Determining Operating Room Capacity at Medisch Spectrum Twente

Determining Operating Room Capacity at Medisch Spectrum Twente

Determining the needed operating room capacity at the ENT department of Medisch Spectrum Twente

Author
Name: Simon Berend Jan Goorden

Supervisors University of Twente
dr. Daniela Guericke
dr. Sebastian Rachuba

Supervisor Medisch Spectrum Twente
Lisanne Penterman M.Sc

Medisch Spectrum Twente
ENT Department
Adress: Koningsstraat 1, 7512 KZ Enschede
Phone: 053 4872000

University of Twente
Faculty of Behavioural, Management and Social Sciences (BMS)
Department of Industrial Engineering & Business Information Systems
PO Box 217, 7500 AE Enschede,
tel. +31(0)534899111

This report was written as a graduation assignment for the Industrial Engineering and Management Science educational program at the University of Twente.

Content

List of Figures

List of Abbreviations

MST	=	Medisch Spectrum Twente
OR	=	Operating Room
POR	=	Polyclinical Operating Room
ENT	=	Ear, Nose, Throat

Management Summary

The main purpose of this study is to determine the needed capacity for operating rooms and polyclinical operating rooms of the ENT department. Resource allocation plays a crucial role in optimizing healthcare utilization and patient flow. Inefficient scheduling can lead to unnecessary costs, low utilization rates, and overall longer completion times which can be considered unbeneficial. The main problem addressed in this study is that the required operating room capacity after reallocation from treatments at operating rooms to polyclinical operating rooms was unknown. The research aim is stated as:

"The aim of this research is to determine the needed capacity in hours for the operating rooms at the ENT department of Medisch Spectrum Twente"

Several methods are used to solve the problem. For the main capacity problems, queueing theory as well as two different heuristics are used. Those heuristics encompass the "Shortest-Processing-Time" and the "Longest-Processing-Time" algorithm. These heuristics make use of historical schedules to determine the needed capacity. To determine which historical period was the most suitable, deseasonalization of demand is used. Queuing theory is the last used method. This method, which is a more mathematical approach is applied but required assumptions which should be taken into account when using the results. To apply queueing theory, data analysis is used to determine the statistical distribution of the provided data.

Certain results from this thesis stand out significantly. As described in the time analysis, 37.5% of the total treatment time at the operating rooms in the current situation can be rescheduled to polyclinical operating rooms. When applying the right scheduling heuristics, the completion time can be decreased by 51.63%, while the average makespan can be decreased by 32.7%. The queueing theory chapter shows a possible operating room utilization decrease of 26.3 percentage points. This decrease causes a total utilization rate of 37.3% in a single polyclinical operating room. These findings imply that reallocating the eligible treatments to polyclinical operating rooms and the use of appropriate scheduling heuristics has the potential to improve the completion time, makespan, and operating room utilization. These findings highlight the benefits of optimizing resource allocation and scheduling.

Recommendations can be made based on these outcomes. The main recommendation is to reserve at least 37.3% of the total capacity of one polyclinical operating room for the reallocated treatment types and decrease the capacity of the treatments at the operating rooms with 37.5 %. Furthermore, the proposed heuristics for polyclinical treatments as well as operating room scheduling could be used to decrease the makespan and the completion time.

Chapter 1. Introduction

1.1. Introduction

This graduation assignment is executed at the ENT (Ear, Nose, Throat) department of Medisch Spectrum Twente (MST) in Enschede. The reason for starting this project includes capacity problems for operating rooms as well as financial benefits. At the moment, Medisch Spectrum Twente faces a high demand which forces it to treat patients at "OCON", "Equipe" or "Flexclinics" - clinics outside of MST used to treat patients - which is more expensive than treating patients inside of the hospital. This assignment gives insight into how much treatment time can be rescheduled from operating rooms to polyclinical operating rooms. This has prompted the need to assess the required capacity of the operating rooms. Next to this insight, this research also includes a description of the current capacity situation at MST.

1.1.2. Medisch Spectrum Twente

MST is a Dutch hospital located in Enschede, belonging to the largest non-academic hospitals in the Netherlands with a license to operate 1070 beds. The coverage area includes approximately 264,000 residents. MST was created in 1990 from the merger of several hospitals in the region of Twente, including "Ziekenzorg", "Sint Joseph Stadsmaten", "St. Bernardusziekenhuis", "Sint-Antoniusziekenhuis", and "Heil der Kranken". The hospital has over 4000 employees, including 240 specialists, and an annual budget of around €350 million. The facilities in Enschede were originally located on two sites, but a new hospital was built on Koningsplein in 2016, and the buildings at Ariënsplein were sold. The hospital has many specializations, including a solvent team, a trauma center, an intensive care unit, and a thoracic center. MST offers secondary as well as limited tertiary care. ENT is a surgical subspeciality within medicine that deals with the surgical and medical management of the condition of the head and neck. The ENT department of MST is concerned with the treatment of disorders of the throat, nose, and ears and more complex disorders such as balance disorders and voice problems. Furthermore, the ENT department at MST provides comprehensive diagnostic and treatment services for various sleep-related breathing disorders such as snoring and sleep apnea.

1.2. Problem Identification

1.2.1. Problem Context

Since the use of the new building in 2016, the fixed charges of MST turned out to be very high. This has mainly to do with the depreciation expenses of the new building and the newly bought technology and equipment for top clinical surgeries. To ward off financial problems, MST must ensure a proper yield and prevent unnecessary expenses at any cost. In the current situation, the ENT department treats patients in operating rooms which could also be treated in polyclinical operating rooms. Rescheduling those patients will decrease unnecessary expenses, and therefore reduce the costs which will contribute to preventing any financial problems. After a discussion with the management of MST's ENT department, the following managerial problem statement could be made:

"There is no insight into the needed capacity after rescheduling treatments from operating rooms to polyclinic rooms, we need to know how much capacity we need to reserve and what the planning looks like"

To be able to easily communicate with the management of the ENT department about the problems, and check if the view of the management on the situation corresponds with the view of the researcher, a problem cluster is made. A problem cluster is highly valuable for this since it depicts the causal relationships between the problems.

1.2.2. Problem Cluster

The problem cluster starts with two problems that are only causes and not consequences. These problems are “ZGT Oral surgery in MST since 2023” and “Needed ENT OR Capacity Unknown”. The first one refers to the given fact that from 2023 onwards, all oral operations that were formerly conducted in hospital ZGT Hengelo are conducted in MST which requires a part of the operating room capacity. The second problem is closely related to the managerial problem and describes that the needed operating room capacity of the ENT department is unknown after rescheduling an eligible amount of patients from operating rooms to polyclinical operating rooms.

The problem “Needed ENT OR Capacity Unknown” has multiple implications. The first implication is that there are too many patients scheduled in operating rooms that are also possible to be treated in cheaper polyclinical operating rooms. This leads to unnecessary use of equipment, which can be considered a waste of money, and “too much operating time scheduled” which causes insufficient operating room capacity available and is also influenced by “ZGT Oral surgery in MST since 2023”. Scheduling too much operating room time results in insufficient operating room capacity available which has several implications:

1. *Longer waitlists* cause patient health deterioration because of patient treatment delays as well as a probable stress increase for the medical staff because of a higher workload. Managing waiting rooms and waiting lists can be considered necessary but not valuable work which is therefore described as “personnel productivity loss”.
2. *Failing to meet production requirements* refers to not treating the required number of patients as set by the contracted insurance companies which results in less reimbursements and an equal amount of costs.
3. *Patients are forced to go to different hospitals* since there are no operating rooms available at Medisch Spectrum Twente.
4. *Insufficient operating time for “top clinical” treatments* –complex treatments which require relatively high operating time - treatments that are considered a priority for MST.

These implications all cause a lower yield which could cause “not being able to pay interest payments”, especially the interest payments for the new building – built in 2016- of MST, “lower operating result”, which is considered as the *action problem* and “failing to meet the financial requirements”, which consist of “Failing to meet the Debt Service Coverage Ratio (DSCR)”, “Failing to meet EBITDA requirements” and “Failing to meet the solvency ratio requirements”. These “requirements” are set by creditors of Medisch Spectrum Twente. Failing to meet the financial requirements will in any case result in intensified financial supervision of Medisch Spectrum Twente

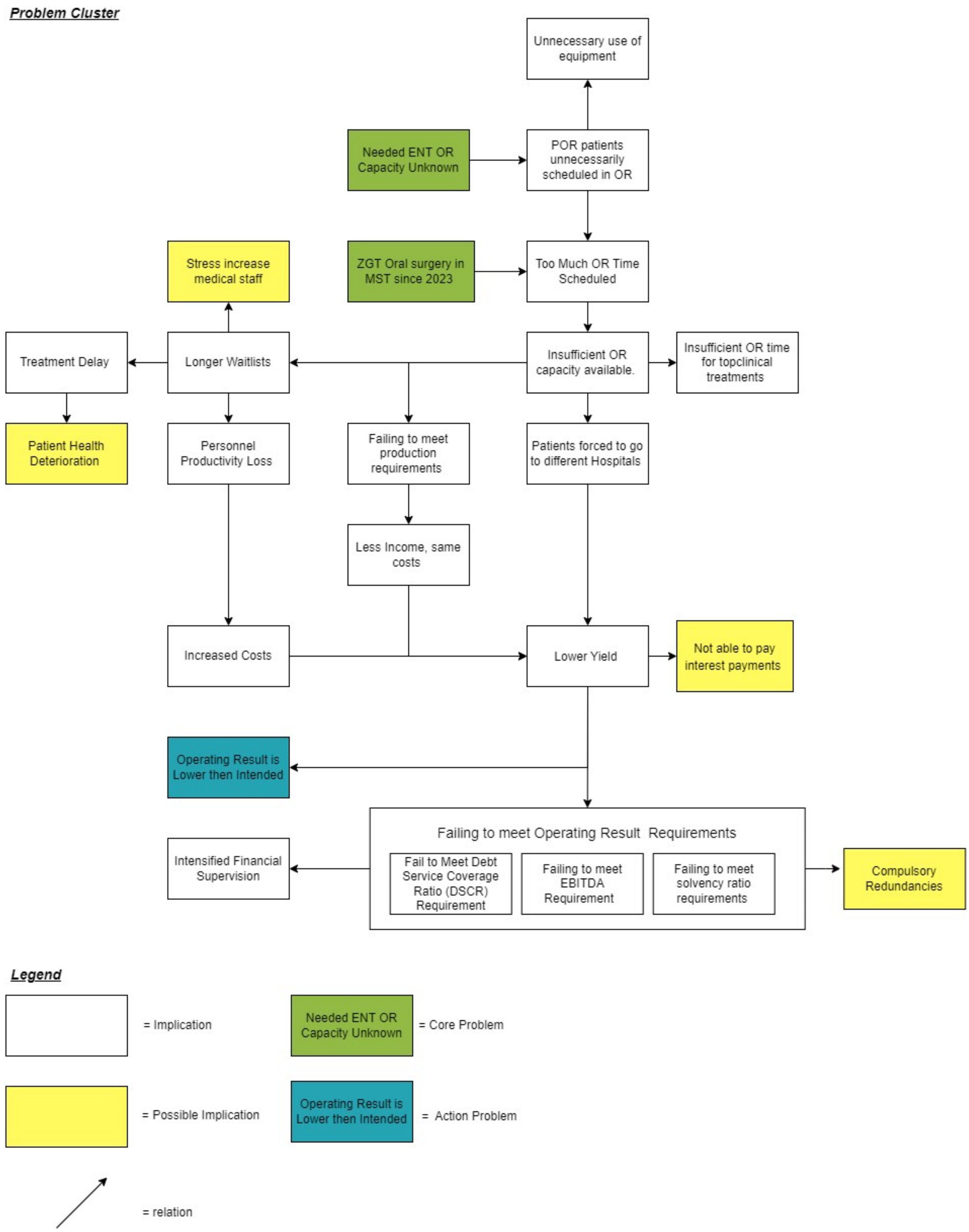

Figure 1: Problem Cluster

1.2.3. Core Problem

To find the core problem, *the five rules of Hans Heerkens* (Heerkens & Winden, 2017) are used. These rules consist of leaving out what is not known, considering causes irrelevant (pneumonia rule), going back in the causal chain to end up with problems that are only cause, and leaving out what cannot be influenced. If all these criteria are fulfilled, the "relevance rule" should be used, which chooses the most "relevant" problem as the core problem in order of money and effort. After going back in the causal chain, two problems "ZGT Oral surgery in MST since 2023" and "Needed ENT OR capacity Unknown" are left. As the first problem is an immutable fact beyond the control of this research, it cannot be considered a core problem. Therefore "Needed ENT OR Capacity Unknown" is the only potential core problem that makes applying the "relevance rule or "pneumonia rule" irrelevant. After identifying the core problem by the problem cluster, the core problem is clearly defined as:

"The needed capacity for operating rooms at the ENT department is unknown after reallocating certain types of treatments from operating rooms to polyclinic operating rooms."

In the current situation, there are types of treatments conducted in operating rooms at MST that can also be conducted in polyclinical operating rooms. These treatments encompass approximately 20% of the total amount of operating room treatments done at the ENT department. Reducing this amount of operating room treatments will decrease the number of treatments and therefore operating time at expensive operating rooms. Presently, a so-called "block plan" is made that schedules a certain amount of time for ENT treatments in an operating room around three months upfront. This time has a preset date, and the specific time of each treatment will be decided later on mainly dependent on the urge of the treatment. The main outcome of solving the core problem is a determination of how much time should be scheduled for the ENT department in those "block plans" after reallocating treatments which will likely result in less needed operating time at the ENT department.

The core problem will be *measured* by the used operating hours for ENT treatments. Since all operating rooms at Medisch Spectrum Twente are shared and no department is allowed – in practice – to operate in two operating rooms a the same time, the only option to measure the number of operating rooms is in hours. Since this research is concerned with determining capacity instead of improving a situation, there are no clear boundaries between the norm and reality. Therefore, the *norm* for this core problem is defined as the "use of fewer treatment hours in operating rooms for the same number of patients at the ENT department". The *reality* can be stated as "the used number of operating hours in the current situation".

1.2.4. Research Aim

After identifying the core problem and action problem as well as taking into account the managerial problem statement, the research aim can be identified. The research aim indicates the purpose of the project and contributes to solving the core problem. The research aim is defined as follows:

"The aim of this research is to determine the needed capacity in hours for the operating rooms at the ENT department of Medisch Spectrum Twente"

1.2.5. Limitations and Data Processing

This research is limited to the data and information provided by MST. Furthermore, the research is based on a single department and may therefore not apply to other departments or hospitals. Also, this research does not consider other factors that may affect the capacity of operating rooms nor does it take into account the preferences of patients and doctors for certain types of surgeries to be performed in operating rooms rather than polyclinical rooms.

