. Particular attention was paid to unusual values, such as zeros or empty fields, to detect potential issues.

Patient tracing and Logical Checks: We observed the simulation's behaviour using visual outputs to check the correctness of the model's logic. The model was paused at specific points, and predictions were made about the expected behaviour of specific patients. These were then compared with the actual simulation outcomes and the intended meaning of the conceptual model. Special attention was given to the sequencing rules, as this method was more complex.

Extreme Condition Testing: We tested the model under extreme scenarios, for example, with excessive patient arrivals or appointment durations, to verify whether it behaved as intended.

Output Inspection: We tested the inputs defined in the conceptual model against the outputs of the simulation model to ensure alignment. Table A.2 and Table A.4 in Appendix A.3 show the calculations of this analysis. Table 5.1 shows that the means of the arrival rates, appointment durations, and the percentage of appointment type 1 do not differ significantly between the simulation and the input data. However, half of the simulation distributions have a significantly lower variance than the input data, suggesting that the simulation may not fully capture reality in this aspect. However, with approximately 350 arrivals per week and 70 per day, we expect that the variances in arrival rates, appointment durations, and appointment type percentages will balance out, making the overall variance acceptable. It is therefore essential that the simulation includes a sufficient number of replications. Lastly, we verified that all intended outputs and intermediate calculations were correctly implemented by manually calculating and comparing them with the simulation results.

	Input mean	Input sd	Simulation mean	Simulation sd	t difference between the means	Significant difference between variances
Walk-in arrivals Monday	35.3	7.24	35.6	6.5	No	No
Walk-in arrivals Tuesday	30.52	6.62	30.33	5.36	No	No
Walk-in arrivals Wednesday	36.43	8.21	36.06	5.76	No	yes
Walk-in arrivals Thursday	28.53	5.92	28.7	5.23	No	No
Walk-in arrivals Friday	28.69	7.2	28.4	5.67	No	No
Emergency arrivals Monday	6.49	2.76	6.06	2.18	No	No
Emergency arrivals Tuesday	6.18	3.12	5.87	2.13	No	Yes
Emergency arrivals Wednesday	6.12	2.91	6.00	2.28	No	No
Emergency arrivals Thursday	5.42	2.75	5.21	2.02	No	yes
Emergency arrivals Friday	6.82	2.96	7.13	2.65	No	No
Inpatient arrivals Monday	10.76	3.99	11.39	3.14	No	No
Inpatient arrivals Tuesday	11.48	3.76	11.18	3.12	No	No
Inpatient arrivals Wednesday	10.21	3.84	9.62	3.03	No	No
Inpatient arrivals Thursday	10.64	4.59	9.89	3.02	No	Yes
Inpatient arrivals Friday	10.33	3.62	10.08	3.13	No	No
Appointment duration 20 min	13.24	7.48	13.12	9.3	No	No
emergency	0.07	0.13	0.07	0.11	No	Yes
Percentage appointment type 1 walk-in	0.06	0.07	0.06	0.04	No	Yes
Percentage appointment type 1 inpatient	0.08	0.12	0.08	0.09	No	Yes
Percentage appointment type 1 outpatient	0.22	0.21	0.22	0.21	No	No

Table 5.1: Verification of mean and variances of the simulation model

5.5.2 Validation

This section confirms that the model accurately represents the real-world system and meets the objectives of the simulation study. Robinson (2014) outlines several validation steps, which are described below.

Conceptual Model Validation: Along with stakeholders, we reviewed the assumptions and simplifications to ensure they are reasonable. We also thoroughly examined the sequencing and priority rules by conducting interviews and observations with multiple laboratory technicians to gain a deeper understanding of the patient selection process. In addition to creating an accurate model, gaining stakeholders' acceptance of the conceptual model is crucial for building their confidence in the model.

Data Validation: We gathered data through a step-by-step approach, beginning with observational studies and interviews to gain an understanding of the process. We then used test patients, for whom we knew all relevant data and processing times, to verify the accuracy of the data. This allowed us to understand what data was available and its accuracy. All gathered data was cross-checked with other available data tables. Additionally, experts reviewed the data to assess its reliability. However, as mentioned earlier, we found that not all the data was registered correctly, which affected its accuracy and validation.

White-box validation: White-box validation ensures that each part of the model accurately represents reality. We focused on timings, such as cycle times, scan durations, and waiting times, and compared them to real-world observations. The flow of patients through service points was also reviewed to ensure it followed real-world logic. The implementation of priority rules, such as selecting the next patient for scanning, was tested. Empirical data verified that the distributions of patient arrivals, scan durations, and waiting times followed realistic patterns. Additionally, domain experts reviewed the simulation's outputs, we compared them with the HIX agenda, and conducted observation studies to confirm that individual components behaved as expected.

Black-Box Validation: Black-box validation determines whether the model represents the real world with sufficient accuracy for its intended purpose (Robinson, 2014). Table 5.2 shows the validation of the simulation model's mean and variance. Table A.3 and Table A.5 in Appendix A.3 show the calculations of this analysis. Using a confidence interval, critical value, and computed F-statistic, we assessed whether there was a significant difference. The percentage of inpatients and the total number of arrivals per day are inputs of the simulation. So we can check if the system is working properly and represents reality. The utilisation and average waiting time are outputs generated by the simulation and are therefore, essential if the simulation represents reality. We concluded that the number of patients arriving per category and the utilisation closely represent the real world. However, the average waiting time and variance are higher in reality than in the simulation model. Together with our stakeholders, we discussed what the cause of this could be. We discovered that we had not accounted for the waiting time for blood results or the time for taking an oral contrast. Since the data does not indicate which patients had to wait for these reasons or how long they had to wait, we were unable to include this information in our model. However, we can still compare the waiting time of the experiments with the initial waiting time of the simulation and see if they increased or decreased.

Experimentation Validation & Solution Validation: It is essential to keep in mind that the experimental procedures are providing results that are accurate for our research (Robinson, 2014). We achieved this by frequently referring back to our research goal while conducting the experiments. All experimental factors, such as run length and repetitions, are carefully controlled. In addition, we ensured that all decisions regarding simulated annealing and searching the solution space were based on literature and tested performance.

5.6 Experiments

Section 4.9 described the experiments that need to be performed to know the best setting for flexible and hybrid CT scanners, and appointment schedules to meet the target access time without negatively affecting other KPIs. This section outlines all the experiments and their methodology.

We use the same random number streams for the distributions in the DES across all experiments and iterations. This ensures that differences in the performance of appointment schedules and priority rules are not due to randomness. However, random numbers are also used in the SA process to select the next appointment schedule to test. It is crucial that when we return to the SA after evaluating a schedule in the DES, we do not restart the same random number stream in the SA. If we did, each next iteration would be identical to the previous one. To avoid this, we base the seed values for selecting the next iteration on the iteration number itself. For example, at iteration 10, we use seed value 10 to determine the next appointment schedule.

	Historical data mean	Hystorical data sd	Simulation mean	Simulation sd	Signifivant difference between means
Percentage of Inpatients	0.06	0.06	0.15	0.04	No
Percentage of Oupatients	0.3	0.3	0.31	0.04	No
Percentage of Emergency patients	0.09	0.09	0.09	0.03	No
Percentage of Walk-in patients	0.46	0.08	0.46	0.06	No
Average utilization	0.58	0.1	0.61	0.08	No
Waiting time Inpatient	6.43	9.78	1.99	1.5	Yes
Waiting time Outpatient	5.23	10.15	2.11	1.47	No
Waiting time Emergency patient	7.13	9.87	1.43	1.3	Yes
Waiting time Walk-in patient	19.17	15.27	6.1	5.78	Yes
Total number of arrivals per day (weekday)	68.94		69.32	7.92	

Table 5.2: Validation of mean and variances of the simulation model

5.6.1 Exp 1: Baseline schedule

In this experiment, we test the performance of the initial appointment schedule and priority rules with the DES, to validate the model and for comparisons with the other experiments.

5.6.2 Exp 2: Baseline schedule, Flexible scanners

In this experiment, we adjust the priority rules of CT3 from walk-in to flexible. This means that the priority rules for picking the next patient change (Figure 5.2). We test this with the initial appointment schedule in the DES.

5.6.3 Exp 3: Baseline schedule, Flexible scanners, Appointments on all scanners (SA)

This experiment also starts with the baseline schedule and has flexible priority rules for all scanners. A flexible CT3 means that outpatient appointments can also be planned on CT3. We will test the performance if we move appointments from CT1 and CT2 to CT3.

Simulated annealing

Our appointment schedule has 825 different places where appointments can be scheduled. This means that if we test all possible variations, we need to test 2^{825} different schedules. This will have a runtime of many days. We will therefore use SA to search the neighbourhood. In SA, the method of generating new iterations is a critical component of the search process. We will randomly select an appointment from CT1 or CT2 and move it to a randomly chosen time slot on CT3, and check the performance in the DES. This strategy enables us to navigate the solution space effectively, improving performance while maintaining the feasibility of the runtime.

Often, simulated annealing employs a swap or two-sided move operator, allowing you to revisit previously explored solutions. In our case, when the simulated annealing accepts a worse solution, it can never go back since we use a one-way (from CT1 and CT2 to CT3) move operator. Zheng et al. (2025) proposed a solution to this problem by restarting the simulated annealing algorithm after a specific period. Studies have shown that the restart mechanism outperforms the original Simulated Annealing algorithm (Zheng et al., 2025). The restart mechanism helps to avoid getting stuck in a local optimum. The restart occurs if the solution fails to find a new best solution after a specified number of iterations (Zheng et al., 2025).

Performance function of appointment schedule

The SA algorithm accepts new iterations based on their performance relative to the best solution found so far. Since we use multiple KPIs to assess performance, we combine them into a single performance function for straightforward comparison.

The primary goal is to minimise the average patient waiting time while meeting target service levels for both waiting time and overtime. When a solution fails to meet service level targets, penalties are added to the average waiting time. These penalties are scaled based on the degree of deviation from the targets. Following the recommendations of Thomas Schneider et al. (2020), who noted that suboptimal penalties can lead to different and suboptimal outcomes, we tested various configurations. Table 5.3 summarises the calibration results.

