



Long-term multi-priority surgery scheduling

A case study for the gynecology department
S. B. Vertregt

Long-term multi-priority surgery scheduling

A case study for the gynecology department

by

S. B. Vertregt

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday June 28, 2018 at 10:00 AM.

Student number: 4255046
Project duration: September 1, 2017 – June 28, 2018
Thesis committee: Prof. dr. ir. K. I. Aardal, TU Delft
Dr. ir. J. T. van Essen, TU Delft, supervisor
Prof. dr. F. W. Jansen, LUMC
Dr. ir. N. M. van de Vrugt, LUMC, supervisor

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Preface

Six years ago, I started my adventure of studying mathematics at the Delft University of Technology. At first, I was disappointed that I was not selected to study medicine, but after 3 weeks of studying, I was very happy that it happened. Now, this adventure is coming to an end and this thesis is the final requirement for the Applied Mathematics Master of Science degree at the Delft University of Technology. This research was performed in collaboration with the Leiden University Medical Center. So indeed, there are more roads that lead to a hospital.

I learned a lot while writing this thesis and I could not have done this without the help of a few people. First of all, I would like to thank my supervisors at LUMC and Delft University of Technology. Maartje and Theresia, thank you for helping me and introducing me to the interesting world of health care optimization. You both motivated me to do the best I can. You took the time to read this report many times and give me lots of feedback. Karen, thank you for stepping in while Theresia was absent. Frank-Willem, thank you for guiding me through the gynecology department and introducing me to the right people.

I want to thank my friends, who are also graduating soon. You helped me through the difficult parts, not by understanding what I was doing, but just by listening and letting me think it through out loud. My parents, thank you for believing in me, supporting me and for all the advice you have given me. Lastly, I want to thank my boyfriend for the way he handled my stress and insecurities, for the many motivational speeches and support.

*S. B. Vertregt
Leiden, June 2018*

Summary

Hospitals have the difficult task to organize their processes more efficiently and effectively. The health care sector is under pressure as the demand is rising and the resources are limited. Therefore, it is important to make the most of the resources available. Operating rooms are among the most critical resources that generate the highest costs for a hospital. For this reason, the planning and scheduling of operating room activities have become major priority for hospitals. In this report, we describe our research performed at the gynecology department of Leiden University Medical Center (LUMC). The goal of this research is to find a scheduling method for the gynecology department so that we maximize the utilization of the operating room time and lengthen the scheduling horizon without increasing the number of rescheduled patients.

In Chapter 2, we give an analysis of the current patient flow within the hospital and gynecology department. The analysis shows that overall the OR demand and OR capacity are well balanced, and thus, it should be possible to make a schedule with reasonable access times for all patients. In 2015 until 2017, the OR utilization was around 80%, which is good. However, the desired access times are not achieved on average and approximately 60% of priority 1 and 2 patients are treated on time. Moreover, patients receive their surgery date one week in advance, which gives them no time to prepare for it.

In Chapter 3, literature is discussed that is related to our problem. We discuss clustering techniques which we can use to form patient types of the surgical procedures such that we can make efficient schedules. We discuss three solution methods; Markov decision process method, protection level method and integer programming method. After consulting the gynecology department, we started with formulating the scheduling problem as a Markov decision process. One reason for this was that the schedulers of the gynecology department already had some sort of protection levels in place and were interested in a new approach.

In Chapter 4, we use both K -means clustering and Ward's hierarchical clustering method to determine patient types. The clustering methods result in 10 patient types with different session times. Each patient type contains surgical procedures with the same priority.

In Chapter 5, we formulate the scheduling problem as an MDP. The MDP formulation has a reduced state space of size $3 * 10^{30}$, which is too large for our computer to compute. We could have used approximate dynamic programming, but decided to try a different method due to inexperience with approximate dynamic programming.

In Chapter 6, the multiple MIP method is introduced. The method consists of three different scheduling approaches for the 3 priority groups. The MIP models, used for scheduling patients, are designed to fill up the days with a bounded number of patients from each patient type. There is no rescheduling of patients after a patient is scheduled.

In Chapter 7, we introduce four scheduling methods that we could use to compare the performance of the multiple MIP method with: First-come-First-Serve method, protection level method, Just-In-Time method and Master Surgical Schedule method. The advantages and disadvantages are discussed for each method. We decided to only look into the protection level method, JIT method and MSS method, as the FCFS does not include patient preference and long term scheduling.

In Chapter 8, the performance of the four methods was measured based on four aspects; the achieved access times, the number of on time scheduled patients, the total number of scheduled patients and the OR utilization. The conclusion is that the multiple MIP method, protection levels and the JIT method are most suited for use, however, all methods have some aspect that they do not perform well enough on.

In Chapter 9, the conclusion of this research project is given. Moreover, recommendation are made for future work and for the gynecology department on how they can improve their OR schedule and how our model can be used and what to expect.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Context	2
1.3	Problem description	2
1.4	Research questions	2
2	Current situation	5
2.1	Processes.	5
2.2	Scheduling process	6
2.3	Data analysis	7
2.3.1	Priority patient groups	7
2.3.2	Demand for the gynecology department	8
2.4	Performance measures	9
2.4.1	Operating room utilization	9
2.4.2	Access time.	9
3	Literature review	11
3.1	Clustering surgical procedures.	12
3.2	Solution methods	12
3.2.1	Protection levels	13
3.2.2	Markov decision process.	13
3.2.3	Integer programming	14
3.3	Conclusion	14
4	Patient types	17
4.1	<i>K</i> -means clustering.	17
4.2	Ward's hierarchical clustering	18
4.3	Conclusion	19
5	Markov decision process	21
5.1	Markov decision process concept	21
5.1.1	State and action sets	21
5.1.2	Rewards and transition probabilities.	22
5.1.3	Decision rule and policies	22
5.2	Formulation of MDP	22
5.3	Conclusion	23
6	Multiple mixed integer program method	25
6.1	Set up of the solution method	25
6.2	Set up of mixed integer program.	28
6.3	Set up of simulation	29
7	More accessible approaches	31
7.1	First come, first serve	31
7.2	Just in time	31
7.3	Protection levels	32
7.4	Master surgical schedule.	32
7.5	Conclusion	33
8	Computational results	35
8.1	Set-up simulation	35
8.1.1	Scheduling constraints	35

8.1.2	Patient arrival	36
8.1.3	Performance measures	36
8.2	Results multiple MIP method	37
8.2.1	Average access time	37
8.2.2	Not scheduled patients	37
8.2.3	Utilization	38
8.3	Results more accessible methods	39
8.3.1	JIT method	39
8.3.2	Protection level method	39
8.3.3	Master surgical schedule method	40
8.4	Conclusion	41
9	Conclusions and recommendations	43
9.1	Conclusion	43
9.2	Recommendations	44
9.2.1	Future work	44
9.2.2	Practical recommendations	44
A	Ward's hierarchical clustering	47
B	Patient arrival rates	49
C	Late scheduled patients	51
	Bibliography	55

Introduction

In this report, we describe our research performed at the gynecology department of Leiden University Medical Center (LUMC). The goal of this research is to find a scheduling method for the gynecology department so that we maximize the utilization of the operating room time and lengthen the scheduling horizon without increasing the number of rescheduled patients.

We develop a scheduling method based on Banditori et al. [3]. Contrary to Banditori et al. [3], we do not incorporate the length of stay in our model and do not make a master surgical schedule, but use the mixed integer linear program to schedule patients once they arrive, while reserving time for other patients that can arrive with higher urgency. Moreover, we implement a longer scheduling horizon and use different models for different patient groups.

In Section 1.1, the motivation for this research is explained. In Section 1.2, we explain more about our stakeholder the LUMC. In Section 1.3, a problem description is given and the research goals are stated. In Section 1.4, we introduce our research questions and give an outline of the rest of the report.

1.1. Motivation

Hospitals have the difficult task to organize their processes more efficiently and effectively. The health care sector is under pressure as the demand is rising and the resources are limited. Therefore, it is important to make the most of the resources available. Operating rooms are among the most critical resources that generate the highest costs for a hospital. For this reason, the planning and scheduling of operating room activities have become major priority for hospitals, see Lamiri et al. [11]. Furthermore, they want to improve their 'customer' services. Long waiting times and cancellation will result in dissatisfied patients.

The goal of this research is to make the scheduling process better for both the hospital as for the patients. By increasing the scheduling horizon, the patients are earlier informed about when their surgical procedure is performed, and by maximizing the utilization, the hospital is making the most of their available resources. When patients are earlier informed, they have more time to prepare for the weeks after the surgical procedure. Some surgical procedures performed by the gynecology department are more invasive than others, which means a patient can need 2 up to 6 weeks to fully recover from the surgical procedure. During these weeks, the patient cannot ride a bike, drive a car or perform heavy (domestic) work (like cleaning or doing groceries). Most importantly, the patient cannot work during this period. Therefore, the patient needs to arrange a replacement at work and help at home. This is difficult to arrange at short notice, thus, planning ahead is required.

Currently, the surgery schedule is made a week in advance. This makes planning ahead for patients not possible now. The gynecology department wants to improve this situation by informing the patients earlier in which month, week or day their surgical procedure is scheduled.

1.2. Context

This research is conducted at the Leiden University Medical Center (LUMC), which was founded in 1996 from a corporation between the Academic Hospital Leiden and the Faculty of Medicine of University Leiden. LUMC is a modern university medical center for research, education and patient care. LUMC aims to provide the highest quality, both in care and attention for patients. Openness and information are key. Besides the daily patient care, the hospital has a function in the education of medicine and biomedical science students. The hospital does not only educate the students in their field of study, but also teaches them about medical scientific research. The interaction of fundamental research and patient care forms the basis of the policy within the hospital.

The gynecology department focusses on the surgical and medical treatment of the female reproductive system and the breasts, including fertility disorders. The department consists of 20 doctors, which are distributed over multiple specialties (fertility, psychology, oncology and general gynecology).

1.3. Problem description

In this research, we only consider elective patients. Emergency patients are being treated in the emergency operating room and we do not need to take them into account with the operating room planning.

In the current scheduling of the operating room time of the gynecology department, unused hours are being observed. One of the causes is the desired access time of different elective patients. While scheduling patients, there is room reserved for patient with higher urgency. However, the desired access times are not met in all cases.

We can separate three groups of elective patients: group 1 has a desired access time of maximum two weeks, group 2 has a desired access time of maximum four weeks and group 3 has a desired access time of three months. If we would make a schedule with a scheduling horizon of 2 months, then the group 3 patient is first to be scheduled in a week in the future (more than four weeks away) and later group 2 and group 1 patients are scheduled around the group 3 patients. As group 1 and 2 have higher urgency, this is not the desired order of scheduling. Therefore, it is important to reserve enough time, while scheduling group 3 patients, for patients of group 1 and 2. However, if we reserve too much time and this is not filled up with group 1 or 2 patients, the utilization is not maximized.

Moreover, the gynecology department wants to take into account the patient preference. Especially some priority 3 patients like to schedule their surgical procedure around special events in their life as the procedure is not urgent. Therefore, the model would be best if patient preference could be incorporated in some manner.

The main difficulty of the scheduling process is to find the balance between maximizing utilization and treating patients within their desired access time.

We have set the following goals for this research;

- Maximize the OR utilization.
- Minimize the cancellation/rescheduling of patients.
- Lengthen the planning horizon.
- Lengthen the time between the moment of known operating date and the operating date itself.
- Treat the priority groups within their set access time.

1.4. Research questions

We have defined the following research questions to structure the report. Each question will be answered in a new chapter.

What is the current situation and the performance of the situation? (Chapter 2)

This chapter is based on interviews we had with the schedulers of the gynecology department in LUMC

and data we received. We will explain current planning methods and define performance measures to evaluate the performance of our method.

What models or solution methods are currently used in literature on similar problems? (Chapter 3)
In this chapter, we discuss previous research done in the department of OR planning and choose a direction for our method. We explain which research papers from our literature review are used and what we contribute with our approach.

How can we define patient types for the scheduling process?(Chapter 4) *In this chapter, we explain two clustering methods to define the patient types that we use in the scheduling process.*

Which models or solution methods are used? (Chapter 5, Chapter 6, Chapter 7)
In these chapters, we discuss mathematical models that can be used to solve our problem. In Chapter 5, an exact method is explained which takes into account all possible future events when making a decision. In Chapter 6, a method is discussed which takes into account the future but in a different manner. The method consists of different models for each priority group. In Chapter 7, more accessible methods are discussed to compare with the method from the previous chapter.

What is the performance of our methods? (Chapter 8)
In this chapter, we compare our solution methods with the defined performance measures. What is the best method for the gynecology department?

What are the conclusions and recommendations for the future? (Chapter 9)
In this chapter, the conclusions of our research are presented and recommendations are given for future research.

2

Current situation

In this chapter, an analysis of the performance of the gynecology department is given. The patient flow through the hospital is explained in Section 2.1. An overview of the current scheduling process of the gynecology department and scheduling restrictions are given in Section 2.2. The operating room data is explained in Section 2.3 and some interesting facts are highlighted. Lastly, the performance measures that we use to test our method are discussed in Section 2.4.

2.1. Processes

In this section, the processes in the hospital involving surgical patients are explained for both elective and emergency patients.

We consider the typical route of an elective surgical patient in the gynecology department through the hospital, who stays longer than a day, see Figure 2.1. A patient's journey through the hospital usually starts at the outpatient clinic. In the outpatient clinic, diagnosis and care are provided. During the diagnostic phase, the doctor can determine whether a surgical procedure is needed for the patient. Then the patient is put on the waiting list. Before the actual surgical procedure is performed, the patient will get a pre-operative screening. On the day of the scheduled surgical procedure, the patient is admitted to the hospital and assigned to a bed on the inpatient ward. The patient is then taken to the operating room, where the surgical procedure will be performed. After the surgical procedure is performed, the patient is taken to the recovery room and once the patient is stable enough, she will return to the patient ward and eventually will be discharged.

The typical flow of an emergency patient is also displayed in Figure 2.1. The journey of an emergency patient is shorter and starts at the emergency room. From the emergency room, they are usually transported straight to the operating room after which they will follow the same route as the elective surgical patients. In the LUMC, there are dedicated emergency operating rooms so that emergency patients can be treated immediately.

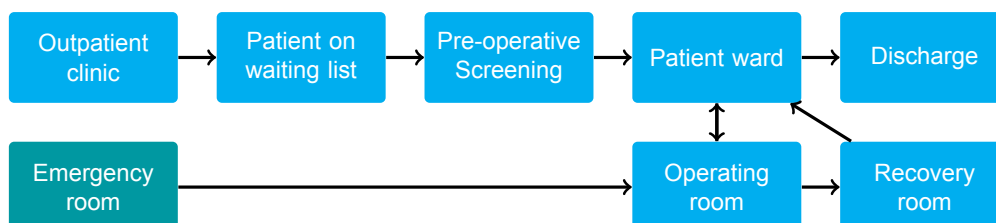


Figure 2.1: Surgical patient flow

A surgical procedure consists of several steps, as displayed in Figure 2.2. The patient is transported from pre-operative holding to the operating room (OR). The moment the patient enters the operating room

is called *operating room in-time*. At that time, the surgeon is not yet in the operating room. First, the anesthetist puts the patient under general anesthesia. This is called *the induction time*. Next, the patient is positioned, equipment is connected and the instruments needed during the surgery are made ready for use. Once the patient is ready, the surgeon is called to perform the surgical procedure. The moment at which the surgeon starts operating is called *start intervention time*. The point when the surgeon closes the wound is called *end intervention time* and at that time, the surgeon leaves the operating room. Then the equipment is disconnected and the anesthetist brings the patient out of general anesthesia. The time the patient is transferred to the recovery area is called *operation out-time*. The *operation out-time* minus *operation in-time* is called the *session time*; this includes the *induction* and the *intervention time*.