1.4. Identification of Research Questions

To structure this research, a main research question, as well as several sub-research questions, are defined. Those "sub-research questions" are subdivided into smaller sub-questions as well. The main research question is defined as:

"How much operating room capacity is needed at the ENT department of Medisch Spectrum Twente after rescheduling the eligible treatments from operating rooms to polyclinical operating rooms?"

The first two sub-questions are used to get a better overview of the current scheduling and capacity situation at MST which is essential for this research and can be highly valuable for determining the challenges for implementing the solution later on. The third question – the literature study – is crucial for the mathematical modeling which will be done in question four. The research questions are described as follows:

1. What is the current scheduling and capacity situation at Medisch Spectrum Twente?
   - *Sub Question 1.1:* "What is the statistical distribution of treatment data of patients treated by the ENT department of Medisch Spectrum Twente?"
   - *Sub Question 1.2:* "What is the capacity of the operating rooms in the current situation?"
   - *Sub Question 1.3:* "How many patients are eligible to be rescheduled from operating rooms to polyclinic operating rooms?"

2. What does the scheduling process at Medisch Spectrum Twente look like?
   - *Sub Question 2.1*: "What does the current scheduling process look like at a polyclinical operating room?"
   - *Sub Question 2.2*: "What does the current scheduling process look like in an operating room?"

3. What are the appropriate models to apply for the capacity problem at MST?
   - *Sub Question 3.1*: "Which techniques are known in literature for determining operating room capacity?
   - *Sub Question 3.2*: "Which heuristics are known in literature for scheduling treatments at operating rooms?"
   - *Sub Question 3.3*: "How can the techniques found in literature be applied to the ENT department at Medisch Spectrum Twente?"

4. How can the capacity question at Medisch Spectrum Twente be modelled?
   - *Sub Question 4.1*: "What is the necessary input data for the mathematical model?"
   - *Sub Question 4.2*: "How can a mathematical model be formulated for the operating room capacity problem at MST?"
   - *Sub Question 4.3*: "What would a historical schedule look like when applying scheduling techniques?"

5. What are the results according to the mathematical model and literature study?
   - *Sub Question 5.1*: "How many operating room hours are needed according to the mathematical model?"
   - *Sub Question 5.2*: "How many operating room hours are needed according to the scheduling heuristics?"
   - *Sub Question 5.3*: "What are the expected challenges for implementing the solutions?"

1.5. Definition of Key Variables

Key variables can be divided into different types. The main categories are independent- and dependent variables. Furthermore, internal- and external variables can be distinguished. Internal variables can be influenced by MST, while external variables cannot be influenced by MST.

1.5.1. Independent variables

Independent variables are factors that influence dependent variables. Table 2 describes the most important independent variables. The first independent variable, personnel capacity, is highly relevant to this problem since decisions about rescheduling treatments from operating rooms to polyclinical operating rooms depend on the amount of available personnel. The second independent variable, the number of available polyclinical operating rooms, is crucial for the last sub-research question. The third independent variable provides insight into the amount of time an operating room at the ENT department is available. The first three independent variables are internal variables since personnel, number of available polyclinical operating rooms and operating rooms, as well as the utilization of (polyclinical) operating rooms can be partly influenced by MST, while reimbursements rates, patient demographics, flow patterns, and volumes cannot.

Independent variable	**Operationalization**
Personnel capacity at (polyclinical)operating rooms (internal)	Number of (polyclinical) operating room FTEs available
Number of (polyclinical)operating rooms available (internal)	The number of (polyclinic) operating rooms available is dependent on time and day.
Utilization of (polyclinical)operating rooms (internal)	Percentage of time that a (polyclinical)operating room is used.
Patient demographics (external)	List of (anonymized) patients with their needed treatment.
Patient flow patterns (external)	The number of patients dependent on time and day.
Reimbursement rates (external)	Amount of money MST gets for each type of treatment
Patient volumes (external)	The umber of patients that need to be treated dependent on time and day.

Table 2: Independent variables

1.5.2. Dependent variables

Dependent variables are always influenced by independent variables and can be considered a consequence of changes in the values of the independent variables. This research considers two main dependent variables that are concerned with the capacity of (polyclinical) operating rooms. The capacity of these rooms is dependent on a variety of factors, particularly on the independent variables described in the previous section. By analyzing the relationship between the independent and dependent variables, this research aims to identify the factors that have the greatest impact on the capacity of (polyclinical) operating rooms.

Dependent	**Operationalization**
Operating rooms Capacity at MST's ENT department.	Measured in hours
The capacity of MST's ENT department at the polyclinical rooms. (internal)	Measured in hours

Table 3: Dependent Variables

1.6. Research Design Validity and Reliability

As with any other research design, this research will likely face reliability and validity issues. Validity can be described as “the extent to which a test measures what we wish to measure” (Cooper & Schindler, 2011) while reliability can be described as “the accuracy and precision of a measurement procedure” (Cooper & Schindler, 2011) which asks the question if the same results can be achieved by using equal research methods with highly comparable circumstances.

1.6.1. Research Design Validity

Cooper & Schindler (2011) identify two types of validity. *External validity* can be stated as “data’s ability to be generalized across persons, settings, and times” and *internal validity* is “the ability of a research instrument to measure what it is purported to measure”.

For this research, numerous threats negatively influence internal validity. However, the most important internal validity threats can be described as *history* and *maturation*. *History* refers in to the situation where the study can be influenced by unforeseen occurrences which could change the conditions of the study and its outcome. In this case, that could be the change of capacity requirements such as changes in the required staff members or the required equipment. *Maturation* refers to the change of dependent variables as time passes by. This could be the change of the capacity requirements over time because of for example a change in patient demographics. For example, history and maturation are hard to avoid. Changes in capacity requirements are primarily determined by external factors which are therefore beyond the hospital’s control and hard to prevent.

The most important external validity threat for this research can be described as sampling bias. Sampling bias refers to the situation where te research participants significantly differ from the real population. In the case of this research, it would refer to data provided MST which is not representative of the “real-life situation”. During the COVID-19 pandemic – as well as a short period after - the number of treatments in MST was lower than during the years before. Therefore, the provided data could differ from the “real-life situation”. The validity issues are hard to prevent. Data use is necessary and all available data includes influences from COVID-19.

1.6.2. Research Design Reliability

For this research, generalizability should be taken into account as part of the reliability. Generalizability refers to the “extent to which the findings of a study can be applied or extended to other settings, populations, or samples beyond the specific conditions of the study” (Kakull & Ganguli, 2012). Since this research is very specific, this generalizability might be a problem. However, there are certain parts that could be used for other studies. The used methods to solve the problem are not specific to the ENT department but could be applied to other departments as well with minor changes. Furthermore, the literature study, as well as the problem-solving approach, are not specific to this assignment at MST. To validate this research, *triangulation* as well as regular meetings with MST supervisors will be used. *Triangulation* consists of “involving more than one source of data to confirm the validity and authenticity of the data, analysis and interpretation” (Saunders, 2019).

Chapter 2. Current Situation Description of the ENT department

In the description of the current situation, the provided data and information is analyzed. The analysis is mainly focused on a statistical test of the treating times at the ENT department and the determination of several descriptive statistics. To do this analysis, decisions had to be made about the eligible treatments for reallocation from operating rooms to polyclinical operating rooms. The description of the current situation is mainly based on provided data by MST in appendixes two and three, and discussion with medical personnel from the ENT department as well as the capacity department.

2.1. Scheduling process overview

To get a better overview of the decision made with regard to the scheduling process, a process model is made. The process model considers the scheduling process of polyclinical treatments as well as treatments in operating rooms.

Before patients undergo treatment at Medisch Spectrum Twente, they usually visit the common practitioner. If the common practitioner considers a specialist treatment necessary, patients are referred to a specialist. This specialist decides if a patient needs a treatment and if so whether the treatment needs to be polyclinical or executed in an operating room. If no treatment is needed, the patient leaves the process. In the case of operating room treatments, the patient is placed on the waiting list first. The treatment time is based on four different factors. Urgency, waiting time, available specialists, and the number of instruments. The urgency of the treatments is separated into different levels. The higher the urgency, the earlier the patient should be scheduled. The patient with the highest level of urgency is scheduled first. If the decision about urgency is made, the waiting time is taken into account. In this case, the patient with the longest waiting time is scheduled first. If those decisions are made, the availability of the required specialists as well as instruments are checked. If both of these are available at the desired time, the patient can be scheduled.

Next to the availability of specialists, instruments, waiting time, and patient urgency, the time of scheduling is determined by the “session day”. The ENT department features a certain amount of session days which is based on historical data and insurance company regulations. Those two aspects are taken into account by the “Marketing & Sales department” which calculates how many session days need to be planned. These calculations are influenced by the expected “cases time” of the treatments, which encompasses the total treatment time. If the sessions as well as the desired time are known, the patient can be treated. The scheduling process works differently in the case of a polyclinical operating room treatment. If the specialist decides on a polyclinical treatment, the scheduling process is less complicated. The ENT specialist schedules a treatment in the consulting room, right after the appointment. When making the treatment, three aspects have to be taken into account. The availability of the room, instruments, and patient at the desired time and date. If the time and date is considered as “suitable”, the treatment is scheduled. If there is no recovery treatment needed, the patient leaves the system immediately.

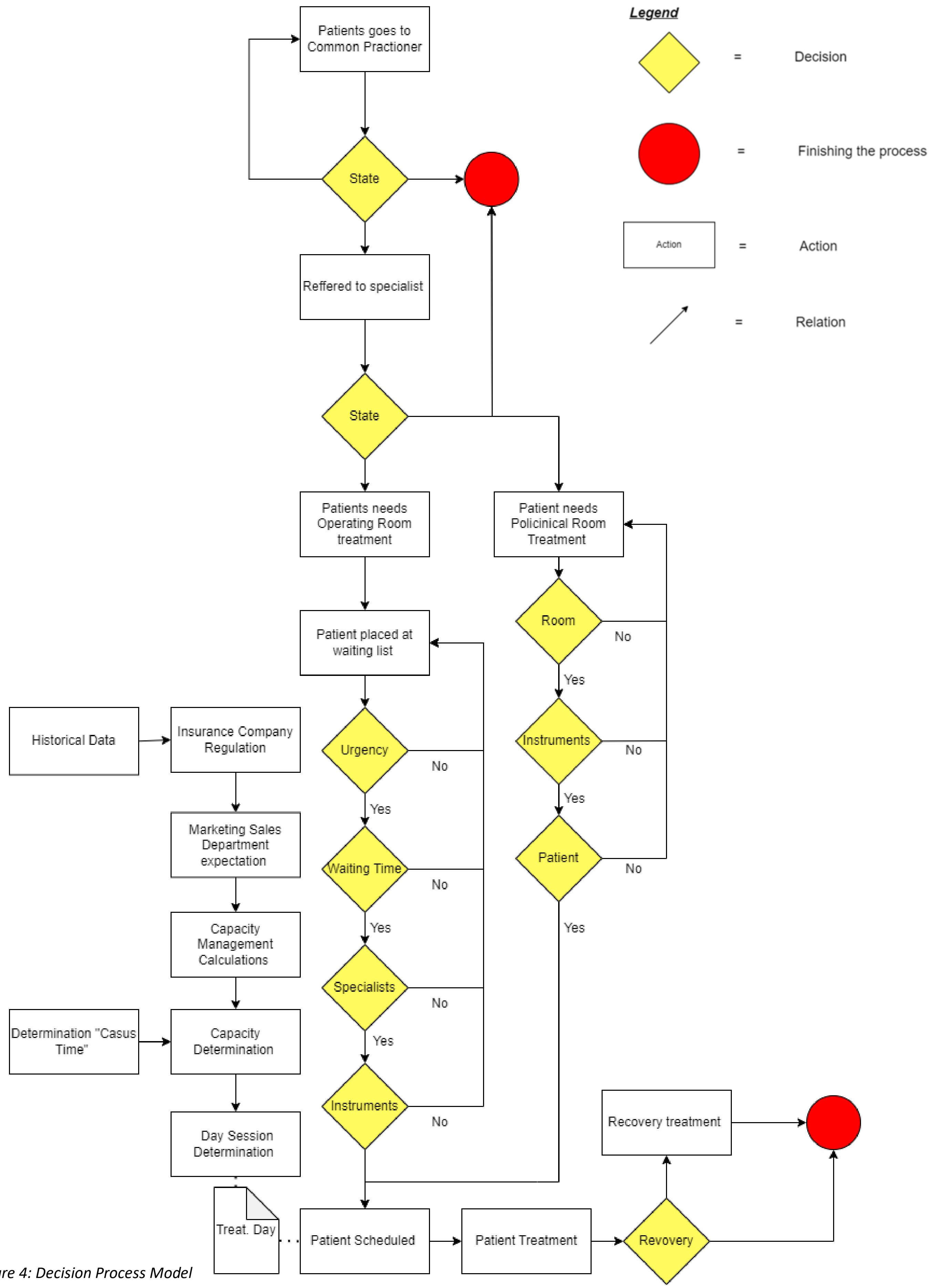

Figure 4: Decision Process Model

Based on the decision process model in Figure 8, several conclusions can be drawn. First of all, if operating room treatment is required, the patient is placed on a waiting list in case of a non-emergency patient. The scheduling of the treatment takes into account four factors. The urgency, the waiting time, the availability of specialists, and the availability of instruments. A combination of waiting time and urgency is used to determine which patient has to be scheduled first. The scheduling process also takes into account the "session days" available for the ENT department. In case of the polyclinical treatments, the process is less complicated. The ENT specialist can schedule the treatment directly after the appointment in the consulting room. Availability of the room, instruments and the patient at the desired time and date are considered before scheduling.

2.2. Data Analysis

The ENT department is concerned with 108 different types of treatments in operating rooms. Those different "types" consist of a distinction between different operations as well as sorts of patients. Next to "regular" patients that are planned to be treated according to the waiting list, the ENT department mainly discusses two types of urgencies. These are "emergency" and "non-emergency". The total number of executed operations in the studied period is 1692.

All operations at the ENT department are executed in three different operating rooms, called "OK01", "OK02" and "OK04". However, in all cases, there is only one operating room used at the same time. During this period, a total of 1692 treatments were executed. From this amount, 45 were executed in December 2021, 1553 in 2022, and 94 in January 2023. Finally, the data gives a clear seasonality indication for December, July, and August. This can only be considered as an indication of the lack of long-time data. However, as can globally be seen in Figure 4, the number of treatments during these months was less than during other months. In 2022, in July, August, and December the number of treatments was 113, 99, and 101. This is less than other months, which have the lowest value of 134.