Penalty	Avg Waiting Room Time	Overtime	SL Emergency	SL Walk-in	SL Inpatient	SL Outpatient	Performance	Acceptance probability
Initial objective	236	79	0.89	0.93	0.95	0.97	280	
1 min per 0,01 SL	241	78	0.89	0.93	0.95	0.97	296	0.96
3 min per 0,01 SL	241	78	0.89	0.93	0.95	0.97	405	0.73
1 min per 0,01 SL	279	125	0.87	0.92	0.96	0.96	441	0.67
3 min per 0,01 SL	279	125	0.87	0.92	0.96	0.96	765	0.30
3 min per 1 min overtime	238	364	0.90	0.95	0.97	0.97	430	0.69
5 min per 1 min overtime	238	364	0.90	0.95	0.97	0.97	558	0.50

Table 5.3: Objective parameter calibration with start temperature

Solutions near the target service level must be accepted with high probability at the initial temperature, while inadequate solutions are rejected. Overtime is critical for hospital operations, so it must be penalised appropriately. We found that a penalty of 1 minute per 0.01 deviation from the service level is effective. For example, a service level of 0.89 (slightly below the target) has a high acceptance probability of 0.96, while a level of 0.87 (somewhat below the target) drops significantly to 0.67. For overtime, a 3-minute penalty still results in a relatively high acceptance probability (0.69). Since overtime is most important for the hospital, we determined that a 5-minute penalty per minute of overtime is most effective. The final performance function, including the penalties, is defined as:

$$\text{Performance Function} = \text{AvgWaitingRoomTime} + P_1 + P_2 + P_3 + P_4 + P_5$$

Where:

$$P_1 \text{ (Penalty 1)} = \begin{cases} (0.90 - \text{SLEmergency}) \times 100 \times 1:00, & \text{if SLEmergency} < 0.90 \\ 0, & \text{otherwise} \end{cases}$$

$$P_2 \text{ (Penalty 2)} = \begin{cases} (0.90 - \text{SLWalkIn}) \times 100 \times 1:00, & \text{if SLWalkIn} < 0.90 \\ 0, & \text{otherwise} \end{cases}$$

$$P_3 \text{ (Penalty 3)} = \begin{cases} (0.90 - \text{SLInpatient}) \times 100 \times 1:00, & \text{if SLInpatient} < 0.90 \\ 0, & \text{otherwise} \end{cases}$$

$$P_4 \text{ (Penalty 4)} = \begin{cases} (0.90 - \text{SLOutpatient}) \times 100 \times 3 = 1:00, & \text{if SLOutpatient} < 0.90 \\ 0, & \text{otherwise} \end{cases}$$

$$P_5 \text{ (Penalty 5)} = \begin{cases} (\text{Overtime} - 5:00) \times 5, & \text{if Overtime} > 5:00 \\ 0, & \text{otherwise} \end{cases}$$

Parameter choices of Simulated annealing

The parameter choices are described below. An optimal combination is crucial for not getting stuck in a local optimum too soon or getting too long computation times.

- **Begin temperature:** The beginning temperature must be high enough so iterations at the beginning are accepted with a relatively high probability, to escape from a local optimum (van Essen et al., 2014). If the temperature is too high, this will result in an unnecessarily high computation time. Thomas Schneider et al. (2020) uses the formula below to calculate the initial temperature, which is similar to the method employed by van Essen et al. (2014). Since we do not know what a realistic decrease is, we will choose 280 seconds since this is the maximum decrease and therefore a safe option to start with. This results in a start temperature of 404. Which means that we want to accept this decrease with a probability of 0.5.

$$Begin\ temperature = \frac{-Maximum\ decrease\ of\ objective\ function}{\ln(0.5)} = \frac{-280}{\ln(0.5)} = 404$$

- **Acceptance probability:** Zheng et al. (2025) and van Essen et al. (2014) use the same method to accept or reject a solution, as can be seen in the formula below. We will always accept a solution if it is better than the best solution before. Otherwise, a solution will be accepted based on the difference between the current performance and the best performance, as well as the temperature. When the difference is small and the temperature is high, there is a high probability that the solution will be accepted. You will therefore easily escape from a local minimum. As the temperature gradually decreases, an increasing number of solutions will be rejected.

$$Acceptance\ probability = \begin{cases} 1, & \text{if } NewObjective - BestObjective < 0 \\ e^{-\frac{\Delta\ Objective}{Temperature}}, & \text{Otherwise} \end{cases}$$

- **Number of iterations per temperature:** van Essen et al. (2014) suggests setting the number of iterations for each temperature equal to the number of neighbour solutions that can be achieved by one swap of the initial solution. In our case, we have 105 appointments that could be moved to CT3 to 260 different time slots. With a high temperature, almost all moves will be accepted, so it makes no sense to have 260 iterations per temperature since CT1 and CT2 will be empty very quickly. We chose to do 80 iterations instead. It still happens that CT1 and CT2 get empty, and no other neighbourhood solutions are possible without a reset. This means that more iterations will be unnecessary at one temperature. The same applies to the number of iterations before a reset. Now, CT1 and CT2 are occasionally empty, but the temperature does not drop too quickly, and the reset occurs at a reasonable rate. This provides a good balance between computation time and searching for neighbourhood solutions.

- **Cooling rate:** The cooling rate makes sure the temperature goes down after the set number of iterations with the formula below. Zheng et al. (2025) says that the best cooling rate depends on the specific case. A high cooling rate results in a high computation time, while a low cooling rate can miss escapes from local minima. Many researchers suggest that a cooling rate of 0.95 is a safe starting point (van Essen et al., 2014). But the optimal value can lie around this.

$$Temperature := Temperature * CoolingRate$$

- **End temperature:** The end temperature must be chosen so that at the end of the procedure, almost no worse solutions are accepted, to convert to a local minimum (van Essen et al., 2014). We used the formula of Thomas Schneider et al. (2020) to compute the end temperature. We decided to accept a negative change in service level, with a 0.005 decrease from 0.9 or an average waiting room time exceeding 60 seconds, to a probability of 0.001. With the formula below, the end temperature will be 8.7.

$$End\ temperature = \frac{Minimum\ nagative\ change}{\ln(0.5)} = \frac{-60}{\ln(0.001)} = 8.7$$

Parameter calibration simulated annealing

Setting	Start Temp.	End Temp.	Iterations per Temp	Cooling rate	Best performance	Last best iteration	Number of iteration
1	404	8.7	60	0.95	210	471	4500
2	404	8.7	80	0.95	212	1766	6000
3	404	8.7	60	0.97	213	1267	7620
4	404	50	80	0.97	201	1253	5520
5	404	8.7	80	0.97	201	1253	10160
6	300	30	60	0.97	213	1857	4560

Table 5.4: Simulated annealing parameter calibration

Zheng et al. (2025) confirms that simulated annealing values significantly impact the algorithm's performance and suggests performing parameter calibration. Table 5.4 shows the parameter calibration that we performed. The first setting is the one indicated by the literature. A higher start temperature was not necessary to test, since almost all solutions would be accepted initially. A lower-end temperature also does not need to be tested, as no worse solutions are accepted anymore in the end. For the other settings, we want to find the best-performing iteration within a reasonable computation time. Setting 4 gives a significantly better performance value than the different settings and has a computation time of 2 hours. This is a reasonable computation time since we do not have to run it weekly or monthly. In a good setting, the number of accepted worse experiments gradually decreases. Figure 5.6 shows that the number of accepted experiments with this setting gradually decreases until almost no worse solutions are accepted. Setting four is also safe since the last best found experiment is long before the latest experiment. This is beneficial since the hospital can utilise alternative seed values and random number streams, which may require more experiments to determine the optimal performance function value.

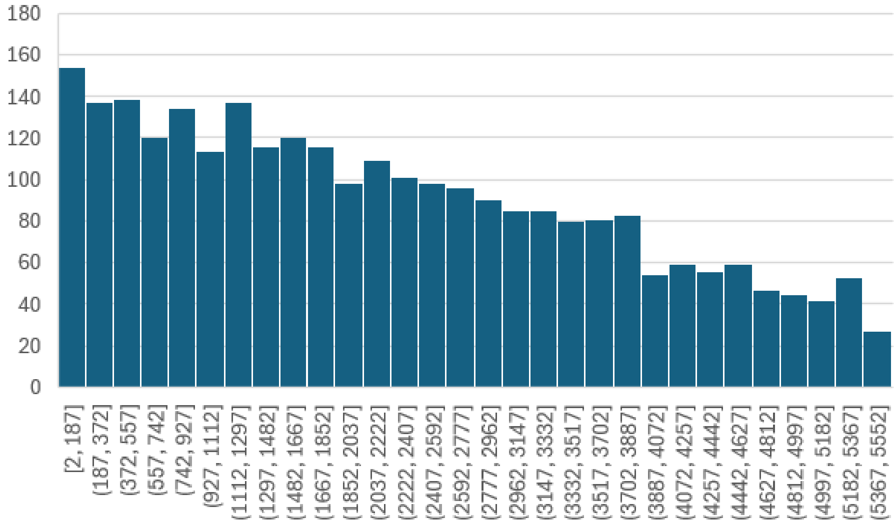

Figure 5.6: Number of worse accepted iterations per 200 iteration

5.6.4 Exp 4: Baseline schedule, Flexible scanners, More appointments with same capacity (SA)

This experiment tests the effect on the baseline schedule when more outpatient appointment slots are created.

Simulated annealing

We will create an extra appointment slot at a random, empty location of one of the CT scanners. We decided with the hospital employees that a maximum of 7 extra scans per day can be added. Harden (2023) states in the Royal College of Radiologists that a 10% increase in scans performed can lead to an increased workload in the rest of the chain. In our case, radiologists already have too high a work pressure, and therefore we do not want an increase in scans of more than 10% (max 7 extra appointment slots per day). We maintained the SA parameters constant, allowing for a reasonable comparison of the different experiments.

Performance function of appointment schedule

The base of the performance function will be the same as in previous experiments, allowing for easy comparison. However, we added a reward for every extra outpatient appointment slot. Through trial and error, just as in the previous experiment, we determined that a reward of 25 seconds would be given if there was an extra appointment opportunity on every weekday, so 5 seconds per weekday. This balance is important because we do not want to violate target service levels too much with extra appointment slots. The final performance function is:

$$\text{Performance Function} = \text{AvgWaitingRoomTime} + P_1 + P_2 + P_3 + P_4 + P_5 - R_6$$

$$R_6 \text{ (Reward 1)} = \begin{cases} (\text{ConsultedOutpatients} - 21) \times 0:25, & \text{Number of consulted outpatients} > 21 \\ 0, & \text{otherwise} \end{cases}$$

5.6.5 Exp 5: Input exp 3, Flexible scanners, More appointments with same capacity (SA)

This experiment is the same as experiment 4, but with the best-performing appointment schedule of experiment 3 as input.

5.6.6 Exp 6: Schedule staff

In experiment 6, we test the performance of the newly designed schedule by the staff with DES.

5.6.7 Exp 7: Schedule staff, Flexible scanners

In this experiment, we test with DES the performance of the schedule designed by the staff with flexible priority rules for all scanners.

5.6.8 Exp 8: Schedule staff, Flexible scanners, Appointments on all scanners (SA)

This experiment is the same as experiment 3, but with the input schedule of the staff.