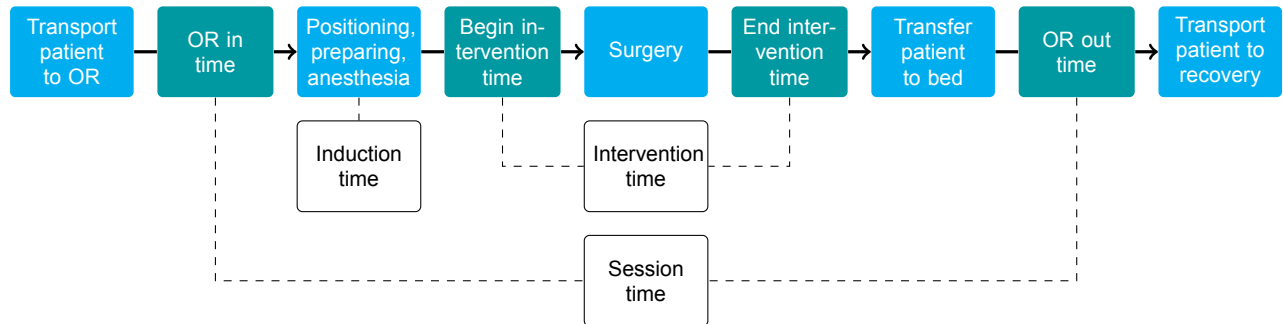


Figure 2.2: Surgical process

In Figure 2.1, we mentioned areas and departments that are important for the gynecology patients. The explanation of what those areas and departments do is stated in the following list;

Outpatient clinic An outpatient clinic is the part of a hospital that is designed for the treatment of outpatients, people with health problems who visit the hospital for diagnosis or treatment, but who do not at that time require a bed or need to be admitted for overnight care.

Pre-operative screening (POS) Pre-operative screening is done by an anesthetist in order to fully characterize the patient's health condition and to make sure that the patient is in an as good as possible condition for the surgical procedure. A brief health check is done and information about the surgical procedure is provided.

Operating room The LUMC has twenty operating rooms which are organized on the basis of a master surgical schedule. Two operating rooms are reserved for incoming emergency patients. This master surgical schedule allocates the operating rooms across the surgical specialties so that every surgical specialty has a number of allocated full days in an operating room to treat their patients. In this schedule, the gynecology department has one operating room for four days each week.

Recovery room The recovery room or post-anesthesia care unit (PACU) is a space that a patient is taken to after surgery to be able to safely regain consciousness after anesthesia and to receive appropriate post-operative care. During a patient's stay in the recovery room, their vital signs are monitored closely as the effects of the anesthesia wear off. The amount of time a patient needs in the recovery room varies depending on the surgical procedure and type of anesthesia used.

Patient ward A patient ward is a medical specialty facility in the hospital where patients get specific care. In the LUMC, there are three types of wards: day clinics, short stay (less than five days) and long stay (more than 5 days).

2.2. Scheduling process

The master surgery schedule of the LUMC allocates the operating rooms to the various surgical specialties. This cyclic schedule is repeated every two weeks. The schedule allocates a number of operating days to the gynecology department, which need to be filled with patients by the department herself.

Two people are responsible for making the patient schedule and they do this each Monday morning. They schedule only the intervention time, based on experience. In busy periods, the head nurse of the

gynecology ward joins the scheduling session to make sure that the number of available beds is not exceeded by the surgical schedule.

The scheduling process is complicated. More than half of the surgical procedures are specialized procedures. The specializations include fertility, pelvic floor, oncology and laparoscopy. There are at most 2 days in the week that an oncology surgeon is operating and oncology patients can be treated. Once every two weeks the surgeon specialized on pelvic floor problems is operating and patients with pelvic floor problems can be treated. Furthermore, the surgeons that are specialized in fertility problems and laparoscopic surgery operate one day every week.

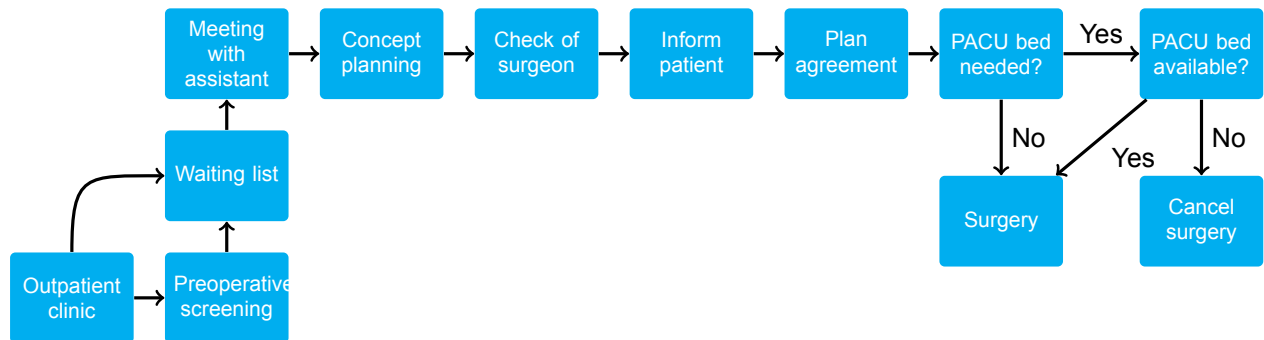


Figure 2.3: Scheduling process

After the patient has seen the surgeon, who has decided a surgical procedure is necessary, the patient sees the physician’s assistant and discusses what days/weeks are not possible for surgery. A week prior to surgery the concept planning is completed and checked by the responsible surgeon. The next day, the physician’s assistant calls the patients, that are planned in the concept planning, to confirm their surgery date. In a few cases an PACU bed is needed for the patient. Then a request for a PACU bed is made on Tuesday. Every Thursday an announcement is made which patients have a PACU bed the next week. If the patient of the gynecology department receives a bed for next week, the patient is called again to confirm. If there is no bed the surgical procedure is cancelled.

2.3. Data analysis

We received anonymous OR patient data from November 2014 until January 2018. The data set has the following properties; patient number, surgery number, OR number, ward name, surgeon name, surgery request date, surgery date, session time, corrected session time, intervention time, diagnosis code and surgical procedure description. The data set contained 1900 entries of surgical procedures performed by the gynecology department. In this section, we give an overview of the data and highlight some interesting facts.

2.3.1. Priority patient groups

The access times are different for all patients at the gynecology department. However, we can distinguish three groups of patients, say priority groups. Priority group 1 has the highest priority and the hospital does not want priority group 1 patients to wait more than 2 weeks to be treated. There are 7 surgical procedures included in priority group 1 and they have all similar session time, on average 60 minutes. Priority group 2 contains 30 surgical procedures which vary in session time between 60 and 300 minutes. The surgical procedures are mostly oncological procedures, and therefore, the hospital does not want priority group 2 patients to wait for more than 4 weeks to be treated. Lastly, priority group 3 contains 56 surgical procedures which vary in session time between 60 and 400 minutes. The patients in priority group 3 need to have surgery, but they do not have an urgent deadline. The hospital wants to have priority group 3 patients treated within 3 months, however, if the patient chooses to wait longer this is also possible.

In Figure 2.4, the percentage of OR minutes used by the different priority groups is shown. We know

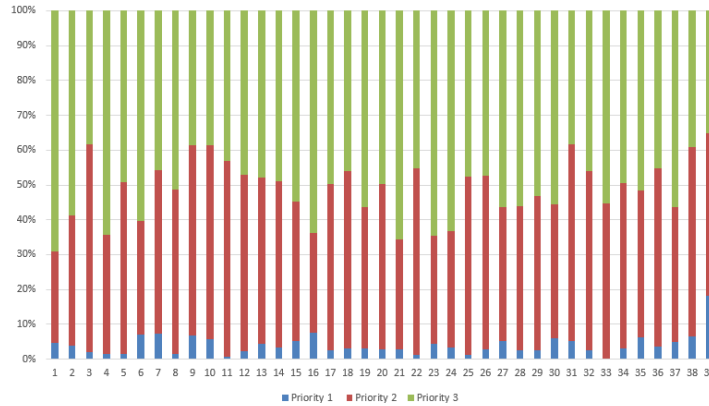


Figure 2.4: Requested minutes percentage of priority 1, 2, 3 per month

that on average priority group 1 requests 5% of the total requested minutes in a month. Priority group 2 requests on average 45% and priority group 3 requests 50% of the total requested minutes in a month.

2.3.2. Demand for the gynecology department

When scheduling patients, it is important to keep in mind what the overall demand is. It is difficult to make a schedule when the patient demand is higher than the assigned OR hours. In this case, the waiting list will increase and patients of priority 3 will have to wait for a longer period. If the patient demand is low, the waiting list will decrease and the patients of priority 3 will have to wait for a shorter period. The gynecology department has approximately 8000 minutes per month of OR time assigned. In Figure 2.5, the requested minutes and the performed minutes are shown. We can see that the overall demand is usually between 6000 and 8000 minutes. However, there are some highs and lows.

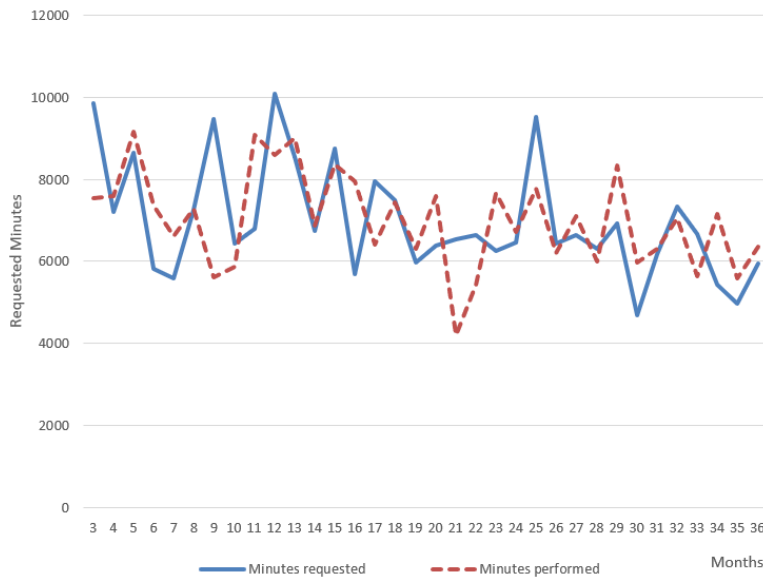


Figure 2.5: Requested and performed surgical procedure minutes per month

It is clear that overall the requested surgical procedures and the assigned OR hours are well balanced and making a schedule with reasonable access times should be possible. A small downward trend is visible, the gynecology department confirmed that the waiting list is decreasing. However the number of oncology patients is not decreasing.

2.4. Performance measures

In this section, we describe the measures we use to evaluate the performance of the system. The problems regarding operating room utilization and under/overtime is discussed in Section 2.4.1. In Section 2.4.2, the access times of patients are discussed.

2.4.1. Operating room utilization

Most of the surgical procedures performed by the gynecology department are carried out in operating room 9. In this section, we look at the utilization of operating room 9 and the minutes of overtime worked in operating room 9. It can be seen in Table 2.1 that the utilization percentage has not improved since 2015. The number of times overtime was needed has dropped slightly, and the average overtime needed, when overtime occurred, has decreased.

Year	2015	2016	2017
Utilization(%)	79%	83%	79%
Number of overtime	30	36	27
Average overtime (min)	57	46	38

Table 2.1: Utilization and overtime in OR 9

2.4.2. Access time

The gynecology department has set targets for the access times for all patients, based on the surgical procedure requested for them. The patients can be split up into 3 priority groups based on the set target. Priority group 1 includes patients who need to be treated within 14 days, priority group 2 includes those who need to be treated within 28 days and the third group includes those who we want to treat within three months. The data shows different access times than the desired access times (Table 2.2, Table 2.3 and Table 2.4).

Table 2.2: Access times of priority 1 quantiles - target is 100% in 14 days

p	0%	5%	25%	50%	75%	95%	100%
p-quantile in days	0	2	6	12	20.5	52.5	89

Table 2.3: Access times of priority 2 quantiles - target is 100% in 28 days

p	0%	5%	25%	50%	75%	95%	100%
p-quantile in days	0	8	17	26	34.25	61.85	148

Table 2.4: Access times of priority 3 quantile - target is 100% in 100 days

p	0%	5%	25%	50%	75%	95%	100%
p-quantile in days	1	8	23	37	61	125.9	368

The conclusion we can draw from these tables is that priority 3 patients are scheduled too soon and priority groups 1 and 2 are scheduled too late. If priority 3 patients were scheduled later, than there could have been enough time for priority 1 and 2 to be treated within their set access time. Now it seems as if priority 3 patients are filling up time that is meant for priority 1 and 2 patients. When discussing the results with the planners from the gynecology department, it was explained that for priority 2 it can sometimes be the case that the request for a surgical procedure has already been entered into the system, while the pre-treatment before surgical procedure is not yet finished and takes a few weeks. Then, the actual target time for the patient is indeed longer than the set target for the priority group. Only 60% of the patients in priority group 1 is treated within 14 days and only 58% of the patients in priority group 2 is treated within 28 days.

3

Literature review

Since the 1950s, the application of operation research to health care has made significant contributions to optimizing health care delivery (Hulshof et al. [10]). Many different topics have been addressed, such as operating room planning, nurse staffing and appointment scheduling. One of the most critical resources is the operating theatre (OT) as it is costly and a high percentage of hospital admissions are linked to surgical interventions (Guerriero and Guido [6]). So a large part of the research is done in the field of OT optimization.

There are several factors that influence the OT scheduling. The resources needed to perform a surgical procedure are personnel (surgeon, nurses, anesthetists) and facilities (equipment, pre-surgical capacity, an operating room, post-surgical capacity and intensive care capacity). The main aim of hospital managers is to secure the optimal utilization of medical resources, ideal access times for patients and to maximize patient flow without incurring additional costs.

Both Hulshof et al. [10] and Guerriero and Guido [6] give a good structured overview of the research that has been done and how operational research can help to improve the OT planning. The taxonomy for resource capacity planning and control is given below in Figure 3.1 as is presented in Hans et al. [7]. Every overview paper describes their own taxonomy, which always includes the three levels of planning derived from operational research, i.e. strategic, tactical and operational.

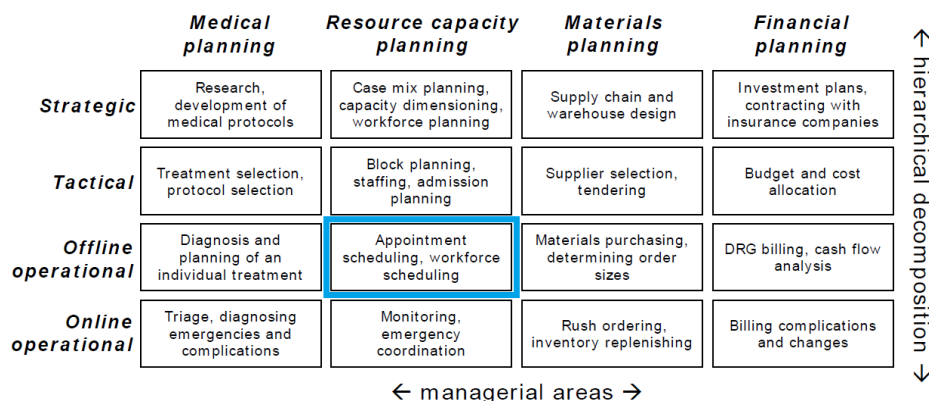


Figure 3.1: Framework for health care planning and control from Hans et al. [7]

The strategic level addresses structural decision making. It involves defining the organization's mission, it has a long planning horizon and it is based on highly aggregated information and forecasts. The tactical level translates strategic planning decisions to guidelines that facilitate operational planning decisions. It addresses the organization of the operations and the execution of the health care delivery process. The operational level involves short-term decision making related to the execution of the health care delivery

process. It follows the tactical blueprints. Execution plans are designed at individual patient level and at individual resource level.