2.2.1. Capacity

MST provides clear statistics about the current situation at the ENT department. During the period 2022 – 2023, there were 608 operations executed at the ENT department. The goal was 564. The operating room occupation was 80.4% which is 4.59 less than the goal of 85%. The goal is stated as 85% because of the "change time" that is needed in between the operations. This "change time" is mainly used for cleaning the operating room and the arrival as well as leaving of patients. This "change time" encompasses – highly dependent on the length of operations – around 15% of the total session time. On average, operations at the ENT department need 8 minutes of change time.

The waiting list includes – at the time of writing – 329 patients which translates to a "working load" of 486 planned operating room hours. At the ENT department, 63.5% of operating room treatments were treated in time which is 16.49% less than intended. Furthermore, during the years 2023 and 2022, a total of 386 operations were executed in operating rooms. The average waiting time was 63.06 days. During 2022 and 2023, there were 35.0% more executed operations than planned. Unfortunately, the ENT department suffers from cancellations. From January 2022 to May 2023, a total of 359 operations were canceled less than 24 hours before the operations which equals 17.2%. Next to the cancellations, the ENT department executes a percentage of the treatments outside of the assigned "session". In 2022, a total of 261.837 minutes were scheduled outside the session which equals 14.1%. Since treatments outside of the assigned session are "overtime", they can be considered non-beneficial.

All operating rooms at Medisch Spectrum Twente are shared between the different departments. In total, 15 operating rooms could be scheduled 8 hours a day with an occupation of 85%. In 2022, the ENT department made treatments for 1553 patients in a total of 1189 hours. In the most compact planning, 149 sessions of 8 hours could be used. However, in 2022, 241 sessions were used to fulfill the demand. Since the capacity of operating rooms is shared among the departments, it is hard to state the "maximum" capacity, since this is highly dependent on the demand for other medical treatments. However, in general, there is only one operating room used for ENT treatments. Therefore, the maximum capacity can be considered as 261 sessions of 8 hours - operations at operating rooms are not executed during the weekends - which can be used for 85%. This means that the "effective hours" can be considered as $261 * 8 * 0{,}85 = 1.774{,}8$. Since the average length of operations in 2022 was 45.95 minutes, the maximum capacity can be considered as $\frac{1{,}774.8 * 60}{45} = 2365.3$ operations a year. Figures 7 and 8 show the number of treated patients in 2022 at polyclinical operating rooms and operating rooms with the week at the x-axis and the number of treatments at the y-axis

[redacted]

2.2.2. Patients that can be reallocated

A significant amount of patients can be rescheduled from operating rooms to polyclinical operating rooms. As of 2023, the ENT department has 109 different types of treatments. From these treatments, 39 treatment types can be rescheduled from operating rooms to polyclinical operating rooms. Those 39 different types of treatments account for 1015 treatments while the other 70 count for 493 treatments. Therefore, 67.3% of the treatments can be rescheduled. However, the treating time of the treatments that are expected to be rescheduled is significantly higher for the treatments that are expected to stay in the operating rooms. The total treatment times of treatments that are expected to be rescheduled in 2022 was 29.210 minutes, which counts for 67,3% of treatments, while the other 32.7 % took 48.98 minutes to complete. So, while 67,3% of the types can be rescheduled, this counts for only 37,4 % of the total treatment time. In Figures 5 and 6, the situation after the reallocation of treatments to polyclinical operating rooms based on the information in Appendix 3 is shown. The vertical ax represents the number of treatments in 2022, while the horizontal axis shows the week number.

2.3. Data Analysis of Treatment Data

To solve the stated assignment as described in the first chapter, Medisch Spectrum Twente provided data about the executed operating room operations at the ENT department. The data consists of execution time, date, type of treatment, emergency situations, and the used operating room from December 2021 to January 2023. Since the provided information of Medisch Spectrum Twente includes a lot of information that could be valuable in the process of mathematical modeling, the main parts of the data is analyzed. Several types of data are analyzed. To obtain a clear overview of the operating time data, the range, minimum, maximum, mean, standard deviation, variance, skewness (a measure of the asymmetry of the distribution of values in a dataset), and kurtosis (a measure of the shape of the distribution of values in a dataset) are determined for each of for three different datasets. The first dataset includes all patients treated at the operating room before reallocation. The second dataset includes the patients that will be treated at a polyclinical operating room after reallocation while the third dataset includes the treatments that will remain in the operating rooms after reallocation.

[redacted]

As can be seen in Table 7, there are differences in the descriptive statistics between the three different datasets. First of all, there is a big difference in the Kurtosis. The kurtosis is the peak sharpness of a frequency distribution curve. A high kurtosis implies that there are more extreme values present in the dataset, resulting in heavier and broader tails in the distribution. In this case, the treatment time of the polyclinical operating room treatments has a relatively high kurtosis, which indicates a relatively big amount of observations far-off the mean value. Secondly, the skewness is important to take into account. The skewness provides information about the shape and symmetry of the data distribution in each dataset. The skewness of the first dataset indicates a moderately positive skewness, where the distribution has a longer right tail and therefore some outliers of the treating time. A skewness of 7.09 indicates a significant positive skewness. This dataset is highly right-skewed, suggesting a pronounced concentration of data on the left side and a long tail on the right side with relatively many outliers of the treating time. The skewness of 1.424 indicates a moderate positive skewness with a small longer right tail suggesting the presence of some outliers.

When analyzing the variance, a few aspects can be considered as important. The datasets that consider all treatment before allocation exhibits a relatively high variance of 4243.6. The range of treatment times is likely to be wide with some patients requiring significantly more or less time treatment compared to the average. In the case of the reallocated treatments at the polyclinical room, there is a significantly less spread of treating times. This could have multiple reasons, but it suggests that the patients in this dataset might have more similar conditions or undergo a more standardized treatment process. The dataset of the operating room treatments after reallocation exhibits the highest variance of 4877, similar to the first dataset.

2.4 Used Techniques to Determine the Capacity

The capacity department at Medisch Spectrum Twente has recently been established. As a result, the number of techniques used to determine the required capacity is limited. In 2023, all decisions on the scheduling days were based on historical data combined with the current waiting list. The marketing & sales department of Medisch Spectrum Twente decides how many treatments are intended to take place during the next calendar year. This depends on three main factors. The forecasts of the treatments are solely based on historical data, the agreements regarding the extent of financial concerns, negotiations with insurance companies, and meetings with medical teams. Hospitals in the Netherlands have a duty of care which means they cannot reject patients because of a lack of insurance. Therefore, Medisch Spectrum Twente makes agreements with all of the medical insurance companies in the Netherlands. After negotiation with the marketing and sales department, a certain amount of treatments, which encompasses all needed actions, are “purchased”. The amount of treatments that are “purchased” depends on the treatments that were done last year, the amount of money the insurance companies offer for a certain type of treatment, and the desired focus of Medisch Spectrum Twente. The desired focus depends on the main strategy of Medisch Spectrum Twente

Chapter 3. Literature Study

This chapter presents the methodology and findings of a comprehensive review of capacity determination techniques and scheduling approaches in healthcare. The literature study followed a systematic order of steps, starting with the definition of a knowledge problem related to capacity planning decisions in healthcare whereafter a literature search was conducted. This chapter highlights the suitability of different mathematical techniques for capacity planning in healthcare as well as scheduling heuristics. The results serve as a foundation for solving the research questions as discussed in chapters five and six.

3.1. Methodology

The execution of the literature study followed a certain order of steps. For the "identification of capacity determination techniques" part, a knowledge problem was defined first. The knowledge problem, which is defined as "which mathematical techniques are suitable for making capacity planning decisions in healthcare" is also one of the research questions of this thesis. To answer this question, a literature search turned out to be needed. This literature search started with defining inclusion and exclusion criteria. The inclusion criteria include esteemed features of the academic articles such as peer reviews, and the used language. The exclusion criteria encompass features that are undesired such as articles that were published a very long time ago, or articles that do not address the topic of capacity planning. The inclusion and exclusion criteria can be found in Appendix 4. Thereafter, the most relevant academic databases are identified. In this case, two general databases are used as well as two specific databases specialized in Industrial Engineering and Operations Research. As general databases, "Scopus" and "Web of Science" are used. Both Scopus and Web of Science include a "citation analysis" option, which can be useful to identify the most cited articles in the area of capacity planning. As specific databases, PUbsONline and IEE Xplore are used.

To determine relevant search terms which could be used to make searches in those databases, "key concepts", "synonyms", "broader terms" and "narrow terms" were identified. After the searches, a search log and a conceptual matrix were made. This conceptual matrix provided a clear overview of the searched literature and the topics which were addressed. Both the conceptual matrix as well as the search matrix can be found in Appendix 4.

3.2. Identification of Capacity Determination Techniques

Several methods are used to solve capacity-related issues in healthcare. Queueing Theory, and Linear Programming are considered suitable methods.

Based on the executed literature study, it is evident that *linear programming* is a widely used approach in the literature relevant to this research. Notably, several studies, such as those conducted by, Sitepu (2018), and Shafaei (2018), have successfully applied linear programming in areas closely related to this research. The study of Sitepu (2018) used linear programming to develop a mathematical optimization model for capacity optimization, which is directly relevant to this research. Their study produced satisfactory results, indicating the effectiveness of linear programming in solving similar problems. Additionally, the study of Shafaei (2018) used linear programming to optimize the planning and scheduling of OR, resulting in valuable insights, despite requiring robust estimations. Given the high degree of relevance and success of these studies, the probability to solve linear programming with computer programming, as well it is reasonable to consider linear programming as a potentially effective approach to solving this research problem.

Although linear programming has advantages, some limitations have to be taken into account. First of all, linear programming relies on linear equations. However, it is not likely that all variables in MST have a linear relationship. Furthermore, using linear programming requires modeling the problem into a model that minimizes costs or maximizes profits. This means that determining and assigning costs to every single treatment is required. However, this approach is not used in MST, which makes the application of linear programming hard.

Queueing theory can be used "to analyze and understand the impact of capacity constraints on system performance" (Hall *et al,* 2012). Queueing models come in two parts, analytical and simulation. While analytical models are "typically represented by formulae" (Hall *et al,* 2012), simulation models "often require specialist software" (Hall *et al,* 2012). To make queuing models easy to work with, "almost all queueing models assume that arrivals occur "at random" or equivalently as a poison process" (Hall *et al,* 2012). This is not the case with appointment-driven scheduling like at the ENT department. However, there are certain ways to make queueing models for appointment-driven systems as described in "Queueing models for appointment-driven systems" by Creemers *et al* (2012). The "ADQ model" provides a clear solution to determine the idle time – also called vacation time – of a system as well as the number of services in a situation where the appointments are determined upfront. Because the application of an ADQ model could turn out as too advanced for this research, looking into M/G/C queues could be a suitable option as well. The use of M/G/ C queues for solving this problem is explained in chapter five. However, assumptions will be necessary. The main assumption, in this case, would be the assumption of Poisson arrivals. Furthermore, several studies were found where queuing theory was applied successfully in capacity planning which gives a good indication that queueing theory in capacity planning is a good method to use. Furthermore, queueing models can be solved analytically as well as computationally which can be considered a clear advantage over other methods.

With this information, the knowledge question is - Which mathematical techniques are suitable for making capacity planning decisions in healthcare? - can be answered. There are several methods known in the literature for making capacity planning decisions in healthcare. There are likely methods known in the literature which are not discussed in this literature review. However, queuing theory, and linear programming are both suitable methods with their advantages and disadvantages. Because of the number of studies that apply queueing theory in capacity planning, the possibility to use the "ADQ model" and the wide variety to solve queueing models, queueing theory is considered the most suitable for this project.

3.3. Identification of Scheduling Techniques

In parallel with queueing theory as capacity determination technique, scheduling The scheduling approach heavily depends on the intended goal. In this context, there are three main goals: minimizing project completion time (minimum completion time), reducing late jobs (minimal lateness), and minimizing total processing time (minimum makespan). Each goal corresponds to a suitable algorithm that is used to optimize scheduling.

3.3.1. Shortest Processing Time

Minimizing the sum of completion times can mainly be done in two ways. Applying the shortest shortest processing time (SPT) to minimize the average completion time (Bobelin, 2016) $\sum_{j=1}^{n} C_j$ or the Weighted Shortest Processing time first rule (WSPT) to minimize $\sum_{j=1}^{n} w_j C_j$ with c as completion time, j as the treatment, n as the total number of treatments, and w as the weight of a specific task. In case of the shortest processing time rule, the scheduling of jobs is based on arranging them in increasing order of their processing time p_j. The weighted shortest processing time minimizes the weighted sum of completion times, there the scheduling of jobs is based on arranging the min increasing other of their ratio $\frac{p_j}{w_j}$ with w_j as weight. The application of this rule involves, next to arranging the to-be-scheduled jobs in increasing order, scheduling as soon as possible when a machine becomes available. The job with the shortest processing time will begin the execution. In a scenario where there is a single machine and all jobs have a ready time of 0, this algorithm proves to minimize the number of jobs in the system, minimizing the average waiting time of jobs from their arrival until processing begins, minimizing the maximum waiting time, and minimizing the average lateness.

3.3.2. Minimizing Lateness

Hodgson-Moore. The Hodgson-Moore algorithm is proven to minimize the number of late jobs when scheduling them on a single machine. "In this algorithm, each job j has a non-negative processing time p_j u and a non-negative due date d_j" (Cheriyan, 2021). The Hodgson-Moore algorithm can be described as follows:

- Step 1: Sort jobs in order of increasing due date d_j;
- Step 2: Start with the scheduled job set $J_0 = \emptyset$ load $\lambda = 0$
- Step 3: For $j = 1, \dots n$, if $\lambda + p_j \leq d_j$, then $J_j = J_{j-1} \cup \{j\}$; $\lambda = \lambda + p_j$; otherwise, let $j_{max} \in J_{j-1} \cup \{j\}$ have the largest processing time; set $J_j = J_{j-1} \cup \frac{\{j\}}{\{j_{max}]}$; $\lambda = \lambda + p_j - p_{jmax}$.

The first step orders the jobs in order of the increased due date, while the second step initializes the schedule set $J)$ to be an empty set and sets the current load λ to 0. The third step encompasses checking for each job if adding the processing time of the job to the current load λ would result in a load that is less than or equal to the job's due date $\lambda + \ p_j \leq d_j$. If that condition is satisfied, the job is scheduled by adding it to the scheduled job J_j and update the current load to $\lambda + p_j$. If the above condition is not satisfied, the job j_{max} has the largest processing time among the jobs in the scheduled job set $J_j = J_{j-1} \cup \{j\}$. Remove j_{max} from the scheduled job set $J_{j-1} \cup \{j\}$. The main idea of this algorithm is to schedule jobs based on their due date, while also taking into account their processing time. If adding a job to the schedule causes the load to exceed its due date, the algorithm removes the job with the largest processing time and replaces it with the new job, as long as this new job can be completed within its due date.

