A SURGICAL PROCEDURE TYPE SCHEDULE FOR GENERAL SURGERY

A robust tactical surgery scheduling approach to manage elective- and semi-urgent patient uncertainty

Author:

R.F. Jacobs

Examination Committee:

Prof. Dr. Ir. E.W. Hans, University of Twente

Ir. N.J. Borgman, University of Twente

A.A.J. van der Zalm, HagaZiekenhuis

UNIVERSITEIT TWENTE.

HagaZiekenhuis

Management Summary

The specialty General Surgery at HagaZiekenhuis (Haga) is struggling to keep access time for elective patients below the national set limits while keeping operating room (OR) availability for semi-urgent patients high. Scheduled elective patients are cancelled at the last moment for semi-urgent and emergency patients that require surgery. An analysis of current General Surgery performance shows an average utilization of 68% in elective surgical schedules with an overtime frequency of 38%. For semi-urgent patients, the probability of access to surgery within a week is currently 47%. For elective patients, Haga achieves a five week access time probability of 37% compared to the national allowed limit of 80%. The access time probability within seven weeks is 75% compared to the national allowed limit of 100%. We identify a number of underlying causes of poor performance and determined that these mainly originate on tactical and operational offline levels. We determine that on these levels, the current scheduling approach fails to properly manage uncertainties related to surgical demand and duration. For example, nearly 86% of the surgery duration estimates that OR planners use when scheduling are off more than 10 minutes. Therefore, our research objective is:

To develop an OR scheduling approach which manages surgical demand and duration uncertainty for elective and semi-urgent patients.

Solution approach

Based on a theoretical framework, we propose a robust cyclic surgical schedule aimed on managing surgical demand and duration uncertainty. To manage this uncertainty, we decompose the solution approach into several steps. We apply a clustering approach proposed by *van Oostrum et al.* [1] as a method to combine individual surgical procedures into homogenous surgical procedure types in terms of duration. This allows us to reduce demand uncertainty through a pooling effect.

To manage semi-urgent demand uncertainty we apply the discrete time slot queuing theory approach presented by *Kortbeek et al* [2]. The queuing model determines the probability of access within a week based on a chosen number of slots. We determine the number of slots to be the weekly demand for semi-urgent patients that we should cover to provide timely access and to prevent the current frequent elective patient cancellations.

With elective and semi-urgent demand input known, we apply a mathematical programming approach with column generation approach based on *van Oostrum et al.* [3] to create a surgical procedure type schedule (SPTS). In this SPTS, we select operating room days (ORDs), which are ORs filled with surgical procedure types from an implicit set and assign these to specific dates and operating rooms. Implicit refers to the fact that we iteratively expand the set with potential ORDs. New ORDs are iteratively generated in a sub-model that offer an improvement to this set. This sub-model incorporates surgical slack and the portfolio effect by described by *Hans et al.* [4] for both Gaussian and log-normal distributed surgical duration procedures to manage overtime probability. For all sub-specialties, we select as many ORDs required to balance elective waiting lists to an acceptable level. The result is a schedule where surgical cases can be planned into the first available surgical procedure type slots that they

are part off. Only semi-urgent procedures should be scheduled in slots reserved for semi-urgent procedure types.

We propose a flexible approach to creating a SPTS that consists of a fixed SPTS and a variable number of monthly add-on ORDs that fluctuates in response to production estimate variations. The number of add-on ORDs should be determined monthly, based on production estimates for each sub-specialty. This approach presents two advantages. The fixed component allows for easy continuation in planning for surgeon planners, staff planners and admission office schedulers to cover the majority of monthly demand. Relative small variations in monthly demand are accounted for with the variable component, which ensures an overall close match of demand and capacity.

Results

We conclude that Haga will benefit from using a SPTS. The size of this benefit vary depending on which management choices are made with relation to overtime probability, semi-urgent slots, opening hours and whether a flexible or static SPTS approach is chosen.

We introduce a default scenario where ORs are opened eight hours a day, only one sub-specialty is allowed per OR and where we reserve 44 semi-urgent patient slots. The default scenario results in an OR utilization of 79%, one week access time probability of 85% and overtime probability of 30%. We also introduce a method where ORs are subdivided into several smaller ORDs of a different capacity. When we add the possibility of multiple ORDs capacity types, utilization may increase up to 85% and less capacity is required for the same throughput. The best results with a 30% probability of overtime can be achieved when implementing five and three hour ORDs in an eight hour day, with an OR utilization of 85%. For practicality, we advise Haga to implement a combination of eight and four hour ORDs that are planned into eight hour days, which result in an OR utilization of 82%.

A monthly minimum of 43 slots is required to stabilize access time for semi-urgent patients, which results in a 82% probability of access within a week. Access time can be improved by reserving more slots, but the marginal access time benefit of each slot decreases. We advise to reserve 44 semi-urgent slots, which results in a 85% probability of access within a week for semi-urgent patients.

Utilization is influenced by the amount of slack that is chosen by management to limit the probability of overtime. With an overtime probability of 40%, the estimated utilization increases to 89%. When we decrease the overtime probability to 20%, the estimated utilization drops to 73%. We recommend to limit the probability of overtime to 30%.

We argue that Haga will also benefit from using the flexible approach towards using a SPTS. On paper, this approach is slightly outperformed with up to 2% by the static approach in terms of utilization. However, we anticipate that the flexible approach offers benefits in dealing with elective demand fluctuations, such a reduced risk of idle time in periods of low elective demand.

Recommendations

Our main recommendation is to implement our flexible SPTS approach consisting of eight and four hour ORDs with monthly 44 semi-urgent slots and a 30% probability of overtime. We expect that the initial period requires closer monitoring to determine a proper balance between production and elective demand. To reduce the current waiting list, we advise to schedule an increased amount of ORDs in the SPTS and evaluate after that. After the two initial months, we advise to schedule a monthly capacity re-allocation meeting as recommended in subsection 6.3.1 to estimate the number of variable add-on ORDs required for the next month. An excel tool is available to determine the number of add-on ORDs required, based on the size of the waiting list.

The SPTS provides an overview of those surgical procedures that can be assigned at each specific date and OR. Available elective procedures may be scheduled on a first come- first served base. The SPTS can be scheduled with available surgical cases for an entire month such that patients can be notified well in advance of their procedure date. Semi-urgent shots should be kept open for semi-urgent patients. We advise to send the schedule to the OR department one week in advance. If semi-urgent patients arrive after this time, these patients can be scheduled “online” in the still available slots. If semi-urgent slots are not filled two days in advance, we advise to schedule an elective patient in the available slot. For each sub-specialty, the added excel tool comes with a list of potential elective procedures that fit in the duration in a semi-urgent slot.

We also provide some general recommendations:

- Increase the quality of data by adapting how data is registered at the OR.
- Implement an OR dashboard to monitor a chosen set of performance indicators.
- Set specific utilization targets for specialties based on their case-mix when dimensioning capacity at strategic level.
- Implement a monthly meeting with specialties to discuss production estimates and capacity re-allocation for the next month.
- Implement a hierarchical structure with responsibilities at OR personnel related to morning preparations of the OR.
- Implement two starting times at the OR to reduce workload for anesthesiologists during morning rush hour.

Management Samenvatting

Het snijdend specialisme Heelkunde binnen het HagaZiekenhuis heeft moeite om de toegangstijd voor wachtlijst patenten onder de nationaal gestelde norm te houden en daarbij operatie kamers (OKs) toegankelijk te houden voor semi-spoed patiënten. Geplande wachtlijst patiënten kunnen op het laatste moment worden afgezegd om plaats te maken voor semi-spoed patiënten. Een prestatie analyse van de Heelkunde laat een gemiddelde OK benutting zien van 68%, met een uitloop frequentie van 38%. De kans op toegang binnen een week voor semi-spoed patiënten is 47%. Vergeleken met de nationale norm van 80% hebben wachtlijst patiënten een kans van 37% om binnen vijf weken toegang te krijgen tot chirurgische ingrepen. Voor zeven weken is dit 75% vergelijken met de nationale norm van 100%. We onderscheiden een aantal onderliggende oorzaken van prestatievermindering die vooral op tactisch en operationeel niveau bestaan. We concluderen dat de huidige planningsaanpak op deze niveau niet in staat is om onzekerheid met betrekking tot ingreep vraag en duur te managen. Een voorbeeld van een onderliggende oorzaak is dat schattingen van de ingreep duur in 86% van de gemeten gevallen met meer dan 10 minuten afwijkt van de realiteit. Daarom stellen we de volgende onderzoeksopdracht voor:

Het ontwikkelen van een OK planning aanpak gericht op het managen van vraag en duur onzekerheid van chirurgische ingrepen van wachtlijst en semi-spoed patiënten.

Aanpak

Aan de hand van een literatuuronderzoek stellen we een robuste cyclische OK planning voor, gericht op het managen van vraag en duur onzekerheid van chirurgische ingrepen. Om dit te bereiken delen we onze aanpak op in stappen. Eerst passen we een cluster methode toe op de zes sub-specialismen binnen Heelkunde om verschillende ingrepen samen te voegen als ingreep typen. Deze methode is beschreven door *van Oostrum et al.* [1] en staat ons toe om onzekerheid in vraag te verminderen door aggregatie.

Om de onzekerheid met betrekking tot de vraag naar semi-spoed ingrepen te managen passen we het discrete tijdvak wachtrijmodel van *Kortbeek et al.* [2] toe. Hierbij nemen we aan dat iedere patiënt binnen een tijdvak met onbepaalde lengte geholpen kan worden. Dit model bepaalt de kans op toegangstijd binnen een week aan de hand van een gekozen aantal vakken dat we reserveren voor semi-spoed patiënten. We bepalen het aantal vakken en stellen dat gelijk aan het aantal semi-spoed patiënten dat we wekelijks moeten reserveren om tijdig toegang te kunnen bieden, en om te voorkomen dat wachtlijst patiënten afgezegd moeten worden. Dit aantal vakken wordt aan de hand van historische observatie onder sub-specialismen van Heelkunde verdeeld, net als de duur van een algemene semi-spoed ingreep binnen ieder specialisme.

Nu de semi-spoed en wachtlijst vraag bekend is, passen we een wiskundig programeer model toe beschreven door *Oostrum et al.* [3] om een schema te maken met chirurgische ingreep typen. In dit schema selecteren we operatie dagen (ODs), gedefinieerd als een dag op de OK die volgepland is met chirurgische ingreep typen, uit een impliciete set met verschillende ODs. Nieuwe ODs worden iteratief gegenereerd door een sub-model en toegevoegd aan de set als

ze een verbetering zijn. Dit sub-model gebruikt slack tijd en het portfolio effect beschreven door *Hans et al.* [4] voor ingrepen met zowel een normaal als log-normaal verdeelde duur. Voor alle sub-specialismen kiezen we het minimale aantal ODs nodig om aan de ingreep type vraag te voldoen. Het resultaat is een schema waarbij ingrepen in het eerst mogelijk beschikbare passende ingreep type vak gepland kunnen worden. Ingrepen mogen alleen in hun eigen sub-specialisme vakken gepland worden, evenals semi-spoed.

We stellen een flexibele methode voor om een ingreep type schema te maken. Deze methode bevat een vast schema en een variabel aantal extra toe te voegen ODs. Het aantal extra ODs wisselt per maand en hangt af van of productie schattingen voor de komende maand groter zijn dan momenteel gepland staat. Deze aanpak met een vaste en variabele component heeft twee voordelen ten opzichte van een volledig statisch schema. Het veruit grootste deel van het schema kan nog steeds herhaaldelijk gebruikt worden en hoeft daarom niet aangepast. Het relatief kleine variabele aantal extra ODs zorgt ervoor dat de capicteit netjes aansluit op de vraag, en voorkomt dat OKs onbenut blijven bij perioden met minder vraag.

Resultaten

We concluderen dat Haga profijt zal hebben van het gebruik van een ingreep type schema. De mate van de winst hangt af van management keuzes die gemaakt worden met betrekking tot de kans op uitloop, het aantal semi-spoed tijdvakken, OK openingsduur en of een flexibele of statische methode wordt gebruikt voor het ingreep type schema.

Het basis scenario resulteert in een OK benutting van 79%, met een kans van 85% op toegang binnen een week voor semi-spoed patiënten en een kans van 30% op uitloop van de OK. Als we de mogelijk van het gebruik van OKs met verschillende duur toestaan kan benutting toenemen tot 85%, waarbij minder OK capaciteteit nodig is. De beste resultaten met 30% kans op uitloop worden behaald met ODs van vijf en drie uur, met een benutting van 85%. Om pragmatische redenen adviseren we Haga een combinatie van acht en vier uur durende ODs die resulteert in een OK benutting van 82%.

Een maandelijks aantal van 43 tijdvakken is nodig om toegangstijd voor semi-spoed patiënten te stabiliseren tot binnen een week met een kans van 83%. Deze kans kan worden vergroot door extra tijdvakken toe te voegen, maar met afnemende winst voor ieder tijdvak. We adviseren Haga om 44 vakken te reserveren, wat resulteert in een kans van 85% op toegang binnen een week.

Benutting wordt beïnvloedt door de hoeveelheid “slack” dat gekozen wordt door management om de kans op uitloop te beperken. Met een kans op uitloop van 40% kan de benutting oplopen tot 89%. Wanneer de kans op uitloop beperken tot 20%, daalt de benutting naar 73%. We adviseren Haga om de kans op uitloop te beperken tot 30%.

We denken dat Haga baat zal hebben bij het gebruiken van de flexibele methode om een ingreep type schema te maken. Op papier presteert leidt aanpak tot 2% OK benutting minder dan de statische aanpak. We denken echter dat de flexibele aanpak voordelen biedt in het inspelen op schommelingen in vraag van wachtlijst patiënten, zoals een verminderde kans op leegstand in perioden met minder vraag.

Aanbevelingen

Onze hoofdaanbeveling is het implementeren van een ingreep type schema via de flexibele methode, met ODs van acht en vier uur en een maandelijks aantal van 44 tijdvakken voor semi-spoed patiënten. We verwachten dat de wachtlijst extra goed in de gaten gehouden moet worden in de eerste twee maanden om tot na te gaan hoe de instroom van nieuwe patiënten beïnvloedt wordt door het nieuwe schema. Om de huidige wachtlijst te verminderen adviseren we om de eerste twee maanden extra ODs te draaien. Na twee maanden adviseren we om een maandelijkse bijenkomst te houden, gericht op het schatten van toekomstige productie en het aantal benodigde extra ODs voor de komende maand. Een excel-tool is beschikbaar om het aantal extra ODs te bepalen aan de hand van een huidige wachtlijst.

Het ingreep type schema biedt een overzicht van welke ingrepen op welke dag ingepland kunnen worden. Wachtlijst patiënten kunnen in het eerst beschikbare vak gepland worden. Het ingreep type schema kan voor een hele maand ingepland worden zodat patiënten tijdig op de hoogte kunnen worden gesteld van hun ingreep. Semi-spoed tijdvakken moeten vrijgehouden worden voor semi-spoed patiënten. We adviseren om een ingeplande week van het schema een week van te voren naar de OK te sturen. Semi-spoed patiënten die hierna nog in beeld komen kunnen in de vrijgehouden tijdvakken gepland worden. Wanneer er twee dagen van tevoren nog geen semi-spoed patiënt in beeld is, kan er getracht worden om alsnog een wachtlijst patiënt in te plannen. Voor ieder sub-specialisme is er een lijst met potentiele verrichten die in een semi-spoed tijdvak past.

Verder hebben we nog enkele algemene aanbevelingen:

- Verbeter de kwaliteit van gegevens huishouding door aan te passen hoe gegevens moeten worden geregistreerd op de OK. Implementeer daarbij een prestatie monitor aan de hand van een gekozen verzameling prestatiemeters.
- Kies specifieke benutting doelen voor ieder specialisme, gebaseerd op hun patiënten-mix, wanneer er jaarlijks capaciteit wordt toegewezen op strategisch niveau aan de hand van productie afspraken.
- Las iedere maand een bijeenkomst in met vertegenwoordigers van alle specialismen om productie schattingen te bespreken voor de komende periode, en of er OK capaciteit herverdeeld kan of moet worden.
- Implementeer een duidelijke structuur met verantwoordelijkheden binnen het OK personeel gericht op de voorbereidingen aan het begin van de dag binnen de OK. Kies voor twee starttijden met tussenpozen van een kwartier om de ochtend werkdruk van anesthesisten te beperken.

Preface

This thesis will mark the end of an eight year adventure at the university of Twente. I started out as a Technical Medicine student, still uncertain of what I really wanted for a career. After I had sought out enough distractions, I realized that I was looking for something different in my studies. One talk with Erwin Hans convinced me to take a leap a faith and start a masters in Industrial Engineering & Management. I loved it from the first course on and have never regretted the transition.

During the first thesis meeting with Erwin, I recall that I wanted elements of queuing theory, linear programming and excel in my research. Looking at the results, I can safely say that Erwin really delivered. It took a long time to develop a clear solution approach and to determine which techniques would be suitable. Dedicating myself to new theory while still unsure of its usability was sometimes difficult and frustrating, and made the first months more difficult than I had anticipated. As the solution approach became more solid, it became increasingly easier to find motivation to put in extra hours. Now that my thesis is finished, I am happy that I had to piece this large puzzle of a thesis together. I hope that it can contribute to Haga and its patients.