5.7 Conclusion

In this chapter, we implemented the solution approach to the CT department of UMC Utrecht. With the help of historical data and literature, we developed empirical and statistical distributions that serve as inputs to our simulation model. We extensively tested whether our simulation model represented the conceptual model. We found no significant difference between the means, but there are some differences in the variances, which could be the cause of the chosen distributions. We also tested whether the simulation model accurately represented the real world and would be usable in practice. Unfortunately, there is a difference between the expected waiting time and the actual waiting time. As described, this is because the simulation model does not include the waiting time for blood tests or the time required to drink fluids. Therefore, the outcome of the waiting time in the simulation model is not reliable. However, a decrease in the ratio can still be utilised. Lastly, the experiments are set up, the simulated annealing parameters are chosen, and the performance functions are tested and chosen to get the most reliable results. We implemented a special simulated annealing algorithm with a restart to escape from a local optimum, as we only use one-way moves. The results are described in the next chapter.

6 Result analysis

In this chapter, we describe the results of the experiments designed in the previous section. For each experimental setting, we identify the best-performing appointment schedules. All experiments are compared to the initial appointment schedule and the target levels, and compared to each other. Table 6.1 presents the outputs of all experiments. The outputs are the averages per day over all replications. The lower the performance function, the better the appointment schedule performs. Figure 6.1 shows the number of appointments per weekday, while Figure 6.2 illustrates the number of appointments per hour and the total per week. We will not refer to these figures and tables repeatedly in the text. Additionally, a sensitivity analysis is conducted in Section 6.10, and the simulated annealing algorithm is validated in Section 6.11. The chapter concludes with a comprehensive summary an conclusion in Section 6.12.

Experiment & best performing iterations	WTEmergency	WTWalkIn	WTInpatient	WTOutpatient	TotalAvgWT	Outpatients	Access time	Overtime	SLEnegency	SL WalkIn	SLInpatient	SL Outpatient	Utilization	Performance
1. Baseline	110	345	162	206	235	21	31	79	0.86	0.95	0.92	0.93	0.62	499
2. Baseline, Flexible	87	411	123	134	236	21	31	80	0.89	0.93	0.95	0.97	0.62	281
3.1 Baseline, Flexible, App. on all scanners	76	346	111	118	201	21	31	179	0.91	0.95	0.96	0.98	0.62	201
3.2 Baseline, Flexible, App. on all scanners	81	341	110	136	203	21	31	240	0.91	0.96	0.96	0.97	0.61	203
3.3 Baseline, Flexible, App. on all scanners	82	318	108	125	204	21	31	195	0.9	0.96	0.96	0.97	0.61	204
4.1 Baseline, Flexible, More appointments	88	464	120	146	274	24.6	25.7	108	0.9	0.92	0.95	0.97	0.65	166
4.2 Baseline, Flexible, More appointments	102	516	143	176	285	28.6	19.8	265	0.89	0.91	0.94	0.95	0.68	167
4.3 Baseline, Flexible, More appointments	89	554	132	157	297	26.8	22.4	168	0.89	0.9	0.94	0.97	0.67	184
5.1 Exp 3, Flexible, More appointments	88	389	119	131	232	24.6	25.7	286	0.9	0.94	0.96	0.97	0.64	124
5.2 Exp 3, Flexible, More appointments	88	392	117	137	237	23.8	26.9	233	0.9	0.94	0.95	0.97	0.64	153
5.3 Exp 3, Flexible, More appointments	90	412	122	149	241	26.6	22.7	239	0.89	0.94	0.95	0.97	0.66	158
6 Schedule staff	152	549	208	329	369	32,0	14.8	415	0.81	0.9	0.89	0.88	0.71	1637
7 Schedule staff, Flexible	121	719	156	198	375	32	14.8	413	0.85	0.85	0.94	0.95	0.71	1549
8.1 Staff, Flexible, App. on all scanners	114	763	149	233	411	32	14.8	297	0.86	0.86	0.94	0.93	0.72	896
8.2 Staff, Flexible, App. on all scanners	112	747	154	230	407	32	14.8	281	0.86	0.85	0.94	0.93	0.72	941
8.3 Staff, Flexible, App. on all scanners	114	717	161	218	388	32	14.8	314	0.86	0.86	0.93	0.94	0.71	955

Far below target | Just below target | Just above target | Far above target | Best performing experiment

Table 6.1: Experiments and best performing (lowest performance function) iterations

6.1 Exp 1: Baseline schedule

The initial appointment schedule has two flexible CT scanners and one dedicated walk-in CT scanner. The performance of this initial schedule is shown in Table 6.1. Notably, the service level for emergency patients (0.86) falls significantly below the target of 0.9, which negatively impacts the performance function (499). In contrast, the other service levels, overtime, and average waiting times across patient categories perform well. The access time is 31 days, as calculated in Section 5.2, which is well above the target of 22.4 days. This experiment is used for simulation validation, verification, and comparisons with other experiments.

6.2 Exp 2: Baseline schedule, Flexible scanners

We aimed to investigate the impact of flexible priority rules on CT scanner 3. To do so, we tested the initial schedule with flexible priority rules for all scanners. Compared to the baseline experiment, the most notable result is that the service level for emergency patients (0.89) is closer to the target of 0.9, positively influencing the performance function. The average waiting times for emergency patients, inpatients, and outpatients also improved, along with the service levels for inpatients and outpatients. Overtime and the overall average waiting time remained unchanged. However, the average waiting time for walk-in patients increased significantly, although their service level still met the target. In conclusion, the overall performance function (281) improved compared to the baseline (499), while the access time remained the same.

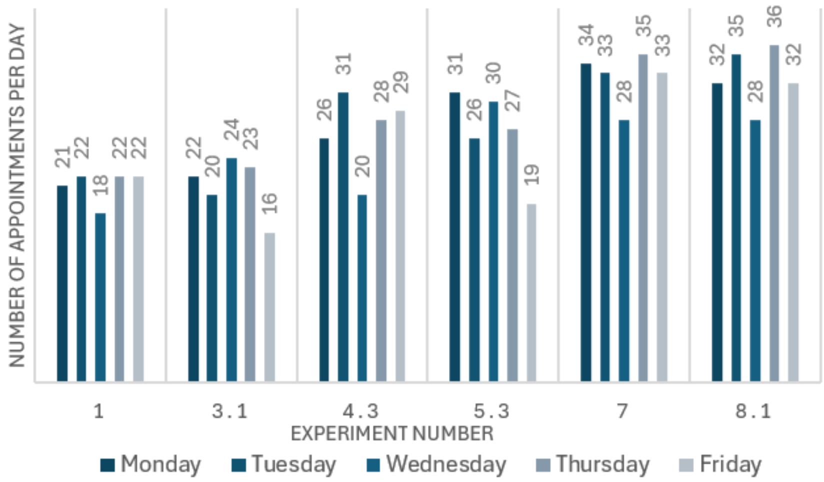

Figure 6.1: The number of appointments per day per experiment

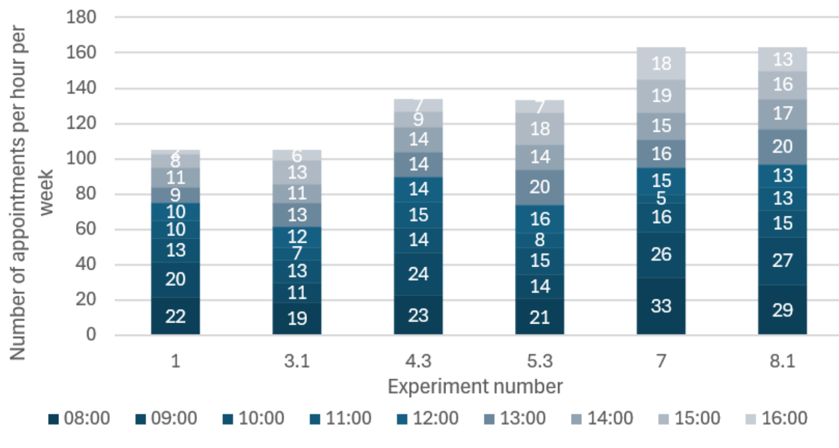

Figure 6.2: The number of appointments per hour and per week per experiment

6.3 Exp 3: Baseline schedule, Flexible scanners, Appointments on all scanners (SA)

This experiment examines the effects of CT scanner 3 operating with flexible priority rules and fixed appointment slots. Table 6.1 shows the results of the three best-performing iterations (Experiments 3.1, 3.2, and 3.3). Since their performance functions are similar, we selected the best iteration based on the overall output. Overtime is a key concern for the hospital, as well as meeting all target levels. Iteration 3.1 achieves the best performance in terms of overtime and meets all targets, making it the most effective configuration in this setting.

In this iteration, 39 out of 58 possible appointments shifted to CT3, indicating that many were not initially scheduled at optimal times. The specific CT scanner used for an appointment does not affect the sequence in which patients are scanned, as all scanners select the next patient from a shared waiting room based on the longest waiting time. In the baseline schedule, 40 appointments were planned in the afternoon, compared to 55 in iteration 3.1 (see Figure 6.2). This shift is logical as observations showed significant idle time in the afternoon, as staff tend to over-reserve time for unplanned patients to avoid working overtime. Additionally, many CT1 and CT2 appointments were at the same time in the baseline schedule, whereas in this experiment, appointments were more evenly distributed. We also see a shift in the number of appointments scheduled per day. For example, more appointments are planned on Wednesday and fewer on Friday (see Figure 6.1). This was unexpected, given that more walk-in patients typically arrive on Wednesdays. However, shifting appointments to the afternoon likely had a greater impact on the objective function than the daily appointment distribution, especially since current capacity exceeds demand.

In conclusion, having flexible priority rules and shifting appointments to a more optimal timing at CT scanner 3 has a big positive effect on the performance function.

6.4 Exp 4: Baseline schedule, Flexible scanners, More appointments with same capacity (SA)

In previous experiments, we found that assigning flexible priority rules to all CT scanners positively impacted the performance function. However, the access time remained unchanged. To address this, we tested a combination of the flexible priority rule configuration with increased appointment slots. This experiment had the initial schedule as input.

4.1, 4.2, and 4.3 in Table 6.1 show the performance values of the three best-performing iterations in this experimental setting. While their performance functions are close, there are notable differences in overtime and access time. Iteration 4.2 provides the most appointment opportunities but results in significantly more overtime and fails to meet the service level target. Iteration 4.1 has the highest access time, yet most of its other performance indicators are better than those of iteration 4.2. Iteration 4.3 achieves an access time of 22.4 days, meeting the target, and outperforms iteration 4.2 on most other indicators. Since our goal is to reduce access time without causing significant adverse effects, iteration 4.3 proves to be the most effective.

Compared to the baseline, this schedule adds 29 appointment slots per week, 11 in the morning and 18 in the afternoon (see Figure 6.2). As previously discussed, it is unsurprising that more slots were added in the afternoon, given the hospital's hesitancy to schedule many appointments during this period. The additional slots are distributed as follows: five on Monday, nine on Tuesday, two on Wednesday, six on Thursday, and seven on Friday, showing an increase across all weekdays. Notably, Wednesday, which is expected to have the highest number of unplanned arrivals, now has the fewest appointment slots (20), which is a logical outcome. Thursday, with 31 slots, has the most, even though it does not have the fewest unplanned arrivals. This likely reflects the efficiency of the baseline schedule's timing on that day, allowing for additional appointments without disruption. Thursday and Friday, which are expected to have the fewest unplanned arrivals, have 28 and 29 slots respectively, aligning with expectations.