3.3.3. Minimizing the Makespan

One effective approach to minimize the makespan is by formulating it as a mixed-integer linear programming model and employing the longest processing time (LPT) algorithm, which has demonstrated its ability to minimize the makespan (Darvish, 2012). Taking into account $C_{\max}$ as the makespan or maximum completion time, N as the number of jobs as an integer, m as the number of machines and p_j as the processing time of job j as integer, with i ranging from 1 to m and j ranging from 1 to N. The mathematical integer programming model is described as follows:

$$\min C_{\max}$$

Subject to:

$$\sum_{j=1}^{N} p_j x_{ij} \leq C_{max} \qquad i = 1, \dots, m$$

$$\sum_{i=1}^{m} x_{ij} = 1 \qquad j = 1, \dots, N$$

$$x_{ij} \in \{0,1\} \qquad i = 1, \dots, m \; j = 1, \dots, N$$

$$C_{max} \geq 0.$$

In this model, the objective function $C_{\max}$ needs to be minimized. The second equation ensures that the load on any machine is equal to or less than $C_{\max}$. The third constraint shows that each job must be assigned to exactly one machine. The fourth constraint describes the decision as binary variables, with x_{ij} taking a value of 1 if job j is assigned to machine i and 0 otherwise. $C_{\max}$ represents the makespan and is constrained to be greater than or equal to zero.

The LPT algorithm plays a vital role in minimizing the makespan by prioritizing smaller jobs towards the end of the schedule, thus facilitating the balancing of machine loads. According to the LPT algorithm, when one of the machines becomes available, the longest job among the remaining jobs N is selected for processing. The next job j, is then scheduled on machine i using the equation $i^* = \operatorname{argmin}\{L_i + p_j : i = 1, \dots, m\}$. The makespan $C_{\max}$ of any feasible solution is determined as: $C_{max} = \max\{\, C_i : i = 1 \,, \dots . m\}$. To apply the longest processing time algorithm, the total number of jobs N should be ordered in non-increasing order of processing time.

The steps of the LPT algorithm follow a specific order. The first step involves sorting N jobs according to the non-increasing order of their processing time. The second step is to select the first job on the ordered list and set this as 1. The last step includes assigning the job j to machine i according to the equation $i^* = \operatorname{argmin}\{L_i + p_j : i = 1, \dots, m\}$ By following these steps, the LPT algorithm systematically assigns jobs to machines while optimizing the makespan.

3.4. Conclusion

In conclusion, the literature study provided suitable capacity determination techniques as well as scheduling techniques. After an extensive analysis of the literature, queueing theory turned out to be the best method for capacity determination. In queueing theory, ADQ models or Poisson arrivals will be used. Linear programming is also addressed, but is not considered as suitable because of the possible lack of linear relationships and the needed allocation of costs to each task. In case of the scheduling techniques, three suitable techniques were identified. The shortest processing time rule, which minimizes the completion time, the longest processing time rule which minimizes the makespan, and the Hodgson-Moore algorithm which minimizes lateness. Since minimizing the completion time and makespan is more applicable in an appointment-driven system, the first two scheduling techniques are used.

Chapter 4. Rescheduling with Heuristics

Efficient scheduling and allocation of treatments in the operating room are crucial for maximizing resources, minimizing patient waiting times, and improving overall healthcare delivery. To optimize the utilization of operating room time, various scheduling rules and strategies have been developed and implemented. Two commonly used scheduling techniques are used. This encompasses the longest processing time rule which prioritizes longer treatments by assigning them earlier in the schedule, which helps to minimize the makespan. The other used scheduling technique is the earliest processing time rule, which prioritizes treatments with shorter treating times to minimize the total completion time. Both techniques will be applied to the situation after reallocation of the treatments to polyclinical operating rooms, from which there are three to take into account. To decide on the period of application of these heuristics, a seasonality analysis was made. The impact of the use will be analyzed which could assist in making informed resource allocation decisions.

4.1. Seasonality Analysis

Two heuristics combined with the necessary input data are used to determine the needed capacity of the operating rooms. The analysis by heuristics makes use of historical data only, which is useful, but has some disadvantages. The use of history is often limited by the available sample size and has a risk of seasonality. Seasonality can clearly be noticed at the ENT department of MST. The number of operations mainly depends on the vacation time of personnel as well as patients. Because the application of heuristics is time-intensive, only a part of the provided sample size can be used. To prevent seasonality, demand is deseasonalized. Deseasonalized demand represents the "demand that would have been observed in the absence of seasonal fluctuations" (Chopra, 2019). When taking into account p for the periodicity, the "number of periods after which the seasonal cycle repeats" (Chopra, 2019), the deseasonalized demand $\overline{D_t}$ for period t can be calculated with:

$$\overline{D_t} = \left[D_{t-\frac{p}{2}} + D_{t+\frac{p}{2}} + \sum_{i=t+1-\frac{p}{2}}^{t-1+\frac{p}{2}} 2D_i \right] * \frac{1}{2}p \tag{6.11}$$

To experience minimal consequences of the seasonality, the period to apply the heuristics on is chosen as the period that is the closest to its own deseasonalized demand. The analyzed data encompasses the operating room treatment data of 2022 subdivided in twelve months. The application of equation 6.11 is not suitable for providing solutions for periods one, two, three and twelve since these calculations require unknown demand values. Since there is a linear relationship between deseasonalized demand $\overline{D_t}$ and time t, the following formula is used to determine the values for the remaining periods:

$$\overline{D_t} = L + Tt \tag{6.12}$$

In this case, L represents the level of deseasonalized demand at Period 0, and T represents the rate of growth of deseasonalized demand which is often referred to as a trend. Using linear regression in Microsoft Excel the L 139.92 is set as And T is set as -1.36. The results are shown in the table.

[redacted]

As can be derived from the table, the percentual difference between the actual demand and the deseasonalized demand is the lowest in May. Next to the deseasonalization, the lack of COVID-19 regulations in May 2022 will result in values that are more comparable to a situation without COVID-19 regulations. Therefore, May 2022 is chosen as a representative month for determining capacity.

4.2. OR Only Scheduling Situation

In the "OR only situation" which equals the situation where no reallocations were made or heuristics were applied in May 2022, 134 treatments were scheduled spread over 16 different days. The total treating time equals 5240 minutes. Of these treatments, 130 were non-emergency patients, while 4 treatments were considered "emergency" patients. The complete schedule is shown in Appendix 1, but to give an overview of the number of operating room treatments in the starting situation taking into account task j and day i, a table is made. The observation of a comparatively low number of treatments scheduled on days 27 and 28 is of significant interest in this study. This decline in scheduled treatments can be attributed to an emergency patient requiring immediate medical attention on the 28th, together with a lengthy operation spanning 259 minutes on the 27th. These examples highlight the complexities involved in predicting and managing healthcare schedules, as unexpected emergencies can disrupt the planned routine.

[redacted]

4.3. OR + POR Scheduling situation

Together with the rescheduling information provided by medical managers at MST, an overview is made in Table 11 to show the number of treatments that can be rescheduled. The number of treatments that would have stayed in the operating room in May 2022 when applying the rescheduling information can be found in the table below. From the total treating time of 5240 minutes, only 3274 minutes of treating remained in the operating room. This counts for 62.5 percent of the original treating time. The remaining operating room treatments look as follows:

Next to the overview of the remaining operating room treatments, an overview can be made of the polyclinic operating room treatments. This overview does not provide an overview of which treatment to schedule in which operating room since all operating rooms are – yet- considered to be equal. As can be seen, 98 of the 134 treatments were rescheduled to polyclinical operating rooms, which counts for 73.1 percent. However, this only counts for 37.5% of the total treatment time. This is as expected since the average treating time for the treatments that are eligible for reallocation in May 2022 equals 26.2 minutes while the average treating time of the treatments that could not be rescheduled equals 92.3 minutes

[redacted]

4.4. Heuristics

As described in the literature study, two heuristics were applied. The heuristics were applied to a "single-machine" in the case of the operating rooms since there is only one operating room available while the operating in a polyclinical setting was applied to multiple machines at the same time since three polyclinical operating rooms are available. All operations were numbered in the order of the original schedule to keep track of the operations, give a concise scheduling overview as well as make it easy to compare different outcomes. In each case, the treatment day is kept the same.

4.4.1. Application

While Chapter 6 focuses on the determination of the required time, the scheduling techniques focus on how to fill this determined time. Two of three techniques – the LPT rule and SPT rule -, as introduced in Chapter 4, are used. Those techniques are used to give a variety of optimization options to the management of MST to have a suitable technique in multiple different situations. The techniques can be applied by parallel machine scheduling and single machine scheduling. Parallel machine scheduling, which can be described as "scheduling n jobs on m parallel machines" (Xing & Zhang, 1999) could only be used for non-emergency patients when considering different days as machines since there is only one available operating room. However, parallel machine scheduling is only possible when there are no severe consequences when patients can be scheduled a few days later than the first possible date. Single machine scheduling considers scheduling n jobs on 1 machine. This can be applied when taking into account one single day instead of multiple in the case of parallel machine scheduling. The chosen techniques to minimize the makespan, completion time, and lateness can be applied to single-machine scheduling as well as parallel-machine scheduling.

In the case of scheduling the treatments for the ENT department, the first step involves making a clear distinction of the treatments, based on the information in Appendix 3. In this case, since the scheduling is executed on the schedules in May 2022, distinctions were made for the treatments during this period. After this, the treatments were ordered in terms of ascending treatment time in case of the shortest processing time rule. Since there is only one operating room, the execution of the two heuristics for the operating rooms is finished when ordering them in the right way. Scheduling the treatments at the polyclinic operating rooms needs more steps since three polyclinical operating rooms have to be taken into account. First, the treatment with the highest priority, so the treatment with the longest processing time when using the LPT rule, and the treatment with the shortest processing time in the SPT rule, is selected. After that, the selected treatment is scheduled at the polyclinical operating room with the lowest amount of treatment scheduled up to the point of scheduling. The schedules below use j and i for treatment j on day i.

4.4.2. Shortest Processing Time Rule

The shortest processing time, which aims to minimize the sum of completion times, can be easily applied by scheduling the treatments with the lowest treating time first. After the reallocation of the treatments, the revised schedule – split up into operations at the operating room in Table 12 and operations that can be executed at polyclinic operating rooms - for the operating room at the ENT department for May 2022, is shown in Table 15. The numbers shown in the table correspond with the order of the original schedule. To make the different methods easily comparable, it is assumed that all operations would start at 8 am, which is the earliest time possible. As can be seen in the table below, the operations that are not allocated were assigned a relatively high number because those treatments were regularly scheduled near the end of the session which is closely associated with the – on average – longer treatment time when taking the indicated aim of Medisch Spectrum Twente into account.

j/i	4	9	10	11	12	13	16	17	18	20	23	24	25	27	28	31
1	13	31	33	44	51	53	72	75	82	93	105	115		129		131
2	14	30	34	45	49		59	74	79	92	95					134
3	15		32		50		56	77								133
4							62	78								
5							57	76								
6							58	73								

Table 12: Operating room schedule after applying SPT heuristic

4.4.3. Longest Processing Time Rule

Since the longest processing time rule tends to schedule treatments with longer treating times first, the outcomes on a single “machine” as it is often mentioned will give an opposite schedule as when applying the shortest processing time rule. Because there is only one operating room available for the ENT department and treatments remain to be scheduled on the same day after reallocation, the scheduling situation for operating rooms after reallocation can be considered as a single-machine problem. The schedule – with the treatment numbers which can be found in appendix one – can be found in the table below.

j/i	4	9	10	11	12	13	16	17	18	20	23	24	25	27	28	31
1	15	30	32	45	50	53	58	73	79	92	105	115		129		133
2	14	31	34	44	49		57	76	82	93	95					134
3	13		33		51		62	78								131
4							56	77								
5							59	74								
6							72	75								

Table 13: Operating room schedule after applying LPT heuristic

The schedule for the reallocated treatments to polyclinical operating rooms when applying the longest processing time rule is significantly different from the polyclinical schedule when applying the shortest processing time rule. Because the longest treatments are scheduled first, which are occasionally longer than the remaining unscheduled treatments together and treatments are scheduled in the polyclinical operating room which will be the first available the schedule ends up with more days where polyclinical operating rooms are scheduled with only one treatment.

4.4.4. Completion Time

The sum of completion times plays a crucial role in the shortest processing time rule, as it aims to minimize this sum. Completion time is referred to the total time required for a task or job to complete its processing. By comparing the completion times, C_j, in the "old" situation without treatment allocation to the "new" situation with the intended treatment allocation, the impact of the allocation on the overall efficiency can be analyzed. In the old situation, the sum of completion times can be calculated with:

$$C_o = \sum_{i=1}^{n} \sum_{k=1}^{m} C_{ik} \tag{5.2}$$

With C_o as the sum of completion times in the "old" situation, n as the number of days, and m as the number of treatments and as C_{ik} as the completion time of treatment k on day i. In the intended situation, the sum of completion times can be calculated with:

$$C_n = \sum_{i=1}^{p} \sum_{k=1}^{q} C_{ik} + \sum_{t=1}^{r} \sum_{u=1}^{s} C_{tu} \tag{5.3}$$

With p as the number of treatment days at operating rooms, q as the number of treatments on a specific day. The values p and q are values after the reallocation of treatments. In this case, C_{tu} provides the completion time of treatment u on day t at a polyclinical operating room while C_{ik}, as in the "old" situation counts for the completion time of treatment k on day i. The values r and s count for the number of treatment days and number of treatments at polyclinical operating rooms after reallocation. The starting time to determine the makespan of the treatments is set as 8:00.

After the use of equations 5.2 and 5.3 on the data of Appendix Two, the total completion time in the "OR-Only" situation – without applying the shortest processing time heuristic or relocated treatments - C_o was determined as 17878 minutes while the completion time of the "OR-POR situation" – with applied shortest processing time heuristics and reallocated treatment - C_n was determined as 8648 minutes. This results in a percentual decrease of completion time of 51.63% after reallocation as well as applying the shortest processing time heuristic. However, since the treatments were spread out over multiple (polyclinical) operating rooms after reallocation the total completion time would have been decreased also without heuristics, which has to be taken into account. The completion time after application of the SPT Rule can be found in tables 13 and 17.