I would like to thank Nardo Borgman for our weekly thesis feedback sessions. He guided me in the right direction whenever I was struggling with different concepts and helped me structure the solution approach. I also would like to thank Arnoud van der Zalm for his input and support. His experience helped in putting data in perspective and getting a clear view of what is really happening at Haga. I would like to thank Erwin Hans for his feedback on the research process and his guidance and support in my overall graduation. I would like to thank all those that had an impact in my life at Twente, at my rowing club, student fraternity and in general. Lastly, I would like to thank my parents and girlfriend for supporting me so long in all my academic and un-academic adventures.

Ruud Jacobs

Contents

1 Introduction

In this thesis we conduct a research on how to improve operating room (OR) performance of General Surgery at the HagaZiekenhuis (Haga). The first chapter provides with a short background introduction of Haga. Chapter 2 provides an analysis of how General Surgery functions, after which the problem and the research objective are explained.

Section 1.1 provides some background information about Haga. Section 1.2 explains the motivation for this research. Section 1.3 explains the research methodology that we use to determine how to improve OR performance.

1.1 Context

In this section, we provide background information and key figures about HagaZiekenhuis and the specialty general surgery.

1.1.1 HagaZiekenhuis

HagaZiekenhuis is one of 28 top-clinical hospitals in the Netherlands, situated in The Hague. It facilitates highly specialized care, education to medical personnel and medical research. Its employees number nearly 3600, of which over 210 are medical specialists. Haga originated in 2004 as a merger between three hospitals situated in The Hague and is still divided over multiple locations, with multiple OR departments. Note that OR department refers to the entire OR complex, OR to a single room and ORs to multiple operating rooms. The location of interest in this thesis is Leyweg which facilitates the OR department for General Surgery. Some key figures from 2014 are detailed in Table 1.

Employees	**3569**
Medical specialists	**317**
Beds	**611**
Admissions	**29728**
Average length of stay per admission	**5**
Single day admissions	**29644**
New outpatient visits	**200412**

Table 1.1: Key figures of Haga in 2014 (Source: Annual report 2014)

1.1.2 General surgery

The specialty General Surgery covers a wide array of medical sub-specializations such as for example vascular-, abdominal-, oncologic- and trauma-surgery. A total of 16 sub-specializations are covered by 13 surgeons. These surgeons perform major surgical procedures in the OR, minor outpatient procedures and outpatient consults. In this thesis, only the major surgical procedures that are performed in ORs are of interest.

General Surgery makes intensive use of the OR department by performing surgical procedures on three types of patients. Emergency patients (1) arrive through an unexpected incident and urgently require care, often immediately. The other patient types are both elective patients. Elective patients with more urgent requirements are a distinctive group called “semi-urgent” (2) patients. The date and time of their surgical procedure is urgent but can still be planned to a various degree. Elective patients (3) that are less urgent are simply called elective patients. There is no clear definition to as what makes a patient elective or semi-urgent, but in general a border of 2 weeks is used.

1.2 Research motivation

The specialty General Surgery at Haga is struggling to keep access time for elective patients below the national set standards while keeping availability for (semi-)urgent patients high. Scheduled elective patients often give way at the last moment for unexpected semi-urgent and emergency patients that require surgery.

Historically, operations research and management science approaches have rarely been used within Haga. However, an increasing number of operations research approaches have recently been introduced by Graduate and PhD- students through cooperation with the University of Twente. These methods have had positive results with complex problems. Haga hopes that an operations research approach will have similar results at General Surgery.

1.3 Research methodology

We would like to maximize the performance benefits of our solution approach. A problem however, is that there are many indicators of OR performance available and many factors that influence performance. How does one identify the key performance indicators and the main contributors to poor OR performance? *Heerkens et al.* [5] proposes a general managerial approach as a research framework to identify problems and determine the right solution approach that we will use as a guideline throughout this thesis.

The general managerial approach demands that we first analyze the current situation and performance to determine of it meets the desired standard. We then identify the underlying causes that lead to the problems that were found in performance through a problem bundle. The problem bundle is a comprehensive cluster of all underlying causes and their interrelations, which enable us to determine the core causes that we want to focus on. This enables us to come up with tailored solution approaches that target the main contributors of poor performance. The framework also provides a research cycle if we lack information at some point in the general managerial approach.

1.4 Research objective & demarcation

Based on the research motivation, we propose our research objective to be:

To develop an OR scheduling approach which manages surgical demand and duration uncertainty for elective and semi-urgent patients.

A key goal of the new planning approach is to manage semi-urgent surgical demand uncertainty robustly in such a way that access time requirements are met without generating online operational chaos and elective patient cancellations. By robust, we refer to probabilistic robust optimization in which we quantify uncertainty in the “true” value of the parameters of interest by probability distribution functions. Similar to semi-urgent demand, our planning approach should also be robust against overtime. Other gains from a new structured approach to scheduling should be increased utilization of resources and a smoother workflow for medical personnel.

We introduce the framework presented by *Hans, et al.* [6] to subdivide planning decisions in four hierarchical levels of planning and four managerial areas. Figure 1.1 visualizes the framework. Our focus of OR scheduling is identified in the framework as the domain of resource capacity planning. Section 2.2 provides an extensive description of the hierarchical levels, and how Haga operates on these levels. We will research a new solution approach to resource allocation on Tactical and Operational offline levels. We found that problems reported at online operational level have their origins mainly in planning decisions at these two levels. We reason that interventions at these levels exert positive influence on performance at online operational level, while interventions at online operational level will result only in minor efficiency gains within a framework of inadequate planning decisions. We will also exclude interventions at strategic level. We anticipate that we are able to increase performance within the current framework of strategic decisions, such as current case-mix and long horizon capacity dimensions.

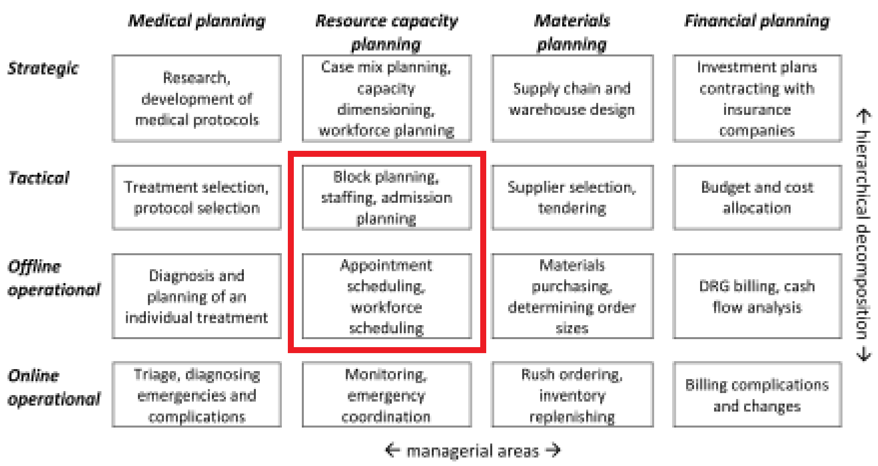

Figure 1.1: Thesis demarcation within the hierarchical framework presented by Hans, et al.

To demarcate this research further, we exclude the observed surgeon consult scheduling problem that leads to the lack of admission control. This means that we will research a new solution approach for the numbers 6 to 13 and 16 of the observed core problems in subsection 2.5.4. We will exclude performance measurement related core problems 19 and 20 in our research since Haga is already developing a performance measurement system in its new data registration system. The research will focus on improving OR performance only within the general surgery specialism on the Leyweg location, which covers both elective surgery care- and emergency care services as shown in the framework above. Any procedures that occur outside of the general OR department at Leyweg, such as small outpatient surgical procedures, are excluded. To scope this research, effects on pre- and postsurgical wards, ambulatory- and inpatient care services are not taken excluded.

1.5 Research questions

Based on the research objective and methodology, the following research questions are formulated. Each research question corresponds to a chapter in this thesis.

- *How is the current scheduling approach organized and how does it perform?*
    - How is the current system organized?
    - How are planning and control of the current system organized?
    - How does the current scheduling process perform?
    - Where should our research focus on?

We describe the current system and the current planning and control approach in chapter 2. We define performance indicators and analyze the current system performance to determine the underlying causes of poor performance. We then introduce our research objective and demarcation.

- *Which operations research techniques could be applied in our solution approach?*

Through literature research we compose a theoretical framework with promising mathematical techniques for our solution approach in chapter 3.

- *How should the organizational intervention be modeled?*

We introduce our solution approach and describe the underlying models and assumptions in chapter 4.

- *How does the proposed intervention perform?*

We determine the performance of our proposed solution approach and conduct additional experiments. We describe the results and compare them with current performance in chapter 5.

- *What are the main findings and what recommendations could be made for implementation?*

The findings of this thesis are summarized and discussed in chapter 6.

2 Analysis of current situation

In this chapter we conduct a study of the current system of General Surgery to chart performance and underlying causes of bad performance.

To understand the context of the performance measurements, first the current system is explained. The current system is divided in 2 components, the OR department and the control component. The Operation Rooms department is described in section 2.1. Understanding of how the OR department works will provide insight in the methods and restrictions encountered when reviewing the control process explained in section 2.2. In section 2.4, we describe some of the demand characteristics of general surgery patients. In section 2.4, we conduct a performance analysis of the current system. In section 0, performance results are analyzed to determine the problem bundle and ultimately the core problems that result in lack of performance. With the core problems charted, section 1.4 describes the research objective and section the scope. Finally, section 2.6 provides a brief summary of this chapter.

2.1 OR Department description

In this section, the OR department is described. To gain an understanding of the OR department, first the general lay-out and staff are described in subsections 2.1.1 and 2.1.2. Subsection 2.1.3 explains the process of a surgical procedure.

2.1.1 OR department lay-out

All of General Surgery's capacity is allocated to operating rooms at the location Leyweg. This OR department was delivered in June 2015 and features 15 rooms divided over two floors. This set-up was designed to separate three different types of surgical procedures, children's, fast-track and "regular". Children procedures are all procedures on patients under the age of 18 and require special facilities and personnel. Fast track procedures "quick" procedures and generally do not involve narcosis. The remaining procedures can be described as "regular". For the General Surgery department, only "regular" procedures are of importance.

The first floor contains the fast-track area for single day admission patients with 2 smaller ORs and a lounge-like holding and recovery ward. These ORs are mainly used for eye-surgery. It also contains a section with 3 ORs and a holding and recovery ward dedicated to children. Due to their high specialization, these 5 ORs are never used by General Surgery. Apart from these 5 dedicated rooms, the first floor does contain 2 "regular rooms" that are used frequently by General Surgery. The second floor contains 8 ORs, that may be used by all the medical specialties, but most rooms do feature some facility that benefits some specialties more than others. For example, an OR may have a drain which is useful for Urology procedures or connection-points for a Heart-Lung machine. General Surgery may perform surgery in any of these 8 ORs on the second floor, or in the 2 "regular" ORs on the first floor.

The second floor also contains the Holding, Recovery and PACU wards that are used by General Surgery. Note that the Holding, Recovery and PACU wards are small and temporary wards surrounding a surgical procedure, and are distinct from the general wards which have much higher capacity and are intended for longer stay. The Holding ward contains 4 beds is and used to check, prepare and hold patients for surgery. The Recovery ward contains 7 beds and is used to monitor patients until they are recovered enough to transport them back to the other wards. The PACU ward contains 4 beds and is used for as a recovery area for patients who require more intensive monitoring. The PACU ward is a separated space in the same room as the Holding and Recovery with its own beds, but shares personnel with the other two wards.

Figure 2.1 depicts a schematic view of the 2nd floor of the OR department. We may observe that 2 ORs in the right corner are larger than the others. These particular ORs are more beneficial to Orthopedics due to their space requirements, but can be used by all specialties.

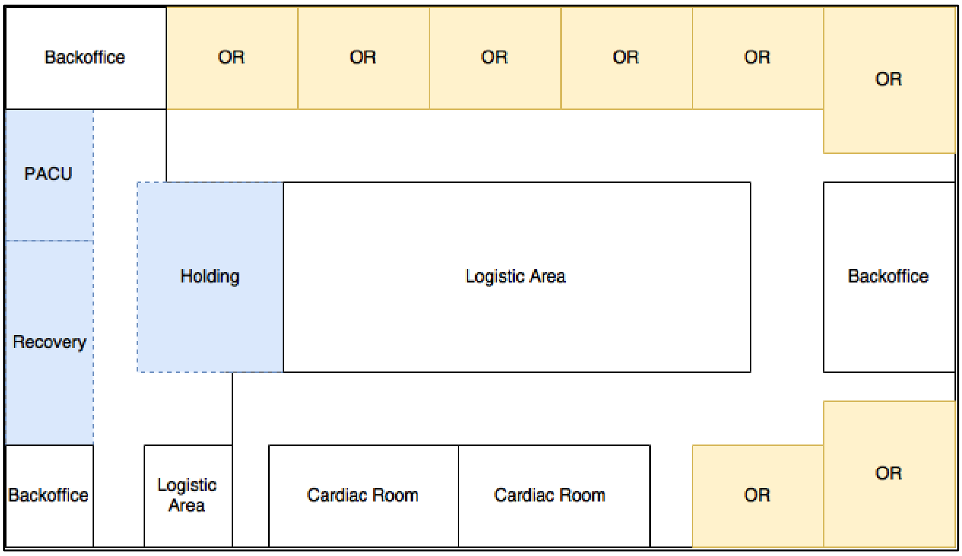

Figure 2.1: Schematic overview of 2nd floor OR Department

General wards

General Surgery has access to several different General Surgery wards and a short-term ward where patients recover after transport from the Holding ward. The General Surgery back office states that there is always capacity for patients but that ward personnel workflow varies.

2.1.2 Personnel

In this subsection we describe the personnel involved with surgical procedures and the planning process, and their responsibilities.

We define groups and assign personnel to these groups based on their tasks and responsibilities. An overview of the groups is visualized in Table 2.1,

Personnel group	**Personnel**
Project manager	Arnoud van der Zalm
Day coordinators	Program-, surgeon- and anesthesia-coordinators
Surgeons	All surgeons
Anesthesiologists	All anesthesiologists
OR personnel	Surgical-and anaesthesiologist assistants, nurses, surgeon-residents, cleaners and support personnel
Ward personnel	Holding- and Recovery- and PACU ward personnel

Admission office	Unitmanager & OR planners

Table 2.1: Overview of involved personnel (2016

Project manager

A key role in the OR planning process is played by project manager who is involved with many planning related subjects. Examples are capacity re-allocation decisions, OR performance measurement, and allocating capacity to specialties and creating the master surgical schedule, which is further explained in section 2.2.

Day coordinators

Three key coordinators within the OR department are the program-, surgeon and anesthesia-coordinator. The program coordinator is responsible for the day activities such as capacity-, schedule- and personnel-management. The program coordinator is involved with any uncertainty or non-clinical issue that may come up during the day, and is the contact for ORs and the admission office. The other two coordinators are mainly involved with the clinical discussion involving new emergency patients, and any clinical challenge in their field that may arise during the day.

Surgeons

Surgeons perform the surgical procedures and supervise the remaining personnel during surgery. They are responsible for anything that happens during the procedure. Surgeons generally switch surgical activities with other activities, such as outpatient consults. The time that surgeons spend at the OR department may vary greatly, unlike most of the personnel involved with surgery.

Anesthesiologists

Anaesthesiologists are responsible for anything narcosis related regarding the patient, which is common for most procedures in the OR department. Each anaesthesiologist supervises 2 ORs at the same time. Anesthetization occurs in the OR just before the surgical procedure. After anesthetization, the patient is monitored by an anaesthesiologist-assistant who is present at the room and the surgeon receives full control over the patient. At the end of the procedure, the anaesthesiologist returns to awake the patient from narcosis. Anaesthesiologists generally work full time and only within the OR department.

OR personnel

The OR personnel consists of a large group such as surgery assistants, anaesthesiologist-assistants, nurses, cleaners, residents and all other support personnel. Assistants are specialized personnel that assist surgeons and doctors with the surgical procedures. Generally, each OR has 2 surgery assistants and 1 anaesthesiologist-assistant that are supervised by either the surgeon or anaesthesiologist. Cleaners are responsible for cleaning the ORs during patient changeovers and at the end of the day. Residents have no acting function but may sometimes perform tasks under supervision of the surgeon. The OR personnel, excluding the residents, generally works only within the OR department. Note that OR personnel is not the same as OR department personnel, which consists of all the personnel at the OR department.

Holding, Recovery and PACU Ward personnel

These wards contain the personnel that is specialized in preparing and recovery of patients before and after surgery. These wards share the same room and personnel, and thus can be seen as one large group. This personnel generally works only within the OR department.

Admission office OR patient planners

The admission office is a different department in another part of the building than the ORs. At the admission office, we may identify the unit manager and patient planners. The unit manager often confers with the earlier described project manager about specialty capacity related issues. There are also two patient planners who are responsible for scheduling surgical cases into available ORs. These planners seldom physically enter the OR department but play an important role in the control process explained in section 2.2.

2.1.3 OR Process

For both elective and the vast majority emergency patients, the Holding ward receives patients before their surgical procedure and performs a time-out. In this time-out, the patient and requirements for the planned procedure are reviewed. Patients often require a range of preparations for surgery, such as blood-tests and suspension of some medication. If a patient is not adequately prepared, the Anesthesiologist may reject him as unfit for surgery or narcosis. For the patient, this may mean a delay varying from hours to days depending on the required preparations. After a successful time-out, the ward prepares the patient for surgery where possible. Holding is not responsible for calling or transporting the patients from their respective wards, and may announce an admission-stop to the program-coordinator of the OR department if their 6 beds are all occupied. The OR rooms call directly to the wards on where the patients are and the wards are responsible for transporting the patients. The first patients of the day are generally ordered a day in advance, so that they arrive at 07:30 at the Holding. The only exceptions to this process are extremely urgent emergency patients in dire need of surgery. For example, patients with a ruptured aorta need to be helped within minutes if they are to have a chance at survival.