In Figure 3.1, the operational level is divided in offline and online. Offline planning reflects the in advance planning of operations. It comprises detailed coordination of the activities regarding current (elective) demand. Online planning reflects the control mechanisms that deal with monitoring the process and reacting to unplanned events. Our research is located in the 'Resource Capacity Planning' - 'Offline operational' block as we want to schedule elective patients further in the future. We do not have to take emergency patients into account, as they are treated in the emergency operating room. Our main resource is the given operating room for which we need to plan the given surgeries from a waiting list of patients.

An important aspect of a good solution method is that is relatively fast in giving a solution. If the running time is long, it is inconvenient for the gynecology department to use it. One approach to make sure the solution method is faster is reducing the problem size. In stead of scheduling specific surgical procedures it is also possible to schedule patient types. A patient type combines multiple surgical procedures based on specific characteristics such as surgery duration, bed occupation or priority level. For example Van Oostrum et al. [15] and Astaraky and Patrick [2] use clustering techniques.

The goal of this thesis is to minimize the idle time of the operating room and the access time of priority 1 and 2 patients, while increasing the planning horizon. We divide this literature review into two parts: the first describes clustering methods and the second part describes solution methods. In Section 3.1, the clustering techniques are reviewed to define patient types. In Section 3.2, three strategies are reviewed that are applicable to this research, namely installing protection levels (Section 3.2.1), representing the problem as a Markov Decision Process (Section 3.2.2) and using integer programming (Section 3.2.3).

3.1. Clustering surgical procedures

Clustering is the task of grouping a set of objects in such a way that objects in the same group or cluster are more similar to each other than to those in other groups or clusters. There are many different algorithms which all have another definition of what a cluster is.

In our research, we want to cluster patients so that we find patient types that can be scheduled in a similar manner. They have, for instance, similar operating, recovery or access time. To get efficient patient types, they need to be constructed in such a way that the variability within one type is low. Patient types that cannot be scheduled repetitively are put together in so-called dummy surgeries. Narrowly defined patient types, with low variability, lead to a large volume of dummy surgeries that reduce the benefits of scheduling with patient types.

In Van Oostrum et al. [15], they propose a method based on Ward's hierarchical cluster method to obtain surgery types. The aim is to construct patient types with a minimal loss of information compared to individual patient case types. The basic principle of the method searches for the best combination of patient types that can be combined into one with minimal costs. This is repeated until only one patient type is left. Then, with the error sum of squares, they calculate the best solution for the patient types, while minimizing the number of dummy surgeries.

In Astaraky and Patrick [2], they use the K -means clustering technique to classify patients into a manageable number of classes. The patients are classified according to their resource consumption. Therefore, within each surgical specialty, surgery length and post-operative length of stay are used as the clustering attributes.

3.2. Solution methods

In this section, we discuss the articles that solve a similar problem as our surgery scheduling problem. The articles can be divided into three solution methods; protection levels, Markov decision process and mixed integer programming.

3.2.1. Protection levels

The main idea behind protection levels is to retain a percentage of the available resource capacity for a certain period in the future in such a way that we are able to schedule patients with a higher priority or shorter desired access time. Planning too few patients in the future will result in an empty operating room schedule, while planning too many patients in the future will result in high priority patients who are not being treated within their desired access time. Both are equally undesirable. The following papers have studied this balance in protection levels;

In Bertsimas and De Boer [5], the problem of optimizing the passenger mix on a single-leg flight is discussed. A simulation-based method for booking-limit calculation is proposed. Starting with any nested booking-limit policy ('expected marginal seat revenue - booking limit' or 'linear programming - booking limit'), they combine a stochastic gradient algorithm and approximate dynamic programming ideas to improve the initial booking limits. The proposed way to calculate booking limits takes into account the stochastic and dynamic nature of the demand and the nested character of booking-limit control in a network environment.

In Vermeulen et al. [16], the focus is on urgencies and preferences for planning of central diagnostic resources, which often are a bottleneck in patient's pathways. They formulate a mixed integer program to compute protection levels. This is done by allocating resource capacity to patient groups. A search method (estimation of Distribution Algorithm) is used to find the best protection levels for a given scenario. Patients are only scheduled in time-slots allocated to their group. The capacity usage is made more efficient with conditional exceptions to the nested capacity allocation, which makes sure that capacity allocated to higher urgencies is also available if its utilization is below a certain threshold. The scheduling solution is a heuristic based on a combination of First-Come-First-Serve (FCFS) and balanced utilization, which counters the negative effects of FCFS.

In Hof [9], the main objective is to minimize the weighted average access time, overtime and idle time of outpatient clinics for the dermatology and urology department. The model determines the expected future demand, which is forecasted with average arrival rates, obtained from historical data. The forecasting can be done in a deterministic or stochastic manner. The decision is made to use forecasting with a stochastic heuristic approach. The demand is approximated by forming a scenario tree, which has a low, medium and high demand scenario. Within each scenario, the demand is approximated with a uniform distribution. The exact amount of capacity to protect each week is calculated by making a simulation of the booking process and counting back the number of patients that were scheduled at what moment. The schedule is then made with the FCFS method.

3.2.2. Markov decision process

A Markov decision process (MDP) presents a mathematical framework for modeling decision making in situations where outcomes are to some extent random, but also under the control of a decision maker. An MDP is a discrete time stochastic control process. At each time step, the process is in a state s and the decision maker can select an action a available at state s . Then, the process responds to the selected action in the next time step which will result into a state s' . The following papers have implemented a Markov decision process to solve a (similar) scheduling problem.

In Wang and Gupta [17], the focus is on the problem of allocating available time-slots to appointment requests to maximize clinic revenue. The paper identifies possible scenarios and suggests implementable heuristics when no simple structure exists. The model describes the clinic's problem of choosing which appointment requests to accept for a particular workday as an MDP with discrete time and a finite horizon. Three models are discussed, which describe the choice probabilities for a patient calling and choosing the j^{th} time slot. They show that for each appointment request with a particular physician, there exists an appointment's threshold beyond which it is not efficient to book additional appointments.

In Patrick et al. [12], a method for dynamically scheduling multi-priority patients to a diagnostic facility is made. The challenge is to allocate available capacity to incoming demand such that waiting time targets are achieved in a cost-effective manner. As it models a diagnostic resource, the state space and action space both represent the number of patients that are or are not yet scheduled. In the results, the model uses a rolling two day horizon and plans thirty days in the future.

In Astaraky and Patrick [2], the waiting time of patients, OT overtime and ward capacity utilization are the main objectives incorporated in the model. The model manages demand by ensuring available capacity for cases that require longer surgical times and longer post-operative stays, while fitting the shorter cases around these longer cases. The model incorporates J surgical specialties, a single bed class, a rolling horizon of length N and the booking decisions are made once a day.

In Barz and Rajaram [4], the elective patient admission control problem for a hospital with multiple constraints is formulated as an MDP. Waiting time and resource capacity utilization are the main objectives incorporated in the model. The hospital provides services for R different resources with capacity c_R . The health state of individual patients is modeled dynamically. The assumption is made that the resource requirements for each patient are available at time of admission.

In most of the cases, the state space is too large to allow for an exact solution. We can resort to approximate dynamic programming to develop heuristics to overcome this problem and to construct a bound to evaluate the quality of the heuristics.

Above described models give a strategy output. The strategy is based on the current waiting list, so a distribution of patients over the different priority groups results in a certain predetermined action.

3.2.3. Integer programming

An optimization problem is called an integer programming problem, when at least one variable is restricted to be integer. The optimization problem is called an integer linear program if all constraints are linear. If some, but not all, decision variables are continuous, the problem is called a mixed-integer programming problem. In the following papers, the optimization problem is formulated as an (mixed) integer (linear) programming problem.

In Adan and Vissers [1], admission profiles are generated per specialty, given a target of patient type throughput and utilization of resources, while satisfying given restrictions. Within a specialty, different types of patients can be distinguished based on their resource requirements. The mix of admissions is an important decision variable for a hospital to manage the workload. Adan and Vissers [1] develop an integer linear program with the objective to minimize the absolute deviation of the realized from the target utilization with a scheduling horizon of one week. The resources taken into account are bed, OR time and nurse availability. The characteristics taken into account are length of stay, pre-operative days, days an IC bed is needed, session time and nursing workload.

In Banditori et al. [3], the objective is to maximize the patient throughput, taking into account the due date of the patients and control of the waiting list. This is done with a combined optimization-simulation approach, which makes the trade-off robust and efficient. A master surgical schedule for one month is made by solving a mixed integer program. The objective of the mixed integer program is to maximize the number of scheduled surgeries and minimizing the penalties resulting from missing due dates and bed mismatches.

3.3. Conclusion

The clustering techniques discussed in Van Oostrum et al. [15] and Astaraky and Patrick [2], can both be applied to our problem. The objective of both the papers is similar to our objective. Therefore, we decided to compare the two different clustering methods to define the patient types. We choose Ward's hierarchical clustering method and K -means clustering method as they are both used often in the literature.

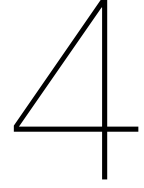
There are differences between the problems introduced in the papers discussed in Section 3.2.1, Section 3.2.2 and Section 3.2.3 and the problem of the gynecology department. First, the gynecology department wants to take the patient preference into account while planning, such that it is possible for the patient to choose a day or week to be treated in. Second, the planning horizon for the gynecology department is about three months, whereas in most papers the planning is done on short term (1-2 weeks). Only Banditori et al. [3] incorporates a planning horizon of 1 month which could be extended. Thus, our model incorporates long term patient arrival. Third, the discussed papers consider diagnostic

facilities, which implies that the treatment time of patients is more or less the same for all patients, while session times of surgical procedures are different for all patients. Concluding, our method is different as it includes a long scheduling horizon, surgical procedures with varying session times and patient preference.

When reviewing the output of the models described in the previous section, they return scheduling rules. In the case of protection levels, these rules are set once for a period and indicate for each period in the future how much capacity may be booked. The MDP results in rules that depend on the waiting list. For each partition of priority groups on the waiting list, a different set of rules is made. The integer programming solutions give scheduling rules that depend on an admissions profile and the rules can in some cases be presented as a schedule. Therefore, protection levels are easier to implement at the gynecology department as the planners have to remember less rules to make the schedule, whereas integer programming solution and MDP gives a more guided solution.

The models have advantages and disadvantages. As stated above, protection levels are easy to use as the planners of the gynecology department will always use the same set of rules, whereas with an Markov decision process the rules change when the waiting list changes. By adapting the rules to the waiting list (demand), we expect that the access time for patients will be lower for a Markov decision process than with protection rules. An integer program solution can be formulated in a block schedule, which makes it easy to use, but the block schedule does need to be refreshed once in a while, which can make it more complicated.

The schedulers of the gynecology department think they already have some sort of protection levels installed and the MDP solution sounds as complex as the integer programming approach. After consulting with the schedulers of the gynecology department, we decided to start with implementing an MDP solution as an MDP gives at every moment the optimal solution. Unfortunately, the implementation of the MDP formulation needed too much memory and , therefore, it was not useable for the gynecology department. Because of this, we also implemented a mixed integer programming method, which we tested via simulation and compared with a more accessible approach.



Patient types

The gynecology department performs more than 90 different surgical procedures. We can cluster these procedures into patient types. Each patient type is a cluster of surgical procedures with the same priority and similar surgical duration. When scheduling the patient types instead of the surgical procedures, the number of possible states in the MDP is restricted, which is beneficial to the runtime and memory usage of the model. When using patient types when solving a MIP, the number of constraints is reduced and the runtime is shorter. In this chapter, we present the method to define the patient types based on priority and session time.

In the literature review, papers are discussed that use clustering to define patient types. Both hierarchical and non-hierarchical clustering methods are used. We use two methods and through comparison decide which clustering method fits best to our problem.

4.1. K -means clustering

K -means clustering is a method that is popular for cluster analysis in data mining. The K -means clustering algorithm used is from Hartigan and Wong [8]. The goal of the method is to find K groups in the data. The data points are clustered based on similarity in one or multiple features. The results of the K -means method are centers of the K clusters and a label for every data point (the label refers to one cluster).

Assume that we have N observations, which need to be separated into K groups. Define n_k as the number of observations contained in cluster k . Each observation has P variables or features. Let x_{ij} be the data for observation i and variable j . Then, the within cluster sum of squares is defined as;

$$WSS_K := \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^P (x_{ij} - c_{jk})^2 \quad (4.1)$$

where c_{jk} is the center value of the j^{th} variable in the k^{th} cluster. The WSS_k is used as the distance measure between the observations within the clusters. If this value is high the distance within the clusters is high. The variation percentage for k clusters is defined as;

$$PV_k := 100 \frac{WSS_k}{WSS_1} \quad (4.2)$$

The variation percentage PV_k is a measure which shows how useful it is to add another cluster. If the variations percentages drops significantly from k to $k + 1$, then clustering the observations in $k + 1$ clusters is better for performance of the clusters. There are 93 distinct surgical procedures that are being performed by the gynecology department, which we want to cluster. The number of clusters is not set, but we do want to cluster surgical procedures with the same priority. We want to get efficient

patient types, so they need to be constructed so that the variability within one type is low. Checking the variability within one type is done with a goodness-of-fit test. The goodness-of-fit criterion used to compare cluster configurations is based on the within-cluster sum of squares, WSS_K .

First, we divide the data into three sets based on the priority of the surgical procedure. Therefore, the number of variables per observation is equal to 1, namely the session time. There are 7 surgical procedures with priority 1, 30 surgical procedures with priority 2, and 56 surgical procedures with priority 3. We applied K -means clustering for different K . The variation percentage within the clusters is given in Table 4.1.

For priority 1, we see that the variation percentage only reduces for four or more clusters. This means

	Priority 1 (7)	Priority 2 (30)	Priority 3 (56)
PV_1	100%	100%	100%
PV_2	100%	24%	30%
PV_3	99%	11%	17%
PV_4	0.8%	5%	8%
PV_5		4%	5%
PV_6			4%

Table 4.1: Variation percentage per priority group and number of clusters

that K -means advises to form four clusters. We see for priority 2 that the variation percentage keeps decreasing significantly until the fifth cluster is added. As the variation percentage of four clusters and five clusters is so similar, we can conclude that four clusters is enough for the 30 surgical procedures of priority 2. Lastly, we see that for the surgical procedures with priority 3 adding a fifth cluster is beneficial to the variation percentage.

4.2. Ward's hierarchical clustering

Ward's hierarchical clustering method is an alternative approach for cluster analysis, see Ward [18]. The method starts with n clusters of size 1 and continues adding clusters together until all observations are together in one cluster. The first time step in adding these clusters is to select two clusters out of the n clusters, which will lead to the least loss of information when adding them together in one cluster.

Adding two different surgical procedures in one patient type together leads to loss of information compared to a situation where both procedure types are individually assigned to a surgery type. This method uses the error sum of squares (ESS) as a measure of the loss of information. Let ESS_c be the error sum of squares of patient type c , which is computed by;

$$ESS_c := \sum_{i=1}^m x_i^2 - \frac{1}{m} \left(\sum_{i=1}^m x_i \right)^2 \quad (4.3)$$

where x_i is the data for observation i and m is the size of cluster c . It measures per cluster how much the observations are distant from the center. Then, the overall ESS is defined as the sum of all ESS_c ;

$$ESS := \sum_{c=1}^K ESS_c \quad (4.4)$$

where K is number of clusters. When adding two clusters together into one, we calculate the ESS_c of this new cluster c . We want to add two clusters together so that the ESS_c is minimal. The overall ESS is used as an overall measure of loss of information when applying clustering. When the ESS is high there is a lot of loss of information.

After adding two clusters together, the remaining $n - 1$ clusters are examined to determine if a third cluster should be added to the first pair or if another pair should be made to make sure that the loss of information is minimal for $n - 2$ clusters. This process continues until all n observations are together in one cluster.