When only considering the operating room, the total completion time decreases from 17878 minutes to 5274, which is a percentual decrease of 70.5%.

Since the LPT is also used in this research, the completion time is determined after applying the LPT rule as well. As can clearly be seen, the total completion time after application of the LPT rule equals 11170 minutes, which is a percentual increase of 29.2% which is a clear indication that the SPT rule contributes to a lower completion time. However, in case of the average makespan for each polyclinical operating room, the percentual decrease was 11.6, -15.5 and 1.7 for the first second and third polyclinical operating room respectively.

Day	4	9	11	12	13	16	17	18	20	23	24	25	27	28	31	Total
OR	1483	1680	817	1872	972	485	2737	1201	1412	1672	1473	897	259	51	867	17878

Table 14: Completion time before reallocation and applying SPT heuristic

j/i	**4**			**9**			**11**			**12**		
1	3	6	1	23	22	17	47	46	41	52		
2	12	10	2	20	19	25	36	40	39			
3	8	4	5	28	26	24	38	43	42			
4	11	7		27	27	21	37	35	48			
5					16							

j/i	**13**			**16**			**18**			**20**		
1	54	55		60	61	65	83	81	84	94		
2				70	71	66	87	89	85			
3				63	67	68	88	91	86			
4				69	64		90	80				

j/i	**23**			**24**			**25**			**28**	**31**
1	23	22	17	116	112	113	119	125	128	130	132
2	20	19	25	114	11	110	120	126	121		
3	28	26	24				123	124	127		
4	27	27	21				122	117			
5		16					118				

Table 15: Polyclinical operating room schedule after applying SPT rule

Day	**4**	**9**	**10**	**11**	**12**	**13**	**16**	**17**	**18**	**20**	**23**	**24**	**25**	**27**	**28**	**31**	**Total**
POR 1	46	113	0	114	48	101	241	0	100	39	141	78	160	0	51	109	3374
POR 2	39	111	0	182	0	63	368	0	142	0	89	116	198	0	0	0	
POR 3	57	75	0	183	0	0	240	0	59	0	79	191	61	0	0	0	
OR	488	247	771	136	442	94	772	895	340	258	227	44	0	129	0	431	5274
Total	718	552	771	446	494	258	1069	895	641	704	532	429	419	129	51	540	8648

Table 16: Completion time after applying SPT Rule

Day	**4**	**9**	**10**	**11**	**12**	**13**	**16**	**17**	**18**	**20**	**23**	**24**	**25**	**27**	**28**	**31**	**Total**
POR 1	71	90	0	231	48	101	502	0	64	39	245	95	124	0	51	109	1770
POR 2	71	89	0	345	0	63	609	0	177	0	142	269	253	0	0	0	2018
POR 3	101	126	0	547	0	0	608	0	223	0	90	144	269	0	0	0	2108
OR	488	247	771	136	442	94	772	895	340	258	227	44	0	129	0	431	5274
Total	731	552	771	1259	490	258	2491	895	804	297	704	552	646	129	51	540	11170

Table 17: Completion time after applying LPT Rule:

4.4.5. Makespan

As described in the literature study, the longest processing time rule is used to minimize the makespan. The makespan is the length of the schedule which is determined by equation 5.4. Minimizing the makespan could be considered valuable since it improves cost-effectiveness by optimizing the utilization of expensive resources and improving the workload distribution of the treatments. The makespan $C_{\max}$ can be calculated with:

$$C_{\max} = \max\{C_i : i = 1, \dots m\} \qquad (5.4)$$

With C_i as the finishing time of treatment i. The outcomes could be considered valuable, however, the relocation to multiple polyclinic operating rooms which decreases the makespan in and of itself has to be taken into account. In this case, to get a complete overview, the maximum makespan for each day will be compared from the "new" situation where reallocations and the longest processing time rule are applied and the "old" situation where no reallocations and heuristics were taken into account.

As can be seen in figure 19, the maximum makespan in the "OR -only" situation, where no heuristics were applied equals 458. In the case of the new situation, where patients were allocated over multiple rooms and the longest processing time rule was applied, the makespan is split up over four different rooms. In the new situation with reallocated treatments and an applied heuristic, the completion times are less than during the "OR-only" situation. The average makespan, when considering the makespan daily instead of monthly, equals 322,6 for the "old" situation while it equals 217,213 during the "new" situation. This is a percentual decrease of 32.7%, which can be considered beneficial. Unlike the average makespan daily, the makespan monthly does not decrease by a significant amount. This is due to one day where no reallocations of treatments were possible which may give a distorted view for the possible decrease in makespan. The makespan in the "OR-only" situation equals 458 while the makespan in the "new" situation equals 401. This is a percentual decrease of 12.5%. The results of the makespan can be found in the tables 18 -21.

When only considering the operating room, the makespan decreases from 458 to 401, which equals a percentual decrease of 12.4%. However, the average makespan decreases from 327,5 to 207,12 which is a percentual decrease of 36,8%.

However, the main makespan decrease likely comes from the rescheduled amount of patients instead of the applied rule. The average of the maximum makespan in May on the different operating rooms equals 212 for the SPT rule, and 199 for the LPT rule. This is a percentual difference of 6.53% which can be considered as a minor decrease. Furthermore, the average makespan after application of the LPT and the SRT where the same except for the second date.

Day	4	9	10	11	12	13	16	17	18	20	23	24	25	27	28	31	MAX
OR	394	397	401	458	337	258	395	300	441	226	336	340	258	259	51	389	458

Table 18: Makespan before applying LPT heuristic and reallocation

Day	4	9	10	11	12	13	16	17	18	20	23	24	25	27	28	31	MAX
POR 1	32	34	0	116	48	63	127	0	64	39	52	95	124	0	51	109	127
POR 2	35	33	0	121	0	101	133	0	57	0	50	105	64	0	0	0	133
POR 3	35	39	0	119	0	0	135	0	57	0	55	96	70	0	0	0	135
OR	284	200	401	92	291	92	248	300	263	187	179	44	0	259	0	280	401
MAX	284	200	401	121	291	101	248	300	263	187	179	105	124	259	51	280	401

Table 19: Makespan After Applying LPT heuristic and reallocation

Day	4	9	10	11	12	13	16	17	18	20	23	24	25	27	28	31	MAX
POR 1	33	41	0	73	48	63	136	0	55	39	70	63	74	0	51	109	136
POR 2	29	42	0	134	0	101	152	0	91	0	39	90	149	0	0	0	152
POR 3	40	32	0	159	0	0	107	0	32	0	48	143	35	0	0	0	159
OR	284	200	401	92	291	92	248	300	263	187	179	44	0	259	0	280	401
MAX	284	200	401	159	291	101	248	300	263	187	179	143	149	259	51	280	401

Table 20: Makespan after applying SPT heuristic and reallocation

j/i	**4**			**9**			**11**			**12**		
1	7	9	11	18	16	28	46	47	48	52		
2	8	4	5	27	24	26	39	38	43			
3	12	1	2	25	19	21	36	42	41			
4		3	6	20	23	22			37			
5						17						

j/i	**13**			**16**			**18**			**20**		
1	54	55		58	57	62	80	90	86	94		
2				72	59	56		91	85			
3				69	63	66		84	88			
4				70	71	65		89	81			
					60	61			87			
									87			

j/i	**23**			**24**			**25**			**28**	**31**
1	99	36	108	110	111	114	117	127	118	130	132
2	98	103	109		112	113		122	123		
3	102	107	100					121	119		
4	101	104						124	125		
5	100							120	128		
	97							126			

Table 21: Polyclinical operating room schedule after applying LPT heuristic and reallocation

4.5. Conclusion

In conclusion, the application of the shortest processing time (SPT) heuristic and treatment reallocation yielded significant improvements in completion time for the analyzed data. The total completion time in the “OR-Only” situation, without applying heuristics or reallocation was determined as 17,878 minutes, whereas it reduced to 8,648 after reallocation and applying the SPT rule resulting in a decrease of 51.63%. However, applying the LPT rule in the situation after reallocation resulted in a decrease of 29.2%, which is an indication that the SPT rule does contribute to a lower completion time, but that the reallocation of patients has a significant influence on the completion time as well.

The examination of makespan, which represents the maximum time taken by any treatment, revealed interesting insights. In the "OR-Only" situation, the makespan equaled 458. However, in the "POR + OR" situation, the makespan was divided among four different rooms. As a result, the average makespan daily decreased significantly from 322.6 units in the "OR-Only" situation to 217.213 units in the "POR + OR" situation, representing a percentual decrease of 32.7%, which is highly beneficial for operational efficiency. However, it essential to mention that the main decrease in the makespan likely stems from the rescheduling of patients rather than the applied heuristic. The average maximum makespan in May 2022 for the different operating rooms was 212 minutes for the SPT rule and 199 minutes when applying the LPT rule which represents a minor percentual difference of 6.53%.

When only considering the operating rooms, 62,5% of the treatment time is still needed after reallocation. When applying the LPT heuristic, the average makespan can decrease by 36.8%, and the maximum makespan by 12.4%. Furthermore, when applying the SPT heuristic, the completion time can be decreased by 70.5%.

To summarize, the findings highlight the significant positive impact of treatment reallocation and the SPT heuristic on completion time and makespan. These strategies contribute to enhanced scheduling efficiency and resource utilization in the analyzed scenario, ultimately leading to improved operational outcomes.

Chapter 5. Determining Required Capacity by Queuing

To determine the needed capacity at the ENT department of Medisch Spectrum Twente, Queuing Theory can be used. Two different queueing models are used. For the current situation, an $M/G/1$ queue turned about to be a suitable system, while the "POR + OR" at MST requires two different queues. In that situation, an $M/G/c$ queue is used for the polyclinic operating room next to a $M/G/1$ queue that is used for the operating room. Both situations are analyzed by the use of several different formulas. This analysis aims to give insight into the capacity usage in the current situation as well as provide a clear starting point for future simulation studies. Calculations can be found in Appendix 1.

5.1. Treating Time Data Analysis

Since the treating time of patients is a significant part of treating operating room patients, the treating time of the executed operations is analyzed. The analysis is done by determining the distribution of the data with a Chi-Square goodness of fit test as well as taking into account descriptive statistics and Q-Q plots. The Chi-Square test determines if a sample of data matches a specific distribution and "can be used to that the hypothesis that observed data follow a particular distribution" (ShierRosie, 2004). Three different types of treating time data are analyzed for the sake of a mathematical model in the later parts of this project. The treatment times of the patients when no treatments are reallocated from operating rooms to polyclinic operating rooms (dataset 1), the treatment times of the reallocated patient to the polyclinic operating rooms (dataset 2), and the remaining patients at operating rooms after reallocation of the eligible patients to polyclinic rooms (dataset 3).

The initial step of this data analysis encompasses generating Q-Q plots of the three different, but highly related data sets. A Q-Q plot is a useful graphical tool to assess the distributional similarity between a given dataset and theoretical distribution. For this data analysis, a Q-Q plot is made for the Normal, Beta, Gamma, and Exponential distribution. For each data set, in the case of the Gamma distribution, the data was almost perfectly aligned with the diagonal line as can be seen in the figures below. This is a clear indication that the data follows a Gamma - $\Gamma(k)$ - distribution. Therefore, the Goodness-of fit test is based on the Gamma distribution. Since all three data sets are tested to the Gamma distribution, the hypothesis is the same for each situation. The quantiles are plotted along the x-axis as the "theoretical quant quantiles" while the sample quantiles are plotted along the y-axis.

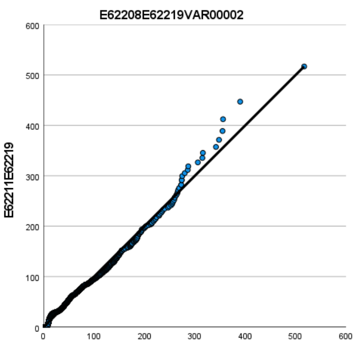

Figure 22: Q-Q Plot Data Set 1

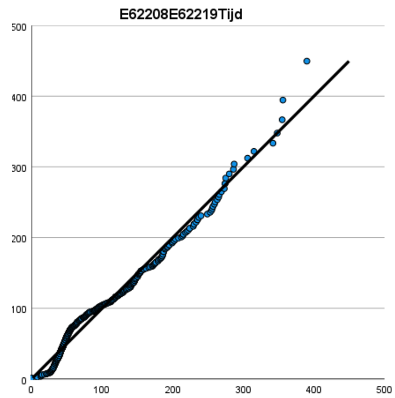

Figure 23: Q-Q Plot Data Set 2

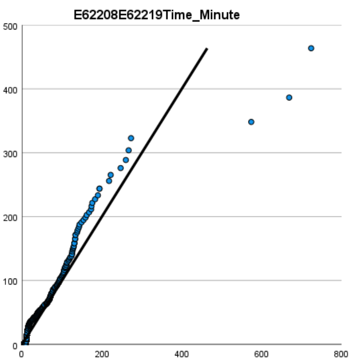

Figure 24: Q-Q Plot Data Set 3

The null hypothesis and the alternative hypothesis can be stated as:

$H_0 =$ The provided data set does not follow a Gamma - $\Gamma(k)$ - distribution.

$H_1 =$ The provided data set does follow a Gamma - $\Gamma(k)$ - distribution.

In this case, the events $N_1, \dots, N_k$ $(k \geq 2)$ have a multinomial distribution. Those events have success rates $p_1, \dots, p_k$. The total number of events is n. The average number of occurrences of the outcome N_i can be described as EN_i which is calculated with $EN_i = np_i$. The Chi-Square value, which is used to measure the degree of association with a statistical distribution and calculated by comparing observed frequencies of events or categories with the expected frequencies can be calculated as follows:

$$X^2 = \sum_{i=1}^{k} \frac{(N_i - E_0N_i)^2}{E_0N_i}$$

The X^2 the test statistic has an approximate Chi-square distribution with $k-1$ degrees of freedom, with k as the number of used bins. Since this test makes use of a large dataset, bins are used to maintain a reliable and workable number of calculations. The optimal, or close to optimal, number of bins can be determined with the rule of thumb $\sqrt{n}$. The mean $\overline{x}$ is determined with $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, and the variance S^2 is determined with: $S^2 = \frac{\sum_{i=1}^{n}(x_i - x_m)^2}{n-1}$. Those statistics, in combination with the mean are shown in table

[redacted]

For a - $\Gamma(k)$ – distribution, an α, and β values need to be determined to determine the expected value of a gamma distribution. α is determined as $\alpha = \overline{x}^2/S^2$ and $\beta = \overline{x}/S^2$. Furthermore, the critical value of the Chi-square test needs to be determined when taking into account a commonly used significance level – α – of 5%. The use of a higher significance level could be too stringent and may result in a higher chance of a type II error, also known as a false negative, which occurs in statistical hypothesis testing when the null hypothesis is erroneously accepted or not rejected, despite it being false. The use of a lower significance level increases the likelihood of false positives and weakens the reliability of the conclusions. Therefore, the 5% significance level is considered the middle ground.