In conclusion, the baseline schedule can accommodate additional appointment slots with limited adverse effects. Most new appointments are scheduled during periods with fewer expected unplanned arrivals. However, this was not always the case and may be influenced by how well-timed the appointments were in the original schedule.

6.5 Exp 5: Input exp 3, Flexible scanners, More appointments with same capacity (SA)

From the previous experiment, we observed that the input appointment schedule significantly influences how many additional appointment slots can be added. Therefore, we tested the scenario using the appointment schedule from iteration 3.1 as input. When comparing iterations 5.1, 5.2, and 5.3 to iteration 3.1, we see that all outputs, except access time, perform slightly worse. However, all target levels are still met. When comparing these results with iterations 4.1, 4.2, and 4.3, we find that iterations 5.1, 5.2, and 5.3 perform better despite offering the same number of appointment slots. Nonetheless, iteration 4.2 remains the one that yields the highest number of appointments, a value not matched in this group. We also compared the performance of iterations 5.1, 5.2, and 5.3 with each other. Iteration 5.3 is the only experiment with an access time of 22.7 days, close to the target access time of 22.4 days. The other performance metrics are nearly identical, and almost all targets are met. Therefore, iteration 5.3 is the most effective within this setting.

This iteration started with 50 appointments in the morning and 55 in the afternoon. In total, 28 additional slots were added, 8 in the morning and 20 in the afternoon, resulting in 58 morning and 75 afternoon appointments (see Figure 6.2). This distribution contrasts significantly with iteration 4.3, which had 76 morning and 58 afternoon appointments. This demonstrates that the initial appointment schedule has a substantial effect on the experiment's outcome and overall performance. Moreover, the number of appointments per day does not always align with the expected number of walk-in arrivals. For example, this schedule allocates 30 appointment slots to Wednesday, even though Wednesday typically sees the highest volume of walk-ins. This may be due to the favourable timing of existing appointments on Wednesday in the initial schedule, which allowed for more additions without negative impact.

In conclusion, the input appointment schedule plays a crucial role in the outcome of the experiment, particularly because we use one-way move operators. Therefore, experimenting with different initial schedules will give a more robust solution.

6.6 Exp 6: Schedule staff

When the senior laboratory technicians saw our data, especially the low utilisation and high access times, they realised something had to change. They did not want to wait until the end of this research to improve and implement. Therefore, they came up with a new appointment schedule and implemented it. They used information such as that many walk-in patients arrive during lunchtime, especially on Wednesdays. It is too early to assess the real-world performance based on historical data, as several months of data are needed to obtain reliable insights. Therefore, we evaluate the performance using our simulation model. The simulation shows that the access time dropped to the lowest level observed across all experiments. This improvement was achieved by adding 11 extra appointment slots per day. Radiologists work with a long working list and therefore have not yet experienced a higher workload. However, we expect that the working lists will increase gradually with this number of appointment slots. As outlined in Section 5.6.4, the radiologist can accommodate a maximum of 7 additional appointments per day. This confirms that adding 11 extra appointments exceeds the feasible workload limit.

Other performance indicators also showed undesirable effects. The average waiting time for all patient types increased and nearly doubled over time. In addition, the service level of emergency patients is far from the target level, and the service level of inpatients and outpatients is also below the target level. The overtime also increased and is confirmed by the laboratory technicians. As a result, the performance function increased significantly. One important improvement is that now the target access times are met.

6.7 Exp 7: Schedule staff, Flexible scanners

Since we discovered that priority rules can significantly influence performance, we tested the effect of a flexible priority rule on CT3. Compared to Experiment 6, the objective value improved, along with the service levels. The average waiting times for emergency patients, inpatients, and outpatients decreased, while their service levels increased. In contrast, the service level and waiting time for walk-in patients decreased and increased, respectively. This pattern is consistent with the behaviour observed in previous experiments involving flexible priority rules.

6.8 Exp 8: Schedule staff, Flexible scanners, Appointments on all scanners (SA)

We also wanted to test the effects of moving appointments to a more optimal timing at CT3 with this schedule as input. Iteration 8.1 performed the best, as its service levels were closer to the target. Compared to experiments 6 and 7, it outperformed in all areas, even meeting the target level for overtime. Most appointments remained on the same day, five were moved to the morning, and several found a more optimal time.

We can conclude that laboratory technicians got better at determining the optimal day and time for appointments with the help of the provided data. However, SA is better at finding the ideal timing for appointments.

6.9 Conclusion experiments

To conclude, iteration 4.3 of experiment 4 has the best overall performance. It is the iteration that meets the target access time while all other KPIs are closely met. This means that it is best to have flexible priority rules for all scanners and have fixed appointments on CT 3. Additionally, on each scanner, hour and day extra appointment are added.

6.10 Sensitivity analysis

Ferrand et al. (2014), just as many other researchers suggest, to perform a sensitivity analysis of the simulation model to test the robustness to change of the model. We will do this by assessing the consequences of changing the input parameters of Experiment 1, the initial schedule. We already tested the simulated annealing with different input appointment schedules. We will adjust the appointment duration distributions, arrival rate distributions, and opening hours.

6.10.1 Distribution 20 minute appointment duration

Our first analysis tested whether the model would work when we adjusted the appointment durations, and it did. Figure 6.3 shows that as the appointment duration increased, the objective value also increased. This is logical since waiting time, overtime, and service levels negatively impact the objective function. From 0% to -10%, the performance function decreased, but decreasing the appointment duration further had little effect. This is because there is less waiting time and overtime to improve.

In conclusion, the simulation model can effectively test different appointment durations. Longer appointment durations significantly affect the performance value, while shorter durations have minimal impact.

6.10.2 Arrival rate Walk-in patients

The model was also able to test variations in walk-in arrivals. Decreasing the number of arrivals had a similar effect to reducing the appointment duration. The performance value increased

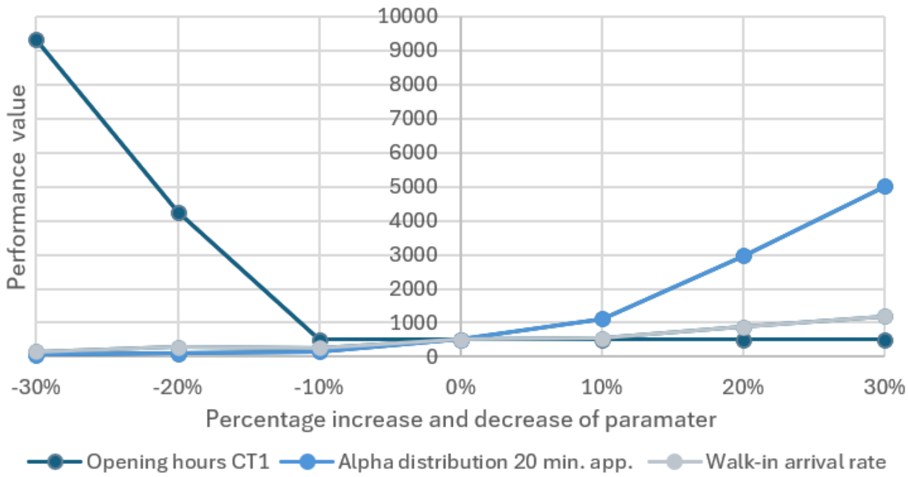

Figure 6.3: sensitivity analysis

as more arrivals were added, but the impact was less significant than increasing appointment duration. This is logical, as the appointment duration applies not only to walk-in patients. With a 10% increase in arrivals, the performance changed slightly. This indicates that the appointment schedule can accommodate more patients without adverse effects.

6.10.3 Opening hours CT1

As expected, increasing the opening hours did not affect the objective value since there was already less overtime. On the other hand, reducing the opening hours of CT1 had a significant impact on waiting time and overtime. However, a 10% reduction in opening hours had minimal effects, suggesting that it could be cost-effective to close CT1 earlier or schedule more patients during the last hour.

6.11 Validation Simulated annealing

As described in Section 4.8, verification and validation are crucial. Many researchers, such as Zheng et al. (2025) and van Essen et al. (2014), focus on validating their models with real-life data. However, they rarely describe the process of verifying their simulation models or simulated annealing. This section outlines the verification steps we took for our simulated annealing.

When we rerun our simulated annealing from the start, it tests the same set of appointment schedules. This ensures that the random number streams for selecting the next experimental setting are implemented correctly. This is important because, if an experimental design with a different input appointment schedule performs better, it should not be attributed to randomness. Additionally, each time a new appointment schedule is selected for testing, this confirms proper implementation, as we aim to test more than just one appointment schedule.

We also tested Experiment 3.1 using a different random number stream for our simulated annealing. Simulated annealing is a metaheuristic and will not necessarily find the optimal solution (Thomas Schneider et al., 2020). Therefore, we expect the best performance to vary with different random number streams. However, the simulated annealing should still yield near-optimal solutions, without extreme outliers. This means the best performance should be closely aligned. Figure 6.4 displays the results of the 10 best performance values from four simulated annealing runs with different random number streams. As shown, the worst best performance (208) and the best best performance (194) differ by only 6%, with no extreme outliers. Since we actively tested with different simulated annealing parameters, we do not expect further improvements from adjusting them. However, there may be potential for improvements in choosing the next neighbour or the restart mechanism. Compared to the baseline schedule (499), the least performing performance (208) still shows a 58% improvement, compared to the best performing performance (194) with a 61% improvement. This indicates that the simulated annealing is stable enough for our purposes.

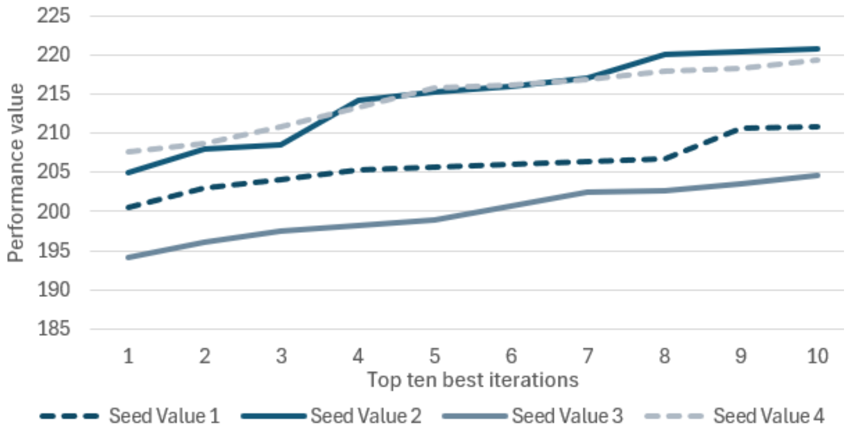

Figure 6.4: Testing simulated annealing with different random number streams

6.12 Conclusion

We were able to add appointments to all hours of the day and still meet the KPIs without changing capacity. First, changing the priority rule to flexible CT scanners positively affected the performance value and all other target levels. The waiting time for walk-in patients and the service level increased but remained within the target. Moving appointments to CT3 at more optimal times also positively affected the performance value. We discovered that it no longer matters which CT scanner the appointments are assigned to, as the priority rules for all scanners are now the same. By adding more appointments to the schedule, we achieved the target access time of 22.4 days. If more appointments are added, the other targets will no longer be met. We found that the input schedule of the simulated annealing had a significant effect on the output, suggesting that different input schedules might perform even better. Laboratory technicians got better at developing schedules with the help of the provided data. However, SA is better at finding the ideal timing of appointments.