Once the OR is available, the patients (if ready for surgery) are picked up from Holding by the OR personnel. At the OR, the patient receives a final time-out check for which both surgeons and anesthesiologists have to be available. Often, the patient is then brought under narcosis after which the surgical procedure starts. When the patient wakes up from narcosis in the room, he is brought for monitoring to either PACU or the Recovery ward. After the criteria of recovery are met, the patient is transferred back to his own ward. Times can be registered in SAP at several moments.

1. When the patient is ordered from his or her ward
2. When the patient arrives at Holding ward
3. When the patient arrives at the OR
4. When the anesthesiologist is finished anaesthetizing
5. When the surgical procedure starts
6. When the surgical procedure is finished
7. When the patient is awakened

8. When the patient leaves the OR
9. When the patient leaves the Recovery ward

The timestamps and the different actions and waiting times that they entail are depicted in Figure 2.2. Unfortunately data registration is optional, which means that data is often lacking. The most consistent registrations are found at numbers 3 and 8.

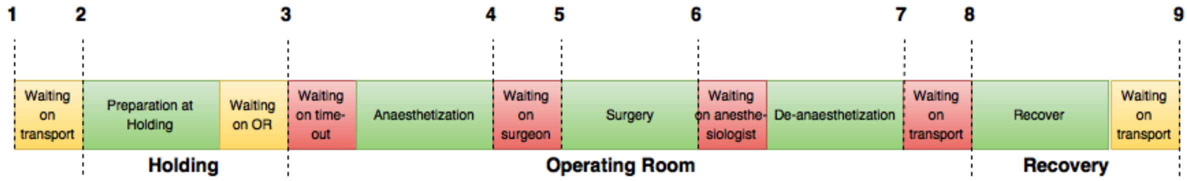

Figure 2.2: Visualization of timestamps

2.2 OR Planning & Control

This section describes the resource capacity planning decisions which control the flow of patients through the OR department described in section 2.1. The framework presented by *Hans, et al.* [6] and introduced in section 1.4 is applied to structure the different resource capacity decisions which are observed.

The entire control process is also visualized in Table 2.2 and Figure 1.1. Table 2.2 briefly describes the demand and supply characteristics, planning horizon and planning criteria on each level. Figure 2.4 visualizes the stakeholders and their decisions on each level of the framework. Each level of the scheduling process is explained in detail in subsections 2.2.1, 2.2.2, 2.2.3 and 2.2.4.

	Demand	**Supply**	**Horizon**	**Optimization Criteria**
Strategic	Production estimates from specialties	Product of ORs, shift duration and weekdays per year	1 year >	Check accuracy of production estimates. Reserve enough capacity per specialty to cover check production estimates
Tactical	Reserved capacity for General Surgery	Product of ORs, shift duration and weekdays per year	1 year >	Smooth allocation of ORs over 4 week cycle MSS
	Surgeons requesting weekly OR capacity	MSS OR Capacity	6 weeks >	Equal distribution of available capacity to surgeons
Operational Offline	Surgery waiting list	Scheduled ORs and surgeons	14 - 3 days >	Schedule semi-urgent patients within maximum access time. Maximize utilization. Schedule elective patients FCFS
Operational Online	Scheduled patients and emergency patients	Scheduled ORs and surgeons	Daily	Minimize access time for emergency patients. Maximize utilization. Minimize make-span.

Table 2.2: Key attributes of control process

2.2.1 Strategic

On the strategic level, Haga planners allocate OR capacity to specialties based on production data and forecasts.

Data regarding surgery frequencies and durations about past years amount of procedures performed and procedure durations are gathered from the hospital system (SAP). Apart from these achieved numbers each surgery performing specialty is interviewed about current perceived bottlenecks and next year's case mix and production forecasts. The gathered forecasts are compared with historical achieved numbers for consistency. For each procedure, the resulting production number is multiplied with its historical average duration to determine the required OR capacity in minutes. This capacity is increased by a factor of 15% to compensate for idle time, and the resulting time is computed to units of OR days. An OR day is defined as exclusive access to an OR for the duration of 8 hours (during weekdays). Each specialty is allocated a number of OR days for the next year based on these calculations. The allocations are adjusted for the total available capacity available. The full distribution can be inspected in Figure 2.3. Next, the number of OR days per week is determined for two staff settings. One is the setting in which the specialism operates within the OR with fully staffed. The other is a reduced "holiday" staffing with some estimated percentage of the normal available staff. Haga distributes the amount of OR days between a fixed number of 41 normal working weeks and about 10 holiday weeks with a reduced staffing. With this distribution, the number of OR days per week are determined for the 2 different situations. The specialty General Surgery receives 12 OR days each week, and a dedicated emergency OR, also known as a flex room, which is staffed by surgery personnel receives (on average) 7,5 days per week.

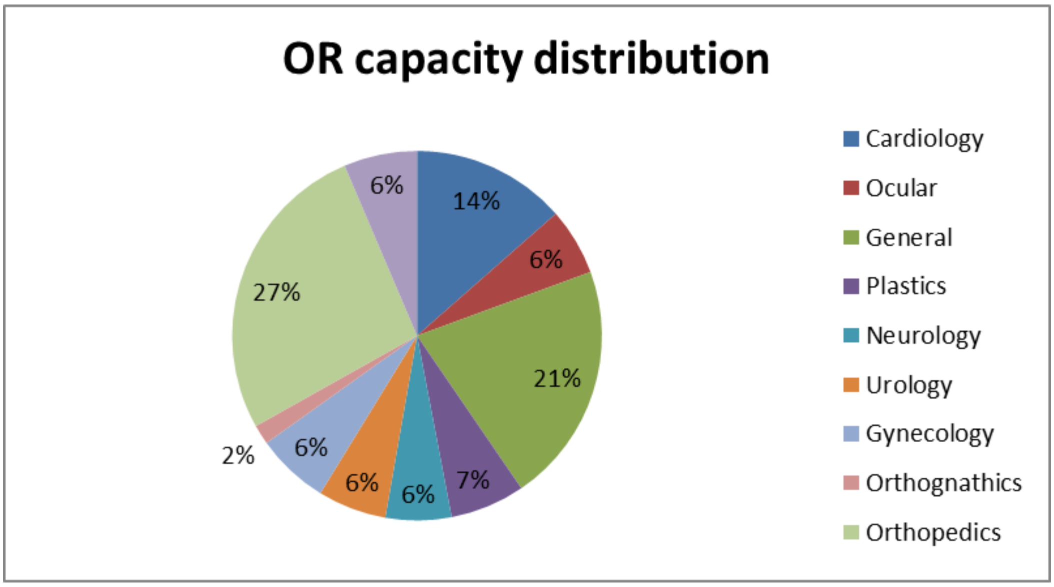

Figure 2.3: Distribution of OR capacity amongst specialties (Source: SAP OR planning, numbers for 2015).

2.2.2 Tactical

On Tactical level, Haga Planners create a four week cycle Master Surgical Schedule (MSS) based on the input from the Strategic level. Also, some background information is given on the methods used to schedule the personnel described in subsection 2.1.2. We differentiate the personnel in 2 groups based on how they are scheduled.

Master Surgical Schedule

A specialty may have received a fractional number of weekly OR days at Strategic level. In such a case, the actual allocation in the MSS is adjusted such that over the four week cycle the average number of allocated OR days equals the earlier calculated fraction. This procedure is done for both the normal week situations and the reduced holiday week situations. While OR planners are aware of period variations in surgery demand, this information is currently disregarded in the design of the block schedule. A collection is made of procedures which are too rare to effectively estimate the demand from but the majority of this information is also disregarded in the scheduling. The normal four week cycle is applied until the complete year is planned. For known holiday periods and days, the adjusted holiday schedule is used. Any resulting discrepancies between the number of actual yearly scheduled OR days and earlier allocated yearly OR days due to such holiday adjustments are manually minimized. This results in a Master Surgical Schedule where on each day all the ORs are allocated to specialties.

Surgeon schedules

General Surgery surgeons are scheduled six weeks in advance by the General Surgery ward staff. Surgeons have fixed weekdays on which they have outpatient consults, educational days and surgery. A surgeon may determine this based on his own preferences. In principle, each surgeon is allocated a full OR day per week, regardless of patient demand. The scheduler makes sure that surgeons of each sub-specialty are on call for emergency patients.

Remaining personnel schedules

The remaining personnel can be described as all the described personnel from subsection 2.1.2, excluding surgeons. They are scheduled one month in advance without any regard patient demand. Personnel are scheduled “by hand” by a scheduler. OR Department Personnel either scheduled in the normal day shift, which runs from 07:30 to 16:30, the support shift, which runs from 09:00 to 18:00 or in one of the evening and night shifts. Each OR is manned by personnel from the day shift, and two support shifts make sure that the day shift personnel get to lunch. The support shifts also take over ORs that are running late, to prevent frequent overtime of the day shifts. The scheduler will make sure that there is an even balance in the amount of shifts each personnel member receives. Within General Surgery, there are no real specializations for the OR personnel. The scheduler will make sure that there is always a certain balance between experienced and inexperienced personnel for any surgical procedure.

Elective patient control

Apart from surgeries, the surgeon also performs in- and outpatient consults. Some of these out-patients are referenced by a general practitioner and result in inflow of elective patients. Currently, there are admission controls in check at the outpatient clinic to influence the inflow of new elective patients. All out-patient consults are accepted and scheduled on a first come-first served base.

2.2.3 Operational Offline

At Operational Offline level, OR patient schedulers at the admission office schedule patients to the allocated capacity in the MSS.

Admission office

When a medical specialist decides that a patient should be planned for surgical operation, his or her personal information, net surgery time and urgency (or maximum access time) are registered in a physical admission form and forwarded to the OR patient schedulers at the admission office. The urgency of a surgical procedure may vary between days to several months. There are no clear definitions for urgency, but in general access time below two weeks is considered "semi-urgent", and smaller than five days is considered "urgent". The definition "net surgery time" defines the duration of time between the first incision of the surgeon, and closing of the entry points. Another definition called "slack" is used to denote all extra time needed for a single surgery, such as preparation of the OR and cleaning afterwards. This slack is a fixed estimate of twenty minutes.

When registered, the patient is automatically placed on a waiting list in SAP. However, OR patient schedulers manually check the admission papers too. Over time, they have accumulated insight in urgency of different surgical procedures. New physical admission files are checked with their corresponding position on the digital waiting list for potential human errors.

The amount of available ORs is known for each day through the MSS. Surgeon availability is planned well in advance by General Surgery schedulers. Initially, patients that are feasible for the scheduled surgeons are scheduled according to urgency. Each procedure is scheduled with the net surgery time registered on the admission form and twenty minute slack time. This process is repeated until the full capacity of the OR is reached. No room is reserved for unexpected elective patients. Patients are not scheduled into flex rooms.

Apart from some general surgical procedures, surgeons can only perform surgery within their own sub-specialization. Often, some semi-urgent patient demand from sub-specializations for which there is no capacity left remains. As a result, OR patient schedulers often try to swap surgeons around in cooperation with the General Surgery personnel schedulers. The resulting offline schedule is updated every day, until it is forwarded to the OR department three days in advance.

Program Coordinator

The program coordinator at the OR department receives the concept schedule three days in advance. The program coordinator may adjust the sequence of the scheduled patients to minimize conversion times between the different procedures, based on her experience. Apart from conversion times, it is preferred that day-admission patients are scheduled early in the day, so that they can recover in time to be discharged on the same day. Patients with a high narcosis risk are preferred in the morning so that their recovery can be monitored by the same anesthesiologist. Any changes in sequence are communicated back to the admission office, so that that the patients can be notified. The program coordinator reviews all the planned surgical procedures of all the specialties and determines the allocation of specialties to the

ORs based on their necessities of all the procedures. The flex room is usually allocated to one of the remaining ORs.

Any urgent patients that arrive within and need to be treated within the 3 day window of the concept schedule are planned in consultation between the program coordinator and the admission office. This may result in overtime or cancellations of less urgent patients since the initial schedule is often filled completely. Sometimes, these patients are not scheduled but told to come back the next day and admit themselves as emergency patients.

2.2.4 Operational Online

At operational online level, the OR concept schedule is executed and the program coordinator allocates emergency patients that arrive during the day. On the day of the surgeries, the program coordinator manages most of the decisions. The program coordinator will for example find the next replacement if a patient is unexpectedly rejected or does not show up. At the end of the day, the program coordinator determines if ORs may finish up or take in patients that were scheduled in an OR that is progressing slow, to minimize the make-span.

Emergency Patients

Emergency patients may either have arrived sometime during the night, or during the day.

For the emergency patients that “arrived” overnight, a meeting is held with the surgeon- and anesthesia-coordinators at 07:30 in which these patients are discussed medically and planned for surgery. If emergency patients are determined fit for surgery, they are scheduled in the flex-room that has its own personnel but no surgeon. The surgeon is determined by the-

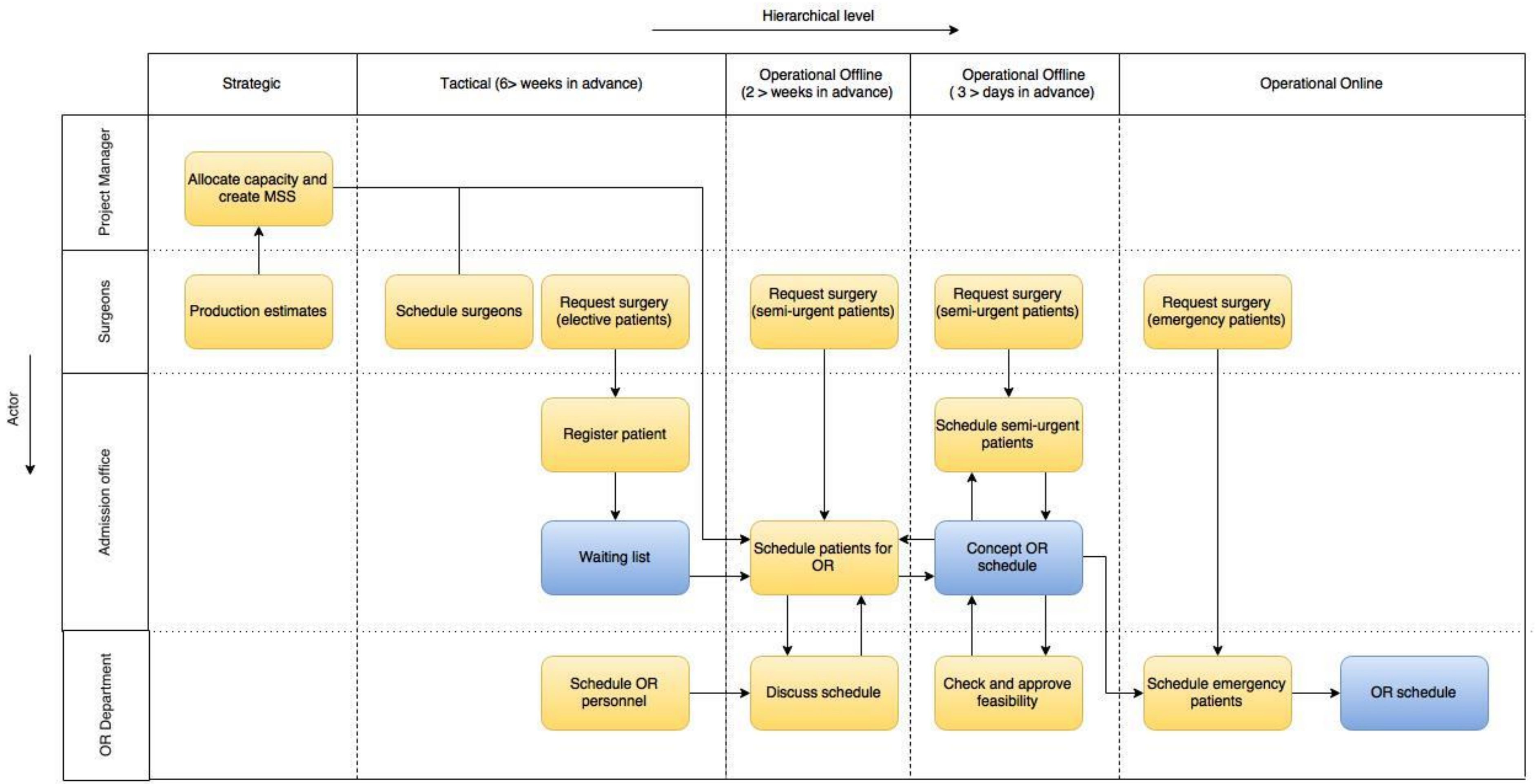

Figure 2.4: Control process visualized

surgeon-coordinator of the day, since each emergency patient may require a specific area of sub-specialization within General Surgery. This means that an emergency patient sometimes has to wait before the right surgeon is available, depending on the schedule of the surgeon. The program-coordinator determines the sequence of the flex room schedule based on the availability of the surgeons.

The urgency of emergency patients that arrive during the day are determined by the surgeon- and anesthesiology-coordinators. Coordinators may decide to perform surgery on the same day on that the patient is scheduled to the flex room. Emergency patients are always initially allocated to the flex room and kept separate from elective patient programs. In some instances, such as multiple high urgent patients or procedures with room specific necessities, the program coordinator may decide to break in the program of an elective OR. A patient may also be deemed urgent or semi-urgent, after which he or she is registered on an admission form and forwarded to the admission office. The patient is then registered as described in subsection 2.2.3.

2.3 Demand characteristics

In this section, we analyze some of the patient demand characteristics. Wee analyze access times and patient distribution and take a deeper look into demand variability.