We divide the surgical procedures by priority in three sets and apply clustering on the three sets. The value of x_i is then equal to the session time of surgical procedure i . We apply Ward's hierarchical clustering method. This process of clustering can be displayed in a so-called dendrogram. The dendrogram for priority 1 surgical procedures is shown in Figure 4.1. The dendrogram shows that surgical procedures 7, 4 and 5 have very similar session times. Also, surgical procedures 1 and 3 are similar and surgical procedures 2 and 6. The ESS_c seen on the vertical axis is the measure of distance between the session times of the surgical procedures.

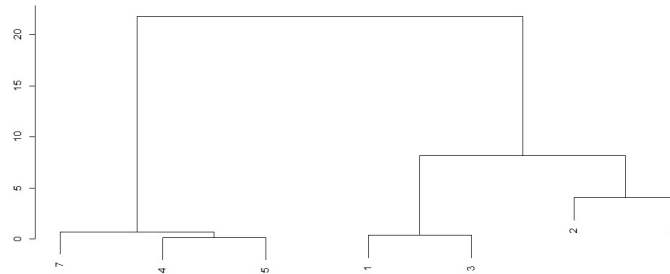


Figure 4.1: Dendrogram for priority 1 with 7 different procedures.

The clustering dendrogram for priority 2 and 3 are shown in Appendix A. When deciding on the number of clusters we look at the overall ESS , if the ESS does not decrease significantly when adding another cluster, the cluster is not necessary for the performance of the clusters. For priority 2, Ward's hierarchical clustering determined four clusters. For priority 3, the clustering determined five clusters. Thus both dendrograms for priority 2 and 3 are cut off around an ESS_c of approximately 100.

4.3. Conclusion

The dendrograms of Ward's hierarchical clustering and variation percentage of the K -means clustering are compared to define the number of clusters for each priority. The clustering methods both gave similar results. For priority group 1, both the dendrogram as the variation percentage resulted in four clusters for the seven surgical procedures of priority 1. However, when looking at the 7 procedures of priority 1, we concluded that the session times of these seven surgical procedures are similar enough to be in one patient type. Priority group 2 contains 30 surgical procedures and is split up into four clusters of size 4, 6, 10 and 10 surgical procedures by the K -means method and Ward's method. We decided to continue with four cluster as the variation percentage is not decreasing drastically when adding a fifth cluster Table 4.1. Priority 3 contains 56 surgical procedures and is split up into five clusters of size 2, 5, 12, 16 and 21 surgical procedures. This is both suggested by K -means and Ward's hierarchical clustering. In Table 4.2, the patient types are listed. The number of surgical procedures in each patient type are given and the number of times these surgical procedures were performed in the last 2 years. The cluster name is based on the priority of the cluster. Cluster 'Prio 3.1' is the first cluster of priority 3 patients. The clusters are not ordered.

Table 4.2: Patient types made with the clustering method

	Average	Occurrences	Surgical
	surgery duration	in data	procedures
Prio 1	55	248	7
Prio 2.1	225	205	12
Prio 2.2	144	70	3
Prio 2.3	79	145	5
Prio 2.4	295	131	5
Prio 3.1	72	449	26
Prio 3.2	125	386	16
Prio 3.3	182	223	12
Prio 3.4	240	20	5
Prio 3.5	374	2	2

5

Markov decision process

A Markov decision process (MDP) presents a mathematical framework for modeling decision making in situations where outcomes are to some extent random, but also under control of a decision maker. In this chapter, we explain what an MDP is and how we implemented an MDP to aid the scheduling decision making. In Section 5.1, the concept of a Markov decision process and necessary conditions for obtaining a good solution are explained. In Section 5.2, we formulate the constraints we have to incorporate into our MDP. Lastly, we will conclude this chapter in Section 5.3. We were not able to make this method work due to the size of the state space.

5.1. Markov decision process concept

An MDP is a discrete time stochastic control process. A decision maker is faced with the problem of influencing the behavior of a probabilistic system as it evolves over time. He/she does this by making decisions or choosing actions. His/her goal is to choose a sequence of actions which causes the system to perform optimally with respect to some predetermined performance criterion. Since the system is ongoing, the state of the system prior to tomorrow's decision depends on today's decision. Consequently, decisions must not be made myopically, but must anticipate the opportunities and costs or rewards associated with future system states (Puterman [13]).

Consider the following example: suppose we should schedule the admission of patients 1, 2, ..., 20 with similar diagnoses at a ward for the upcoming five days. Each patient stays for at least one day, and 30% of the patients needs to stay an additional day. Each day, we should decide which patients to admit such that the probability that there are enough beds available is above a certain threshold and all patients are admitted after five days. Then, the phases of the system are the days 1, 2, ..., 5, the state of the system is the number of free beds, and our action is the number of patients we schedule on the current day (Van de Vrugt [14]). More formally, at each time step, the process is in state s and the decision maker can select an action a that is available at state s . Then, the process responds to the selected action in the next time step, which results into state s' .

A decision making problem can be formulated as an MDP in multiple manners. The challenge is to find the specific formulation so that the runtime is low and solution is good. A good solution for our problem is a scheduling method that has a scheduling horizon of three months and that can schedule patients with high certainty. We do not want to disappoint a patient by cancelling their planned surgical procedure on short notice.

5.1.1. State and action sets

Decisions made at certain points in time are referred to as *decision points*. The set of decision points is discrete, which entails that decisions are made at all decision points. At each decision point, the system occupies a *state*. The set of possible system states is denoted by S . If the decision maker observes the

system in state $s \in S$, he may choose *action* a from the set of allowable actions in state s , A_s .

When scheduling patients on different days in the future, the decision made at each decision point in time could be when a patient with a request for a specific surgical procedure is scheduled. However, it is also possible to incorporate a choice to delay the decision, so that we do not schedule the patient yet.

There are many choices in state and action sets. The state s can be a vector where each element s_i represents the number of minutes of scheduled surgeries on day i or the number of patients scheduled on day i or it can be a matrix where each element s_{ij} represents the number of patients of type j scheduled on day i . However, a different approach is also possible where you only keep track of the number of patient requests that have not had their surgical procedure. The choice what to incorporate in the state depends completely on which information you want to incorporate in the decision making process. The action a , when scheduling patients, is a change in the state s by adding a patient to the state. The available actions depend on the state, for instance, the state can have fully booked weeks where the new arrived patient cannot be scheduled. The actions can also depend on the decision points, for instance, when there are planned decision points in time. Then, at each decision point there can be multiple newly arrived patients. The action would then be to schedule multiple arrived patients, which results in a bigger change between the original state and the new state in comparison to adding one patient at each decision point.

5.1.2. Rewards and transition probabilities

As a result of choosing action $a \in A_s$ in state s at decision point t , the decision maker receives a reward $r_t(s, a)$ and the system state s' at the next decision point is determined by the probability distribution $p_t(s'|s, a)$. This probability distribution represents the probability of from state s to s' with action a . This probability $p_t(s'|s, a)$ is called the transition probability.

The reward may be regarded as income then it has positive value, but it may also be regarded as costs then it has negative value. In general, it is immaterial how the reward is accrued during the period. However, it is required that its value or expected value is known before choosing an action, and that it is not affected by future actions. The reward might depend on the priority of the patient that is scheduled, on the capacity utilization of the day that is scheduled or on the access time of the patient.

The collection of the decision set, the state set, the action set, the reward function and the transition probabilities is called a Markov decision process.

5.1.3. Decision rule and policies

A decision rule prescribes a procedure for action selection in each state at a specified decision point. Decision rules range in generality from deterministic Markovian to randomized history dependent, depending on how they incorporate past information and how they select actions. We use a deterministic Markovian decision rule, which is a decision rule that depends on previous system states and actions only through the current state of the system and it chooses an action with certainty.

A policy, plan or strategy specifies the decision rule to be used at all decision points. It provides the decision maker with a prescription for action selection under any possible future system or history. A policy is a sequence of decision rules.

5.2. Formulation of MDP

We have explained the definition of an MDP, the different choices in formulation of an MDP and formed the patient types. In this section, we present the chosen MDP formulation. The formulation is dependent on the number of clusters and on the decision making process. The aspects that are used to make a decision should be put in the state definition.

We defined the decision points to be the points in time when a new patient arrives. We want the MDP to take into account the arriving patients for the upcoming weeks and therefore, the number of decision points is finite.

Decision points:

$$T = \{1, 2, \dots, N\}, \quad N \leq \infty. \quad (5.1)$$

The state space is defined as a vector, which contains the patient type (p) of the new arrived patient and all previously booked patients split up per patient type per week. Priority 1 patients can only be booked in the first two weeks, so only the first two weeks are taken into account for this patient type. Priority 2 patients can only be booked in the first four weeks, so only these weeks will be taken into account for the patient types of priority 2. Thus, each cluster of priority 1 has two elements in the state vector (a_1, a_2), each cluster i of priority 2 has four elements in the state vector (bi_1, bi_2, bi_3, bi_4) and each cluster i of priority 3 has 12 elements in the state vector ($ci_1, ci_2, \dots, ci_{12}$). Therefore, the length of the state vector is therefore 79 ($1 + 2 + 4 * 4 + 12 * 5$) elements.

States:

$$s = [p, a_1, a_2, b1_1, b1_2, \dots, b4_4, c1_1, c1_2, \dots, c5_{12}] \quad (5.2)$$

$$S = \{s_1, s_2, \dots, s_M\}, \quad M \leq \infty. \quad (5.3)$$

The possible actions are scheduling the new patient with the given priority in one of the weeks in the planning horizon. For an arriving priority 2.2 patients, the possible actions are given by Equation (5.5). So for a priority 2.2 patient the possible weeks are 1, 2, 3, 4. For now, we assume that every arriving patient fits in the schedule within their set access time.

Actions:

$$\text{if } p = 1 \quad \text{then } A_{p=1} = \{1, 2\} \quad (5.4)$$

$$\text{if } p = 2.i \quad \text{then } A_{p=2.i} = \{1, 2, 3, 4\} \quad (5.5)$$

$$\text{if } p = 3.i \quad \text{then } A_{p=3.i} = \{1, 2, 3, \dots, 12\} \quad (5.6)$$

The transition probabilities from state s to s' at decision points t are defined as the probability of an arriving patient from cluster i at time point t . We assume that arrival rates are independent from previous arrived patients as we expect to receive the same amount of patients each week. The reward function takes into account the utilization rate of the chosen week at state s' in the booking horizon. The weekly capacity is denoted by C and the session time for the different clusters i of priority p are denoted with dp_i . As we want to maximize the utilization, we minimize the reward to make sure that the scheduled patients fill up the capacity.

Transition probabilities:

$$p_t(s|s') = \mathbb{P}(p = i) \quad (5.7)$$

Rewards:

$$R(s_i, w_j) = C - d1_1 * a_j + d2_1 * b1_j + d2_2 * b2_j + d2_3 * b3_j + d2_4 * b4_j + d3_1 * c1_j + \dots + d3_5 * c5_j \quad (5.8)$$

5.3. Conclusion

The data shows that on average 50 new patients arrive each month: 5 patients with priority 1, 16 patients with priority 2, and 29 patients with priority 3. The planning horizon spans approximately 3 months, thus, on average 150 patients arrive in total. These patients are divided over 10 clusters within the three priority groups. In Table B.2, the maximum number of arriving patients over one month are shown. The MDP should take into account all the possible scenarios of arriving patients, thus, also the extreme situations. The most extreme situation would be that over one month 10 patients of patient type 1 arrive, 13 of patient type 2.1, 5 of patient type 2.2, 13 of patient type 2.3, 9 of patient type 2.4, 18 of patient type 3.1, 15 of patient type 3.2, 15 of patient type 3.3, 3 of patient type 3.4 and 2 of patient type 3.5. To determine the number of states, we need to calculate how many different ways there are to divide the patients over the weeks for the total horizon. The number of ways to divide 10 patients or less of patient type

1 over two weeks is 66. The number of ways to divide 13 patients or less of patient type 2.1 over four weeks is 1049. The number of ways to divide 5 patients or less of patient type 2.2 over four weeks is 96. Multiplying all the combinations of all patient types together, results into a state space in the order of 10^{42} states.

To reduce the state space and with that the computation time of the model, we defined some restrictions and feasibility checks on the state space. The feasibility checks are overall capacity, oncology capacity per week and a maximum number of performed surgeries of each patient type possible in one week. To reduce the number of states even more, we discarded the unlikely booking schedules. Unlikely booking schedules are booking schedules with almost empty first two weeks or schedules with very full later weeks. We do not expect to see these schedules, as our reward is higher when we schedule patients earlier and because we have enough capacity such that priority 3 patients do not fill up the weeks further in the future. Also, we use a discount factor. A discount factor is a scalar between zero and one, which measures the value at time t of a unit reward received at time $t + 1$. When an MDP is solved with a 0.99-discounted reward, the future possibilities are taken into account while making a decision at time t . Therefore, the MDP will not fill up a week further in the future, before the weeks earlier are filled up.

Unfortunately, this reduction is at most of a factor 10^{10} . Thus, the number of states is still in order of 10^{32} , which is too large to compute the policy on Intel(R) Core(TM) i5-6200U CPU with 8 GB RAM. As we want to make a decision based on the priority of the new patient and on the utilization rate of the booking schedule, there is in our opinion no other way of formulating the state space. However, there are other possibilities of surpassing this problem. In the literature, we found multiple examples of this, for instance in Astaraky and Patrick [2], Patrick et al. [12], Zhu et al. [19] and Barz and Rajaram [4]. One solution that is often used in literature is approximate dynamic programming. In Patrick et al. [12], they solve, instead of the MDP, the equivalent linear program through approximate dynamic programming, and in Barz and Rajaram [4], they use approximate dynamic programming to construct heuristics to solve their scheduling problem. In Zhu et al. [19], the feasibility of using stochastic simulation methods for the solution of a large-scale Markov decision process model of online patient admissions scheduling is demonstrated. Stochastic simulation methods allow for the problem size to be scaled by a factor of almost 10 in the action space and exponentially in the state space. However, we decided to use a different solution method and not pursue the Markov decision process combined with approximate dynamic programming any longer. The reason for this is that we do not have enough experience with MDP and how to transform an MDP with approximate dynamic programming.

6

Multiple mixed integer program method

An optimization problem is called a mixed integer programming problem (MIP), when some, but not all, variables are restricted to be integer. The variable connected to the scheduling of patients is of course integer, but other variables do not need to be integer.

The set up of the solution method is explained in Section 6.1. The MIP that is used by the solution method is explained in Section 6.2. In Section 6.3, the setup of the simulation process is explained, which we use to check if our solution works.

6.1. Set up of the solution method

The scheduling process is different for all three patient priority groups. In this section, we explain the different ways of scheduling and why we think we need different ways of scheduling.

According to their priorities, we would like to schedule patients of priority 1 first, then priority 2 patients and finally priority 3 patients. This would require that all patients have the same scheduling horizon, as is the case currently in the hospital. However, the hospital wishes to schedule priority 3 patients with a longer horizon. Thus, if we would make a schedule with a scheduling horizon of 2 months, then a priority 3 patient is the first to be scheduled in weeks 5 up to 10 in the future as a priority 3 patient has no urgency and is scheduled further in the future. Later priority 2 and priority 1 patients are scheduled around the priority 3 patients. As priority 1 and 2 patients have higher urgency, they are scheduled on a shorter horizon than priority 3 patients. This is not the desired order of scheduling with respect to patients' priorities, but necessary if we want to inform priority 3 patients earlier about the scheduled date. It could happen that we have already scheduled too many priority 3 patients, so that a priority 1 patient does not fit. Or it could happen that we did not schedule enough priority 3 patients and the OR utilization is low. Therefore, it is important to reserve the right amount of OR time for priority 1 and 2 patients.