Through the sensitivity analysis of the base schedule, we found that the performance value deteriorated exponentially with longer appointment durations. Therefore, the hospital should monitor this change closely. With this schedule, expanding the opening hours has no effect. However, if the opening hours of one CT scanner decrease by 20% due to, for example, a breakdown, the overtime and waiting times will increase significantly. A 10% increase or decrease in walk-in patients will have a minimal effect on the performances.

7 Implementation and recommendations

This chapter will answer the research question, "How can the proposed appointment schedule be effectively implemented?". Section 7.1 will give the implementation theory and Section 7.2 explain the stakeholders involved. Followed by the implementation actions and further recommendations in Section 7.3 and Section 7.4.

7.1 Implementation theory

Change management is essential to help employees accept and embrace the changes required to implement the new appointment schedules and priority rules. Chowdhury and Chandra Shil (2022) outlines three key steps necessary for successful change. First, it is important to reduce the forces that maintain the organisation's current behaviour. We achieved this by involving employees at all levels in the decision-making process regarding the research focus. We dedicated considerable time to informing them about the current situation and requesting their feedback on the problems they encountered. Together, we identified the first problem to address. This combination of involvement and data about the current situation helped open them to change.

The next step is to prepare employees for the change. We engaged them in designing the new appointment schedule, ensuring they trusted the data we used and regularly sought their input. The decision to use discrete event simulation in plant simulation to test new appointment schedules was key in building trust in the model. The visualised system not only highlighted the issues with the current system but also demonstrated the new system's performance. This allowed employees to think critically and reason in new ways.

The final step is implementing the change and ensuring its acceptance to prevent a gradual return to the old system. This requires changes to working policies, norms, and structures. Employees' willingness to change is evident in the fact that they have already developed and implemented a new appointment schedule.

7.2 Stakeholders

The primary stakeholders in this research are the integral capacity management team and the head of radiology. It was crucial to keep them involved and address their needs throughout the process. The head of the CT scan department also needed to gain trust in this project. As he has a close relationship with all employees in the department, his influence is key in determining how much time they can allocate to the project. Ultimately, the head of radiology and CT scans will decide on implementation. Administration employees, laboratory technicians, and other departments must also contribute to implementing and adapting their work policies. For this process to proceed smoothly, these employees must be open to change and trust the process. We achieved this by involving them in key meetings, allowing them to validate the model and data, and regularly seeking their input.

7.3 Actions

The new appointment schedule and priority rules can be implemented with the following description.

Implementation of recommended priority rules and appointment schedule

The first step of implementation is to inform all employees about the new appointment schedule and priority rules. Currently, appointment slots are assigned to specific departments, such as cardiology or certain scan types. There is currently insufficient data to reassign large numbers of appointment slots to these patient groups based solely on data. Since we now have more appointment slots available, we recommend maintaining the current distribution across departments and using the additional slots as flexible appointments. As noted by Hulshof et al. (2012), too many reserved slots can lead to idle time, while too few can increase access times for specific patient groups.

The second step involves updating the priority sequencing rules. CT3 employees should now follow the same sequencing rules as those used in CT1 and CT2. After this, the new appointment schedule should be implemented in HIX. Once implemented, administrative staff across departments can view and fill the new appointment slots. The central administration of the radiology department should then review existing appointments to ensure they are scheduled correctly and make adjustments if necessary. This process will require time and effort, making it essential that the appointment schedule is not changed frequently or at short notice.

The final and most critical step for successful adaptation is learning from experience, as highlighted by Chowdhury and Chandra Shil (2022). The hospital now has access to data that can be used to monitor changes and identify opportunities for further improvement and optimisation.

Appointment schedule simulation

We used the simulation model to test the priority rules and develop new appointment schedules. This model should be reused whenever key parameters, such as capacity, appointment duration, or arrival rates, change, to determine the appointment schedule that best fits the updated data. It can also be used to evaluate the performance of specific proposed schedules.

7.4 Recommendations

Followed by the implementations, we also have recommendations for further improvements in aligning capacity with demand. We came up with some of them early in the research during the problem formulation, but they have not been worked out because of the scope. Some of them are direct results from the research, and others are results from the whole change process. We placed them in order, based on ease of implementation and potential contribution.

Direct conclusions from research:

1. **Improve data quality.** As is clear now, the data quality was not always good and useful. We recommend adjusting work processes, such as closing the patient file when treatment is finished, to enhance the accuracy of appointment duration data. Also, distinguishing between expected durations for different patients and scan types can improve demand forecasting.

2. **Implement a uniform working method across all outpatient departments.** Ensure consistent use of the "wish date" (the date entered by the physician so administration knows when to schedule the appointment), rather than interpreting it as the preferred or latest possible appointment date. Departments should also follow the same guidelines for referring patients to walk-in appointments and using working and waiting lists. This consistency will improve future demand forecasting.

3. **Determine the optimal balance between scheduled and unscheduled appointments.** Currently, no guidelines exist for this balance. Finding the optimal mix can improve utilisation, access times, waiting times, and more (Hulshof et al., 2012).

4. **Conduct a performance peer review.** With the available data, it is now possible to benchmark performance against other hospitals and learn from their practices.

5. **Balance capacity across departments.** Since laboratory technicians work across multiple departments, it is essential to balance their capacity with overall hospital demand.

6. **Include the children's hospital CT scanner.** This scanner currently has low utilisation and could be used for adult appointments to improve efficiency.

Indirect conclusions from research:

1. **Create a dashboard to visualise data and current performance.** This research advises on a strategic level and a tactical level. But as described, the tactical and operational levels are just as important and the next step to tackle. The data is currently stored in large Excel sheets, making it difficult to monitor. A dashboard would provide clearer insights and make it easier to align capacity with demand on a tactical and operational level.

2. **Introduce a weekly tactical planning meeting to discuss dashboard insights.** The data that is now available is not useful without translating it into knowledge and actions. The dashboard might show that many appointment slots remain unfilled. In that case, the hospital could increase demand by opening slots to external clinics. Or they see a decrease in the capacity of employees, and can improve this with the help of the simulation model. With these insights, staff scheduling can shift from supply-driven to demand-driven, helping address one extra root cause of long access times.

3. **Use the dashboard for real-time operational planning.** Staff can use it to detect unusually long waiting room queues and take action, such as rescheduling to prevent overtime or assigning walk-in patients to preferred time slots. On top of that, they have control over which patient group has to wait the longest, making it essential to keep track of whether this still matches the vision and steer on this. Additionally, identifying and prioritising short appointments can help improve customer service.

4. **Perform simulation for other image modalities.** We proved that with the simulation study we can improve the appointment schedules. So it would be valuable to also apply this to MRI and X-ray.

7.5 Conclusion

To conclude, the theory of implementation and change highlights its importance. It is essential to keep stakeholders engaged, open to change, and confident in the results. The hospital is now well-positioned to implement the new appointment schedule, priority rules, and simulation model. In addition, several recommendations have been provided to support future changes and successful implementation.

8 Conclusion

In this thesis, a new appointment schedule and priority sequencing rules were developed and evaluated through simulation to better align demand with capacity. The process began with an analysis of the current state and a literature review. Insights gained from the theoretical research, data analysis, and expert opinions guided the formulation of a proposed simulation model. Additionally, an implementation plan was created. This chapter addresses the last research question: 'What conclusions and recommendations can be drawn from the research conducted?'. Section 8.1 explains the conclusion of this research. Section 8.2 lists the practical and scientific contributions. The chapter ends with the limitations and suggestions for future research in Section 8.3.

8.1 Conclusion

The research began in response to multiple capacity issues at the hospital. Together with stakeholders, we identified the most critical problem to address first: the access time for scheduled CT scan appointments was longer than agreed upon with health insurers. Our analysis revealed that this was primarily due to an appointment schedule that did not align with patient demand. This led us to formulate our main research question: *"How can the appointment schedule for CT scans at UMC Utrecht be optimised to better align with patient demand to achieve the target access times?"*. We provided the department with many valuable data and analyses. Because of inadequate registration, we had to manually improve the dataset and remove inconsistencies.

Literature recommends using slot-based scheduling in environments with high levels of unplanned arrivals and the need to reduce access time. This method, allocating fixed slots with set start and end times to specific patient groups, is particularly suitable for decentralised administration. Additionally, we found that using a cyclic appointment schedule is most effective for improving both access time for scheduled patients and waiting time for walk-in patients. To design and evaluate these appointment schedules and sequencing rules, we use a combination of discrete event simulation and simulated annealing. The visualised model in Plant Simulation proved valuable for engaging stakeholders and increasing the likelihood of successful implementation. The literature has not reached a consensus on whether flexible or hybrid rooms are more effective in a system with many unplanned arrivals and focus on improving access time and overtime. That is why we tested this in our simulation as well.

The simulation model showed that changing the priority rules to make CT scanner 3 flexible, and allowing it to accommodate scheduled appointments, reduced the performance function by 60%. We also found that laboratory technicians can improve appointment schedules using available data, but identifying the optimal timing remains challenging. With the best performing schedule generated with the SA with more appointment slots, flexible priority rules, and appointments on CT3, almost all service-level targets were met, and average appointment durations decreased. On top of this, the access time dropped from 31 days to the target of 22.4 days.

To answer our main research question, a simulation model can optimise the appointment schedules for CT scans at UMC Utrecht to better align with patient demand and achieve target access times without changing capacity. This model improves the timing and number of scheduled appointments and should be supported by adjusting the sequencing rules of CT scanner 3 to a flexible priority approach.

8.2 Practical and scientific contribution

8.2.1 Practical contribution

This research has made several practical contributions. First, it brought attention to numerous issues and potential improvements related to capacity and demand, and it successfully encouraged openness to change among employees. In addition, a substantial amount of data is now available to evaluate the department's performance, for example, they now have insight into actual access times, waiting times, and utilisation. Previously, the department had no KPIS or target levels, but we developed them together. As a result, the department can now more easily identify weaknesses and areas for improvement. With the simulation model we developed, they can test the impact of adjusting parameters such as capacity and appointment duration on the appointment schedule. We created a new appointment schedule and sequencing rule that enables the department to meet both the target access time and other key performance indicators. Finally, we compiled a list of additional potential implementations to help better align capacity with demand.