The test statistic highly depends on the degrees of freedom, and therefore on the number of used bins. These bins are different for each of the three datasets. The degrees of freedom for datasets 1,2, and 3 are 29, 32, and 17 respectively. The chi-square values for those degrees of freedom with a significance level of 5% are respectively 42.556, 55.758, and 36.42. For each of the tests exceeding the value means a failure of the rejection of the H_0 hypothesis, which results – in this case – in the failure of the assumption that the provided data follows a gamma distribution. To determine the test statistic X^2, a specific plan is used. Firstly, the number of occurrences in a specific bin range, which consists of minutes of treating time, is determined. After that, the number of occurrences in that specific bin for a cumulative gamma distribution is calculated. Since the gamma distribution is continuous, the cumulative gamma distribution is determined with:

$$F(x; \alpha, \beta) = \int_{0}^{x} f(t, \alpha, \beta) dt$$

With gamma distribution

$$f(t, \alpha, \beta) = t^{\alpha-1} * \frac{\exp\left(-\frac{t}{\beta}\right)}{\Gamma(\alpha)\beta^{\alpha}}$$

Since the cumulative distribution only takes into account one experiment, the cumulative distribution is multiplied by the number of observed values to obtain the intended values. After determining the cumulative distribution, the probability distribution is derived by subtracting the cumulative values of two consecutive bins. By considering the occurrences in each bin and the values of the probability distribution, X^2 can be calculated, where N_i represents the actual occurrences and E_0 N_i represents the values of the probability distribution. After summing the values for each bin and each dataset, the following X^2 values were obtained: 24.13, 52.25, and 229.18. As the chi-square values were calculated as 42.556, 55.758, and 36.42, the H_0 hypotheses can be rejected for dataset one and dataset two, but not for dataset three. This implies that, with a significance level of 5%, it can be assumed that the treatment times of patients when no treatments are reallocated, and the treatment times of the reallocated patients to the polyclinic operating rooms follow a gamma distribution. However, the remaining patients in the operating rooms after reallocating the eligible patients to polyclinic rooms do not follow a gamma distribution with an α of 5%.

5.2. Queueing Model

To determine the required capacity at MST, an $M/G/1$ queue in combination with a $M/G/C$ queues could be used. This method requires an Poisson arrival and a "general" service time distribution. The current situation could be modelled with a single $M/G/1$ queue, since there is only one operating room, while the new situation has to be modelled with a $M/G/1$ and a $M/G/3$ queue. The new situation consists of one operating room and three polyclinical operating rooms. This queueing system can give some valuable results, which might be considered as approximations since several assumptions are made. Firstly, the $M/G/1$ queue allows for making a distinction between emergency patients and non-emergency patients by using the Pollaczek-Khintchine formula for average waiting time and queue length. Furthermore, the possibility to determine utilization rates could be considered as useful when determining capacity.

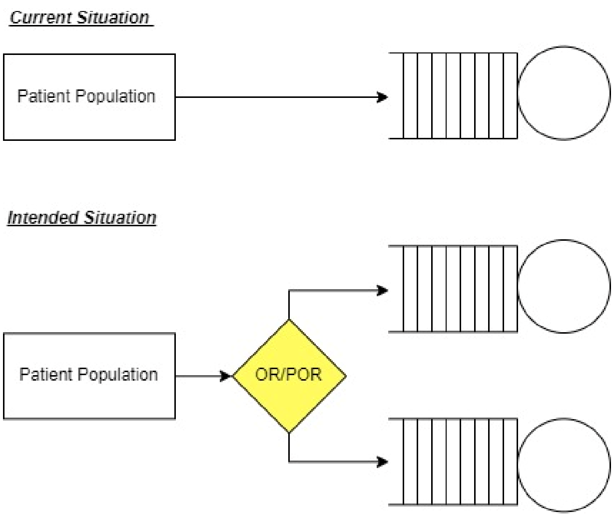

Figure 26: Global view of queuing situation

In the current situation, the "patient population" consists of the patients that are treated in operating rooms. The patients are placed on the waiting list which is considered as the queue, and after the treatment, the patient leaves the system. After reallocating types of treatments from operating rooms to polyclinical operating rooms which is described as the "intended situation", there are two possible routes for the patients. A polyclinical treatment or operating room treatment. In the current situation, patients are placed on a waiting list which is considered as a queue, and after the treatment, the patient leaves the system. Comparing the two situations will allow for determining the idle time of the ENT operating room in each situation.

5.3. Methodology

5.3.1. Necessary input data

To accurately model a queueing situation using a queueing model, several crucial input data points are required. One of the most fundamental input data is the *arrival rate*, which represents the rate at which patients enter the system. In the context of the ENT department, there are three arrival rates to take into account. The first arrival rate λ_1 encompasses the patients for treatment at the operating before reallocation decisions, the second arrival rate λ_2 includes the arrival of patients to operating rooms after reallocation. The third and final arrival rate λ_3, the arrival of patients to polyclinical operating rooms after reallocation can easily be derived from the first two arrival rates. Furthermore, the arrival rates can be easily split up into "emergency arrival" and "non-emergency" since Poisson processes can be split by splitting the arrival rate. arrival. Even though treatments at the ENT department use an appointment-driven system, it is assumed that the arrival rates follow a Poisson distribution. The use of a deterministic arrival process drastically increases the level to a point that is far out of the scope of this research. The arrival rates are determined by dividing the total patients - while taking into account the arrival type - by the number of days on which the certain operating room was available. The following arrival rates with the number referring to the corresponding data set as described in chapter 6.1 and e and n corresponding to the type of patient where e means emergency patient and n means non-emergency patient – calculations are provided in appendix 2 - were found:

$$\lambda_1 = 7.72 \qquad \lambda_{1e} = 0.33 \qquad \lambda_{1n} = 7.39$$
$$\lambda_2 = 2.62 \qquad \lambda_{2e} = 0.1459 \qquad \lambda_{2n} = 2.474$$
$$\lambda_3 = 6.83 \qquad \lambda_{3e} = 0.245 \qquad \lambda_{3n} = 6.58$$

The *service times*, which are extensively analyzed in Chapter 6.1 by a goodness-of-fit test, can be split up in the same three types as the arrival rates. As the goodness-of-fit test proved, the service time when no treatments are reallocated from operating rooms to polyclinic operating rooms as well the service time of the reallocated patients to the polyclinic operating rooms is gamma distributed when taking into account an α of 5%. However, the service time of the patients that remain in the operating rooms after reallocation does not follow a perfect gamma distribution when taking into account an α of 5%. However, the Q-Q plot clearly shows a pattern that indicates a Gamma distribution. Since queueing models require a service time distribution, it is assumed that the service time follows a gamma distribution in all three situations. Since the used equations which will be introduced in chapter 6.2, require service rates instead of service times, service rates will be determined by dividing the number of available treatment hours. The service rates $\mathbb{E}[B_i]$ which use the same numbering system as the arrival rates– calculations are provided in Appendix Two – as follows:

$$E[B_1] = 10.08 \qquad E[B_{1e}] = 6.6 \qquad E[B_{1n}] = 10.32$$
$$E[B_2] = 5.20 \qquad E[B_{2e}] = 5.1279 \qquad E[B_{2n}] = 7.06$$
$$E[B_3] = 18.3 \qquad E[B_{3e}] = 5.98 \qquad E[B_{3n}] = 19.82$$

The *number of servers* differs for the situation before reallocation and the situation after reallocation differs. Before reallocation, only one operating room was available. After reallocation, one operating room and three polyclinical operating rooms are available in the most favorable situation. To give a comprehensive overview of both situations, a transition diagram is shown below.

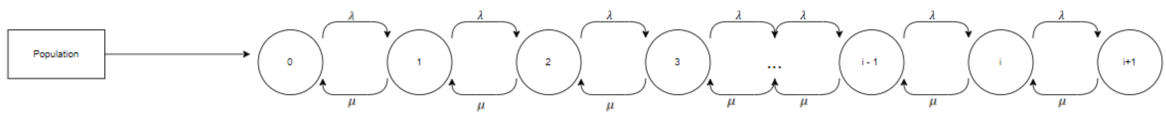

Figure 27: Queueing system "OR only" situation

The "current" situation is shown in Figure 6.1. In this situation, there is a patient population with an arrival rate λ_1, which is split up in emergency and non-emergency patients. This model does not include a maximum, because of the assumption of an infinite waiting room. The service rate μ corresponds to the service rates $E[B_i]$ which are determined before. The model clearly shows one available server, which corresponds to the situation of one operating room.

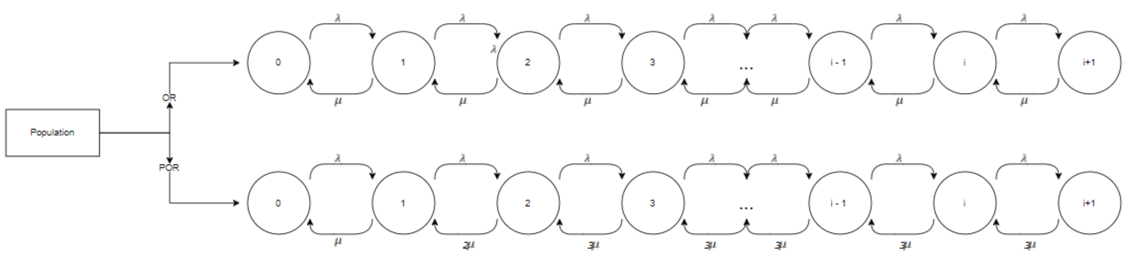

Figure 28: Queueing system "POR + OR" situation

The "intended" situation is shown in figure 6.2. In this situation, the initial population is split up into "polyclinical" patients, or "POR" as shown in the figure, and "operating room" patients, or "OR" as shown in the figure. These two patients have a separate arrival as well as service time as shown in this paragraph before. This situation features four different rooms consisting of one operating room and three polyclinical operating rooms which could, in theory, be used at the same time.

Finally, the *buffer size* and the *queue discipline* have to be taken into account. In the case of the ENT department at Medisch Spectrum Twente, there is – in practice - no maximum buffer size. Patients wait most of the time at home, and there is no fixed number of waiting rooms. The used queue discipline is a first come first served (FCFS) system. Although an appointment-driven system is used, the patients that are not considered emergency patients will be scheduled based on the waiting time. Therefore, First-Come First-Served is considered to be the most suitable queuing discipline.

5.3.2. Operating Room Capacity Determination Method

In the situation without rescheduling, the "OR only situation", which encompasses one single operating room, Poisson arrival rates λ_1, λ_{1e} , λ_{1n} and Gamma service rates $E[B_1]$, $E[B_{1e}]$, $E[B_{1n}]$ can be modeled in the Kendall notation as a $M/G/1$ queue where M stands for Poisson arrivals, G for a generalized service time distribution – in this case the Gamma distribution – and 1 for the number of "servers" which is, in this case, the number of operating rooms. Furthermore, it is assumed that there is an infinite waiting room. The standard deviation of the gamma distribution for the patient treatment time is denoted as $\sigma(B)$. In queueing theory, utilization is used as a measure of productivity. In healthcare, utilization is often referred to as occupancy level. The percentage of time that an operating room is used can be determined by determining the utilization rate, which is highly useful in determining capacity. In an $M/G/1$ queue, the server utilization ρ with $\rho < 1$ can be defined using Little's law as (Boucherie, Braaksma, & Tijms, 2022, p. 365):

$$\rho_i = \lambda_i E[B_i] \tag{6.1}$$

Where i is the type of patient as described in paragraph 6.1. Since this server utilization can be described as the fraction of the time that the server is busy, it can also be described as the fraction of the patients who must wait. (Boucherie, Braaksma, & Tijms, 2022). Next to the utilization rate, the average queue length and waiting time could be useful for relative comparison to the intended situation although Medisch Spectrum Twente uses an appointment-driven system. The *Pollaczek-Khintchine* formula for the average queue length L_q in and $M/G/1$ Queues is as follows (Boucherie, Braaksma, & Tijms, 2022, p. 365):

$$L_q = \frac{1}{2}(1 + c_B^2) * \frac{\rho_i^2}{1 - \rho_i} \tag{6.2}$$

Since the ENT department uses an appointment-driven system for non-emergency patients, the average queue length L_q is mainly relevant for emergency patients where appointments do not have to be taken into account. However, the average queue length could be used as a comparison method between the two situations. The average waiting time - which is closely related to the average queue length because of Little's Law - W_q is defined as follows (Boucherie, Braaksma, & Tijms, 2022, p. 365):

$$W_q = \frac{1}{2}(1 + c_B^2) * \frac{\rho_i E[B_i]}{1 - \rho_i} \tag{6.3}$$

"The coefficient of variation c_B, which is determined as $\frac{\sigma(B_i)}{E[B_i]}$, is used to assess the relative variability of a dataset" (Boucherie, Braaksma, & Tijms, 2022, p. 365). A high coefficient of variation – bigger than one since the standard deviation is bigger than the mean - indicates that the data points in the dataset have a high degree of dispersion or variability relative to the mean, while a low coefficient of variation indicates that the data points are relatively close to the mean and have low variability. A high coefficient of variation will contribute to a higher average waiting time as well as a higher average queue length (Boucherie, Braaksma, & Tijms, 2022).

Since the ENT department makes distinctions between emergency patients and non-emergency patients, a simple rule can be used to determine the waiting time for each of those two types. Using the arrival- and service rates as calculated in Appendix 2, the standard deviation $\sigma(B_i)$ of the service times and the assumption that emergency patients are given priority over non-emergency patients but the treatment of non-emergency patients cannot be interrupted when an emergency patient arrives, the average waiting time for an emergency patient $W_q^{(i)}$ can be determined as (Boucherie, Braaksma, & Tijms, 2022, p.369):

$$W_q^{(e)} = \frac{\lambda_e E[B_e^2] + \lambda_n E[B_n^2]}{2(1 - \lambda_e E[B_e])} \tag{6.4}$$

In the same way, the average waiting time for non-emergency patients $W_q^{(n)}$ can be defined as (Boucherie, Braaksma, & Tijms, 2022, p.369):

$$W_q^{(n)} = \frac{\lambda_e E[B_e^2] + \lambda_n E[B_n^2]}{2(1 - \lambda E[B_e])(1 - \lambda_e E[B_e] - \lambda_n E[B_n])} \tag{6.5}$$

However, formula 6.5 could give a distorted view when applied to an appointment-driven system. Little's law can be applied to determine the average queue length, which is mainly useful for emergency patients (Boucherie, Braaksma, & Tijms, 2022, p.369):

$$L_q^{(i)} = \lambda_i W_q^{(i)} \tag{6.6}$$

5.3.3. Polyclinical Operating Room Capacity Determination Method

As stated before, the intended situation has to deal with two different queueing situations. One situation encompasses the same operating room queue as in the current situation, but with a different arrival- and service rate, while the other queuing situation consists of the treatments that were eligible for reallocation with three servers. For the treatments in the operating room, the same method will be used as presented in paragraph 6.2.1. Therefore, the operating room part of the intended situation will directly be analyzed instead of presenting a new method.