Patient distribution

We analyse the requested and recorded access times from December 2012 to December 2015 to obtain a better understanding of the patient characteristics of General Surgery. We introduce two different variants of access time. Requested access time is the maximum stated access time by the surgeon on a surgery request. Notice that emergency patients are often not placed on the waiting list and hence do not have a registered requested access time. We therefore introduce the recorded access time as the realized number of days between the request and the actual surgery. This is technically also a measure of performance, but we require it to approximate the number of emergency patients. Both request and recorded access times may vary per patient for the same procedure, due to varying urgency.

Figure 2.5 shows a histogram of elective requested access times from a sample of 11159 patients. Access time in days is plotted horizontally against the recorded frequency. Notice that patients with higher access times than two months are not included to keep the image comprehensive, but are added in statistics as “elective”. While Haga does not use clear distinctive categories of urgency some natural levels of urgency for elective patients can be identified from the peaks in frequency. Noticeable is that surgeons seem to prefer certain fixed time domains for their access times, such as one to four weeks.

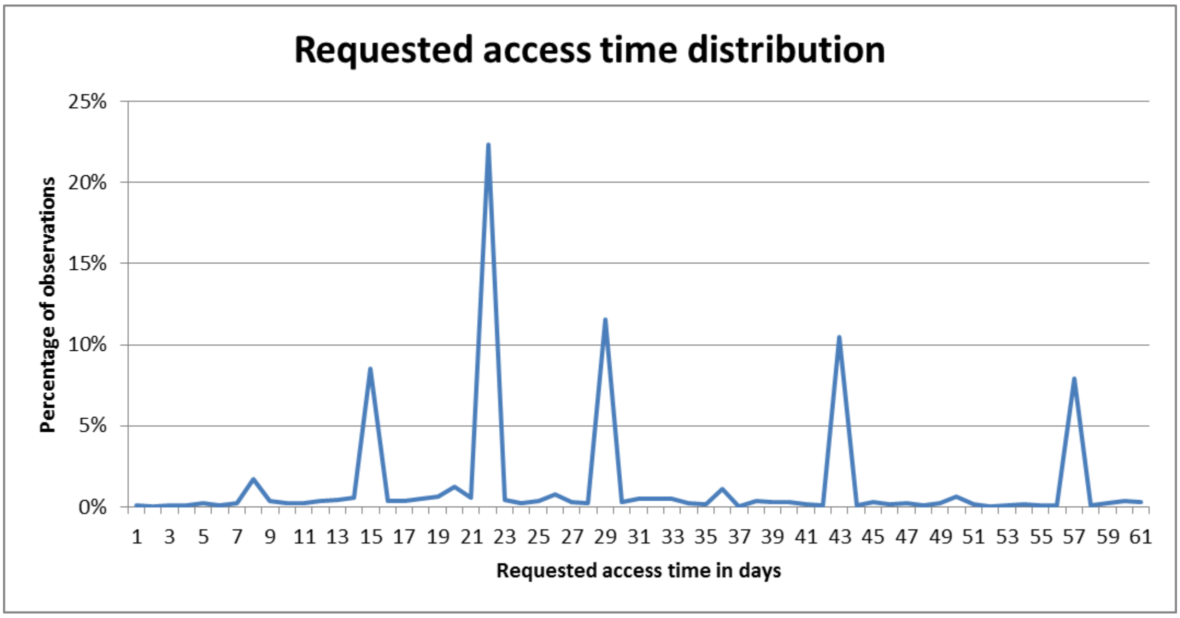

Figure 2.5: Requested access time up to 60 days for elective General Surgery patients (Source: SAP, data between 03-13 and 10-15).

Figure 2.6 shows a histogram of recorded access times of 11159 patients with a realized access time of less than two months. Noticeable is the large peak for patients with access times within one day, which can be attributed to emergency and semi-urgent patients.

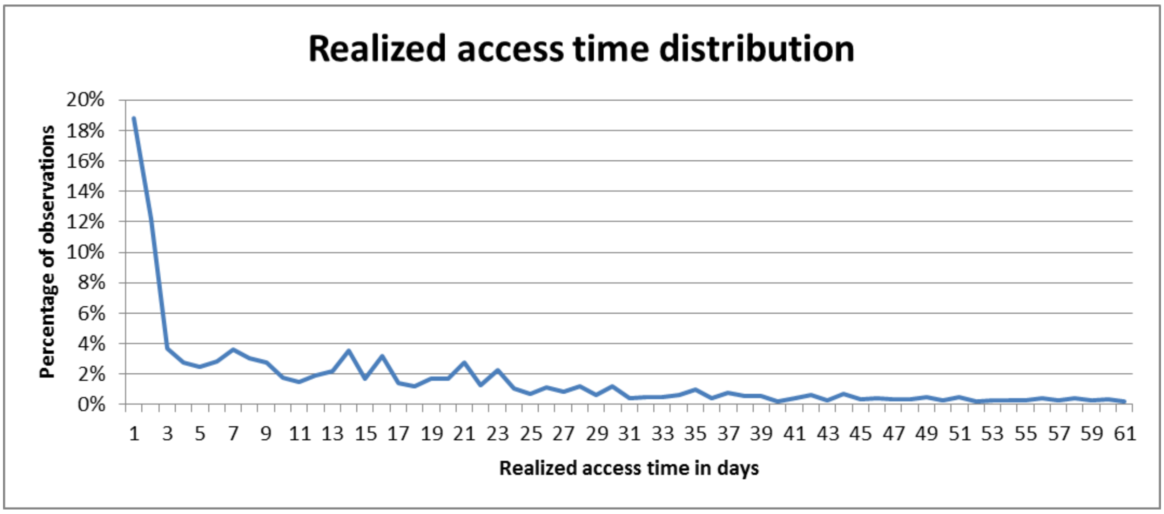

Figure 2.6: Recorded access time up to 60 days for General Surgery patients (Source: SAP, data between 03-13 and 10-15).

While there is a policy for registering patients as emergency at Haga, it is not strictly enforced. It is difficult to determine whether patients with an access time of one day were emergency patients who arrived after office hours and could not be helped, or were simply less urgent. From the peaks around seven and fourteen day access times it can be reasoned that many urgent patients undergo surgery right before their requested access times expire. Since cancellations are not recorded at Haga, it is impossible to determine any statistics about cancellations and resulting waiting times.

Figure 2.7 shows a distribution made from the recorded access time of 9957 patients on which surgery was performed between June and October. The boundaries between patient types were selected based on access time patterns in Figure 2.5 and Figure 2.6.

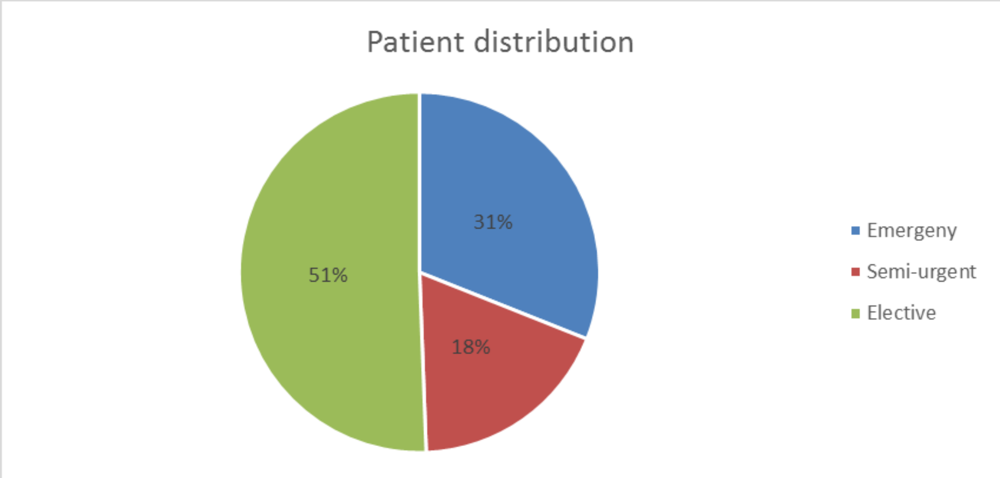

Figure 2.7: Distribution of urgency of patients of General Surgery (Source: SAP, data between 06-15 and 11-15).

We can easily see that half of the patients are either semi-urgent or emergency and arrive within two weeks prior to the surgical procedure. Approximately 21% of General Surgery patients is even scheduled online (within the 3 day period) on top of the initial schedule by the program coordinator. It can be easily reasoned that if the elective ORs are completely filled with known patients three days in advance at operational offline, planning such a large group of extra patients on top of the schedule will cause a lot of strain.

Apart from the urgency distribution, we also review the distribution of surgical procedure types. This provides us some insight in the repetitiveness of surgical procedures at General Surgery. Figure 2.8 shows a distribution of the frequency of surgical procedures. When assuming the current planning cycle of four weeks at Haga, it seems that 86% of the executed procedures are repetitive in nature and performed at least once every period.

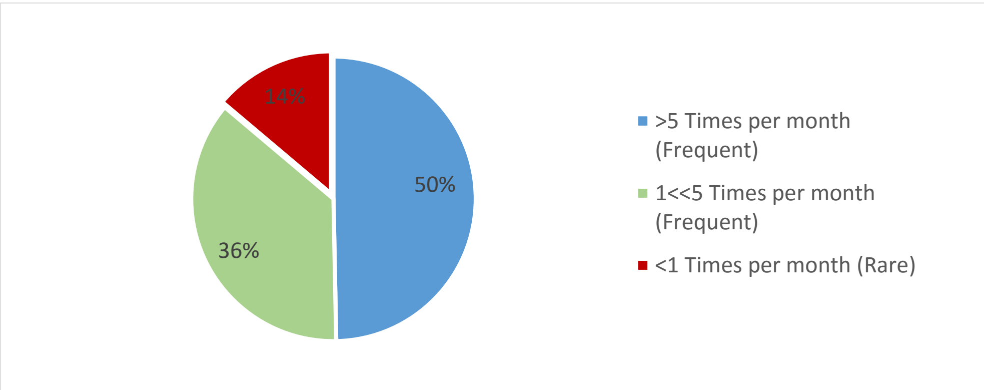

Figure 2.8: Distribution of frequency of surgical procedures of General Surgery (Source: SAP, data between 06-15 and 11-15).)

Demand variability

We can examine the arrival statistics per month in 2014 to determine demand variability between sub-specialties of General Surgery. We also examine the average emergency patient demand per day of the week. Insight in demand variability helps understand demand and supply unbalances since Haga mainly works with fixed capacity allocation based on surgeon preferences.

To illustrate the variability, Figure 2.9 depicts the amount of patients that request surgical procedures of the sub-specialties Traumatology and Vascular surgery. We notice that there is a lot of variability per month between these sub-specialties. In-hospital factors that cause variability such as reduced staffing in February, August and December may affect different sub-specializations differently. They result in a reduction of both surgery and patients consults. An external factor noticed by surgeons is that patients prefer to postpone surgery till after holiday periods.

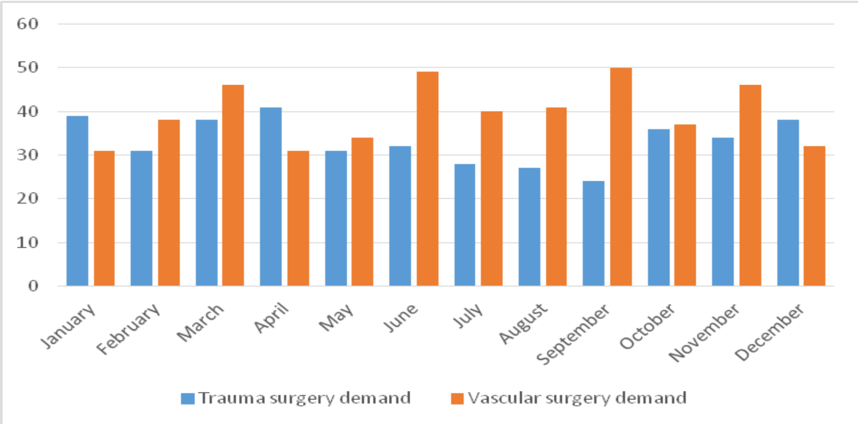

Figure 2.9: Average monthly demand of Traumatology and Vascular surgery (Source: SAP, data from 12-12 to 12-15)

We can also observe variability within the arrival statistics of emergency patients. Table 2.3 depicts the average number of arrivals of emergency patients per day of the week. These numbers might be biased since Haga does not record the first visit of Emergency patients, only the date a surgical procedure was performed. In reality, demand on Monday may be even further increased since some of the patients that arrive in the weekend have to wait until Monday for surgery. Note that there is possibly also variability between arrival statistics on, for example, Mondays between spring and summer.

Weekday of…	Q1	Q2	Q3	Q4
Sunday	2.7	3.4	3.3	2.8
Monday	3.8	3.4	4.3	3.8
Tuesday	3.6	3.9	4.3	3.7
Wednesday	3.9	3.7	4.1	4.4
Thursday	3.8	3.7	4.0	3.7
Friday	4.5	3.9	4.6	4.3
Saturday	2.8	3.0	3.0	2.8

Table 2.3: Average arrival of emergency patients during weekdays (Source: SAP, data between 12-12 and 12-15).

2.4 Performance analysis

This section analyses the performance of the current system. Subsection 2.4.1 provides some insight into access time performance. In subsection 2.4.2, we analyse the utilization of the current system in depth to determine core problems that undermine performance.

2.4.1 Access time

In this subsection, we study the realized achieved access time of patients to gain insight in General Surgery performance. We recall realized access time to be the number of days between the first request for a surgical procedure and the date of the actual procedure.

General Surgery has both a lot of different procedures and large variation in access time per procedure. For most surgery requests, the surgeon determines the maximum allowed access time that is then used when scheduling the patients at operational offline level. We would like to determine whether patients gain access within the maximum time window requested by their surgeon. Since most emergency patients do not have a recorded allowed access time, we will only focus on semi-urgent and elective patients. We define that any procedures with a requested between one and seven days are semi-urgent and that those with requested access times of 8 days and upward are elective.

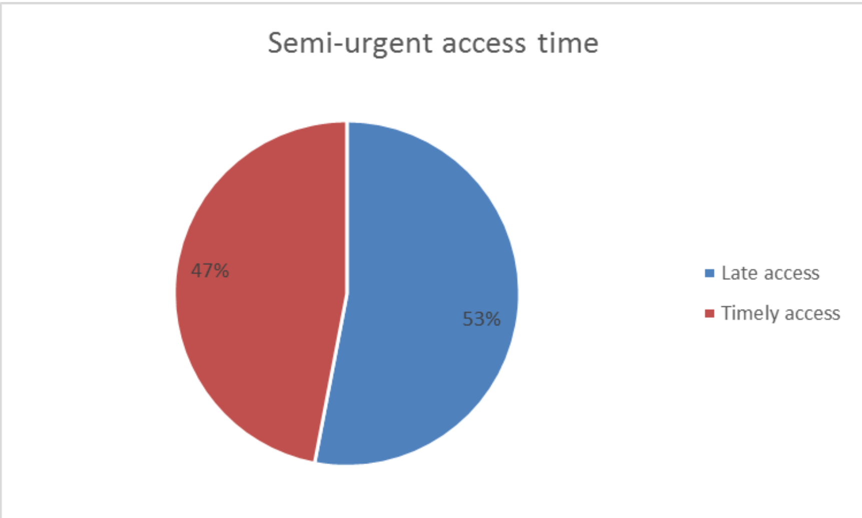

Figure 2.10: Access time performance for semi-urgent patients (Source: SAP, data between 03-13 and 10-15).

Figure 2.10 presents the access time performance for semi-urgent patients. We can observe that 53% of semi-urgent patients, which we define as those with a requested access of within a week, cannot be helped timely. To gain more insight in access time performance, we analyze the access times for elective patients. Since these patients receive a large maximum access time from their surgeons, we may assume that there is little medical urgency for this type of patients. Differently said, we may assume that elective patients have equal priorities that make them suitable for comparison.

For elective patients, we find that 60% did not receive access within the timeframe requested by their surgeon. In 2012, the Nederlandse Zorg authoriteit (Dutch Healthcare Authority) determined that the national maximum allowed access time for elective surgery should be 7 weeks, of which 80% of patients should have an access time of 5 weeks [7]. We compare elective access performance against theses limits and find that only 37% of all elective patients receives access within five weeks, and 75% within seven weeks. Figure 2.11 presents an overview of the results. It can be seen that the required access times are not met. Some of the requested access times by surgeons exceed the seven week limit, which may occur on the request of a patient. Around 40% of the surgeon requested elective access times are not realized. For those requested times within seven weeks, we can see that 25% of patients does not receive timely access. At five weeks, we can see that rather than 80%, only 37% of patients has received access.

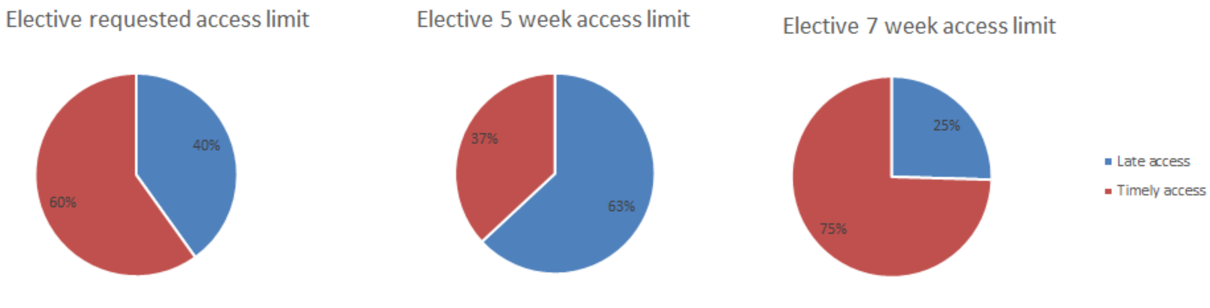

Figure 2.11: Access time performance of elective patients (Source: SAP, data between 03-13 and 10-15).