Our solution method uses an MIP to schedule patients with priority 3. The MIP is solved once every few weeks to determine a master surgical schedule (MSS) for the upcoming 3 months. The MSS reserves OR time for priority 1, 2 and 3 patients. When a patient of priority 3 arrives, the patient is scheduled with the use of the MSS: the patient is scheduled in an available slot that is reserved for her type in the MSS. If there are more priority 3 patients arriving of type i and not enough time slots reserved for patient type i in the MSS, the MIP is solved again to obtain a time slot for all patients of type i . The new schedule found serves again as an MSS for a few weeks. The new schedule takes into account the patients that are already scheduled. The minutes assigned to a scheduled patient are not available to any other arriving patients. The new schedule will look different from the previous schedule as the input for the model has changed. The available capacity has changed due to the previous scheduled patients. A schedule made by the MIP could look like Figure 6.1 and Table 6.1, both the table and the figure show the same schedule. Table 6.1 shows that patients of type 3.1 can be scheduled on day 8, 15 and 25. Patients of type 3.1 can only be treated on Wednesday and Friday and we can see this pattern in the

schedule. Figure 6.1 shows that every week there is time reserved for priority 1 (yellow) and priority 2 (blue) patients.

Table 6.1: Number of patients scheduled per day by the MIP for priority 3 patients

Day number	2	3	4	5	7	8	9	10	12	13	14	15	17	18	19	20	22	23	24	25
prio1			3		1		1			1			2	1					3	
prio2.1			1									1							1	
prio2.2												1								
prio2.3																				
prio2.4					1		1			1					1					
prio3.1						5						5								5
prio3.2	3			3				3		3										
prio3.3									2								2	2		
prio3.4														1						
prio3.5		1														1				

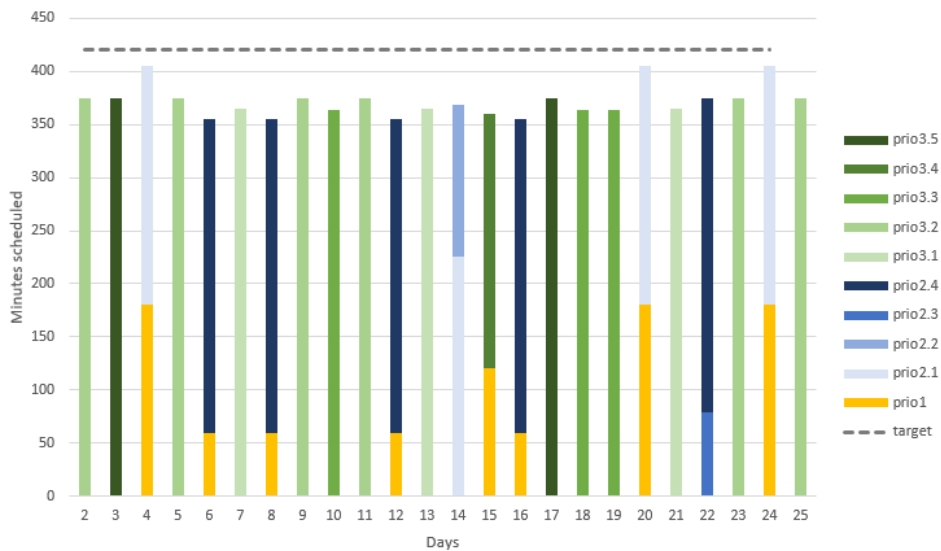


Figure 6.1: Scheduled patients by MIP for priority 3 patients

A priority 2 patient is also scheduled using a MIP. Every day when one or more priority 2 patients arrive we solve the MIP. The MIP for priority 2 patients is different from the MIP used to schedule priority 3 patients. The MIP for priority 2 patients makes a schedule with priority 1 and 2 patients. The MIP reserves OR time for priority 1 patients and other priority 2 patients, while making a schedule for a specific priority 2 patient type. The MIP does not reserve time for patients of priority 3, as we assume that they have already been scheduled with the previous MIP and when they arrive they can wait for four weeks or more, which is outside the scheduling horizon of priority 2 patients. After scheduling priority 3 patients, the remaining available OR time in the first four weeks is thus reserved for priority 1 and 2 patients. The schedule could look like Figure 6.2 and Table 6.2, again both the table and the figure show the same schedule. The schedule in Figure 6.2 and Table 6.2 was made while no priority 3 patients were scheduled. So the available capacity on all days was the same. All priority 2 procedures are part of the oncology specialization, which implies that priority 2 patients can only be scheduled on Thursdays and every other Tuesday. This pattern is also very clear in the schedule. However, it is not realistic to assume all 'oncology' days will be free of other priority 3 patients, as priority 3 patients use more than half of all capacity.

The dotted line is 90% of the total capacity available on surgery days. This is the maximum amount of time we are allowed to use to ensure that the schedule stays feasible when one surgery has a small delay. The goal of the schedule is to fill up the days up to 80 - 85%.

Suppose, as example, we have to schedule a patient of type 2.2. Then, there are two days on which this is possible, namely 14 and 19. Scheduling the patient on a certain day can be done with a policy or left up to scheduler. As some surgical procedures are only done during a specific part of the menstrual cycle, we are not sure if a policy can be implemented. However, we have implemented first-come-first-serve (FCFS), just-in-time (JIT) and a best-fit policy to test our method. The FCFS policy assigns the first available day to the patient, so in case of the example FCFS would choose day 14. The JIT policy would assign the last available day in the schedule that is no later than the desired access time of the patient, in case of the example JIT would choose day 19. The best-fit policy would assign the day that is filled up the most so that the schedule would result in more full days and so that patients with a longer session time can still be scheduled on the empty days. In the example, best-fit would choose the day with the least available capacity left.

Table 6.2: Number of patients scheduled per day by the MIP for priority 2 patients

Day number	4	7	9	14	17	18	19	24
prio1					1	1		1
prio2.1							1	
prio2.2				2			1	
prio2.3	1	1	1	1				
prio2.4	1	1	1		1			1

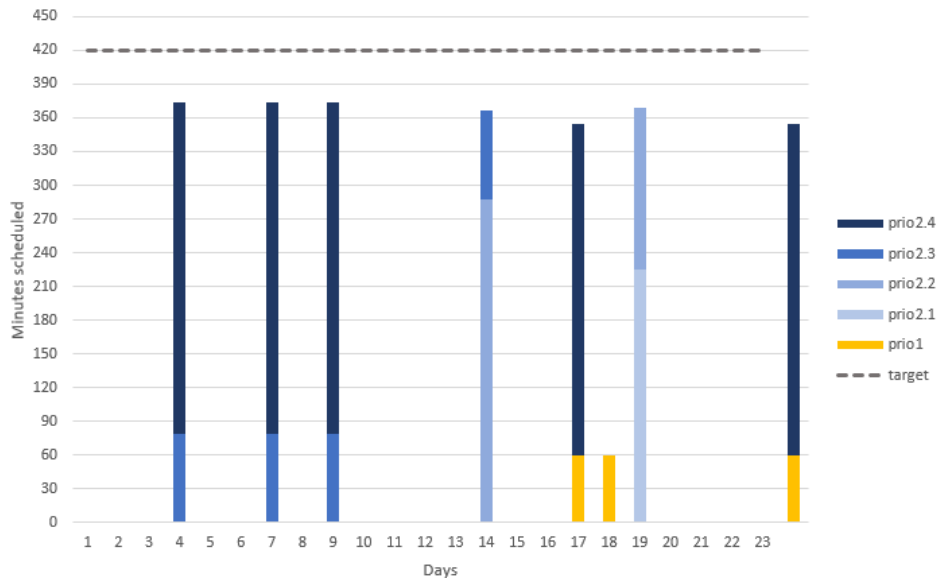


Figure 6.2: Scheduled patients by MIP for priority 2 patients

Lastly, a priority 1 patient is scheduled without the use of a model. A priority 1 patient needs to have surgery as soon as possible, therefore, a patient of priority 1 receives the first available time slot. It is important that during the scheduling of priority 2 and 3 patients, there is room left to schedule the priority 1 patients. If there is no room left to schedule a priority 1 patient, other patients need to be rescheduled, which is undesirable. As discussed above, both the MIP of priority 2 and 3 patients reserve OR time for priority 1 patients, so as long as this time is not more than 2 weeks in the future, the scheduling of a patient with priority 1 is not a problem. As priority 1 patients are scheduled without a model, all remaining available time can be used to schedule a priority 1 patient, thus priority 1 patients can be scheduled over the time slots of priority 2 and 3. When a scheduler would use our solution method, then each day he or she would start with scheduling patients of priority 1 and after that priority 2 and 3 patients would be scheduled.

The reason we split up the solution method in three parts is that the three priority groups have different needs. Priority 1 patients want to be treated as soon as possible and their OR session time is small. Pri-

riority 2 patients also want to be treated as soon as possible, but their session time varies a lot. Therefore, they can not fill up holes in the schedule and planning is needed. Priority 3 patients can postpone their surgery and like to plan their surgical procedure so that the potential revalidation period is convenient for them and does not interfere with big events in their lives.

6.2. Set up of mixed integer program

In this section, we explain in detail the MIP that is used to schedule patients of priority 2 and 3. The MIP determines an optimal schedule given the scheduling constraints and the exact date for a priority 2 or 3 patient is chosen with a scheduling policy.

The MIP is similar to the one described in Banditori et al. [3]. The MIP uses the available capacity of the upcoming days, some patient scheduling restrictions and patient types that need to be scheduled. The scheduling constraints are specialization of the surgeon, daily capacity and number of patients that need to be scheduled. The idea is that by taking into account dummy patients with higher or equal urgency, the model is robust and there is enough time for high urgency patients which request surgery in the near future. The patient with high urgency are priority 1 and 2 patients.

The MIP uses the sets, parameters and variables listed below. The input for the MIP are the sets, the parameters and the horizon length in days and in weeks. The value of β is set by the hospital and the value of α is set by us. The value of α is the minimum utilization percentage so that we are satisfied with a certain day. The value of α is always smaller than the value of β , because the utilization percentage can never be more than β . If we set α too small, the schedule will not be filled up enough to handle all requests of patients. If we set α too close to β , then the scheduling process can become too difficult as there are only a few combinations of surgical procedures session times that sum up to the right number of minutes.

Horizon	Days	D
	Weeks s.t. $D^w \subseteq D$	W
Sets	Patient types	K
	Oncology patient types	$O \subseteq K$
	Non-oncology Tuesday	$D^{nt} \subseteq D$
	Oncology Tuesday	$D^{ot} \subseteq D$
	Wednesday	$D^w \subseteq D$
	Thursday	$D^t \subseteq D$
	Friday	$D^f \subseteq D$
Parameters	Capacity on day $d \in D$	C_d
	Average session time for patient type $k \in K$	p_k
	Maximum number of surgeries of type $k \in K$ each week	\overline{T}_k
	Minimum number of surgeries of type $k \in K$ each week	\underline{T}_k
	Maximum number of surgeries of type $k \in K$ over the horizon	\overline{TH}_k
	Minimum number of surgeries of type $k \in K$ over the horizon	\underline{TH}_k
	Minimum utilization percentage to receive a bonus	α
	Maximum utilization percentage defined by the hospital	β
Variables	The number of patients of type $k \in K$ scheduled on day $d \in D$	x_{kd}
	Bonus variable for full day scheduling	y_d

We know that many surgical procedures require specialized surgeons. The patient types were made without the knowledge of these exact specializations. However, we ended up with good clusters with respect to the specializations. We know that patient type 1, 3.2 and 3.5 can be scheduled on all days with

capacity. Patient type 2.1 up to 2.4 can be scheduled every other Tuesday and every Thursday. Patient type 3.1 can be scheduled on Wednesday and Friday. Patient type 3.3 can be scheduled on every other Tuesday (non-oncology), Wednesday and Thursday. Patient type 3.4 can only be scheduled on Wednesday.

The MIP is formulated in the following way;

$$\max \sum_d y_d \quad (6.1)$$

$$\sum_k p_k x_{kd} \leq \beta C_d \quad \forall d \in D \quad (6.2)$$

$$x_{kd} = 0 \quad \forall d \notin D^t \cup D^{ot}, k \in O \quad (6.3)$$

$$x_{kd} = 0 \quad \forall d \notin D^w \cup D^f, k = 3.1 \quad (6.4)$$

$$x_{kd} = 0 \quad \forall d \notin D^{no} \cup D^w \cup O^t, k = 3.3 \quad (6.5)$$

$$x_{kd} = 0 \quad \forall d \notin D^w, k = 3.4 \quad (6.6)$$

$$\sum_d x_{kd} \geq \underline{TH}_k \quad \forall k \in K \quad (6.7)$$

$$\sum_d x_{kd} \leq \overline{TH}_k \quad \forall k \in K \quad (6.8)$$

$$\sum_{d \in (D^w)} x_{kd} \geq \underline{T}_k \quad \forall k \in K, w \in W \quad (6.9)$$

$$\sum_{d \in (D^w)} x_{kd} \leq \overline{T}_k \quad \forall k \in K, w \in W \quad (6.10)$$

$$\sum_k x_{kd} \geq \alpha C_d y_d \quad \forall d \in O^d \quad (6.11)$$

$$y_d \in \{0, 1\} \quad (6.12)$$

$$x_{kd} \in \mathbb{N} \quad (6.13)$$

The main objective (6.1) is to maximize the bonus. The bonus variable is binary and only equal to one if day d has a capacity utilization of at least α , else the bonus is equal to zero. Constraint (6.2) ensures that the scheduled capacity is maximally the available capacity. In fact, we multiply with β , because this is the maximal capacity we may schedule. This is done to reduce the number of rescheduled patients and times overtime is needed, when there is a delay in the schedule. Constraint (6.3) up to (6.6) ensure that the surgical procedure for priority 2 and patient type 3.1, 3.3 and 3.4 has the same specialization as the operating surgeon. As surgeons have set operating days, there are set days these patients can be treated. The remaining patient types can be treated on every day with sufficient capacity. Constraints (6.7) up to (6.10) enforce the number of patients that need to be scheduled at least and at most. The first two constraints bound the number of patients scheduled over the horizon and the last two over each week. This is done to reserve time for other patient types with equal or higher urgency. By not setting the exact number of patients that need to be scheduled, the model can determine the best schedule if it had choice in patient arrival. Since we do not know the exact arrival numbers of patient types, this is a good way to find the best day to schedule a certain patient. Constraint 6.11 gives the bonus variable y_d a value, which we try to maximize. If the scheduled capacity is more than or equal to α , the bonus variable is equal to 1, else it is zero. We do this to reduce the number of half used days, as we can return empty days to the OR department and then that day can be fully used by another department.

6.3. Set up of simulation

To test the solution method, we use simulation. In this section, we explain the simulation process. The simulation must mimic the whole scheduling process of arriving patients and assigning them to a day

in the future to receive their surgical procedure. We choose to model the arrival of patient types with a Poisson arrival process. The Poisson arrival process has a constant arrival rate. The arrival rate is defined as the expected number of arrivals per time unit.

The arrival rate in the simulation is defined as the expected number of patients arriving per day. Each day in the simulation, we sample from different Poisson distributions to determine how many patients of each patient type arrive that day.

After computing the number of arriving patients, we start with scheduling patients that have arrived. First, we schedule priority 1 patients, then the priority 2 patients, and lastly, the priority 3 patients. The scheduling of priority 1 patients is done without a model, but with the scheduling policy FCFS. The scheduling of priority 2 and 3 patients is done by first solving the mixed integer program. If the run of the model does not give a day to schedule the arrived patient, we rerun the model. Using a lower minimum number of surgical procedures of type k over the weeks and horizon $(\underline{TH}_k, \underline{T}_k)$. This is done to make sure that the model is feasible. Then we find a suitable time slot among the reserved time slots with a scheduling policy. We implemented first-fit, just-in-time and best-fit. We also looked at a random-fit policy, as some surgical procedures are only performed during a specific part of the menstrual cycle. This cycle is different for all women, therefore we randomly choose a day from the schedule made by the model.