8.2.2 Scientific contribution

First, we structured the steps necessary to create an appointment schedule accommodating numerous unplanned arrivals while focusing on access time, waiting time, and overtime. While many studies concentrate on one or two of these aspects, our approach aims to address all of them. We developed two frameworks to help researchers identify the best method suited to their priorities. The first framework, shown in Figure 3.1, assists in selecting the appropriate level of detail when aiming to reduce access time in the appointment schedule. The second framework, illustrated in Figure 3.2, helps determine the most effective method for designing an appointment schedule based on prioritising specific patient types, requirements, and limitations.

The literature offers limited guidance on choosing between a flexible or hybrid policy when dealing with a significant group of walk-in patients. Studies suggest that flexible rules improve utilisation, increase overtime, and reduce waiting times for emergency patients. In contrast, hybrid rules optimise access and waiting times across diverse patient groups. In our findings, the flexible priority rule resulted in a decrease in emergency waiting times and an increase in overtime. We did not observe a change in utilisation when we only altered the priority rules and access time. However, with flexible priority rules, the average waiting time decreased significantly, freeing up additional space for appointments. While literature suggests that hybrid rules would reduce average waiting times, our results showed that the flexible priority rule produced the most significant reduction in the objective. It positively impacted access time, service levels, and average waiting time, although walk-in waiting time and overtime saw a slight increase. Based on our case study, we can not conclude that flexible priority rules work best for all hospitals, since we cannot test all dependencies with one case study.

8.3 Limitations and future research

This research has certain limitations because of its defined complexity, scope, assumptions, simplifications and limited data. As directions for future research, we present suggestions for addressing these limitations and other potential areas for further investigation.

8.3.1 Simulation model limitations

The simulation model relies on various assumptions and simplifications to ensure feasibility and usability. We carefully selected them based on stakeholder input and literature, but they still introduce limitations in how the system is defined.

1. **Not all patient behaviour is explicitly modelled:** Patient punctuality, cancellations, and no-shows are not modelled. For example, we did not model patient rescheduling after cancellations or no-shows. However, variability is included through stochastic arrival and service time distributions. These simplifications may influence the accuracy of waiting time and access time predictions, particularly for outpatient appointments.

2. **Scan durations are only split up in two distributions:** Although scan duration distributions are based on empirical data and fitted using statistical methods, they do not capture different scan protocols or patient types. For instance, complex procedures or newer scanner technologies may require different processing times, which are not separately modelled. However, we do think that with the large number of scans, the appointment durations will balance themselves out, but we did not test the effect on waiting times.

3. **We assume consistent staff availability and scanner uptime:** In practice, staff shortages and machine downtime can introduce unmodelled fluctuations in capacity and can result in longer waiting times and overtime.

4. **Demand forecasted based on historical data:** It does not have to be true that future demand will be the same as history. This assumption could result in other performances than in practice. With higher-quality data, more advanced forecasting methods can be applied. For example, outpatient clinic schedules could be used to predict walk-in patient arrivals based on referral rates per appointment type. This would also allow better demand adjustments to match capacity and the use of alternative appointment schedules during periods of low or high demand.

8.3.2 Simulated annealing limitations

The simulated annealing algorithm has several limitations and opportunities for future research:

1. **No reheated temperature:** Currently, the temperature decreases after a fixed number of iterations. However, when multiple new best solutions are found, staying longer at a temperature may be beneficial. Additionally, if the algorithm becomes stuck in a local optimum, reheating could help it escape.

2. **Not returned to one of the best schedules or a random one:** When the algorithm gets stuck in a local optimum, simulated annealing restarts with the initial appointment schedule. However, it might be more effective to try improving one of the best solutions found so far instead of a random schedule.

3. **Testing different stopping criteria:** To reduce running time, it may be helpful to stop the simulation if no new best solution is found after a specific number of iterations.

4. **No smart move operators implemented:** Keep track of move operators that perform well and apply those more frequently in subsequent experiments.

5. **No tabu lists used:** Save the "bad" moves and skip them for several iterations to avoid revisiting suboptimal solutions.

6. **Potentials in optimising the restart mechanism:** Our current simulated annealing uses one-way operators, meaning that if the best solution is achieved with 50 moves, a new best solution likely requires at least 50 additional moves. As a result, after a restart, a worse objective must be accepted at least 50 times before a better solution can be found. Reducing the frequency of restarts as the temperature decreases may improve the process.

7. **Only one move operator is used :** This makes the quality of the initial schedule particularly important. Incorporating a combination of move, swap operators, and adding or removing appointments could lessen the dependence on the initial schedule. Furthermore,

this approach would allow for a broader search space, increasing the likelihood of escaping local optima and finding better solutions.

8.3.3 Experiment limitations

Although the experiments were carefully designed in collaboration with stakeholders, certain limitations remain in how the scenarios were defined, which may affect the generalisability of the findings.

1. **A limited number of experimental settings are tested:** The experiments were chosen in collaboration with stakeholders and based on expected impact. However, there may be other promising configurations or combinations that were not explored. For example, experimenting with planning more walk-ins or communicating the current waiting time.

2. **Not tested the optimal balance between scheduled and unscheduled patients:** We assumed that we could not influence the decision of planning a patient or not. The right balance between planned and unplanned arrivals ensures both access and waiting times are optimised. Currently, there is no clear guideline on how many patients can arrive unplanned. Further research should develop methods to help determine this optimal balance.

3. **Simulated annealing algorithm uses a specific set of parameters and a one-way move operator:** They are calibrated and tested, but still limit the exploration of solutions, and often optimal solutions are not found (Thomas Schneider et al., 2020). In practice, alternative metaheuristics or mixed-integer linear programming might uncover better-performing solutions. When they are both in place, the pros and cons can be better compared.

8.3.4 Robustness of results

To assess the reliability of the simulation outcomes, it is essential to reflect on the robustness of the results.

1. **The simulation results are based on a one-week cycle time:** This captures day-to-day variation, but does not account for yearly variation such as holiday periods and seasonal demand shifts. As a result, the model's conclusions are most valid for regular weeks and may not apply to atypical scenarios. However, demand and capacity can be changed, and atypical weeks can be tested as well.

2. **The performance is analysed with replications:** This provides stable estimates, but may hide outliers. Further robustness checks, such as worst-case scenario testing, would be beneficial.

3. **Sensitivity to small changes in parameters:** Small changes in opening hours of scans or arrival rates hardly influence the performance and the model still works. However, small changes in the appointment duration distributions do affect the performance and therefore might lack the stability to draw robust conclusions.

4. **Variance in input data may not fully represent reality:** Some of the fitted distributions, especially those for appointment durations, do not perfectly reflect real-world variance. For example, delays caused by blood tests or pre-scan preparation are not included. This means that while the model captures relative changes well, absolute outcomes should be interpreted with caution. For these waiting time outputs, the model is best used for comparing scenarios, rather than for producing exact forecasts.

5. **Simulated annealing results vary across random seeds:** Because simulated annealing is a metaheuristic, its outcome depends on the random number stream used to explore the solution space. We observed some variation in the best performance across different seeds, but no extreme outliers. Since different seeds can produce different appointment schedules, it does not make sense to report a single average performance value. Instead, we report the best result per run and ensure it is comparable to other top results. These findings confirm that simulated annealing provides consistent and near-optimal solutions for our purpose.

6. **Decision on flexible scanners based on one case study:** Our results suggest that flexible priority rules are preferable when unplanned arrivals are high and the focus is on access time, overtime, and waiting time. However, this conclusion is based on a single case. Future research should explore this in different departments or hospitals to validate and generalise the findings.

A Appendix

A.1 Simulation input values

CT1	Monday	Tuesday	Wednesday	Thursday	Friday
8:00	60	60	60	20	60
8:20				20	
8:40				20	
9:00	20	60	60	60	60
10:00			20	20	
10:20	20	20	20		
10:30				20	20
10:40		20	20		
10:50					20
11:00	20	20			
11:20					20
11:30		20	20		
11:40					20
12:00	20	20	20	20	
12:10					20
12:30	20	20	20	20	20
13:00	20	20	20	20	20
13:30	20	20	20	20	20
14:00				20	
14:30	20	20	20		20
14:50		20			
15:00	20		20	20	
15:10		20			
16:00	20	20			

Figure A.1: Initial block schedule CT1, appointment slots in minutes

Average arrivals per hour	Between-group sum of squares	Within-group sum of squares	Mean square between	Mean square within	f statistic	p-value	Significant difference
0,740998839	1,328928641	161,5465023	0,33223216	0,807732511	0,411314582	0,800376263	No significance difference
0,939140534	6,055953764	177,7979094	1,513988441	0,888989547	1,703044143	0,150744156	No significance difference
1,169594981	11,18931416	181,3641686	2,797328539	0,906820843	3,084764273	0,017108937	Significant different
1,432280204	24,54139717	228,0025307	6,135349293	1,140012653	5,38182561	0,000389381	Significant different
0,362253194	5,917714728	61,18931475	1,479428682	0,305946574	4,835578526	0,000961158	Significant different
0,492104503	7,220126335	79,68339224	1,805031584	0,398416961	4,530508888	0,001591884	Significant different
2,626552091	67,79443107	410,3304578	16,94860777	2,051652289	8,26095526	3,47128E-06	Significant different
2,419101224	121,565605	357,9656552	30,39140125	1,789828276	16,98006544	5,28733E-12	Significant different
0,657875105	6,503213998	119,6055217	1,6258035	0,598027609	2,718609436	0,030900323	Significant different
0,237340302	2,123289641	40,83284195	0,53082241	0,20416421	2,599977787	0,037363251	Significant different

Figure A.2: Significant difference in number of arrivals per hour per day of the week

	Monday	08:00	09:00	10:00	11:00	12:00	13:00	14:00
Monday	Correlation coefficient	-0.0005	-0.1558	0.1136	0.1050	0.1934	0.1513	-0.0076
	t-statistic	0.0006	0.1978	0.1438	0.1328	0.2466	0.1920	0.0095
Tuesday	Correlation coefficient	0.0501	-0.1296	0.0410	-0.0903	0.0027	0.0228	0.0543
	t-statistic	0.0633	0.1642	0.0517	0.1141	0.0034	0.0287	0.0685
Wednesday	Correlation coefficient	0.0623	0.0736	0.0612	0.2568	0.1009	0.0099	-0.0929
	t-statistic	0.0787	0.0930	0.0773	0.3300	0.1276	0.0125	0.1174
Thursday	Correlation coefficient	-0.2304	0.2136	0.0681	0.1780	0.1321	0.0107	-0.1696
	t-statistic	0.2949	0.2729	0.0860	0.2265	0.1675	0.0135	0.2156
Friday	Correlation coefficient	0.2175	0.1148	0.1114	0.0969	0.0319	0.1176	0.0022
	t-statistic	0.2778	0.1453	0.1401	0.1226	0.0403	0.1489	0.0028