Figure 2.12Figure 2.12 depicts the descriptive access times statistics of the 6 most frequent occurring elective surgical procedures of General Surgery. It may be noted that both the average and median access time for 5 out of these 6 procedures exceed the maximum allowed standard greatly. With the access time for the most common procedures known, we have a probable cause to determine whether this results from a lack of performance at the OR department.

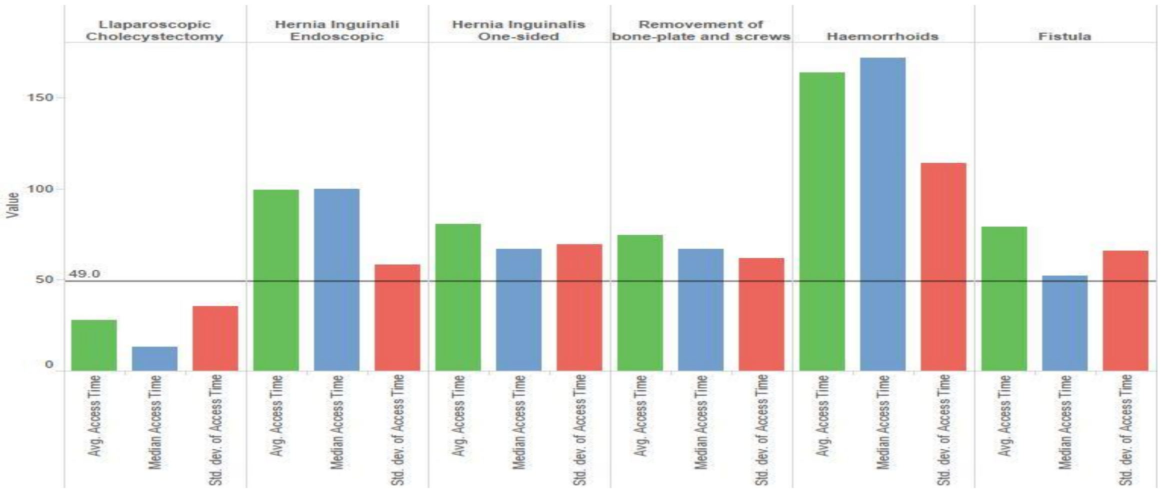

Figure 2.12: Average, Median and standard deviation of access time in days of the 6 most common elective procedures of general surgery (Source: SAP, data between 06-15 and 11-15).)

2.4.2 OR performance indicators

In this subsection, we define and determine OR performance indicators of general surgery.

Utilization is frequently used as a key performance indicator in literature, but is inherently influenced by management decisions such as case-mix and willingness to accept overtime [8].

Additional performance indicators should complement utilization to present a true and representative picture of performance [9]. We will discuss utilization performance, after which we introduce other performance indicators.

Utilization

Weak performance in terms of production may in general result from either lack of capacity or underutilization of existing capacity. If underutilization occurs, it means that production can be increased without adding additional capacity. Because there is no standardized definition of utilization in Healthcare or at Haga it first has to be defined in detail. Hence, we define utilization as the duration that a patient occupies the OR divided during the time that "regular" OR procedures are scheduled. Occupation of an OR by a patient is defined as the time between time-registration points 3 to 8 in Figure 2.2.

Due to frequent missing time registrations of OR-in or OR-out timestamps, a lot of OR utilization would have to be estimated. To create a well representing picture of utilization, we only take days on which the entire OR program of General Surgery was performed in the original planned ORs into account. We chose this method of filtering because General Surgery has the tendency to bring in surgeons that break into ORs of other specialties. This makes it impossible to determine a reasonable capacity and also does not fairly represent the utilization of the "authentic" capacity of General Surgery. Our method prevents misrepresentation of planning performance. Break-ins result in higher utilization but do not necessarily signify good planning. Utilization was also defined under the following conditions:

1. Only days on which surgery was performed in an equal amount of ORs as was planned originally by the admission office were taken into account
2. There were no General Surgery break-ins into other specialty ORs
3. Occupation is defined as (OR out - OR in) time registration in SAP
4. Procedures take place on office days, during the normal OR day shift (07:30 – 16:30)
5. All the procedures that took place in any of the ORs have both OR arrival and departure time registrations
6. There is no overlap in procedures in any of the ORs

$$Utilization = \frac{\sum OR\ occupation}{Planned\ capacity\ of\ OR} \tag{2.1}$$

This means that a patient if a patient enters the OR at 16:00 and leaves the OR at 17:00, this patient only adds 30 minutes to the utilized capacity. If General Surgery performs 15 surgeries in 4 ORs in one day, and one of the surgical procedures lacks an exit time registration, the entire day is excluded from the utilization calculations. Figure 2.13 shows that after rejecting days with unusable data registrations and days where OR break-ins occurred, 30% of the initial amount of days remains usable to determine utilization. Break-ins into elective programs can be observed in nearly 57% out of usable registered days. This fact reveals some of the unrest and constant challenges at operational live level.

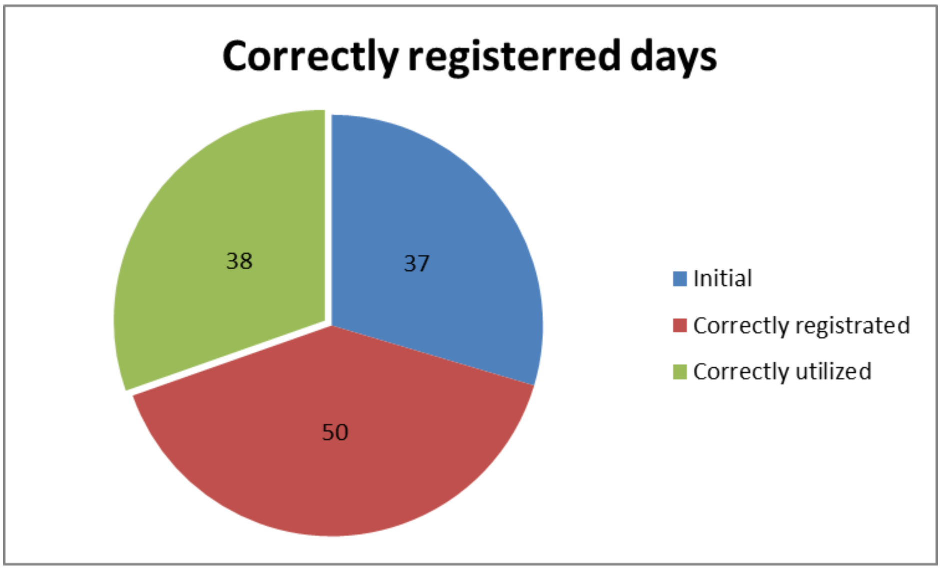

Figure 2.13: Usable days for utilization analysis (Source: SAP, data between 06-15 and 11-15, N= 40 days).

Using the remaining usable 30% of days, the utilization can be calculated for each day and OR averages are taken to determine utilization per day. Note that the days that were used for utilization are not perfectly consecutive since there may be days missing in between the used days.

With the sample and utilization defined, we may calculate utilization for two situations. The first instance only takes elective (non-urgent and semi-urgent) patients and capacity into account. Figure 2.14 depicts the utilization of the elective program. This presents a sense of how General Surgery is performing in their elective program. General Surgery emergency break-ins into elective ORs are not taken into account when calculating the elective utility, which might affect utilization. We later adjust for emergency patients when calculating total utilization in the other situation. The elective program utilization varies per day with an average utilization of 68%. This means that over two third of capacity is actively used to perform elective surgical procedures.

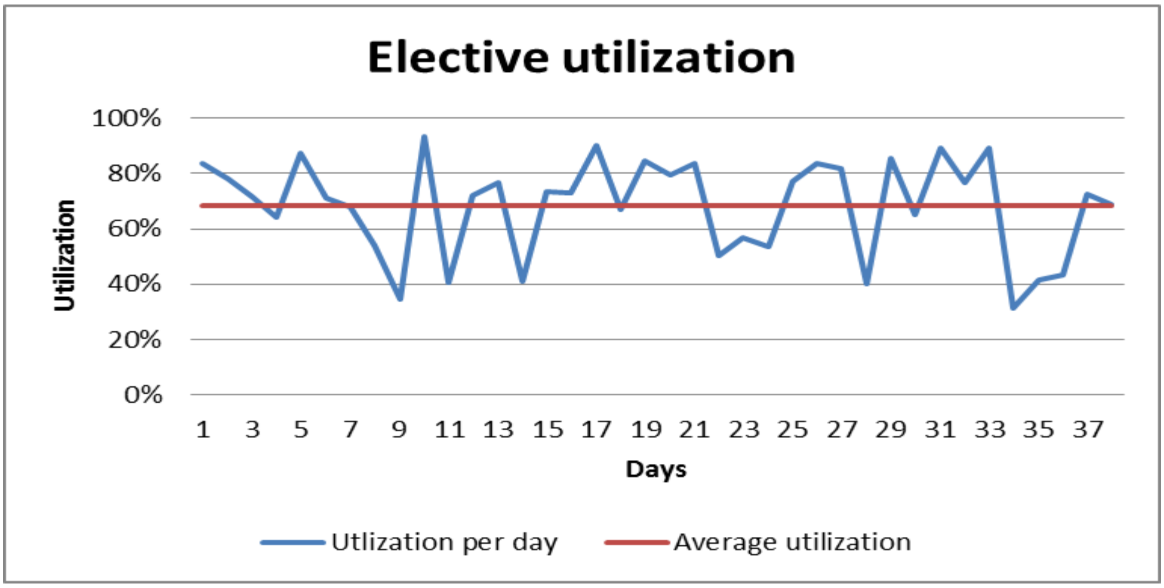

Figure 2.14: Utilization per day of elective General Surgery (Source: SAP, data between 06-15 and 11-15, N= 40 days).

Since there is often some overlap in emergency and elective patients being treated in the same room, we calculate utilization for both elective and emergency patients. We determine the flex room capacity for each day and add this to the earlier determined elective capacity. Figure 2.15 depicts the resulting utilization, with an average of 58%. This drop in utilization compared to the elective program may be explained by the fact that flex rooms are often kept available for emergency patients, which affects total utilization.

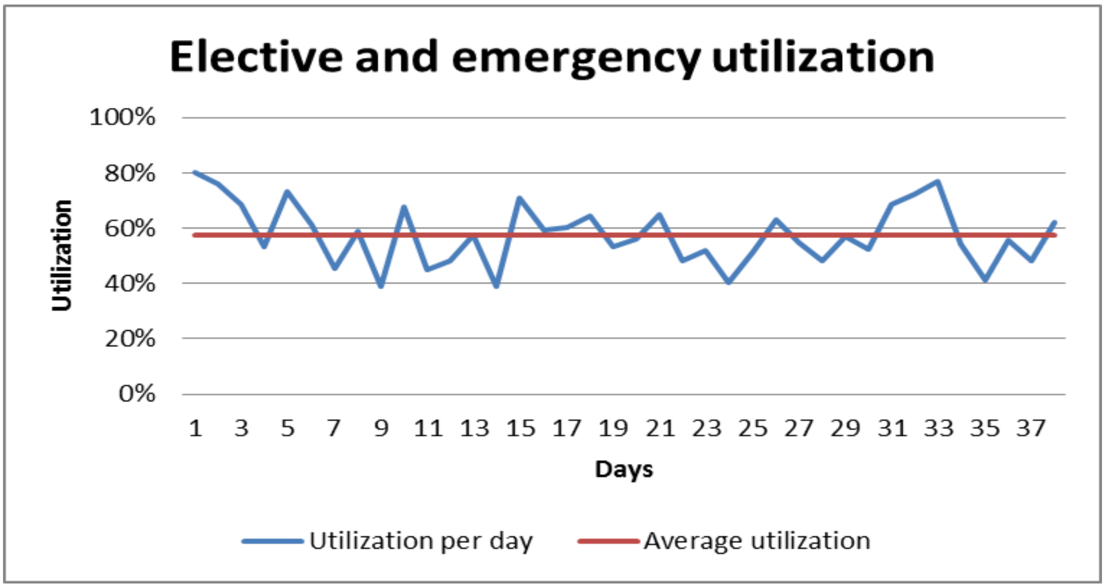

Figure 2.15: Utilization per day of elective and emergency General Surgery (Source: SAP, data between 06-15 and 11-15, N= 40 days).

We compare utilization with a benchmark study performed by Healthcare consultant Plexus in 2009 to determine whether OR capacity of General Surgery is underutilized. In this study, utilization was measured at 45 hospitals that resulted in an average of 78%, with a best practice of 90% [10]. Compared to the national average, General Surgery at Haga is underperforming by 10%. Apart from sub-standard averages it may be noted from the graphs that there is a lot of variability in the utilization, a sign of inefficiency.

With utilization known, a lot of about the different aspects that determine the performance are still unclear. Utilization is the result of a combination of actions, some of which are counted as utilized time and some as idle time. Apart from gathering a single statistic on utilization, it is useful to visualize which of these actions may be an underlying cause of poor utilization. Figure 2.16 depicts the different actions that occur within a day at a single OR. Utilized time is depicted as the white blocks while “unused” capacity is depicted as colored blocks. It is useful to determine the performance of each “type” of block to insight in the causes of poor performance.

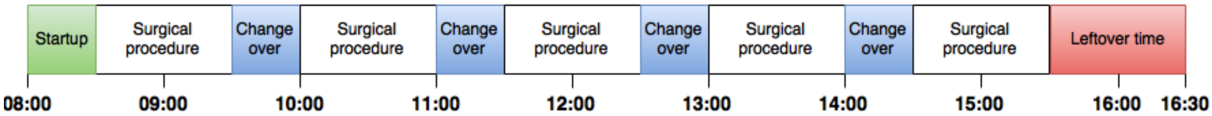

Figure 2.16: Utilization of an OR

In the next part of section 2.4, Starting times, Surgical Procedure durations, Change-over times and Finishing times are analyzed further to determine the causes of this underutilization.

Starting Time

We define the starting time of an OR as the moment that a patient enters the OR and is ready for induction. At that time, surgeons, OR personnel and anesthesiologists are expected to be in the room. Delayed morning starts of the surgical programs may lead to OR underutilization at the start of the day [11]. Figure 2.17 displays the registered arrival times of patients at the Holding ward and at ORs. The 319 patients in this graph were all part of the elective schedule, and only the first patients of the day are taken into account. The OR department has a soft target start of surgery time at 08:00, but it can be seen that 77% arrives at the OR only after 08:05. About 47% arrives only after 08:15. Noticeable are the multiple smaller peaks at Holding arrival and the corresponding peaks at OR arrival. We may notice that patients frequently arrive at the Holding after the target time of 07:30. The Holding is not responsible to pick up patients at the ward and has no insight in which patients are scheduled when for surgery. Patient wards may either forget to transport patients to Holding or are too busy in the morning. Personnel of the OR will find out at some point in the morning, and will call the ward. This causes a delay in the entire OR program for the rest of the day.

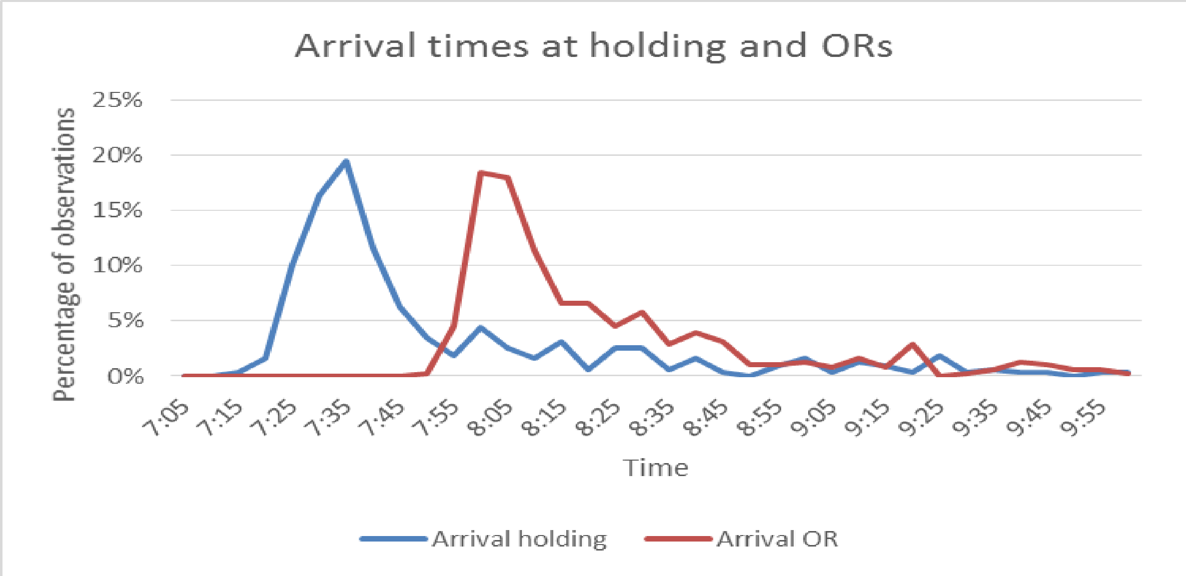

Figure 2.17: Arrival times at Holding and ORs (Source: SAP, data between 06-15 and 11-15, N=319).