7

More accessible approaches

It is important for the gynecology department that the method is accessible such that using the method does not take up more time. A heuristic approach is any approach used to solve problems that uses a practical method that is not guaranteed to be optimal. The advantage is that it works sufficient and fast. Usually, a heuristic approach is logical and it is easy to understand how and why it works. Although, for our method described in Chapter 6, it is more difficult to understand how and why it works. In this chapter, we explain heuristic methods that we can use to compare our MIP simulation model to and we discuss the advantages and disadvantages.

7.1. First come, first serve

The easiest approach is first-come-first-serve. The idea is that the first patient to arrive is the next one to get treated. For example, suppose any patient arrives on June 1st. The current schedule is filled up until June 7. Then, the newly arrived patient is scheduled on June 8, no matter the priority of the patient. This method works well when the requested OR time is no more than the available OR time and the waiting line is not longer than the access time desired for the priority group with the highest urgency.

An advantage of this scheduling policy is that it is easy to implement and use by the gynecology department. It is also possible to implement long term scheduling for priority 3 if the amount of requests is not too high. A disadvantage is that it only works in certain cases. This scheduling policy does not work when the requested OR time is more than the available capacity over a certain period. Moreover, if the requests of priority 3 patients suddenly rise for a period of time and we are scheduling priority 3 patients on the long term, then this method also does not work. This is because the patients of priority 1 and 2 could have to wait for a longer period than their desired access time which would make rescheduling necessary. Moreover, patient preference is not incorporated in the scheduling policy.

7.2. Just in time

Another simple approach is the just-in-time scheduling policy. The idea is that each patient is scheduled just before their desired access time. For example, suppose a patient of priority 1 arrives on June 1st. The patient should be treated before June 15 as the desired access time of a priority 1 patient is 2 weeks. The just-in-time scheduling policy would schedule this patient on June 14, if there is enough available capacity. However, if June 14 is already filled up with scheduled patients, the patient would be scheduled a day earlier (June 13). This process continues until a day is found with enough available capacity.

An advantage of this scheduling policy is that it is easy to implement and use by the gynecology department. Moreover, by using the entire access time period, more patients should be able to be scheduled on time. A disadvantage of this scheduling policy is that the schedule is not optimized such that the

days in the schedule are not filled up to 80%. Thus, it could happen that when interchanging 2 patients on consecutive days, an extra patient could be added. This is possible when two patients with different session times are interchanged and by this change one day is more filled up and the other is less filled up so that a new patient can be added. Of course, the scheduling constraints are taken into account with this interchange. Moreover, patient preference is not incorporated in this scheduling policy.

7.3. Protection levels

A more extensive approach is protection levels. The idea is that the capacity per week is separated over the different priority groups. Each priority group can fill up a percentage of the capacity per week. We do not split it up per patient type as each patient type in a priority group has the same urgency. Patients of priority i can be scheduled in a certain week as long as the used capacity percentage for priority i is not exceeded in that week. The capacity percentages would be based on the request percentages discussed in Section 2.3.1.

The protection levels are based upon the demand for each priority group. In Figure 2.4, it is visible that approximately 5% of all requested surgery minutes is demanded by priority 1 patients. Priority 2 patients request 45% of all requested surgery minutes and priority 3 patients request the remaining 50%. In one week, the gynecology department has an OR capacity of four days, thus, 1680 surgery minutes. This results into 840 minutes for priority 3 patients, 756 minutes for priority 2 patients and 84 minutes for priority 1 patients.

For example, suppose we want to schedule a patient of type 2.3. Then we need 144 minutes available in a week, see Table 4.2. Per week, priority 2 patients can fill up 756 minutes. Then, we look for a week in which less than 756–144 minutes is scheduled for priority 2.

This approach can be extended with some rules to make the schedule more flexible. For instance, a high priority patient can be scheduled using capacity of a lower priority group. Moreover, a lower priority patient can be scheduled using capacity of a higher priority group if the scheduled capacity of the higher priority group is below a certain percentage.

An advantage of this policy is that the scheduler has a lot more freedom with this approach. It is possible for the patient to choose a certain week and/or day if the scheduling constraints allow the choice, i.e., available capacity and surgeon specialty. A disadvantage of this freedom is the difficult extra task to determine the best schedule. When scheduling a patient on a certain day, the consequences for other patients should be taken into account.

7.4. Master surgical schedule

The master surgical schedule approach is based on the MIP formulated in Section 6.2. The idea is that the MIP is solved for the whole scheduling horizon and all arriving patients are scheduled with the schedule made by the MIP. The schedule made by the MIP is used as a master surgical schedule. For example, a part of the master surgical schedule could look like Table 7.1. Suppose a patient of type 3.3 arrives on June 1st and the MSS in Table 7.1 is used that day. Then, the scheduler can decide to schedule this patient on June 19, June 23, June 28 and July 4th. Suppose the scheduler decides to schedule this patient on July 4th. Then, when another patient of type 3.3 arrives that week, it is not possible to schedule this new patient on July 4th. This patient can only be scheduled on June 19, June 23 and June 28.

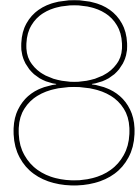
The advantage of this scheduling policy is that it is possible for the patient to choose a date, however, with the restriction that a patient can only choose from the dates for her patient type listed in the master surgical schedule. The disadvantage of this scheduling policy is that the MSS could reserve time for patients that do not arrive before that day. This disadvantage could be resolved by implementing some flexible scheduling rules. For instance, the scheduler can use reserved OR time for priority 3 for patients with higher priority on the short term, when the time slot for the priority 3 patient is not used on the long term. Moreover, we can also implement a scheduling policy to determine the best date for the patient instead of leaving the decision to the scheduler and patient preference.

Table 7.1: Master surgical schedule made by the model formulated in Section 6.2

Day	17	18	19	20	22	23	24	25	27	28	29	30	32	33	34	35	37	38	39	40
prio1		1		1			2	1		1		1		3				3		
prio2.1	1		1												1		1			
prio2.2	1																			
prio2.3									1								1			
prio2.4							1		1		1								1	
prio3.1		5		3	5	1		5				1	5			2		3		4
prio3.2				1		1					1	2				2			1	1
prio3.3			1			1				2					1					
prio3.4														1						

7.5. Conclusion

We will not use first-come-first-serve because of three reasons; the current waiting list of patients is too long, patient preference is not incorporated and the scheduling is not done in a long term manner. Therefore, we think that the FCFS approach is not a good alternative for the gynecology department. We are interested in the other three approaches, as they could be a good alternative for our multiple MIP method. The master surgical schedule approach is interesting as it compares the set up of our advanced method with multiple MIP models for the different patient groups to one single MIP model. We can then clearly see if the advanced set up is necessary or that one MIP model is enough to improve the schedule. The protection level and JIT approaches do not use any model, but are easy scheduling policies. It is interesting to check if our MIP method works better or if the MIP method needs specific improvements, to achieve better results.



Computational results

In this chapter, we use simulation to compare our multiple MIP scheduling method with the more accessible methods discussed in the previous chapter. We explain the specific set-up used for the simulation and which important decisions we have made. Moreover, we answer the following questions: Is our multiple MIP scheduling method better than the more accessible methods? Can the more accessible methods approximate the schedule of our multiple MIP scheduling method? Or do all methods have different advantages?

8.1. Set-up simulation

Each method is tested on the same set of 250 different scenarios of arriving patients. There are more than 100000 different scenarios, we could use to test our method. With relative precision, we calculated that testing 250 scenarios is enough for 95% confidence interval and an error margin of 6%. Thus, 250 scenarios is enough to draw conclusions about the performance of the different methods. Each scenario consists of 115 days on which patients arrive. We do not include weekend days in the simulation, but only working days. So each week contains 5 days and each month contains 22 days. When simulating, we start from a completely empty schedule. The first 4 weeks are used as warmup period for the method and then the simulation has a full scheduling horizon of 4.5 months to test the scheduling methods. We think that this is long enough to mimic the reality and to see how well the method performs on the long run, as the gynecology department is currently scheduling 2 weeks in advance, 4.5 months is a big step.

In Section 8.1.1, we explain which data we used for the input parameters of our multiple MIP model. We explain how we calculated the patient arrival rates for all the different patient types in Section 8.1.2. In Section 8.1.3, the performance indicators are explained.

8.1.1. Scheduling constraints

The values for \overline{TH}_k , \underline{TH}_k , \overline{T}_k and \underline{T}_k used in Constraint (6.7) up to Constraint (6.10) have not been given. Possible values for the maximum and minimum number of surgeries of type k over the horizon are given in Table 8.1. We used our data set with the OR data from 2015 up to 2017 to calculate the distribution of arriving patients. In Table 8.1, the minimum, maximum and quartiles are given for the number of patients arriving per month. We decided to use the maximum number of requests and first quartile as maximum and minimum, respectively. \overline{TH}_k and \underline{TH}_k are calculated by multiplying the number of months in the horizon with the values in Table 8.1. \overline{T}_k and \underline{T}_k are calculated by dividing the values in Table 8.1 by the number of weeks in a month rounded up. In expectation, each week the number of arriving patients is the same, independent of the number of patients that have already arrived. Therefore, we do not adapt these values during the scheduling process.

During the first simulation of our multiple MIP method, we noticed that the lower bound \underline{T}_k was not

Table 8.1: Demand in number of patients per month per cluster

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	0	0	2	0	0	3	2	0	0
1st quartile	3	3	1	5	3	8	7	5	0	0
2nd quartile	4	5	2	7	4	10	9	7	0	0
3rd quartile	6	7	3	9	5	13	12	8	1	0
maximum	10	13	5	13	9	18	15	15	3	2

necessary. The weekly lower bound even made the model infeasible in some cases, as the weeks fill up each time a new patient arrives. When rerunning the model, the available capacity in some weeks was too low to fill up that week with the minimal number of patient types set by the weekly lower bound. Therefore, we decided to set the weekly lower bound to zero for all patient types.

8.1.2. Patient arrival

We explained that we used a Poisson arrival process to mimic the patient arrival process. For each patient type, we calculated a different arrival rate λ_k . This arrival rate λ was calculated by dividing the median (second quartile) of arriving patients per month by the number of working days in a month (22). This resulted in the following arrival rates (Table 8.2). These arrival rates result in on average 253 arriving patients in 115 days.

Table 8.2: Demand in patients per working day

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
Median	4	5	2	7	4	10	9	7	0	0
Arrival rate λ_k	0.18	0.23	0.09	0.32	0.18	0.45	0.41	0.32	0	0

Table 8.3: Average patient arrival over the horizon of 115 days

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
Average patient arrival	21	27	10	37	22	52	37	47	0	0

This results in the average number of arriving patients over 115 days during the simulation process shown in Table 8.3. Both Table 8.2 show that patient type 3.4 and 3.5 do not arrive monthly. Therefore, Table 8.3 also shows that type 3.4 and 3.5 do not arrive over the horizon. In the data set from November 2014 to January 2018, there are 20 patients of type 3.4 and 3 patients of type 3.5. Therefore, we changed the arrival rate λ_k for both patient types to 0.002, such that during some scenario's one or two patients of type 3.4 and/or 3.5 arrive.

8.1.3. Performance measures

In Section 2.4, we explained that the OR utilization and the access times are the most important performance indicators. We incorporated this also in our simulation. The OR utilization is calculated after one simulation of 115 days is completed. This is done by dividing the booked capacity by the initial available capacity for each day in the simulation that have a booked capacity larger than zero. This is done to not put a penalty on the empty days in the schedule as they could be given back to the OR board and used by another department. We calculate the average access time by summing the access times for each patient type and dividing it by the number of arrived patients of that type. Because of this, we do not keep track of the number of late days, but we do keep track of the number of late booked patients.

Moreover, we look at the number of booked and not-booked patients per simulation. Unfortunately, we found out that some methods have difficulty to schedule all patients at arrival. It is interesting which

types of patients are most difficult and how often they do not get scheduled.

The gynecology department aims for an OR-utilization of at least 75%. Moreover, The method should be able to schedule 95% of all arriving patients and the desired access times should be achieved on average for priority 1 and 2 patients. If we find a method which satisfies these three conditions, we consider the method to be suitable for the gynecology department.

8.2. Results multiple MIP method

Our multiple MIP method can be implemented with different scheduling policies. We decided to test 3 different scheduling policies. First fit chooses the first day in the schedule made by the MIP that has capacity reserved for the patient type that needs to be scheduled. Best fit chooses the day on which the previous scheduled capacity is the highest and on which capacity is reserved for the patient type that needs to be scheduled. As we want to fill up the days as much as possible, this could be a good policy. Random fit chooses random the n^{th} day in the schedule on which capacity is reserved for the patient type that needs to be scheduled. As we do not know if implementing a policy can be done with all extra patient constraints that we do not know of, we are interested in how our method performs without a smart scheduling policy.

8.2.1. Average access time

The multiple MIP method works quite good concerning the desired access times (see Table 8.4). In Table 8.4, the average access times per patient type is given over all 250 scenarios. Recall, priority 1 patients have a desired access time of 10 working days, priority 2 patients have a desired access time of 20 working days and priority 3 patients have a desired access time of 65 working days. As can be seen in Table 8.4, only patient type 2.2 and 2.3 do not meet their desired access time on average. Their average access time is larger than 20 working days, thus, either more than half of the patients are scheduled late or, if patient are scheduled late, their access time is significant larger than those 20 working days. When using the multiple MIP method with the first fit scheduling policy, we see that on

Table 8.4: Average access time in days and average number of late scheduled patients per patient type

Patient type	Average access time first fit	Average access time best fit	Average access time random fit	Average late scheduled first fit	Average late scheduled best fit	Average late scheduled random fit
Prio 1	6.064	5.748	6.08	0	0	0
Prio 2.1	20.608	20.964	19.692	16	16	16
Prio 2.2	24.656	23.708	26.088	8	7	8
Prio 2.3	21.76	23.932	22.856	22	25	24
Prio 2.4	14.6	13.304	14.688	8	8	9
Prio 3.1	36.196	42.14	37.128	3	15	7
Prio 3.2	32.808	39.532	36.736	2	8	6
Prio 3.3	37.248	38.744	40.928	1	7	9
Prio 3.4	36.188	38.660	36.151	0	0	0
Prio 3.5	33.265	33.388	32.571	0	0	0

average 64% of the arriving patients is scheduled on time, i.e. before their desired access time. When using the best fit scheduling policy, on average only 57% of all arriving patients is scheduled on time and with random fit on average 60% of all arriving patients is scheduled on time. This is not a high percentage and when we started this research we set our standards higher. More information on the number of late scheduled patients per patient type can be seen in Appendix C.

8.2.2. Not scheduled patients

Unfortunately, our multiple MIP method is not able to schedule all patients in each scenario presented. In theory, it is possible to not schedule a patient, but in reality, we do not want this to happen. On average,

the number of patients not scheduled by the first fit scheduling policy is 32 patients. The average number of patients not scheduled by the best fit or random fit scheduling policy is 21 patients. When comparing the three scheduling policies, Tables 8.5 to 8.7, we clearly see differences. When using the first fit scheduling policy, in each scenario there was at least one patient of type 3.2 that was not scheduled. If we compare the first fit scheduling policy with the best fit scheduling policy, we see that priority 3 patients are all scheduled with best fit, but there are more patients of type 2.4 left unscheduled. If we compare best fit with random fit, then we see that the number of patients and the patient types are similar or the same. Concluding, we suspect that the capacity for priority 2 is too low as all scheduling policies have trouble with scheduling priority 2 patients and the scheduling policies best fit and random fit perform better concerning the average number of not scheduled patients.