The $M/G/C$ queue arises by dropping the assumption of an exponentially distributed service time in the $M/M/C$ queue, which is the case in this situation. There is an assumption of a Poisson-distributed arrival rate and a proven gamma-distributed service rate. In case, there are three available polyclinical operating rooms, which makes the $M/G/C$ queue an $M/G/3$ queue. The utilization rate ρ is now defined as (Boucherie, Braaksma, & Tijms, 2022, p.374):

$$\rho = \frac{\lambda \mathbb{E}[B]}{c} \tag{6.7}$$

As in the current situation, the assumption of $\rho < 1$ is used. Next to the utilization rate, the percentage of customers that must "wait in line" P_w is used. This equals the average rate of emergency patients that will find all the servers busy and therefore have to weigh. A good approximation for this percentage of customers is given by the $P_w(\exp)$ which is the percentage of customers that will find all the servers busy in an $M/M/C$ queue. $P_w(\exp)$ is determined by the Erlang-C formula $C(c, \rho)$ (Boucherie, Braaksma, & Tijms, 2022, p.371):

$$C(c, \rho) = \frac{(c\rho)^c}{c!\,(1-\rho)} \left[\frac{(c\rho)^c}{c!\,(1-\rho)} + \sum_{j=0}^{c-1} \frac{(c\rho)^j}{j!} \right]^{-1} \tag{6.8}$$

The utilization factor ρ represents the ratio of the average arrival rate λ to the product of the number of servers C and the average service rate per server μ. It measures the fraction of time the servers are busy serving customers. On the other hand, the probability of waiting for service $P_w(\exp)$ is calculated using the exponential waiting time formula. This formula takes into account the arrival rate λ, the average service time $1/\mu$, and the number of servers C to calculate the probability of waiting for service in the system. In an M/G/C queue, the relationship between the utilization factor ρ and the probability of waiting for service $P_w(\exp)$ is complex and not directly equal.

5.3.4. Argumentation of Assumptions

As stated in section 5.3.1, Poisson arrivals are used for the appointment driven system. There are three criteria for the application of a Poisson process. These criteria include that the events are independent of each other, that the average events per time period is constant, and that event cannot occur at the same time. The ENT department makes a distinction between emergency patients and non-emergency patients. It can be reasonably argued that the arrivals of emergency patients follow a Poisson distribution. First of all because of their unpredictable nature which makes the arrivals of the non-emergency patients independent of one another. The second criteria, that requires the arrival of patients per time period as constant, is fulfilled this case as can be seen in figure 28. In no particular time period, the number of emergency patients highly deviates from the average number of emergency patients per month. The average number of emergency patients per month equals 5. Furthermore, the probability of two emergency patients arriving at exactly the same time is insignificant.

[redacted]

However, the arrivals of non-emergency patients is different. When treatments are scheduled, the timing and sequence of each treatment are determined in advance. The scheduled treatments in this context are not random but rather deterministic, as they are planned upfront. However, patient-specific factors could introduce randomness. Patients may cancel or reschedule their appointments due to unforeseen circumstances or changes in their condition. Furthermore, since emergency patients are available at MST, unforeseen events or emergencies can occur that require the rescheduling or rearrangement of appointment which introduces a certain degree of randomness. The Poisson process can accurately capture this behavior. As shown in chapter 4, the average events per time period is relatively constant, which suits the second criteria. Besides, it allows analyzation of the system's performance with – among others – service utilization, waiting times and queue length. Finally, it can be argued that the arrivals in an appointment-driven systems have a memory-less property like a Poisson process. In case of an appointment-driven system, the timing of arrivals could be independent of past arrivals. In this context, it means that the likelihood of a patient arriving within a specific time period remains constant, regardless of how much time has passed since the last arrival.

In summary, despite the deterministic nature of the appointment-driven system, the Poisson process remains useful since several criteria of the Poisson arrival assumption have been met, and the analysis made possible by the assumption may be useful. However, when taking into account the results of the analysis, one should take into account the consequences of the model. A main consequence of the use of Poisson arrivals instead of deterministic arrivals is the increase of waiting times and queue lengths - as can be seen in equation 6.2 and 6.3 - since the randomness implies a higher coefficient of variation. Furthermore, it can be questioned whether to what extend the queue lengths for non-emergency patients are useful.

5.4. Results

When analyzing the outcomes of the *situation without reallocations* - detailed calculations provided in Appendix Two- of the provided method in paragraph 5.3, several aspects stand out. First of all, the utilization rate of the operating room, under the assumption that the ENT department can use a complete shift from 8 to 4 on every day where treatments are scheduled, equals 0.766. This is a clear indication that there are no more treatments scheduled than the operating room can handle. Secondly, the treatments in the current situation have a relatively high coefficient of variation, which negatively influences the waiting time and therefore the queue length. This emerges as a prominent factor in the formulas for the general average queue length L_q, which turned out to be 2.03, and the general average waiting time W_q which turned out to be 223.7. Both the average queue length and the average waiting time increased because of a relatively high coefficient of variation. Both the average waiting time and average queue length are, next to the general situation, split up for the two types of patients. The average waiting time for emergency patients under the assumptions as stated in paragraph 6.2.1 turned out to be 54.95 minutes while the average waiting time for non-emergency patients – which can be considered as less relevant because of the appointment-driven system – turned out to be 253.42 minutes. This corresponds to the average waiting time of 223.7 minutes. The average queue length, which is derived from the average waiting time by Little's Law, turned out to be 0.0373 emergency patients and 3.901 non-emergency patients

When looking at the calculations of the treatments that *remained in the operating rooms after reallocation*, a few aspects are important to take into account. First of all, a higher variability in the treating times, which turned out to have a less significant influence because of the longer treating time. Furthermore, the utilization rate equals 0.50308 of the operating room, compared to 0.766 in the situation without reallocations. This is a significant difference that might have been expected to be even bigger when looking at the results in Chapter 5. However, since many treatments could be rescheduled, some days completely disappeared – appendix 1 for calculations - from the operating room schedule, and are therefore left out of the calculations because it is likely that those days will be filled up treating patients from other departments. The lower utilization rate influenced the queue length, which decreased by 80.2 percent, and the average general waiting time which decreased by 39.91 percent. Furthermore, the average waiting time for emergency patients decreased from 54.95 minutes to 37.25 minutes which equals a decrease of 32.2 percent. The waiting time for non-emergency patients faced a percentual decrease of 70.42 percent. This all leads to an average amount of 0.01 emergency patients waiting in line instead of 0.037 in the current situation.

The *reallocated treatments to the polyclinical operating rooms* cover an additional utilization of 12.4% of the total polyclinical operating room capacity or an additional utilization of 37.3% of one polyclinical operating room. Therefore, according to the queueing theory outcomes, it is advised that *at least* 37.3% of the maximum capacity of one polyclinical operating room is kept available for reallocated treatments. When considering no other polyclinical treatments, the percentage of patients – without an appointment-driven system – that has to wait for an appointment after arrival equals 1.42%.

Several results of the queuing analysis can be compared to the outcomes of the application of heuristics. First of all, the utilization rate in the “OR-Only” situation is 76.6%, which is comparable to the utilization rate of the ENT department at MST. The utilization rate at the ENT department equals – at the time of writing – 80.4%. The utilization rate after the reallocation of treatments equals 50.3%, which is comparable to the outcome of the approach by heuristics. According to the heuristics, 37.5% of the treating time is scheduled in polyclinical operating rooms, while the utilization rate according to the queueing analysis decreases by 26.3%. Since the outcomes of the queueing analysis and the heuristic approach is comparable, the heuristics are considered to b reasonable

5.5. Conclusion

The queueing model can answer the question of how much capacity is needed after reallocation. Instead of the utilization of 76.6% before reallocation, only 50.3% utilization is needed in the situation after reallocation in the operating room, which is a decrease of 26.3%. Furthermore, 17 days of treatment are completely excluded from the calculations, which could decrease the utilization as well. The patients that were formerly scheduled on the operating now take into account 37.3% of the utilization of one polyclinical operating room which translates to 12.4% of the total polyclinic operating room capacity. Furthermore, it became clear that after the application of the reallocations, the average waiting time for emergency patients drastically decreased which can be considered a positive side effect of the reallocation.

Chapter 6. Discussion

The aim of this research is stated as “to determine the needed capacity in hours for operating room sat the ENT department of Medisch Spectrum Twente”. The main findings are all related to the needed capacity after the reallocation of the eligible treatments to polyclinical operating rooms. As can be seen in Chapter 4.3, 37.5% of the total treatment time at the operating rooms in the “OR + POR” situation can be rescheduled to polyclinical operating rooms. When applying the right scheduling heuristics, the completion time can be decreased by 51.63%, while the average makespan can be decreased by 32.7%. The queueing theory analysis shows a possible operating room utilization decrease of 26.3 percentage points. The decrease in utilization can be considered good because the non-utilized operating room time can be used for other departments at MST. This decrease causes a total utilization rate of 37.3% in a single polyclinical operating room. These findings imply that reallocating the eligible treatments to polyclinical operating rooms and the use of appropriate scheduling heuristics has the potential to improve the completion time, makespan, and operating room utilization. These findings highlight the benefits of optimizing resource allocation and scheduling. Furthermore, the average waiting time decreased by 32.2 percent, which is a significant decrease.

Next to the main findings about operating room capacity, other results can be considered as valuable. As shown in Chapter 5, the treating time of treatments at the ENT department of MST follows a Gamma distribution before reallocation of treatments as well as after reallocation. Furthermore, the deseasonalization of the treatments can be useful for further decisions. As shown in Chapter 4, the month of May has the least percentual difference between its deseasonalized demand and the real demand, which makes it useful to use. The implications of this research are significant in determining the needed capacity for operating rooms at the ENT department of MST. The main findings indicate that reallocation of eligible treatments to polyclinical operating rooms and using appropriate scheduling heuristics can lead to improvement in completion time, makespan, operating room utilization, and average waiting time for emergency patients.

The majority of the results as presented in this study follow the expectations upfront. The results in the chapter “Rescheduling with Heuristics” shows a significant decrease in makespan and completion time after using the SPT and LPT heuristics, which was expected after the literature study. Furthermore, the seasonality analysis does not give unexpected results as well. The outcomes of the treating time data analysis – treating times follow a Gamma distribution - was somewhat unforeseen. Partly because of these results, M/G/C queues could be used to model the “OR-Only” as well as the “OR + POR” situation. However, these results can be explained. The gamma distribution is a relatively flexible distribution because of a shape parameter and a scale parameter which makes multiple datasets fit in the Gamma distribution.

Although the use of M/G/C queues is defended in section 5.3.1, a different approach might have been suitable as well. As discussed in the literature study an ADQ model with time-limited service, which is created for appointment-driven systems like treatments at MST, is suitable. However, due to the complexity of the ADQ model, this model is considered to b out of the scope of this research. The use of this method could be useful for further study. Further research could be useful to enhance the comprehension of reallocating treatments to polyclinical operating rooms and optimizing scheduling heuristics. Exploring diverse scheduling heuristics and comparing their efficacy would enable the identification of more efficient strategies. Additionally, a simulation study could be executed based on the queueing model to reduce the number of assumptions and gain a higher degree of certainty in the results.

When considering the used heuristics, a few aspects have to be taken into account. The heuristic approach a could improve completion time and minimize waiting time, however, the schedules of hospitals rely on their patients, which might make the application of those heuristics difficult. Besides, since the applied techniques were based on historical treatments, the possibility of a different situation has to be taken into account when using those techniques. However, the used techniques turned out to make a severe contribution to the possible improvement of the schedules at the ENT department of Medisch Spectrum Twente.

Next to the results, the limitations of this research are worth mentioning. The methodological choices were constrained by techniques that are feasible for a bachelor thesis in Industrial Engineering & Management Science. Furthermore, the used treatment data partly contained influence from the COVID-19 pandemic which could give a distorted view of the outcomes. Besides, the decision on which patients to reallocate to polyclinical operating rooms was based on a study by Medisch Spectrum Twente in 2020 which has to be taken into account for valid interpretations. Also, the assumptions made in the queueing analysis have to be taken into account when interpreting the results. One of the main assumptions, the assumption of Poisson arrivals which is used in chapter five, is questionable to be right and should be taken into account when using the results.

Lastly, recommendations can be made based on the outcomes. The recommendation for polyclinical operating rooms is to reserve at least 37.3% of the total capacity of one polyclinical operating room for the reallocated treatment types and use the proposed heuristics for polyclinical treatments as well as operating room treatments to decrease waiting time and the maximum completion time. The recommendation for operating room capacity is to decrease the capacity in the current situation with 37.5 percent.

Chapter 7. Conclusion

This thesis identified the needed operating room capacity after reallocating the eligible treatment types to polyclinical operating rooms. The outcomes of this main research aim resulted in the needed capacity of the operating rooms. Based on the used heuristics and queueing theory, the needed capacity is known. Furthermore, by the application of the SPT and LPT heuristics, the makespan and completion time can be decreased significantly.

Next to the main research aim, other research questions are taken into account as well. Firstly, a clear overview of the current situation at MST is executed which encompasses a data analysis and a scheduling process overview. The scheduling process overview shows that schedules for operating rooms are mainly based on the waiting time, the availability of specialists, urgency, and the availability of instruments after making the block-plan schedule. The data analysis provided clear insights into main statistics such as the current capacity. In the literature review, suitable methods for determining capacity are discussed. In this thesis, queueing theory and two different heuristics, the SPT heuristic and LPT heuristic, are used. The capacity problem is modeled by two different queueing models which use the input data that is determined in the analysis of the current situation as well as the data analysis of the treating times. The historical schedule for operating rooms as well as polyclinical operating rooms is made with the use of heuristics.