A factor that enables errors such as late patients at the Holding is a lack of clear responsibility at the OR department personnel. There is no clear leader within the OR staff, responsibility is shared by the entire staff, hence no one feels really responsible to motivate preparations at the start of the day. Anesthesiologists carry authority with the staff during anesthetization but do not really feel responsible for the progress of the day schedule. The surgeon typically carries authority once he arrives, but surgeons may arrive late on purpose because they have experienced starting delay in the past. This in turn, increases the lack of authority and responsibility at OR personnel in the morning. Another factor is that there are conflicting definitions of what the actual OR starting time should be.

Apart from review Holding and OR arrival times, we review the time between arrival at the ORs and the actual start of surgery. After arrival, a patient first has to be anesthetized by the anesthesiologist. Figure 2.18 depicts the registered times of arrival at ORs and the start of surgery. The 371 patients in this graph were all part of the elective schedule, and only the first patients of the day are taken into account. The sample size differs from Figure 2.17 since arrival times at the holding are often not registered. Noticeable is the width of the curves between the arrival at OR times and the start of surgery. The explanation lies in the sudden high demand on a limited capacity of anesthesiologists. Each anesthesiologist has to work 2 ORs, meaning that if ORs both want to start at 08:00, one of them has to wait till the anesthesiologist is done at the other.

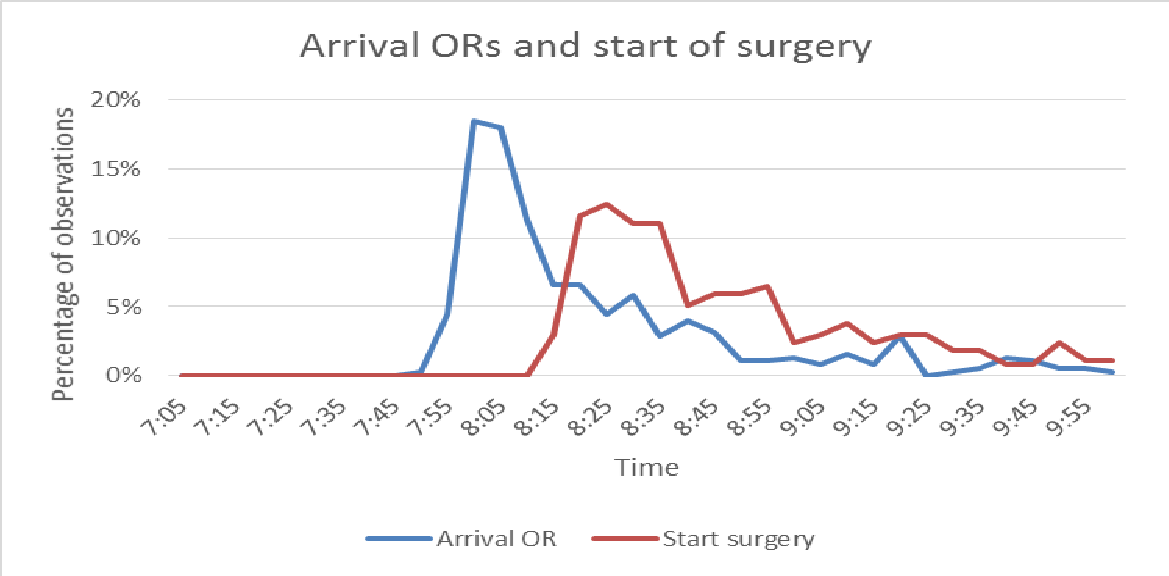

Figure 2.18: Arrival times at OR and start of surgery (Source: SAP, data between 06-15 and 11-15, N= 371).

Planning Time

The time that is scheduled for a procedure is based on estimations of the surgeon, that is often based on a variant of the average time a surgeon spends on that procedure, sometimes based on the patient characteristics. In Figure 2.19, the difference between realized time in the OR and time that was scheduled is depicted for 12998 recorded surgical procedures within General Surgery. Large deviation from the red target line (0 minute difference) can be observed. One issue with this system is that an average estimation does not takes variation in consideration which tends to be large for many procedures. Also, the surgeon that makes the estimation is often not the surgeon that performs the actual surgery and differences in experience many cause deviations in average procedure times. One more factor is there is no clear definition between all the stakeholders which actions are factored in the estimated time. Some surgeons estimate only the net surgical time, others take the anesthetization into account. These factors lead to a difference of at least 10 minutes between planned an realized time in at least 86% of all the surgical procedures.

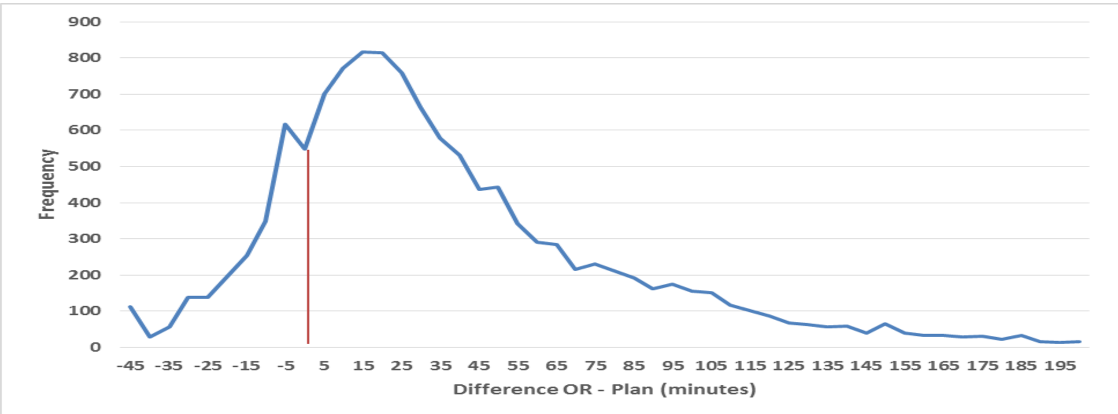

Figure 2.19: Difference in planned and actual OR times (Source: SAP, data between 06-15 and 11-15, N=10040).

Changeover Time

To determine whether capacity is underutilized between surgeries we determine the change-over times between surgeries. Changeover times in minutes of 2058 surgical procedures are depicted in Figure 2.20. Procedures were only taken into account if they are during office hours and if they are not the first procedure of the day. The Admission office uses an estimation of twenty minutes of changeover time between surgical procedures. It may be noticed that the actual changeover time often deviates greatly from the twenty minute estimate and is smaller in 76% of the cases. While this does not necessarily lead to underutilization in practice, it does show that using an average is not an accurate method in scheduling.

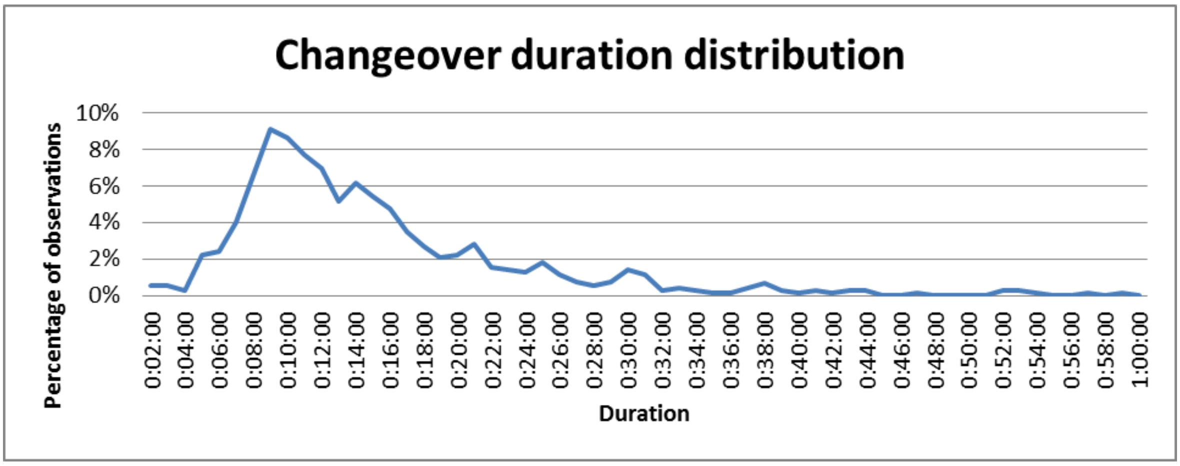

Figure 2.20: Changeover times between surgical procedures (Source: SAP, data between 06-15 and 11-15, N=777).

Figure 2.18 depicts another visualization of changeover times. The two-directional changeovers between two surgical procedures are graphed in the a-axis, in descending order of frequency of occurrence. Changeover duration in minutes can be read from the left y-axis. The cumulative share of total number of changeovers can be read from the right y-axis. The top 90% of changeovers in terms of frequency is included in the graph. We can identify some changeovers that either result in startling low or high changeover durations. This may also be the result of poor registration or intention as we cannot case-wise determine the reasons behind changeover durations. However, the aggregated data allows for the interesting suggestion that the sum of changeover durations in a surgical schedule could be manipulated through scheduling choices.

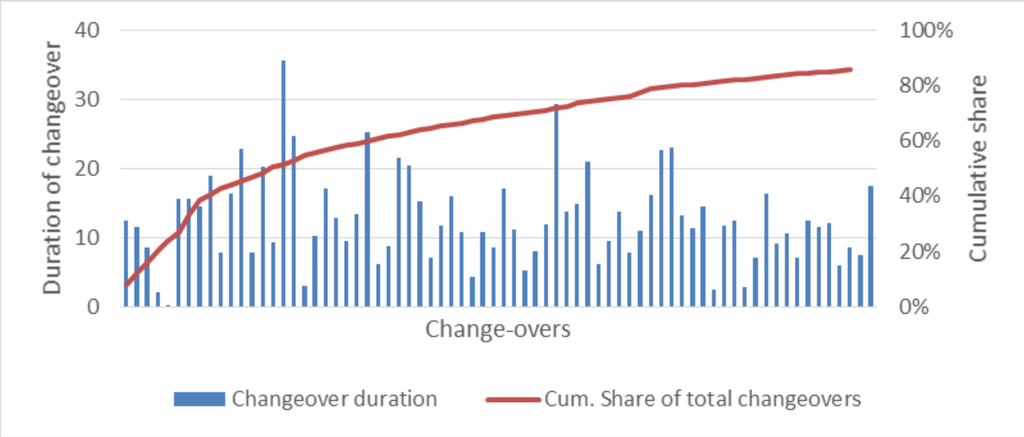

Figure 2.21: Changeover duration and cumulative frequency (Source: SAP, data between 07-15 and 11-15, N= 1105).

Finish Time

To determine whether capacity is underutilized at the end of an OR program, we create an overview of the finish times. Finish time is defined as the time that a last patient of a continuous elective schedule leaves an OR. Schedules in which a break-in by an emergency patient has occurred are also taken into account, though only elective patients are recognized as "last patient". Figure 2.22 depicts the finish time of ORs with a sample size of 98 ORs, which were only taken from the same selection as which was used in determining the utilization. The red reference line depicts the first soft target finish time of 16:00, which is a comfortable finish time for the OR personnel. The green finish line depicts the second soft finish time at 16:30, after which OR personnel is working overtimes. The admission office strives to finish all ORs within the two target times. It is easily seen that OR personnel is often working overtime, with 38% almost a third of the cases. These finish times put the utilization of 58% in perspective. If more capacity between office hours could be utilized, OR personnel would not have to work in overtime this frequently. At the same time, in 23% of the times some capacity is left unused at the end of the day because an OR finishes before 16:00. The finish times are a clear indicator of the inability to estimate the several times associated with the surgical schedule accurately.

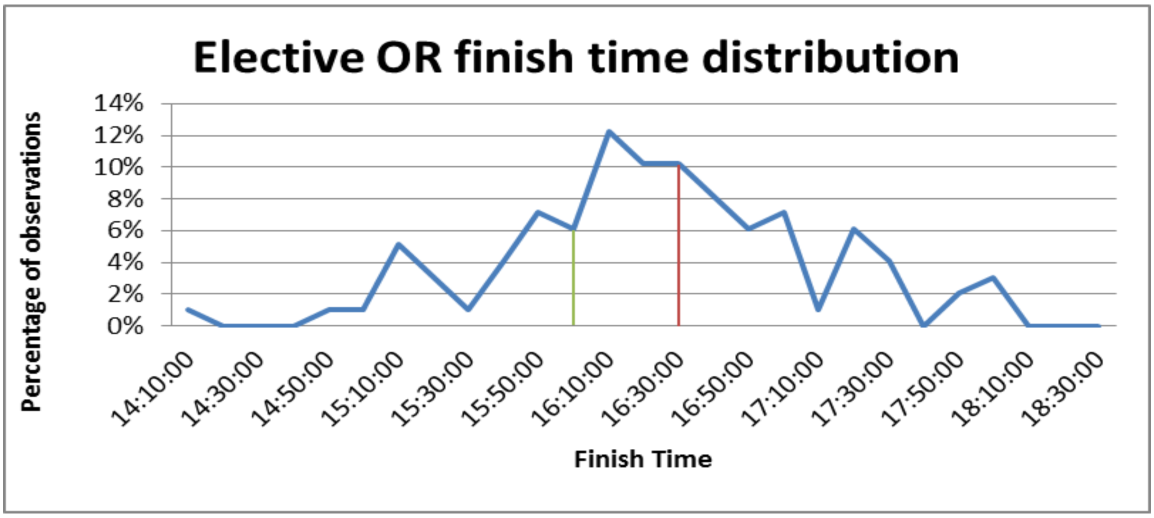

Figure 2.22: Overtime of OR schedules at General Surgery. The red line indicates the first finish target, the green line indicates the second target (Source: SAP, data between 06-15 and 11-15, N=98).

2.5 Problem Identification

This section creates the problem bundle consisting of all the observed problems within section 2.1, 2.2 and 2.4. First, subsection 2.5.1 introduces the stakeholders and their perceived problems. Subsection 2.5.2 describes managerial deficiencies in the control process. All the perceived and observed problems create the problem bundle presented in subsection 2.5.3. Subsection 2.5.4 derives the core problems from the problem bundle.

2.5.1 Stakeholders

In this subsection, we describe several stakeholders that are involved in some way with the system are distinguished and the problems they perceive.

Admission office

The admission office wants to make sure there is enough capacity for (semi-)urgent patients. They want to plan as many elective patients as possible while keeping track of urgency and capacity. The admission office often has to notify a lot of patients very shortly before the surgical procedure about cancellations. They also invest a lot of time in trying to swap sub-specialized surgeons and find extra capacity to schedule all the semi-urgent patients in feasible ORs within an appropriate period.

Surgeons

Surgeons want to be able to schedule both elective and urgent patients within a reasonable horizon. Often, a specialist wants to operate within a preferred sub-area of their specialization and preferably on patients that they have personally met in their consults. Surgeons perceive a lot of chaos in the OR department. To deal with the stochastic arrivals of (semi-) urgent patients with differing sub-specialized demands, they sometimes need to re-adjust their OR schedule short in advance and re-prepare their ORs several times per day. Surgeons perceive that delays in the surgery progress causes patients to be cancelled at the end of the day.

Patients

Patients prefer to be operated on by the same surgeon as they have met during consult. Elective patients have to deal with long access time. Often, they are notified only very shortly in advance of their surgery date and time. Unfortunately, a lot of less urgent patients also rescheduled at the very last moment, sometimes only hours in advance.

OR Department Personnel

The OR department consists of a large group of OR personnel such as planning & coordinating staff, surgery assistants, anaesthesiologist-assistants, nurses, cleaners and all other support personnel. OR personnel often works late hours to finish all scheduled patients. OR Department personnel also perceivers that surgeons may abuse their power by overstating the urgency of emergency patients. A surgeon may push such a patient through the OR at the end of the day, causing overtime for the OR department personnel and the surgeon.

Anaesthesiologists

An issue that anaesthesiologists describe is that sometimes patients need to be rejected at the very last moment if not prepared correctly for surgery. They experience the same frustrations from surgeons overstating the urgency of emergency patients as OR Department personnel do.

2.5.2 Managerial deficiencies

In this subsection, we identify several managerial deficiencies in the control process at Haga. Such poorly addressed planning functions at tactical level lead to time consuming resource allocation challenges at operational level.

Poor capacity requirement prediction

Each year the required capacity for specialties is estimated on a strategic level. Surgical procedure amounts and the average time of each procedure are combined to determine a total required amount of capacity. This capacity is increased by 15% to compensate for utilization. It can easily be observed that a 15% increase would amount to an 86% utility rate, which is unrealistically high. Apart from high utilization estimates, the project manager does consider the fact that General Surgery also performs surgical procedures during the nights and weekends. It is simply assumed that all procedures are within office hours. Though this compensates the high utilization factor, the overall method is still prone to inaccuracy. Research shows that capacity dimensioning at strategic level strongly influences OR performance [12], [13].

Lack of flexibility towards temporary demand changes

Lack of periodic flexibility can be observed on tactical level. Currently, capacity is allocated in a fixed four week cyclic MSS. This MSS is annually made based on yearly demand data and predictions and takes capacity reductions due to holidays into account. However, demand fluctuations that might occur within a yearly period are not measured and hence not also taken into account when creating the MSS. For elective surgical procedures, average demand for a specialty may fluctuate per seasonal or monthly period. For emergency patients, average demand fluctuations may be observed within a single week. Failure to cope with such demand fluctuations may lead to resource allocation inefficiencies [12].