Table 8.5: Number of not scheduled patients with first fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
min	0	0	0	0	0	0	1	0	0	0
1st quartile	0	4	0	0	7	2	5	0	0	0
2nd quartile	0	6	1	0	9	4	8	1	0	0
3rd quartile	0	9	2	2	12	6.75	11	2	0	0
max	1	16	8	10	20	17	17	9	0	0

Table 8.6: Number of not scheduled patients with best fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
min	0	0	0	0	0	0	0	0	0	0
1st quartile	0	4	1	0	8	0	0	0	0	0
2nd quartile	0	6	1	1	11	0	0	0	0	0
3rd quartile	0	9	2	2	14	0	0	0	0	0
max	0	17	8	7	19	0	0	0	0	0

Table 8.7: Number of not scheduled patients by random fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
min	0	0	0	0	0	0	0	0	0	0
1st quartile	0	6	0	1	7	0	0	0	0	0
2nd quartile	0	7	0	2	10	0	0	0	0	0
3rd quartile	0	10	1	3	12	0	0	0	0	0
max	1	18	10	12	23	0	0	0	0	0

8.2.3. Utilization

The utilization of the available OR time for the three different scheduling policies is 74% for first fit policy, 70% for best fit policy and 76% for random fit policy. We did not expect that the random fit policy would perform this well. However, when seeing that the number of not scheduled patients is lower for the random fit policy it explains partly that it performs better. The difference between best fit and random fit is, although they have the same number of not scheduled patients, that best fit has less empty days. This is not expected as best fit chooses the day that was already filled up the most such that the new arrived patient can be added.

8.3. Results more accessible methods

The aim of this research project was to improve the OR schedule of the gynecology department. If our multiple MIP method can improve the current situation, it could be that an easier and more accessible approach can give the same or better results. The usability is important for the schedulers. So if a more accessible approach can obtain the same result, the gynecology department will prefer that approach over our multiple MIP method. In Section 8.3.1, the performance of the just-in-time scheduling approach is given. In Section 8.3.2, the performance of the protection level approach is given. In Section 8.3.3, the performance of the master surgical schedule (MSS) approach is given.

8.3.1. JIT method

The JIT scheduling method is the fastest and the most accessible method that we use to compare our multiple MIP method with. The average utilization percentage is 71%, which is similar to the multiple MIP method. The average access times for the patient types are given in Table 8.8. Again, we see that the patients of type 2.2 and 2.3 are on average not scheduled on time as their average access time is more than 20 working days. On average, 74% of all arriving patients is scheduled on time. For more information on the late scheduled patients per patient type, see Appendix C.

The JIT approach is also not able to schedule all patients (see Table 8.9). On average 24 patients are not scheduled, which is 10% of all arriving patients. There are some exceptions, but the not scheduled patients are patients of priority 2. After seeing these results, we calculated the demand of priority 2 and confirmed our suspicions that there is not enough capacity for priority 2 available and that it also is used too much by priority 3 patients.

Thus, JIT performs better concerning on time scheduling, but a little lower on utilization. JIT is overall not better or worse than the multiple MIP method.

Table 8.8: Average access time in days and number of late scheduled patients per patient type with JIT approach

Priority	Priority 1		Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5	
Average access time	9.1	20.2	22.0	23.0	12.2	56.7	59.7	59.2	51.9	58.0	
Average late scheduled	0	14	5	18	6	0	0	0	0	0	

Table 8.9: Number of not scheduled patients by JIT approach

Priority	Priority 1		Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5	
min	0	6	1	4	1	0	0	0	0	0	
1st quartile	0	11	4	14	5	0	0	0	0	0	
2nd quartile	0	14	5	18	6	0	0	0	0	0	
3rd quartile	0	16	7	22	8	0	0	0	0	0	
max	1	25	13	35	13	3	0	0	1	0	

8.3.2. Protection level method

The protection level method is also a very fast method, but takes a bit more effort from the schedulers as they need to keep track of the protection levels. The protection levels are based upon the demand for each priority group. In Figure 2.4, it is visible that approximately 5% of all requested surgery minutes is demanded by priority 1 patients. Priority 2 patients request 45% of all requested surgery minutes and priority 3 patients request the remaining 50%. In one week the gynecology department has an OR capacity of 4 four days, thus, 1680 surgery minutes. This results into 840 minutes for priority 3 patients, 756 minutes for priority 2 patients and 84 minutes for priority 1 patients. The simulation uses the first fit policy to schedule the patients. However, we made the restriction that priority 3 patients could not be

scheduled in the first four weeks after their arrival date. This was done to make the long term scheduling possible.

On average, this method performs very well with respect to the OR utilization with a percentage of 78%. However, the desired access times are not met for any of the priority 2 patient types (see Table 8.10). On average, 75% of all arriving patients is scheduled on time. On average, 95% of the late booked patients is priority 2. As 45% percent of all arriving patients is priority 2, we can conclude that more than half of all arriving patients of priority 2 is scheduled too late, which is not acceptable for the gynecology department. For more information on late scheduled patients per patient type, see Appendix C.

Table 8.10: Average access time per patient type with protection levels approach

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
Average access time	7.8	35.8	26.2	21.6	35.9	22.8	21.7	22.8	24.6	23.3
Average late scheduled	3	20	7	19	17	0	0	0	0	0

However, the protection level approach is very good in scheduling patients (see Table 8.11). On average 4 patients are not booked, thus 99% of all patients are scheduled on average. This is much better than the multiple MIP method and the JIT approach.

Table 8.11: Number of not scheduled patients with protection levels approach and first fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
min	0	0	0	0	0	0	0	0	0	0
1st quartile	0	0	0	0	0	0	0	0	0	0
2nd quartile	0	1	0	0	0	0	0	0	0	0
3rd quartile	0	3	0	0	2.75	0	0	0	0	0
max	0	9	2	3	9	0	0	0	0	0

8.3.3. Master surgical schedule method

The master surgical schedule method (MSS method) can be implemented with different scheduling policies. We have implemented first fit and best fit. Both scheduling policies perform similar. On average, 63% of all arriving patient is scheduled on time with first fit and only 59% with best fit. This is lower than almost all other methods. More information on the number of late scheduled patients per patient type can be seen in Appendix C.

Table 8.12: Average access time per patient type with Master surgical schedule approach

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
Average access time First fit	20.9	34.7	29.2	35.1	22.9	30.1	41.6	23.1	40.1	34.8
Average late scheduled First fit	16	19	7	28	11	1	2	0	0	0
Average access time Best fit	11.0	21.9	21.1	23.3	21.1	33.0	43.0	25.5	40.9	34.8
Average late scheduled Best fit	12	15	6	24	12	2	3	0	0	0

Both scheduling policies are able to schedule all patients in some scenarios, however the first fit policy is clearly better at this than the best fit method. In Table 8.13, we see that the first fit policy performs well in scheduling patients, as for all patient types except type 1 and 3.1, the median is zero. On average, the MSS method with the first fit policy is able to schedule 96% of all arriving patients. In Table 8.14,

this is quite different. In 75% of the scenarios, the patients of type 1 up to 3.1 are not all scheduled. On average, MSS with the best fit policy is able to schedule 88% of all arriving patients.

Table 8.13: Number of not scheduled patients with the MSS approach and first fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	0	0	0	0	0	0	0	0	0
1st quartile	0	0	0	0	0	0	0	0	0	0
2nd quartile	0	1	0	0	0	2	0	0	0	0
3rd quartile	0	3	0	2	3	5	0	0	0	0
maximum	6	14	7	12	10	16	12	0	2	2

Table 8.14: Number of not scheduled patients with MSS approach and best fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	0	0	0	0	0	0	0	0	0
1st quartile	3.25	4	1	4	1.25	1	0	0	0	0
2nd quartile	5	6	2	7	4	4	0	0	0	0
3rd quartile	8	9	3	10	6	7	0	0	0	0
maximum	15	20	10	20	12	16	12	0	2	2

However, if we look at utilization percentages, the scheduling policies both perform similar. The first fit policy has an OR utilization of 78% and the best fit policy has an OR utilization of 77%. As the best fit policy scheduled less patients than the first fit policy, this indicates that the schedule made with best fit, has more empty days than the schedule made with first fit.

8.4. Conclusion

We compared all methods on four aspects of their performance. The number of patient types the method was able to schedule within their desired access time, the number on time scheduled patients, the number of scheduled patients and the OR utilization. We make a distinction between the number of on time scheduled patient types and the number of on time scheduled patients. If a method does not schedule a certain patient type on time, it could be that the session time and priority combination is difficult for the method and the method should be adjusted to perform better. But, if a method has trouble with a lot of the patient types then we can conclude that the method is not fit for solving this problem. The percentage on time scheduled patients gives more insight in the overall performance. In Table 8.15, all results are given for the different scheduling methods and scheduling policies. The column 'Access time achieved' represents the number of patient types that have on average a desired access time. There are 10 patient types, thus, the maximum score is 10. The multiple MIP method performs quite well concerning the desired access times of the patient types, although the percentage of on time scheduled patients is not high enough. This indicates that a lot of patients are scheduled around the desired access time, but possibly a few days late or a few patients are scheduled really late. The multiple MIP method is good at scheduling patients, however, it needs to be improved to really work for the gynecology department in terms of scheduling patients of priority 2 on time. After calculating the demand, we concluded that the overall capacity the gynecology department has, is enough, but there should be more capacity just for priority 2 patients.

The JIT method has overall a good performance. The number of on time scheduled patients is high and the number of achieved access times is also high. The number of scheduled patients and the OR utilization could be improved, but are not the worst compared to the other methods.

The protection level method has the highest scores in terms of on time scheduling, total scheduled and utilization percentage. However, the access times of patient types 2.1 and 2.4 are almost twice the

desired amount of days. So, the protection levels are clearly not optimal and should be adjusted or a different scheduling policy should be implemented.

The MSS method with either scheduling policy has a low number of achieved desired access times and the overall on time scheduling is, thus, also low. Therefore, we think that this method is not usable for the gynecology department. The different patient types are evenly spread over the different weeks in the MSS, however, the method is not flexible enough to adjust to a small change in the patient arrival.

In conclusion, the multiple MIP method, protection levels or the Just-In-Time method are suited for use at the gynecology department, but all should be improved before implementation at the gynecology department is possible.

Table 8.15: Summary of stated results

	Access time achieved	Scheduled on time	Scheduled	Utilization percentage
Multiple MIP method First fit	7 out 10	64%	87%	74%
Multiple MIP method Best fit	7 out 10	58%	92%	70%
Multiple MIP method Random fit	8 out 10	61%	92%	76%
Just-In-Time method	7 out 10	74%	90%	71%
Protection level method First fit	6 out 10	75%	99%	78%
MSS method with first fit	6 out 10	63%	96%	78%
MSS method with best fit	5 out 10	59%	88%	77%

9

Conclusions and recommendations

In this chapter, the conclusion of our research is presented and recommendations for future research and the gynecology department are given. In Section 9.1, the main conclusions of this research project are given. In Section 9.2, we give our main recommendations about future research and how the gynecology department could improve their OR schedule.

9.1. Conclusion

Analysis shows that overall the OR demand and OR capacity are well balanced, and thus, it should be possible to make a schedule with reasonable access times for all patients. In 2015 until 2017, the OR utilization was around 80%, which is good. However, the desired access times are not achieved on average and approximately 60% of priority 1 and 2 patients are treated on time. Moreover, patients receive their surgery date one week in advance, which gives them no time to prepare for it.

We use both K -means clustering and Ward's hierarchical clustering method to determine patient types. The clustering methods result in 10 patient types with different session times. Each patient type contains surgical procedures with the same priority. This is convenient for scheduling as the patients within a patient type have the same priority, so the scheduled time slots are only split up per patient type and not also per priority.

We formulate the problem as an MDP. The MDP formulation has a reduced state space of approximately size $3 * 10^{30}$, which is too large for our computer to compute. We could have used approximate dynamic programming, but decided to try a different method due to inexperience with approximate dynamic programming.

We define a method that uses multiple MIP models to solve the problem. The method consists of three different scheduling approaches for the 3 priority groups. The MIP models, used when scheduling patients, are designed to fill up the days with a bounded number of patients from each patient type per week. There is no rescheduling of patients after a patient is scheduled. To test the performance of this multiple MIP method, we use three methods: protection level method, JIT method and MSS method.

The performance was measured based on four aspects; achieved access times, number on time scheduled patients, the total number of scheduled patients and the OR utilization. The conclusion is that the multiple MIP method, protection levels and the JIT method are most suited for use, however, all methods have some aspect that they do not perform well enough on. Thus, there is some more research needed before it is possible to use any method for the gynecology department.

9.2. Recommendations

The conclusion from Chapter 8 was that all methods were not immediately ready to be used by the gynecology department. We think that by adding additional rules or making small changes, these methods could work in practice and perform better. The largest part of this research project was spent on the MDP method and the multiple MIP method. Thus, the more accessible methods could easily be improved. In Section 9.2.1, we discuss how these models could be improved. In Section 9.2.2, we discuss how these models should be used in practice and what can be expected from the models.

9.2.1. Future work

The protection level method does not work well concerning the access times. Patients of type 2.1 and 2.4 have an average access time of 35 days instead of 20 days. We think that by giving more capacity to priority 2 patients and less to priority 3, the problem with the access times could be solved. Another possibility is to add a rule that patients with higher urgency can be scheduled on capacity of lower urgency, if this capacity is still available. As priority 3 patients are scheduled on the long term, unused capacity in the first few weeks can be used by priority 1 and 2 patients without elongating the access times of priority 3 patients.

The Just-In-Time (JIT) method is not able to schedule enough patients. On average, 10% of all arriving patients is not scheduled with the JIT method. Most of the not scheduled patients are priority 2 patients. This happens as priority 3 patients can be scheduled on specific days that priority 2 can be scheduled. A solution could be to postpone priority 3 patients even further and put a maximum on the scheduled priority 3 capacity per week. This would be a combination of the JIT and protection level methods. Another solution would be to reserve one day each week specifically for priority 2 patients.

The multiple MIP method has not the best score concerning on time scheduling, but did perform well on average access time. This could indicate that patients are scheduled a little over their desired access time. Moreover, the number of not scheduled patients needs to be improved. We expect that adding more oncology capacity and rearranging the schedule of the surgeons, could improve the number of not scheduled patients. Moreover, the weekly and horizon thresholds (T_k , $\overline{T_k}$, $\overline{TH_k}$ and $\overline{TH_k}$) could be improved further and the schedules created by the MIP could be made fuller. Currently, the objective of the MIP is not to use every day with available capacity, thus the schedule made with the MIP model for priority 3 patients does not use every available day when the horizon threshold $\overline{TH_k}$ is met. With these changes, we suspect that all days could be used and it would be easier to achieve the desired access times.

9.2.2. Practical recommendations

The methods are designed and tested to be used under normal circumstances. Normal circumstances meaning no OR reduction days on short notice or holiday periods. If there are sudden changes in the OR schedule from the OR board (OR reduction) and the gynecology receives a reduction for next week, these models are not the answer.

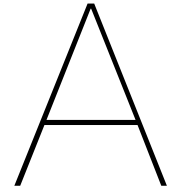
The multiple MIP method, JIT method and the protection level methods are designed to schedule patients up to three months in advance. The patients are scheduled on a certain day by these methods, but the scheduler has to determine the order in which the surgical procedures will be performed. The multiple MIP method optimizes the available capacity, thus the remaining capacity on each day is minimal, which results in more filled up days. However, using the multiple MIP method will take more effort than the other two methods. The JIT and protection level methods do not optimize the available capacity, thus, it could be possible that rearranging the scheduled patients over a week could cause enough consecutive capacity so that another patient can be scheduled when using JIT or protection levels.

It is possible for the three methods to take reduction weeks during holiday periods into account, when the reduction is known 2 to 3 months in advance. However, the multiple MIP method and the protection levels method need some adjustments to do this. The MIP needs different thresholds (T_k , $\overline{T_k}$, $\overline{TH_k}$ and $\overline{TH_k}$) such that the MIP stays feasible in those weeks. The protection level method needs different

protection levels for reduction weeks. As many patients go on holiday during these weeks, elective patients with low urgency will not be scheduled for surgery. Therefore, the reserved capacity for priority 3 patients in the reduction weeks can be lower. Moreover, surgeons also go on holiday, and therefore, some procedures can not be performed during certain weeks.

We have not tested the schedules made by the methods on unexpected delays during the OR-day, which could result in rescheduling of patients. Our main goal was to make a feasible schedule that schedules the patients within their desired access time such that all days are filled up or empty. If a lot of last minute changes need to be made, further research should investigate how this could be incorporated.