Table A.1: Correlation between the arrivals per hour

A.2 Appointment duration chi-squared tests

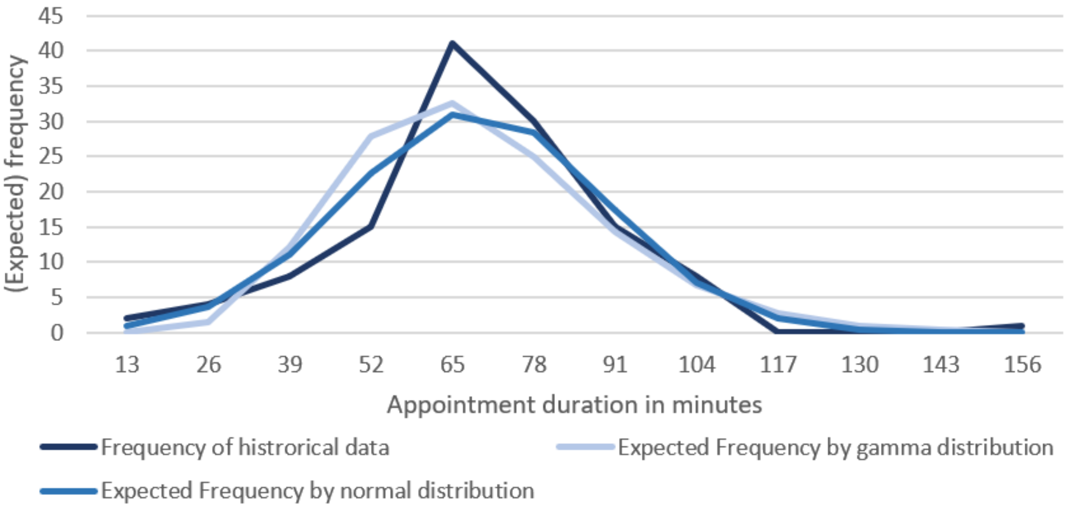

Figure A.3: Histogram distributions 60 minute appointment duration

BinNr	CumProb	Bins	Frequency	Expected frequency	Error
1	0,06	5,25	1159	824	136,2
2	0,12	6,7	558	824	85,9
3	0,18	7,8	654	824	35,1
4	0,24	8,8	713	824	15,0
5	0,29	9,8	741	824	8,4
6	0,35	10,7	817	824	0,1
7	0,41	11,6	827	824	0,0
8	0,47	12,5	864	824	1,9
9	0,53	13,5	881	824	3,9
10	0,59	14,5	798	824	0,8
11	0,65	15,6	747	824	7,2
12	0,71	16,9	675	824	26,9
13	0,76	18,3	1173	824	147,8
14	0,82	20,1	960	824	22,4
15	0,88	22,4	725	824	11,9
16	0,94	26,2	920	824	11,2
17	1,00	More	796	824	1,0

Figure A.4: Chi-squared tests Gamma distribution 20 minute appointment duration

BinNr	CumProb	Bins	Frequency	Expected frequency	Error
1	0,08	33,62	8	10,3	0,5
2	0,17	39,7	6	10,3	1,8
3	0,25	44,4	1	10,3	8,4
4	0,33	48,6	8	10,3	0,5
5	0,42	52,6	8	10,3	0,5
6	0,50	56,6	13	10,3	0,7
7	0,58	60,9	15	10,3	2,1
8	0,67	65,5	14	10,3	1,3
9	0,75	70,9	16	10,3	3,1
10	0,83	77,8	11	10,3	0,0
11	0,92	88,3	13	10,3	0,7
12	1,00	More	10	10,3	0,0

Figure A.5: Chi-squared tests Gamma distribution 60 minute appointment duration

BinNr	CumProb	Bins	Frequency	Expected frequency	Error
1	0,08	31,08	8	10,3	0,5
2	0,17	39,4	6	10,3	1,8
3	0,25	45,3	2	10,3	6,7
4	0,33	50,2	11	10,3	0,0
5	0,42	54,7	10	10,3	0,0
6	0,50	58,9	13	10,3	0,7
7	0,58	63,1	18	10,3	5,7
8	0,67	67,6	10	10,3	0,0
9	0,75	72,5	16	10,3	3,1
10	0,83	78,4	6	10,3	1,8
11	0,92	86,7	10	10,3	0,0
12	1,00	More	13	10,3	0,7

Figure A.6: Chi-squared tests Normal distribution 60 minute appointment duration

A.3 Verification and validation of the simulation model

	Input mean	Input sd	Simulation mean	Simulation sd	t critical value	Upper confidence interval	Lower confidence interval	Signifivant difference between the means
Walk-in arrivals Monday	35,30	7,24	35,30	5,88	2,02	-2,44	2,44	No
Walk-in arrivals Tuesday	30,50	6,62	30,90	5,88	2,02	-2,67	1,87	No
Walk-in arrivals Wednesday	36,40	8,21	36,80	6,57	2,02	-3,17	2,37	No
Walk-in arrivals Thursday	28,50	5,92	28,60	5,35	2,02	-2,14	1,94	No
Walk-in arrivals Friday	28,70	7,20	29,00	4,98	2,02	-2,67	2,07	No
Emergency arrivals Monday	6,50	2,76	6,40	2,28	2,02	-0,83	1,03	No
Emergency arrivals Tuesday	6,20	3,12	6,00	2,37	2,02	-0,84	1,24	No
Emergency arrivals Wednesday	6,10	2,91	5,80	2,53	2,02	-0,69	1,29	No
Emergency arrivals Thursday	5,40	2,75	5,30	2,18	2,02	-0,82	1,02	No
Emergency arrivals Friday	6,80	2,96	7,00	2,64	2,02	-1,22	0,82	No
Inpatient arrivals Monday	10,80	3,99	10,70	3,09	2,02	-1,24	1,44	No
Inpatient arrivals Tuesday	11,50	3,76	11,90	3,26	2,02	-1,68	0,88	No
Inpatient arrivals Wednesday	10,30	3,84	10,30	3,01	2,02	-1,29	1,29	No
Inpatient arrivals Thursday	10,70	4,59	10,60	3,28	2,02	-1,42	1,62	No
Inpatient arrivals Friday	10,40	3,62	10,00	3,13	2,02	-0,84	1,64	No
Appointment duration 20 min	13,24	7,48	13,12	9,30	1,96	-2,60	2,84	No
Percentage appointment type 1 emergency	0,07	0,13	0,07	0,11	1,98	-0,04	0,04	No
Percentage appointment type 1 walk-in	0,06	0,07	0,06	0,04	1,98	-0,02	0,02	No
Percentage appointment type 1 inpatient	0,08	0,12	0,08	0,09	1,98	-0,04	0,04	No
Percentage appointment type 1 outpatient	0,22	0,21	0,22	0,21	1,98	-0,07	0,07	No

Table A.2: Verification of means conceptual model

	Input mean	Input sd	Simulation mean	Simulation sd	t critical value	Upper confidence interval	Lower confidence interval	Signifivant difference between means
Percentage of Inpatients	0,15	0,06	0,15	0,04	2,02	-0,02	0,02	No
Percentage of Oupatients	0,30	0,30	0,31	0,04	2,02	-0,10	0,08	No
patients	0,09	0,09	0,09	0,03	2,02	-0,02	0,03	No
Percentage of Walk-in patients	0,46	0,08	0,46	0,06	2,02	-0,02	0,03	No
Average utilization	0,58	0,10	0,61	0,08	2,02	-0,06	0,00	No
Waiting time Inpatient	6,43	9,78	1,99	1,50	2,02	1,42	7,47	Yes
Waiting time Outpatient	5,23	10,15	2,11	1,47	2,02	-0,02	6,25	No
Waiting time Emergency patient	7,13	9,87	1,43	1,30	2,02	2,65	8,75	Yes
Waiting time Walk-in patient	19,17	15,27	6,10	5,78	2,02	8,27	17,87	Yes
Total number of arrivals per day (weekday)	68,94		69,32	7,92				

Table A.3: Validation mean simulation model

	Input mean	Input sd	Simulation mean	Simulation sd	Input df	Simulation df	Lower critical value	Uper critical value	Computed f statistic	Significant difference between variances
Walk-in arrivals Monday	35,30	7,24	35,30	5,88	42	142	0.591	1,58	1,52	No
Walk-in arrivals Tuesday	30,50	6,62	30,90	5,88	42	142	0.591	1,58	1,27	No
Walk-in arrivals Wednesday	36,40	8,21	36,80	6,57	42	142	0.591	1,58	1,56	No
Walk-in arrivals Thursday	28,50	5,92	28,60	5,35	42	142	0.591	1,58	1,22	No
Walk-in arrivals Friday	28,70	7,20	29,00	4,98	42	142	0.591	1,58	2,09	Yes
Emergency arrivals Monday	6,50	2,76	6,40	2,28	42	142	0.591	1,58	1,47	No
Emergency arrivals Tuesday	6,20	3,12	6,00	2,37	42	142	0.591	1,58	1,73	Yes
Emergency arrivals Wednesday	6,10	2,91	5,80	2,53	42	142	0.591	1,58	1,32	No
Emergency arrivals Thursday	5,40	2,75	5,30	2,18	42	142	0.591	1,58	1,59	No
Emergency arrivals Friday	6,80	2,96	7,00	2,64	42	142	0.591	1,58	1,26	No
Inpatient arrivals Monday	10,80	3,99	10,70	3,09	42	142	0.591	1,58	1,67	Yes
Inpatient arrivals Tuesday	11,50	3,76	11,90	3,26	42	142	0.591	1,58	1,33	No
Inpatient arrivals Wednesday	10,30	3,84	10,30	3,01	42	142	0.591	1,58	1,63	Yes
Inpatient arrivals Thursday	10,70	4,59	10,60	3,28	42	142	0.591	1,58	1,96	Yes
Inpatient arrivals Friday	10,40	3,62	10,00	3,13	42	142	0.591	1,58	1,34	No
Appointment duration 20 min	13,24	7,48	13,12	9,30	1000	1000	0,88	1,13	0,65	Yes
Percentage appointment type 1 emergency	0,07	0,13	0,07	0,11	209	999	0,80	1,23	1,45	Yes
Percentage appointment type 1 walk-in	0,06	0,07	0,06	0,04	209	999	0,80	1,23	2,74	Yes
Percentage appointment type 1 inpatient	0,08	0,12	0,08	0,09	209	999	0,80	1,23	1,67	Yes
Percentage appointment type 1 outpatient	0,22	0,21	0,22	0,21	209	999	0,80	1,23	1,03	No