Lack of capacity re-allocation

Lack of unused capacity re-allocation can be observed on tactical level. A medical specialists may return OR capacity due to a number of reasons, such as lack of surgeons due to a medical congress or lack of patient demand. Rules about returning OR days are made between de project managers that create the MSS and the medical specialists. Medical specialists may return OR days down to ninety days in advance, after which the project manager will make an effort to swap these OR days with other specialties. After this period, there is no guarantee that other specialties have time and surgeons available to use the returned OR capacity (a specialty “loses” the capacity). The project manager involved with capacity re-allocation estimates that about 75% of returned capacity is returned in time to make swaps, and that 90% of returned capacity was actually used by other specialties. However, since there are no control methods in check to monitor utilization of OR time, these numbers may deviate. Research by *Dexter et al.* shows that effective capacity re-allocation has positive results on OR performance [14].

Lack of admission Control

Lack of admission control can be observed on tactical level in 2 different stages.

The first stage address the allocation of surgeons between OR surgical procedures and surgical consults in the clinic. It can be reasoned that this allocation influences direct inward patient flow towards surgical consults and thus, indirectly towards the OR department. More surgical consults mean more patients on the waiting list for surgery, and more OR surgical procedures means less surgical consults, hence less patients on the waiting list. Currently, the available OR capacity is evenly distributed between the different surgeons of General Surgery. Surgeons provide the scheduler with their availability and receive one day per week in the OR. This fixed capacity based on surgeon preferences is often unbalanced compared to the patient demand per sub-specialization. There is also no admission control that balances access times between (semi-) urgent and elective patients, or surgical procedures with a deterministic or stochastic nature. It can be observed that this unbalance between surgeon capacity and patient demand causes a lot of problems on lower levels of the control process, for example when scheduling patients on operational offline level. A research by *Oostrum et al.* [3] shows that balancing resource allocation to surgeons based on demand on tactical level has positive results on OR performance, such as utilization.

On the second stage, the different types of patients are scheduled for a consult with a surgeon. Lack of admission control here refers to the fact that not all consults result in planning a surgical procedure. Often, the first consult is an introduction consult and some consults are only for check-ups. By adjusting the ratio of the type of consults that are planned with surgeons, the periodical inflow of elective patients could be controlled. Currently, consults are planned on first come- first served base that result in little control on patient inflow to the waiting list.

Inappropriate patient scheduling approaches

Several inappropriate approaches can be noticed when patients are scheduled at operational offline level. We will list the five foremost problems:

First, there is no clear patient scheduling algorithm used at Haga. At operational level, patients are loosely scheduled on first-come, first-served while checking if patients are scheduled within their medical maximum access-time. Currently, Admission office schedulers hope to create a time efficient schedule by hand where computer algorithms can compare thousands of combinations from several different scheduling approaches within seconds. A research by *Dexter et al.* [15] shows that the scheduling approach may have large impacts on OR performance.

Second, schedulers at the admission office create a concept schedule for the OR department three days in advance. In this schedule, available ORs are filled with known waiting elective patients. From section 2.4 it can be observed that nearly a quarter of all patients that need access to the OR will arrive in the final three days before the surgical procedure. Due to lack of insight in the demand distribution of these patients, they are not taken into account when creating the concept schedule. This creates a lot of challenges in scheduling these patients, and results in a lot of last minute changes in the schedule, cancellations, break-ins and overtime.

Third, the estimated times for required surgical procedure are inaccurate. This inaccuracy stems from three reasons. First, surgeons and schedulers have conflicting definitions about which actions are included in the estimation. Second, surgeons are optimistic in their estimations. This often results in an estimated time of their 10 "best" performances. Procedures that took extra time for some reason are seen as "exceptions" and not taken into account. Third, using an average as estimation inherently ignores the variability of a surgical procedure. For example, when we assumr that surgical procedure times are gaussian distributed, this would mean that half of the procedures would take longer than estimated. For log-normal, his number would be even greater. General Surgery procedures seem to have a lot of variability, and lack of admission control ensures that all elective OR programs have a large share of variability. Research shows that managing surgical duration variability will result in positive effects on OR performance [4], [16].

Fourth, there is no clear standard for buffer-time or slack that an OR should have. Patient schedulers have an available capacity in minutes and try to fill this capacity as much as possible. Due to the discrete surgical procedure times, this often results in some amount of capacity at the "end of the day" that is either unused or overbooked. This decision lies with the patient scheduler, and the program coordinator at the OR has to approve it. Patient schedulers may "overbook" ORs on purpose, if they feel that leftover capacity is otherwise wasted. This overbooking is not driven by insight in no-show statistics, but by necessity to allocate semi-urgent patients within the required access time. This tendency has a large potential to cause overtime [4], [17].

Fifth, for changeover time between patients a standard of 20 minutes is used. Changeover times are not pooled and there is no insight in the actual realization. The deviation from this estimation can be observed in section 2.4. A research by *Dexter et al.* [18] shows that determining an accurate change-over period results in positive effects on OR performance.

Lack of coherence

Lack of coherence can be observed between vertical levels of the planning process and between stakeholders.

First, there is a lack of vertical interaction between hierarchical levels at Haga. Daily performance is not measured at the OR department. There are no agreed upon indicators such as utilization, throughput and overtime to measure performance. Hence, there is no upward vertical feedback towards offline and tactical level that resource planning decisions should be adjusted. Only production numbers are checked yearly at strategic level when capacity is distributed between specialties. This lack of short-term feedback leads to lack of accountability for medical specialties and inhibits the drive for performance improvements. The lack of insight in OR performance at higher levels leads to absence of objectives and standards for the OR department. This leads to a lack of necessity to measure performance such as throughput, utilization, overtime and waiting time at the OR department. This further hampers the drive for improvement.

Second, there is a lack of coherence between different stakeholders. Stakeholders have conflicting interests that can lead to a lot of frustration. Surgeons work the OR once a week and are used to working evenings. Their interest lies with performing surgery on all the scheduled patients, and they are not bothered by overtimes. This frustrates OR personnel who work the OR every day, especially since they experience overtime frequent with different surgeons. In turn, surgeons are frustrated when patients are cancelled to prevent overtime (if OR personnel is not available) and when OR personnel is changed (day shift to support shift) during surgical procedures. These conflicting interests lead to frustrations because adjustments have to be made “online” to deal with unexpected delays.

2.5.3 Problem Bundle

We cluster all the perceived and observed problems into a problem bundle to identify the core problems and their relations. Figure 2.23 depicts the problem bundle. We can observe that problems at the OR, such as unrest, overtime and underutilization are the result of both problems that rose earlier during the planning process (time consuming capacity challenges at operational offline level) and inefficient practice at the OR department at operational live level. The underlying root causes, or core problems, of the observed problems are depicted in yellow boxes.

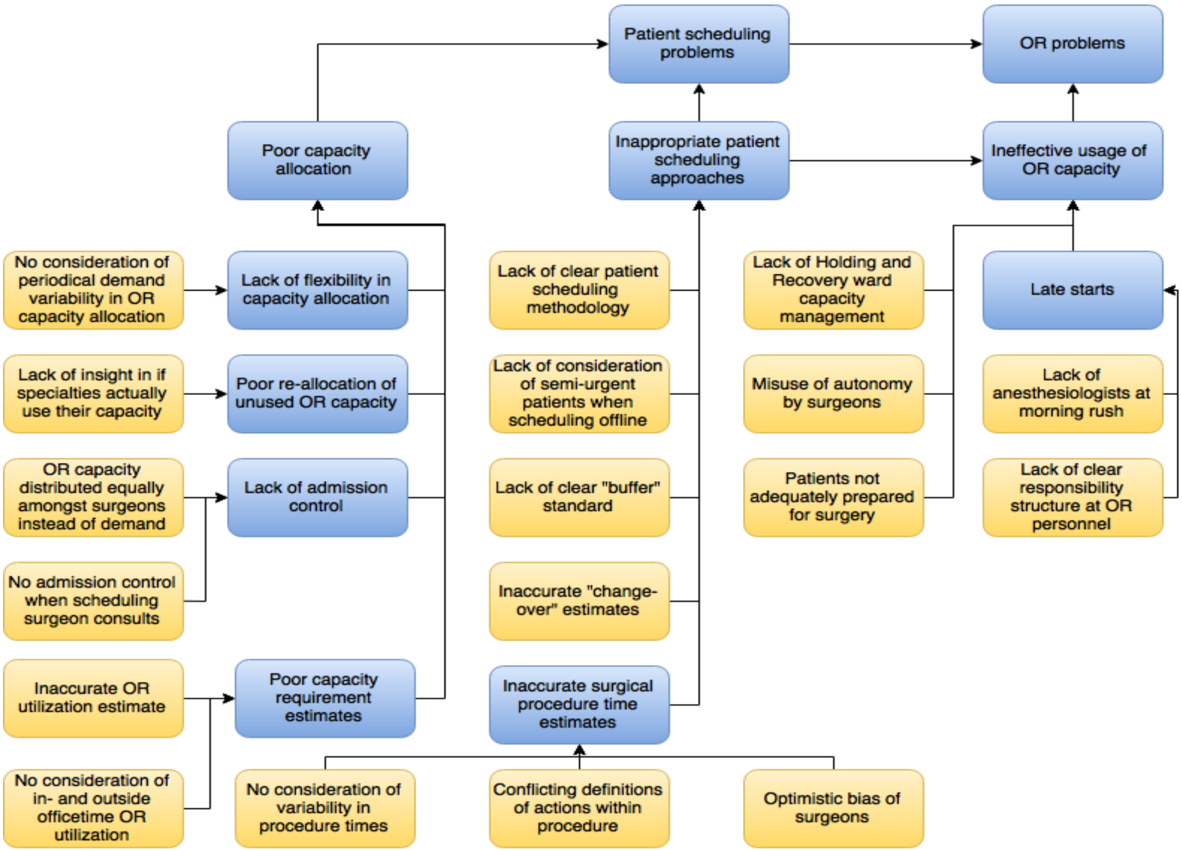

Figure 2.23: Problembundle

Performance measurement

Figure 2.23 lacks one of the observed core problems in subsection 2.5.2. We can observe lack of upward vertical interaction due to the lack of performance measurement. This lack of performance measurement is increased by the poor data management and registration at Haga, which makes it difficult to determine short term performance. This status quo is preserved by a lack of clear performance and objectives set for the OR department, which are the result of a lack of insight on higher levels. Figure 2.24 depicts the interdependency of these problems.

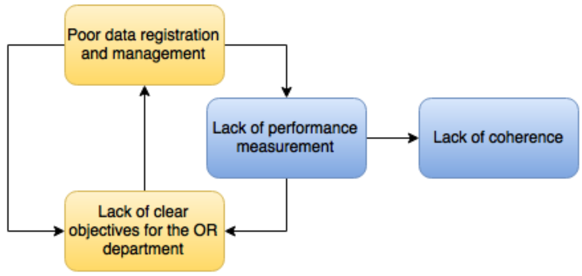

Figure 2.24: Lack of performance measurement

The reason that these observed problems are not visualized in the problem bundle is because they are the "hidden" root cause behind virtually all the core problems. The problem bundle itself was found simply by measuring the outcome of resource planning decisions at Haga.

2.5.4 Core Problems

When reviewing the problem bundle from subsection 2.5.3, we can identify several core problems that lead to bad performance:

1. Lack of clarity in responsibilities and targets at the start of the day
2. Lack of anesthesiologists to serve patients simultaneously at the start of the day
3. Misuse of emergency patient autonomy by surgeons
4. Patients are inadequately prepared for surgery when they arrive at Holding
5. Delays at holding due to lack of Holding capacity management
6. Inaccurate estimates for surgical procedures due to biased estimates
7. Inaccurate estimates for surgical procedures due to lack of consideration of variability
8. Inaccurate estimates for surgical procedures due to conflicting definitions of what the estimate entails
9. Inaccurate estimates for changeover times
10. Lack of clear standards for "buffer" time
11. Lack of consideration of un-arrived "semi-urgent" patients when scheduling
12. Lack of a clear planning method for scheduling patients
13. Poor OR admission control due to capacity allocation based on the preferences of surgeons instead of demand
14. Poor waiting list admission control due to lack of control methods when planning surgical consults
15. Lack of capacity balancing through capacity re-allocation due to lack of insight in whether surgeons will utilize their capacity
16. Lack of capacity balancing through periodical capacity allocation variation
17. Poor capacity requirement predictions due to inaccurate utilization estimate

18. Poor capacity requirement predictions due to inconsideration of the ratio of surgical procedures in- and outside of office hours
19. Lack of performance measurement due to poor data registration and management
20. Lack of performance measurement due to lack of performance objectives for the OR department

The core problems related to data- and performance management are not listed in this overview. We consider these to be the most underlying core problems and a first step to improving performance. We do not include them in our research since Haga is planning to implement a new IT system that includes a performance dashboard.

2.6 Summary

Through a review of the current situation and a performance analysis, we identified the core problems of poor performance and scoped our research. Core problems related to surgical demand uncertainty are:

- No consideration of expected “semi-urgent” patients when scheduling
- Lack of a clear planning method for scheduling patients
- Poor OR admission control due to capacity allocation based on the preferences of surgeons instead of demand
- Lack of capacity balancing through periodical capacity allocation variation

Core problems related to surgical duration uncertainty are:

- Inaccurate estimates for surgical procedures due to biased estimates
- Inaccurate estimates for surgical procedures due to lack of consideration of variability
- Inaccurate estimates for surgical procedures due to conflicting definitions of what the estimate entails
- Inaccurate estimates for changeover times
- Lack of clear standards for “buffer” time

We will conduct a literature research in the next chapter to gain insight into potential solution approaches.

3 Theoretical Framework

In this chapter, we perform a literature study of organizational interventions in healthcare to create a theoretical framework of potential solution approaches.

Our goal is to find relevant operations research and management science oriented solution approaches thath could be applied to the problems we identified in chapter 2. In recent years, operations research has been increasingly more applied in healthcare. *Cardoen et al.* [19] performed a literature review on this subject, while *Hulshof et al.* [20] presented an extensive taxonomy. Both reviews provide an excellent review of available publications and will be the starting base of our research. Apart from these reviews and general literature search engines, we use the ORchestra bibliography provided by CHOIR. Due to the scope of our thesis, we will restrict our research to relevant interventions on the tactical and offline operational level in the surgical care domain. Key search terms related to operations research on these levels are: *linear, mathematical programming, stochastic, deterministic, uncertainty, queuing theory, simulation, heuristics, algorithm, MSS, block-scheduling, capacity, clustering, semi-urgent, elective, operating room, operating theatre, scheduling, surgical care services.*

We will use the structure provided by *Hulshof et al.* to present our findings. Section 3.1 describes interventions that deal with planning a MSS at a tactical level. Section 3.2 describes interventions that deal with operating room scheduling at the operational level.

3.1 Tactical level

In this section, we discuss several organizational interventions found in literature related to master surgical schedule planning.

On a tactical level, planning decisions traditionally revolve around the creation of an master surgical schedule. Various methods of creating a MSS can be found in literature, often aimed at achieving specific goals or performance. In our literature study, we examine the different goals and methods that were used and their results. The process of creating a MSS can be divided in three subsequent steps. In the first step, different types of surgical procedures are defined and often grouped by some criteria to simplify the scheduling process. In the second step, OR department capacity and other resources are allocated to the patient groups. A wide variety of methods, models and criteria are used to drive these allocation decisions. In the third step, allocated capacity is assigned to specific dates, times, surgeons and ORs to create a MSS. The first step of patient group identification is discussed in subsection 3.1.1, subsection 3.1.2 discusses the second step of time subdivision and subsection 3.1.3 the third step of block scheduling.

3.1.1 Patient group identification

In the majority of literature, patients are grouped together by urgency and/ or (sub-) specialization, where capacity is reserved for urgent patient groups. A paper from *Van Houdenhoven et al.* [21] demonstrates efficiency gains through the portfolio effect when clustering elective surgeries within a specialization in a fixed amount of bins based on their variability and allocating capacity based on bin demand and utilization targets [22]. The portfolio effect was first described by *Markowitz et al.* [23] and is extensively used in operations research for risk pooling. Clustering also allows allocation of capacity to rare surgical procedures through demand pooling. A paper from *Van Oostrum et al.* [1] presents a hierarchical clustering method for elective surgical cases based on surgical case demand. This paper also describes the trade-between loss of information and efficiency when varying bin size and bin quantity. The method could be used as a suitable alternative for a fixed-bins method who *Van Houdenhoven et al.* propose.

3.1.2 Capacity allocation

Most literature shows that capacity allocation should be demand driven to balance access times and increase throughput, and that managing demand uncertainty is key to performance increase [21,22]. A paper by *Gupta et al.* [26] presents a mathematical programming model that allocates OR capacity to various sub-specializations within the specialty General Surgery. The paper is cost oriented and focuses on balancing the trade-off between the benefit of contribution and the cost of excess allocation per time-unit of OR capacity. In his paper, *Van Oostrum et al*. [27] uses several mathematical programming models to create a MSS in two phases. In the first phase, operating room days (ORDs) are selected from an implicit set, that is iteratively increased by generating improving ORDs. In the second phase, the MSS is created by assigning ORDs to specific ORs during the planning cycle with the objective of minimizing the maximum bed requirement during any day of the cycle. *Van Oostrum et al.* also finds that it is better not to allocate capacity to rare surgical procedures. *Zonderland et al.* [28] present a queuing theory model that can be used to determine expected semi-urgent patient access times with a given arrival rate and capacity, and *Kortbeek et al.* [2] uses presents a discrete time slot queuing model to anticipate unplanned arrivals in an outpatient clinic. Queuing modeling is interesting since it models “knock-off” and “overspill” effects from decisions in one period into subsequent periods. One key practice of that effect at general surgery is cancelling an elective surgery at the last minute to free capacity for a more urgent patient. That cancelled patient is then often re-admitted as a semi-urgent patient as compensation for the inconvenience, which results in the cancellation of another elective patient

A paper from *Adan et al.* [29] reviews varying ranges of slack and bin-interchange flexibility in allocating block time to patient groups to determine an optimal strategy in dealing with uncertain demand. Slack is additional reserved OR capacity to deal with uncertainty in surgical duration. These strategies were derived from an earlier research by *Dellaert et al* [30] but expanded to allow for uncertainty. In their research, *Adan et al.* show a clear trade-off between utilization against service level and overtime by adjusting the amount of slack. *Vissers et al.* [31] provides a similar insight through a simulation study of four extreme MSS strategies that focus on different objectives such as maximize resource utility, minimize access times and staff preferences. A paper from *Bowers et al.* [25] also presents a simulation study that examines the trade-off between utilization and service level in reserving capacity for emergency patients. One limitation of the research of both *Vissers et al.* and *Bowers et al.* is that only one type of surgery is taken into account.