Lastly, we would like to recommend the schedulers to gain more insight in their waiting list. For instance, by calculating the requested minutes that are on the waiting list or the requested minutes per priority, patients can be clearer informed about the waiting times. Moreover, it is more clear when low urgency patients can be scheduled earlier as they prefer and when scheduling low urgency patients earlier will cause rescheduling.



Ward's hierarchical clustering

The Ward's method starts with n clusters of size 1 and is adding surgical procedures together in clusters until all observations are together in one cluster of size n . Each time two surgical procedure clusters are merged there is a loss of information. The method uses the error sum of squares (ESS) as measure for the loss of information. The Y-axis represents the value of this distance metric between clusters, we used the euclidean distance metric.

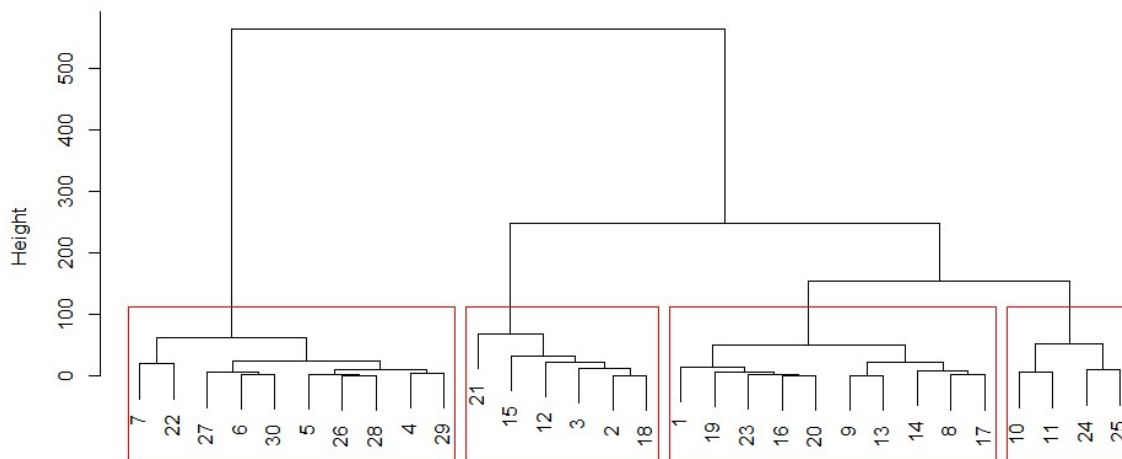


Figure A.1: Wards clustering for priority 2 patients

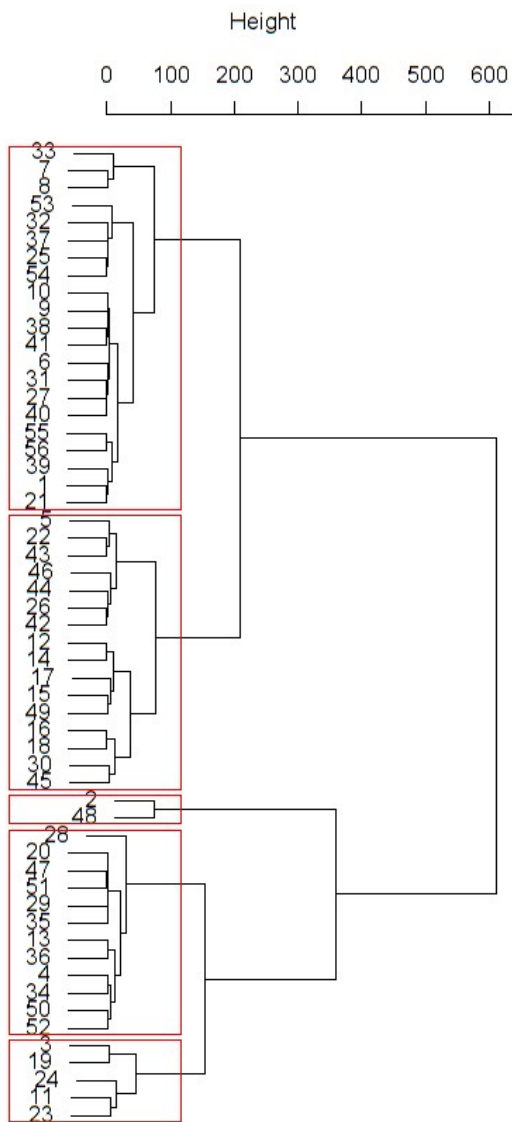


Figure A.2: Wards clustering for priority 3 patients

B

Patient arrival rates

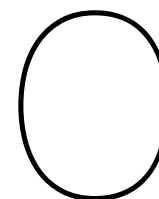
The following two tables contain the number of arriving patients per month per patient type. For example, in February 2015 3 patients of type 1 arrived and 10 patients of type 3.1 arrived, among others. In the last rows of Table B.2, the average, median and maximum number of arrived patients are shown per patient type and lastly, the average demand per day is given. We use the last row for the simulations as the arrival rate of patients. The months November 2014 and January 2018 are not complete in the data set, therefore the total number of arriving patients is lower than in the other months and we did not include these months in any of the calculations.

Table B.1: Number of arriving patients per patient type per month/year

Year - Month	Total	1	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	3.5
2014 - 11	21	2	1	0	3	0	9	4	1	1	0
2014 - 12	75	7	7	5	9	4	14	13	15	1	0
2015 - 1	65	4	11	3	6	8	13	14	6	0	0
2015 - 2	51	3	3	3	4	5	8	12	9	2	2
2015 - 3	60	2	6	2	7	9	14	10	7	2	0
2015 - 4	47	6	3	3	4	3	10	13	5	0	0
2015 - 5	45	8	7	1	4	2	15	4	3	1	0
2015 - 6	53	2	4	4	7	5	14	10	7	0	0
2015 - 7	70	10	6	4	9	8	14	12	7	0	0
2015 - 8	47	7	8	1	10	3	7	7	4	0	0
2015 - 9	45	1	8	1	10	4	8	7	5	1	0
2015 - 10	63	4	13	1	11	3	12	3	10	3	0
2015 - 11	58	6	12	4	7	2	4	15	7	0	0
2015 - 12	48	3	6	3	5	4	11	11	5	0	0
2016 - 1	67	8	6	1	9	5	17	12	7	2	0
2016 - 2	46	8	0	1	5	5	11	7	8	1	0
2016 - 3	54	4	10	1	3	4	12	9	10	1	0
2016 - 4	49	4	8	1	5	6	8	10	6	1	0
2016 - 5	42	3	3	2	9	2	8	7	7	1	0
2016 - 6	43	4	3	0	9	7	6	7	7	0	0
2016 - 7	55	3	2	3	8	2	18	14	4	1	0
2016 - 8	47	1	4	5	5	5	9	12	6	0	0
2016 - 9	53	7	2	2	4	3	15	14	6	0	0
2016 - 10	52	4	3	2	8	3	10	15	7	0	0
2016 - 11	62	2	5	3	12	8	13	8	10	1	0
2016 - 12	51	4	6	3	7	4	16	5	6	0	0

Table B.2: Number of arriving patients per patient type per month/year

Year - Month	Total	1	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	3.5
2017 - 1	50	6	5	3	7	3	2	13	9	1	0
2017 - 2	49	4	7	1	5	2	10	15	5	0	0
2017 - 3	52	3	4	2	13	4	6	8	11	1	0
2017 - 4	42	6	3	3	4	2	11	8	5	0	0
2017 - 5	47	6	6	4	8	3	8	7	5	0	0
2017 - 6	55	4	8	3	12	4	8	7	8	0	0
2017 - 7	42	0	5	2	2	6	8	9	10	0	0
2017 - 8	42	3	3	1	6	4	7	9	7	0	0
2017 - 9	37	5	7	0	6	2	3	9	3	1	0
2017 - 10	48	3	3	4	6	7	9	11	3	0	0
2017 - 11	34	4	5	3	5	0	5	5	5	0	1
2017 - 12	20	2	2	0	2	3	0	5	2	0	0
2018 - 1	5	2	0	0	0	1	1	1	0	0	0
Total average	50,43	4,35	5,51	2,30	6,84	4,16	9,84	9,64	6,68	0,57	0,08
Total median	49	4	5	2	7	4	10	9	7	0	0
Maximum number	75	10	13	5	13	9	18	15	15	3	2
Demand per day	1.63	0.13	0.17	0.07	0.23	0.13	0.33	0.3	0.23	0	0



Late scheduled patients

The following tables give more information about the number of late scheduled patients. A patient is scheduled late if the patient has to wait longer than their desired access time. The desired access time for priority 1 patients is 10 working days, for priority 2 patients 20 working days and for priority 3 patients 65 working days. All tables show the number of patients that are minimally and maximally scheduled late and the quartiles are given.

In case we are using the multiple MIP method and first fit policy, we can conclude from Table C.1, that there are always priority 2 patients that are scheduled too late, which could indicate that the multiple MIP method does not reserve enough time for priority 2 patients. We can conclude from Table C.2 and comparing the table with Table C.1, that the best fit scheduling policy does not perform better than the first fit scheduling policy. When using the best fit scheduling policy, there are also always priority 3 patients that are scheduled too late. This is not as bad as scheduling priority 2 patients late, but does indicate that something is not working with this policy. When comparing Tables C.1 to C.3, we can conclude that concerning late scheduling of patients the first fit policy is performing best.

The JIT method has the most difficulty with scheduling patient type 2.1 and 2.4 on time (Table C.4). Priority 1 and 3 patients are always scheduled on time, when they are scheduled. Patients of type 2.1 and 2.4 have a session time of respectively 225 and 295 minutes. To schedule these patients, we can not fill up holes in the schedule, but need half a day or more. As JIT does not optimize the used capacity it is possible that the remaining capacity is not enough to help a patient of type 2.1 or 2.4.

The protection level method has difficulty with scheduling patients of priority 1 and 2 (Table C.5). As this method does not optimize the used capacity, it is possible that the remaining capacity is not enough to schedule a patient of priority 1 or 2. Therefore these patients are scheduled late. It is also possible that the set protection levels are too low for priority 1 and 2 such that they have to move to the next week, when there is still capacity left.

The MSS method has difficulty with scheduling all patients on time (Tables C.6 and C.7). This indicates that the MSS is not working and that it reserves too much capacity for patients that have not (yet) arrived or the OR utilization of the MSS is not high enough to handle the load.

Table C.1: Number of late scheduled patients with multiple MIP method and first fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	4	2	5	2	0	0	0	0	0
1st quartile	0	13	6	17	7	2	1	0	0	0
2nd quartile	0	16	7	21	8	3	2	1	0	0
3rd quartile	0	18	9	26	10	4	3	2	0	0
maximum	2	26	17	40	16	11	8	6	2	0

Table C.2: Number of late scheduled patients with multiple MIP method and best fit approach

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	9	1	10	2	2	1	1	0	0
1st quartile	0	14	5	21	6	10	6	5	0	0
2nd quartile	0	16	7	25	8	14	8	7	0	0
3rd quartile	0	19	9	29	9	19	10	9	0	0
maximum	3	25	15	43	15	33	20	17	1	0

Table C.3: Number of late scheduled patients with multiple MIP method and random fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	6	2	12	2	1	0	1	0	0
1st quartile	0	13	6	21	7	5	4	7	0	0
2nd quartile	0	16	8	24	9	7	5	9	0	0
3rd quartile	0	18	10	28	11	9	7	11	0	0
maximum	4	25	17	38	16	17	16	25	2	0

Table C.4: Number of late scheduled patients with JIT approach

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	0	0	0	1	0	0	0	0	0
1st quartile	0	5	0	0	9	0	0	0	0	0
2nd quartile	0	8	1	1	12	0	0	0	0	0
3rd quartile	0	10.75	3	3	14	0	0	0	0	0
maximum	0	20	10	15	23	0	0	0	0	0

Table C.5: Number of late scheduled patients with protection levels approach and first fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	8	0	0	8	0	0	0	0	0
1st quartile	1	18	5	12	15	0	0	0	0	0
2nd quartile	3	20	7	19.5	17	0	0	0	0	0
3rd quartile	5	23	9	25	19	0	0	0	0	0
maximum	16	30	17	38	24	0	0	0	0	0

Table C.6: Number of late booked with MSS approach and first fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	0	0	1	0	0	0	0	0	0
1st quartile	11	17	4	24	6	0	0	0	0	0
2nd quartile	16	20	7	29	11.5	1	0	0	0	0
3rd quartile	21	22	10	33	15	2	2	0	0	0
maximum	28	26	12	37	20	7	27	0	1	1

Table C.7: Number of late booked with MSS approach and best fit policy

Priority	Priority 1	Priority 2				Priority 3				
Cluster number	1	1	2	3	4	1	2	3	4	5
minimum	0	7	2	13	3	0	0	0	0	0
1st quartile	11	14	4	21	10	1	0	0	0	0
2nd quartile	12	16	5	24	12	2	1	0	0	0
3rd quartile	14	17	7	26	14	3	3.75	0	0	0
maximum	17	21	10	32	18	9	28	1	1	1

Bibliography

- [1] I.J.B.F. Adan and J.M.H. Vissers. Patient mix optimisation in hospital admission planning: a case study. *International Journal of Operations & Production Management*, 22(4):445–461, 2002.
- [2] D Astaraky and J Patrick. A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. *European Journal of Operational Research*, 245(1):309–319, 2015.
- [3] C Banditori, P Cappanera, and F Visintin. A combined optimization-simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics*, 24(2):155–187, 2013. ISSN 14716798. doi: 10.1093/imaman/dps033.
- [4] C Barz and K Rajaram. Elective patient admission and scheduling under multiple resource constraints. *Production and Operations Management*, 24(12):1907–1930, 2015.
- [5] D Bertsimas and S De Boer. Simulation-based booking limits for airline revenue management. *Operations Research*, 53(1):90–106, 2005.
- [6] F Guerriero and R Guido. Operational research in the management of the operating theatre: A survey. *Health Care Management Science*, 14(1):89–114, 2011.
- [7] E W Hans, M van Houdenhoven, and P J H Hulshof. *A framework for health care planning and control*. Memorandum / Department of Applied Mathematics. Department of Applied Mathematics, University of Twente, 2 2011.
- [8] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1):100, 1979.
- [9] I Hof. Scheduling all patients within their desired access time by determining a reservation level for the consultation hours, 2017.
- [10] P J H Hulshof, N Kortbeek, R J Boucherie, E W Hans, and P J M Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems*, 1(2):129–175, 2012.
- [11] M Lamiri, X Xie, A Dolgui, and F Grimaud. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185(3): 1026–1037, 2008.
- [12] J Patrick, M Puterman, and M Queyranne. Dynamic multi-priority patient scheduling for a diagnostic resource. *INFORMS*, 5(4):7–24, 2014.
- [13] M L Puterman. *Markov decision processes : discrete stochastic dynamic programming*. Wiley, 1994. ISBN 0471619779 9780471619772.
- [14] N M Van de Vrugt. Efficient healthcare logistics with a human touch, 7 2016.
- [15] J M Van Oostrum, T Parlevliet, A P M Wagelmans, and G Kazemier. A method for clustering surgical cases to allow master surgical scheduling. *INFORMS*, 49(4):37, 2008.
- [16] I. B. Vermeulen, S. M. Bohte, P. A.N. Bosman, S. G. Elkhuisen, P. J.M. Bakker, and J. A. La Poutré. Optimization of online patient scheduling with urgencies and preferences. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5651 LNAI:71–80, 2009.

- [17] L Wang and D Gupta. Revenue management for a primary-care clinic in the presence of patient choice. *Operations Research*, 56(3):576–592, 2008.
- [18] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.
- [19] G Zhu, D Lizotte, and J Hoey. Scalable approximate policies for Markov decision process models of hospital elective admissions. *Artificial Intelligence in Medicine*, 61(1):21–34, 2014.