Table A.4: Verification of variances conceptual model

	Input mean	Input sd	Simulation mean	Simulation sd	Input df	Simulation df	Lower critical value	Uper critical value	Computed f statistic	Significant difference between variances
Percentage of Inpatients	0,15	0,06	0,15	0,04	42,00	54,00	0,56	1,76	2,24	No
Percentage of Oupatients	0,30	0,30	0,31	0,04	42,00	54,00	0,56	1,76	48,02	No
patients	0,09	0,09	0,09	0,03	42,00	54,00	0,56	1,76	8,13	yes
Percentage of Walk-in patients	0,46	0,08	0,46	0,06	42,00	54,00	0,56	1,76	1,85	No
Average utilization	0,58	0,10	0,61	0,08	42,00	54,00	0,56	1,76	1,44	No
Waiting time Inpatient	6,43	9,78	1,99	1,50	42,00	54,00	0,56	1,76	42,79	No
Waiting time Outpatient	5,23	10,15	2,11	1,47	42,00	54,00	0,56	1,76	47,41	Yes
Waiting time Emergency patient	7,13	9,87	1,43	1,30	42,00	54,00	0,56	1,76	57,97	No
Waiting time Walk-in patient	19,17	15,27	6,10	5,78	42,00	54,00	0,56	1,76	6,98	yes
Total number of arrivals per day (weekday)	68,94		69,32	7,92						

Table A.5: Validation variance simulation model

A.4 Statistical tests experiments

BinNr	CumProb	Bins	Frequency	Expected frequency	Error
1	0,08	31,08	8	10,3	0,5
2	0,17	39,4	6	10,3	1,8
3	0,25	45,3	2	10,3	6,7
4	0,33	50,2	11	10,3	0,0
5	0,42	54,7	10	10,3	0,0
6	0,50	58,9	13	10,3	0,7
7	0,58	63,1	18	10,3	5,7
8	0,67	67,6	10	10,3	0,0
9	0,75	72,5	16	10,3	3,1
10	0,83	78,4	6	10,3	1,8
11	0,92	86,7	10	10,3	0,0
12	1,00	More	13	10,3	0,7

Table A.6: Statistical difference objective experiment 2 flexible or hybrid CT scanner

References

Bhattacharjee, P., & Ray, P. K. (2016). Simulation modelling and analysis of appointment system performance for multiple classes of patients in a hospital: A case study. *Operations Research for Health Care*, *8*, 71–84. https://doi.org/10.1016/j.orhc.2015.07.005

Bikker, I. A., Kortbeek, N., van Os, R. M., & Boucherie, R. J. (2015). Reducing access times for radiation treatment by aligning the doctor's schemes. *Operations Research for Health Care*, *7*, 111–121. https://doi.org/10.1016/j.orhc.2015.06.005

Cardoen, B., Demeulemeester, E., & Van Der Hoeven, J. (2010). On the use of planning models in the operating theatre: Results of a survey in Flanders. *International Journal of Health Planning and Management*, *25*(4), 400–414. https://doi.org/10.1002/hpm.1027

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *12*(4), 519–549.

Chowdhury, A., & Chandra Shil, N. (2022). Understanding change management in organizational context: revisiting literature. *Management and Entrepreneurship: Trends of Development*, *1*(19), 28–43. https://doi.org/10.26661/2522-1566/2022-1/19-03

Creemers, S., & Lambrecht, M. (2008). Healthcare queueing models. *Katholieke Universiteit Leuven, Facultiy of Business and Economics*, 1–48. https://www.cromso.com/PDF/KBI_0804.pdf

CZ. (2024). Wachttijden in de zorg. https://www.cz.nl/vergoedingen/wachttijden#:~:text=Diagnose%3A%20binnen%204%20weken%20na,de%20polikliniek%20is%206%20weken.

Deglise-Hawkinson, J., Helm, J., Huschka, T., Kaufman, D., & Oyen, M. V. (2018). A Capacity Allocation Planning Model for Integrated Care and Access Management. *Production and Operations Management*, *27*(12), 2270–2290. https://doi.org/10.1111/poms.12941

Ferrand, Y., Magazine, M., & Rao, U. (2010). Comparing two operating-room-allocation policies for elective and emergency surgeries. *Proceedings - Winter Simulation Conference*, 2364–2374. https://doi.org/10.1109/WSC.2010.5678933

Ferrand, Y. B., Magazine, M. J., & Rao, U. S. (2014). Partially Flexible Operating Rooms for Elective and Emergency Surgeries. *Decision Sciences*, *45*(5), 819–847. https://doi.org/10.1111/deci.12096

Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions (Institute of Industrial Engineers)*, *40*(9), 800–819. https://doi.org/10.1080/07408170802165880

Hans, E. W., Van Houdenhoven, M., & Hulshof, P. J. (2012). A framework for healthcare planning and control. *International Series in Operations Research and Management Science*, *168*, 303–320. https://doi.org/10.1007/978-1-4614-1734-7{_}12

Hans, E. W., & Vanberkel, P. T. (2011). Chapter 5 Operating Theatre Planning and Scheduling. (2011).

Harden, S. (2023). Turnaround times – what are we seeing?

Hassin, R., & Mendel, S. (2008). Scheduling Arrivals to Queues : A Single-Server Model with No-Shows. *54*(3), 565–572. https://doi.org/10.1287/mnsc.1070.0802

Heerkens, H., & Van Winden, A. (2017). *Solving managerial problems systematically.*

Huh, W. T., Liu, N., & Truong, V. A. (2013). Multiresource allocation scheduling in dynamic environments. *Manufacturing and Service Operations Management*, *15*(2), 280–291. https://doi.org/10.1287/msom.1120.0415

Hulshof, P. J., Kortbeek, N., Boucherie, R. J., Hans, E. W., & Bakker, P. J. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems*, *1*(2), 129–175. https://doi.org/10.1057/hs.2012.18

Interchange. (2021). Top 10 grootste ziekenhuizen van Nederland. https://www.inter-change.nl/nieuws/ziekenhuizen/top-10-grootste-ziekenhuizen-van-nederland/

Joustra, P. E., Kolfin, R., Van Dijk, N. M., Koning, C. C., & Bakker, P. J. (2012). Reduce fluctuations in capacity to improve the accessibility of radiotherapy treatment cost-effectively. *Flexible Services and Manufacturing Journal*, *24*(4), 448–464. https://doi.org/10.1007/s10696-011-9119-y

Kolisch, R., & Sickinger, S. (2007). Providing radiology health care services to stochastic demand of different customer classes, 375–395. https://doi.org/10.1007/s00291-007-0116-1

Kortbeek, N., Zonderland, M. E., Braaksma, A., & Hans, E. W. (2014). Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals. *Performance Evaluation*, *80*, 5–26. https://doi.org/10.1016/j.peva.2014.06.003

Law, A. M. (2013). *Simulation Modeling and Analysis, FIFTH EDITION*. www.averill-law.com

Ma, X., Sauré, A., Puterman, M. L., Taylor, M., & Tyldesley, S. (2016). Capacity planning and appointment scheduling for new patient oncology consults. *Health Care Management Science*, *19*(4), 347–361. https://doi.org/10.1007/s10729-015-9331-5

Reilly, T., Marathe, B., & Fries, B. (1978). A Delay-Scheduling Model for Patients Using a Walk-In Clinic. *2*(4), 303–313.

Robinson, S. (2014). *Simulation The Practice of Model Development and Use* (2nd editio). ISBN%20978%E2%80%931%E2%80%93137%E2%80%9332802%E2%80%931

Srinivas, S., & Ravindran, A. R. (2020). Designing schedule configuration of a hybrid appointment system for a two-stage outpatient clinic with multiple servers. *Health Care Management Science*, *23*(3), 360–386. https://doi.org/10.1007/s10729-019-09501-4

Tancrez, J. S., Roland, B., Cordier, J. P., & Riane, F. (2009). How stochasticity and emergencies disrupt the surgical schedule. *Studies in Computational Intelligence*, *189*, 221–239. https://doi.org/10.1007/978-3-642-00179-6{_}14

Theisens, H. (2021). *Climbing the mountain - Lean Green Belt*.

Thomas Schneider, A. J., Theresia van Essen, J., Carlier, M., & Hans, E. W. (2020). Scheduling surgery groups considering multiple downstream resources. *European Journal of Operational Research*, *282*(2), 741–752. https://doi.org/10.1016/j.ejor.2019.09.029

UMC Utrecht. (2024). Van apotheker tot zenuwarts. https://www.umcutrecht.nl/nl/van-apotheker-tot-zenuwarts

Van Riet, C., & Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, *7*, 52–69. https://doi.org/10.1016/j.orhc.2015.05.005

van Essen, J. T., Bosch, J. M., Hans, E. W., van Houdenhoven, M., & Hurink, J. L. (2014). Reducing the number of required beds by rearranging the OR-schedule. *OR Spectrum*, *36*(3), 585–605. https://doi.org/10.1007/s00291-013-0323-x

Vermeulen, I. B., Bohte, S. M., & Elkhuizen, S. G. (2009). Adaptive resource allocation for efficient patient scheduling §. https://doi.org/10.1016/j.artmed.2008.07.019

Vieira, B. (2020). *Logistical Optimization of Radiotherapy Treatments*. https://research.utwente.nl/en/publications/logistical-optimization-of-radiotherapy-treatments

Winston, W. (2003). Operations Research: Applications and Algorithms. *Technometrics*, *30*(3), 1059–1113. https://doi.org/10.1080/00401706.1988.10488426

Yan, C., Tang, J., & Jiang, B. (2014). Sequential Appointment Scheduling Considering Walk-In Patients. *2014*.

Zheng, C., Li, Z., Zhang, Z., Zhang, L., & Tang, Q. (2025). *Multi-objective restarted simulated annealing algorithm for assembly line balancing problem with collaborative robots considering ergonomics risks*. Springer US. https://doi.org/10.1007/s10696-025-09590-0

Zomer, V. (2022). Increasing the number of outpatients treated within the access time norm by implementing a blueprint schedule. *University of Twente*. https://essay.utwente.nl/93697/

Zonderland, M. E., Boer, F., Boucherie, R. J., De Roode, A., & Van Kleef, J. W. (2009). Redesign of a university hospital preanesthesia evaluation clinic using a queuing theory approach. *Anesthesia and Analgesia*, *109*(5), 1612–1621. https://doi.org/10.1213/ANE.0b013e3181b921e7

Zonderland, M. E., Boucherie, R. J., Hans, E. W., & Kortbeek, N. (2021). *Handboek of Healthcare Logistics* (Vol. 302). https://doi.org/10.1007/978-3-030-60212-3{_}2

Zorgkaart Nederland. (2024). Hoe worden de wachttijden berekend? https://www.zorgkaartnederland.nl/content/hoe-worden-de-wachttijden-berekend