Several aspects stand out in this thesis. First of all, according to the heuristic approach, 37.5% of the total treatment time in the "OR-Only" – the current situation – can be rescheduled to polyclinical operating rooms. When applying the right heuristic, around a 51.63 decrease in completion time and a 32.7 decrease in the average makespan is possible. The queuing theory analysis indicates a potential 26.3 percentage point decrease in operating room utilization, resulting in the utilization of 37.3% in a single polyclinical operating room. When executing the reallocations, it is advised to reserve at least 37.3% of the capacity of one polyclinical operating room to prevent not being able to treat the required number of patients.

The outcomes as well as the methodologies can have a clear impact at MST. The outcomes as described can be applied when the moment of reallocating the treatments arrives since the required capacity is determined. Besides, the seasonality analysis and the data analysis could be used in studies that require data from the ENT department. Furthermore, the used method could, when using the right data and adjustments be of value for other departments. If MST decides to use these results, the used data in Appendix 3 has to be taken into account. The presented results could be improved by the execution of a simulation study. Further research could focus on the production of such a model taking into account the used methods and analysis since simulation allows for the incorporation of real-world complexity and variability that is difficult to capture in queueing theory models.

References

Bobelin, L. (2016, December). Shortest Processing Time First Algorithm for Hadoop. *Cyber Security and Cloud Computing*, pp. 199-123.

Boucherie, R., Braaksma, A., & Tijms, H. (2022). *Operations Research Introduction to Models and Methods.* Singapore: World Scientific Publishing.

Cheriyan, J., & Skutella, M. (2021). *A simple proof of the Moore-Hodgson Algorithm for minimizing the number of late jobs.* Cornell Univeristy Press.

Chopra, S. (2019). *Supply Chain Management.* Harlow: Pearson Education Limited.

Cooper, D., & Schindler, P. (2011). *Business Research Methods.* New York: McGraw-Hill.

Creemers, S., & Lambrecht, M. (2010, October 15). Queueing models for appointment-driven systems. *Annals of Operations Research*, pp. 155-172.

Darvish, S. (2012, October 4). Minimizing the total tardiness and makespan in an open shop scheduling problem with sequence-dependent setup times. *Journal of Industrial engineering International*, pp. 1-13.

Green, L. (2017). *Queueing theory and Modeling.* 2019: Colombia University Graduate School of Business.

Hall, R. (2012). *Handbook of Healthcare System Scheduling.* New York: Springer.

Heerkens. (2015, Januari 1). MPSM Types of Problems. Enschede: ITC E-Learning.

Heerkens, H., & Winden, A. (2017). *Solving Managerial Problems Systematically.* Enschede: Noordhoff Uitgevers.

Kakull, W., & Ganguli, M. (2012, June 5). Generalizability. *Neurology*, pp. 1886-1891.

Kenton, W. (2020, November 30). *Seasonality: What It Means in Business and Economics, Examples*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/seasonality.asp

Saunders, M. (2019). *Research Methods for Business Students.* New York: Pearson.

Shier, R. (2004). *Chi-squared goodness of fit test.* Mathematics Learning Support Centre.

Sitepu, S. (2016). International Conference on Operational Research. *Optimization Model for Capacity Management and* (pp. 1-8). Gothenburg: ECS.

Xing, W., & Zhang, J. (1999, December 13). Parallel machine scheduling with splitting jobs. *Discrete Applied Mathematics*, p. 259.

Appendix 1. Queueing Theory Calculations

Arrival Rate

The current situation arrival rate λ_1 is based on historical data from December 2021 to January 2023. During these months, there were 219 treatment days scheduled for the ENT department in a single operating room. In this case – except for emergencies – the treatments were executed during shifts of 8 hours. The total number of treated patients equals 1692. Therefore, the arrived patients per day equals $1692/219$ which equals 7.21 arrived patients each day. Furthermore, since there are only 8 hours available each day, the arrival rate for each hour equals $7.21/8$ which equals 0.9657 arrived patients per hour. The Poisson process can be split by multiplying the process type by the probability of occurrence (Boucherie, Braaksma, & Tijms, 2022). Since from the 1692 patients 73 are identified as emergency patients, the probability of arrival to be an emergency patient equals $73/1692$ which equals 4.31% which leads to a percentage of a non-emergency arrival of 95.69%. This leads to an emergency arrival rate of $\lambda_{1e} = 4.31\% * \lambda_1 = 0.33$ patients per day or 0.042 patiens per hour. The non-emergency arrival rate of $\lambda_{ne} = 95.69\% * \lambda_1 = 7.39$ patients per day or 0.9235 patients per hour.

The total number of treatments in a polyclinical operating room after reallocation equals 1114. In this case, the number of days equals 163 since not all days include treatments that are eligible for reallocation. Therefore, the arrival rate λ_2 equals $1114/163 = 6.83$ patients a day. This equals an arrival rate λ_2 of 0.85429 per hour. In the case of the reallocated polyclinical treatments, 40 turned out to be emergency patients. Therefore, the probability of a random patient being an emergency patient equals $40/1114$ which equals 3.59%. Therefore, the percentage of non-emergency patients equals 96.41%. This leads to an emergency arrival rate of $\lambda_{2e} = 3.59\% * 6.83 = 0.245$ per day or 0.031 per hour and a non-emergency arrival rate of $\lambda_{2n} = 6.83 * 96.41\% = 6.5848$ patients per day or 0.823 patients per hour.

The total number of treatments in an operating room after reallocation equals 531. These treatments are spread out over 202 days. Therefore, the arrival rate per day λ_3 equals $531/202 = 2.62$ or an arrival λ_3 of 0.329 per hour. In this case, 28 patients turned out to be emergency patients while 503 patients turned out to be non-emergency patients. Therefore, the probability of a random patient being an emergency patient equals 28/503 which equals 5.57% which means that the probability of a non-emergency patient equals 94.43%. Therefore, the arrival rate λ_{3e} of emergency patients equal $\lambda_3 * 5.57\% = 0.1459$ and the arrival rate λ_{3n} equals $\lambda_3 * 94.43\% = 2.474$.

Service Rate

The average treating time of patients before reallocation – which follows a gamma distribution – equals 47.6253 minutes. Therefore, the treating rate $E[B_1]$ of patients equals 1.26 for each hour which equals a treating rate – considering an 8-hour shift - of 10.0787 per day. In the same situation, the average treating time for non-emergency patients without patient reallocation equals 46.5 minutes which translates to a treatment rate of $E[B_{1n}] = 1.29$ treatments per hour or 10.32 patients per day. The treating time for emergency patients equals 72.69 which gives an hourly treating rate $E[B_{1e}]$ of 0.825 and a daily treatment rate of 6.6.

The average treating time of patients at polyclinical operating rooms after reallocation – which follows a gamma distribution – equals 26.22 minutes which translates to a treatment rate $E[B_2]$ of 2.29 patients per hour or 18.3 patients per hour when considering a single polyclinical operating room. In the same, way, the average emergency treatment time equals 80.225 which translates to a treatment rate $E[B_{2e}]$ of 0.7479 per hour or 5.98 per day. The non-emergency treatments have an average treatment time of 24.20 which translates to a treatment rate of $E[B_{2n}]$ of 2.47 patients per hour or 19.82 patients per day when taking into account one polyclinical operating room.

The average treating time of patients at operating rooms after reallocation - which does not follow a gamma distribution when taking into account an α of 5% but has clear indications of a gamma distribution when taking into account the goodness-of-fit test and the Q-Q plot – equals 92.258 minutes which translates to a treatment rate $E[B_3]$ of 0.65 per hour or 5.203 treatments per day. In case of the emergency patients, the average treatment time equals 68.03 minutes which translates to a treatment rate $E[B_{3n}]$ of 0.8819 per hour or 7.06 per day. The non-emergency treatments have an average treatment time of 93.6 which translates to an hourly treatment rate of 0.631 and a daily treatment rate $E[B_{3e}]$ of 5.1279.

OR Only Situation Calculations

$\sigma(B_c) = 65.12$ minutes.

$c_B = \frac{65.12}{47.62} = 1.3675$

$\rho = \frac{7.72}{10.88} = 0.766$

$L_q = \frac{1}{2}(1 + 1.3675^2) * \frac{0.766^2}{1-0.766} = 2.03$

$W_q = \frac{1}{2}(1 + 1.3675^2) * \frac{0.766*47.62}{1-0.766} = 223.7$ minutes.

The arrival rate of emergency patients per minute $= \frac{0.33}{8*60} = 0{,}000688$

The arrival rate of non-emergency patients per minute $= \frac{7.39}{8*60} = 0{,}015396$

$E[B_e^2] =$ 19410.10 minutes

$E[B_n^2] =$ 5927.59 minutes.

$W_q^{(e)} = \frac{0{,}000688*19410+{,}015396*5927.59}{2(1-0{,}000688*72.69)} = 54.95$ minutes.

$W_q^{(n)} = \frac{0{,}000688*19410+0{,}015396*5927.59}{2(1-0{,}000688*72.69)(1-0{,}000688*72.69-0{,}015396*47.62)} = 253.42$ minutes.

$L_q^{(e)} = 54.95 * 0{,}000688 = 0.0373$ patients.

$L_q^{(n)} = 253.42 * 0{,}015396 = 3.901$ patient

Intended Situation Calculations

Operating Room

$\sigma(B_c) = 69.7744$ minutes.

$c_B = \frac{69.7744}{92.258} = 0.75630$

$\rho = \frac{2.62}{5.20} = 0.5038$

$L_q = \frac{1}{2}(1 + 0.75630^2) * \frac{0.5038^2}{1-0.5038} = 0.40205$

$W_q = \frac{1}{2}(1 + 1.3675^2) * \frac{0.5038*92.258}{1-0.5038} = 134.42$ minutes.

The arrival rate of emergency patients per minute $= \frac{0.1459}{8*60} = 0{,}000304$

The arrival rate of non-emergency patients per minute $= \frac{2.474}{8*60} = 0{,}005154$

$E[B_e^2] =$ 9767,11 minutes

$E[B_n^2] =$ 13581,12 minutes.

$W_q^{(e)} = \frac{0{,}000304*9767{,}11+0{,}005154*13581{,}12}{2(1-0{,}000304*68.03\ )} = 37{,}25359$ minutes

$W_q^{(n)} = \frac{0{,}000304*9767{,}11+0{,}005154*13581{,}12}{2(1-0{,}000304*68.03\ )(1-0{,}000304*68.03\ -0{,}005154*93.6)} = 74{,}9713$ minutes.

$L_q^{(e)} = 37{,}25359 * 0{,}000304 = 0{,}011325$ patients.

$L_q^{(n)} = 74{,}9713 \ * 0{,}005154 = 0{,}386402$ patients.

Polyclinical Operating Room

$\rho = \frac{\frac{6.83}{18.3}}{3} = 0{,}124$

$c\rho = \frac{6.83}{18.3} = 0{,}373$

$C(c,\rho) = \frac{(3*0{,}124)^3}{3!(1-0.124)}\left[\frac{(3*0.124)^3}{3!(1-0.124)} + 1 + (3*0.124) + \frac{(3*0.124)^2}{2!}\right] = 1.42\%$

Appendix 2: Treatments in May 2022

[redacted]

[redacted]

[redacted]

Appendix 3: Treatments Outside of Operating Room

[redacted]

[redacted]

[redacted]

Appendix 4. Systematic Literature Review

Inclusion and Exclusion Criteria

Inclusion Criteria	Argumentation
The article addresses the topic of capacity optimization	Since the research is about capacity optimization, the sources will be valuable if it addresses capacity optimization as well.
The article includes a clear research method and research design.	If it contains a clear research design, it will bring
The article is written in English or Dutch.	If it is not written in English or Dutch, I won't be able to read it.
The article is peer-reviewed	Peer-review secures a high quality of research
The article is published in an academic journal.	If the article is published in an academic journal, it is highly likely that the research has a high quality.

Exclusion Criteria	Argumentation
The article is published more then 20 years ago	If the article is published more then 20 years ago, it may not include important methods that were first used less then 20 years ago.
The article does not address the topic of capacity optimization	If the article does not address the topic of capacity optimization, it is very unlikely that it is useful for the research.
The authors do not hold a masters degree	If the authors do not hold a masters degree, they might not be familiar with techniques for capacity optimization.

Search Matrix

To determine relevant search terms, I identified relevant search terms from the knowledge problem, as well as related, broader and narrower terms. These terms are shown in the search matrix. All these terms could be useful for the search later on.

Key Concepts	Related Terms/Synonyms	Broader terms	Narrow terms
Capacity	Space, quantity, scope	Operations, Volume, Resources, Efficiency	"Operating Room Capacity", "Production Capacity", "Storage Capacity"
Hospital	Clinic, emergency room, hospice, Operating Room, Surgical Theatre	Healthcare, Medical Center, Clinic, Medical Institution	"Public Hospital", "Academic Hospital", "Private Hospital", "Surgical Hospital"
Mathematic*	Calculation, quantitative	Mathematical Sciences, Mathematical Modelling, Mathematical Research	"Optimization Theory", "Statistics", "Dynamical Systems", Algebra

Optimization	Development, increment,	Operations Research, Management Science, Industrial Engineering, Heuristics	“Linear Programming, Dynamic programming, Nonlinear Programming, Stochastic Optimization
Planning	Outlining	Health Planning, Project Planning,	Production Scheduling, Site Planning, Project Management Planning

Concept Matrix

Concepts / **References**	**Total concepts per source**	**Nonlinear Programming**	**Stochastic Modelling**	**Capacity Allocation**	**Linear Programming**	**Simulation**	**Heuristics**	**Algorithms**	**Binary Programming**	**Target Ratio Methods**	**Dynamic Programming**	**Queueing Theory**
Total sources per concept		2	3	2	8	3	5	5	5	1	2	1
Kokangul (2008)	2	X	X									
Burdett *et al.* (2023)	1			X								
Mahaffey *et al.* (2003)	1				X							
Masmoudi *et al.* (2021)	10	X	X	X	X	X	X	X	X		X	X
Keyhanian *et al.* (2018)	2							X	X			
Saadouli *et al* (2010)	1					X						
Chaussalet *et al.* (2007)	2							X		X		
Gartner *et al.* (2023)	1								X			
Shafaei & Mozdgir (2018)	2				X				X			
Sitepu *et al.* (2018)	1				X							
Xiaoqiang & Min (2020)	2		X		X							
Amir *et al.* (2021)	1						X					
Kang *et al.* (2020)	2							X	X			
Pasandideh *et al.* (2019)	3				X		X	X				
Li & Ya (2022)	2						X				X	
Hans *et al* (2015)	2				X		X					
Hall et al. (2012)	2					X						X
Rachuba *et al.* (2021)	2				X	X						

UNIVERSITY
OF TWENTE.

MST