3.1.3 Block scheduling

Common criteria when creating a MSS are surgeon preferences, staff- and OR capacities and patient demand [32]. We can distinguish cyclic and non-cyclic block schedules, or MSSs, in literature. Cyclic block schedules are most common and may greatly reduce planning efforts in environments with repetitive procedures [33]. Non-cyclic block schedules can be adjusted more easily to seasonal demand variations but are more unpredictable to surgeon planners and downstream resources than a cyclic schedule. Mathematical programming is a frequently described method to create MSSs in the literature. *Dellaert et al.* [30] presents an MSS integer linear programing approach to minimize weighted deviations from utilization goals of the OR, staff and IC wards. *Van Oostrum et al.* [27] present a minimax mathematical program to create an adaptation of a MSS that levels the maximum bed occupancy on any day based on recorded average length of stay of surgical procedures. *Beliën et al.* [32] construct an initial MSS using mathematical programming, but then apply a simulated annealing algorithm to level bed occupancy. Instead of average length of stay, their local search allows empirical distributions as input.

3.2 Offline operational level

In this section, we discuss several organizational interventions found in literature related to surgical case scheduling into the MSS at offline operational level.

On an offline operational level, planning decisions mainly concern the scheduling of surgical cases into the MSS. Numerous objectives and methods to schedule surgical cases can be found in literature. The scheduling process can be decomposed in three subsequent steps. In the first step, the planned duration of a surgical procedure is determined. In the second step, a specific date and OR are assigned to the procedure. In the third step, the sequence and start times of surgical procedures within an OR is determined. Subsection 3.2.1 discusses the first step of surgical case duration estimation, subsection 3.2.2 the second step of assigning dates and ORs to cases, subsection 3.2.3 the third step of surgical case sequencing.

3.2.1 Procedure duration

A common method reported in literature is that surgeons estimate the surgical duration for each patient based on patient characteristics. This method is currently also in practice at Haga. Papers by *Hans et al.* and *Houdenhoven et al.* [12,18] introduce another method to estimate surgical case duration. They assume a Gaussian l distribution fit and take empirical historical averages and standard deviations into account when determining surgical procedure duration. All the standard deviations are bundled together to create slack with the benefits of the portfolio effect. The precise amount of slack on a total OR day schedule is determined by the scheduled surgical duration variabilities and aimed probability of overtime of the schedule. One positive aspect of this method is that it takes variability in surgery durations and change-overs into account. *Bosch et al.* [34] presents a method to linearize exponential overtime probability that makes it feasible for implementation in mathematical programs. *Strum et al.* [35] compares distributions for case-duration estimation and finds that a log-normal distribution is more accurate to estimate surgery duration compared to a Gaussian distribution. In turn, *Stepaniak et al.* [36] compares different log-normal distributions and finds that the 3-parametric log-normal distribution has the best predictive case duration results. One drawback of using a log-normal distribution is that its open nature prevents the use of pooled slack, which is employed in the earlier reported portfolio effect. A solution to this drawback is presented by *Van Oostrum et al.* [27] with the Fenton-Wilkinson approach as a method to estimate the distribution of the sum of log-normal distributed variables, that he then uses to estimate the portfolio effect. A paper by *Olivares et al.* [17] presents the news vendor model to determine capacity for individual surgical procedures, based on OR cost characteristics of reserving too much or too little time.

3.2.2 Assigning dates and ORs

Hans et al. [9] presents a constructive algorithm where surgeries are scheduled based on the resulting sum of historical surgery averages and slack. Their goal is to maximize utilization by filling available capacity as possible, while minimizing slack. *Hans et al.* also demonstrate a strong performance of Simulated Annealing as local search improvement algorithm. *Dexter et al.* [37] attempt to maximize utilization through an algorithm where multiple procedures are scheduled simultaneously. Several papers present methods with other goals than maximizing utilization. A paper by *Testi et al.* [12] presents a mathematical programming model to schedule patients based on medical priority instead of surgical duration. Unfortunately, the model has some limitations such as no uncertainty in arrival rate or surgical duration, no emergency patients and excess demand. A paper by *Min et al.* [38] presents a dynamic programming model to schedule patients based on medical priority. In this model, a trade-off between surgery overtime and postponement costs result in a MSS. A drawback of this model is that it assumes all surgery types and durations to be identically distributed, and relaxation of this assumption increases the model complexity dramatically.

3.2.3 Surgical case sequencing

Sequencing by preference of surgeons is the most practiced method encountered in literature. Surgeons may prefer complicated procedures first so that they can monitor their recovery during the rest of the day [39]. *Van Oostrum et al.* [40] presents an interesting method of sequencing to smooth surgical procedure starting times over all ORs on a particular day. This allows for emergency patients to break into an OR (during a changeover) and eliminates the need for dedicated emergency rooms. However, simulation shows that a large set of ORs on a particular day is required to really gain good results, while general surgery only has three or four per day. Other sequences focus on performance by attempting to minimalize the make-span for a given set of planned surgeries in an OR. A paper by *Kwak et al.* [41] finds that the classic longest processing time first (LPTF) algorithm outperforms most other sequences, such as random and FCFS. A paper by *Denton et al.* [42] finds that sequencing procedures on decreasing variance outperforms LPTF in turn.

3.3 Summary

In this section, we will summarize the results from the literature study.

Several papers provide promising interventions. The queuing theory methods proposed by *Zonderland et al.* and *Kortbeek et al.* seem good representations of the "kock-off" effect that occurs frequently between semi-urgent and elective patients at general surgery. Elective capacity allocation and semi-urgent capacity reservation should be demand driven and effectively incorporated in the MSS together. Clustering techniques may prove useful if surgical procedure demand is too scarce to schedule. Mathematical programming seems a robust method to incorporate these elements together in a MSS. Non-cyclic MSSs may handle demand variations better, but cyclic MSSs provide practical benefits for OR schedulers.

Adding slack adds robustness against overtime and incorporating the portfolio effect reduces the required amount of slack through variance pooling. A 3 parametric log-normal distribution reflects surgical procedure duration more accurately than 2-parametric and Gaussian distributions. Since surgical case scheduling is currently done manually, interventions should be implemented on a tactical level as much as possible.

4 Solution approach

In this chapter we propose a robust solution approach to our research objective. Our solution approach is founded on groundwork from earlier chapters. Core causes of poor performance are identified in chapter 2, potential interventions available in literature are reviewed in chapter 3.

We will describe our solution approach over several sections. In section 4.1, we introduce the solution approach as a conceptual model to express the correct interpretation of our model. In section 4.1, we discuss how data was gathered and adapted to be suitable as input for our model. In section 4.3, we provide a detailed description of the complete solution approach and technical models. We briefly summarize this chapter in section 0.

4.1 Conceptual model

We propose a robust cyclic surgical schedule aimed on managing surgical demand and duration uncertainty based on the research of *van Oostrum et al.* [3] , *van Houdenhoven et al.* [22] and *Kortbeek et al.* [2]. We believe that managing procedure demand and duration uncertainty will result in both timely access time to patients and economic use of operating rooms. While non-cyclic surgical schedules have advantages towards managing demand variability, they greatly complicate OR staff scheduling manageability.

The framework presented by *Hans et al.* distinguishes a strategic, tactical, operational offline and operational online level of scheduling [20]. Our solution approach applies extensive data driven admission planning at tactical level to manage uncertainty. Capacity requirements for surgical procedures are determined in such way that patient access times are controlled, overtime probabilities limited and efficiency maximized. The result is a tactical framework that describes OR schedules exactly which surgeons and procedure types to schedule at date. It also describes for each surgeon how much capacity to reserve for the accommodation of semi-urgent procedures. We define this tactical framework as a surgical procedure type schedule (SPTS) instead of the conventional master surgical schedule (MSS), as it describes clusters with procedure types instead of defined surgical procedures. The SPTS describes exactly how much of each surgical procedure type should be scheduled in a specific OR at a specific day. It can be viewed as a very specific version of a conventional MSS. Our solution approach will greatly reduce complexity that currently resides at operational levels. The distinction between the current scheduling approach and proposed solution approach will become clear in the rest of this section. A visual comparison in accordance with the framework of *Hans et al.* is made in Figure 4.1. We propose our SPTS to have a 4 week cycle, determined 3 months in advance. The SPTS can be filled by the admission office planners with surgical cases over a 1 week planning horizon.

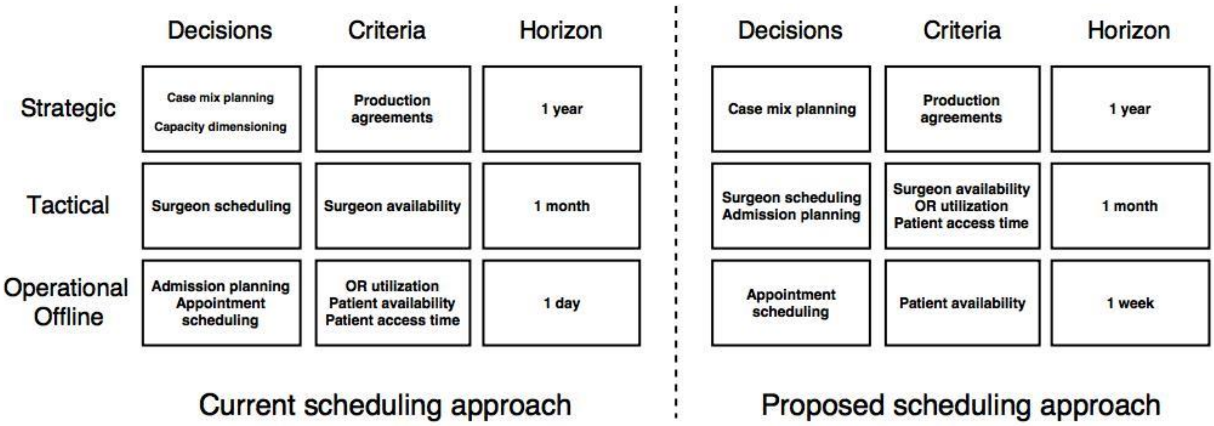

Figure 4.1: Comparison of scheduling approaches

4.2 Model input

In this section, we discuss data gathering and the underlying assumptions of our models. Our data was gathered through interviews with the stakeholders described in 2.5.1 and through Haga registration system SAP. Specifically, we used the OR statistics and admission office data of all patients that underwent surgery between 03-13 and 10-15. We identify three main patient groups as input for the solution approach. We recall the three different levels of urgency:

- Elective patients require access within 7 weeks. They are scheduled operational offline.
- Semi-urgent patients require access within a week. They are scheduled online.
- Emergency patients require access within a day, some may require immediate access. They too are scheduled online.

Each patient may undergo one of nearly 350 unique surgical procedures, performed by one of 15 surgeons. All of these procedures may be categorized in one of 6 surgical areas:

- General
- Traumatology
- Cardiology
- Pulmonology
- Abdominal & Gastro-intestinal
- Oncology

Surgeons are sub-specialized in one or more areas and cannot perform procedures outside their area of sub-specialization. General procedures are relatively simple procedures that can be performed by any surgeon. An uncommon feature at Haga is that there is no surgeon specific demand within a sub-specialty; any patient is allowed to be served by any surgeon. This uncommon feature allows us to pool demand amongst surgeons, which greatly reduces surgeon specific variability. We assume that surgeons are capable to perform all the surgical procedures within their area of sub-specialization. This assumption is confirmed to hold reasonably well for Haga. We therefore drop the notion of allocating surgical procedures to different surgeons entirely, and will allocate to sub-specializations instead.

Our thesis solely focuses on OR scheduling during office hours. Similar to our definition for utilization in section 2.4.2, this requires us to identify which surgical procedures should not be performed during office hour demand. Our solution approach will "include" patients from with all levels urgency of the current case mix at Haga, and employ the current 4 week cycle used at Haga. We ignore procedures performed before 07:00 after 20:00. We will assume that these procedures were not caused by online management problems with the scheduling approach, but for clinical reasons. For the same reason procedures outside of weekdays are also excluded. We define the arrival of a patient as the date at which a patient is registered in the hospital registration system SAP. We exclude any online operational issues that might occur in our solution approach, such as no-shows, late-shows and late-starts. We will also exclude emergency patients from our SPTS, since common practice at Haga is to reserve at least one flexroom per day. For Haga, it is unfeasible to schedule any patients other than emergency in these ORs. To achieve good performance, the method proposed by *van Oostrum et al.* requires a larger number of ORs than we have available [40].

We assume that no specialized staff is required to take into account when scheduling. Similarly, we assume that surgical procedures require no special resources or equipment. We assume that all operating rooms are identical and suitable for general surgery and that operating room availability is unbounded. Operating room capacity that is left at the end of a surgical schedule is considered "waste". These assumptions hold well in reality for Haga.

4.3 Technical models

In this section, we discuss how we allocate demand to the patient sets we identified in section 4.2 and we construct a SPTS in several steps. In subsection 4.3.1 we present a clustering approach as proposed by *van Oostrum et al.* [1] as a method to combine individual surgical procedures into homogenous surgical procedure types. This allows us to reduce uncertainty by pooling demand. In subsection 4.3.2 we will present the queuing theory approach presented by *Kortbeek et al.* [2] as a method to predict weekly required capacity for semi-urgent patients. The queuing model will determine the number of slots required to provide timely access to nearly 90% of semi-urgent patients. We then consider semi-urgent patients to be "just another elective surgical procedure type" with the determined number of slots as its demand. In subsection 4.3.3 we introduce a mathematical programming approach based on *van Oostrum et al.* [27] to allocate capacity to all surgical procedure types and to construct a SPTS. In subsection 4.3.4 we discuss a flexible approach to implement the SPTS.

4.3.1 Clustering

We have already defined 6 surgical sub-specializations, but still have a large number of unique surgical procedures within each sub-specialty. Estimating periodic demand for each unique procedure is possible but each estimation would be subjected to some amount of periodic variability. With nearly 350 unique procedures the total resulting variability could erode the robustness of the demand estimation. We solve this issue by clustering surgical procedures together as surgical procedure types as demonstrated by *van Oostrum et al.* [1]. This allows us to pool demand of individual surgical procedures which reduces uncertainty. In order to economically allocate OR resources we need to ensure that surgical procedure type clusters are homogenous in terms of expected and variability. We therefore perform an analysis to determine whether we can exclude very rare procedures. For each procedure, we determine the share of the total number of performed procedures and exclude those procedures that together make up less than 5% of the total number. Figure 4.2 visualizes surgical procedures graphed against their cumulative share of total production.

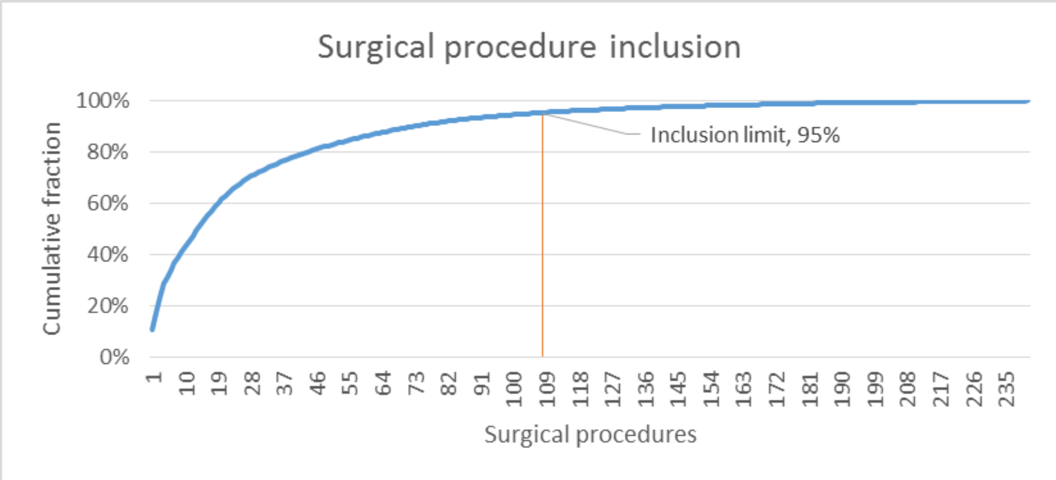

Figure 4.2: Cumulative share of surgical procedures sorted in descending frequency, graphed as blue. We exclude 133 surgical procedures that together make up less than 5% of all performed procedures. The remaining 108 procedures are used for clustering (Source: SAP, data taken between 03-13 and 10-15).

Our clustering approach can be described as follows and is subsequently applied to each sub-specialty. We use the determined log-normal parameters as input variables to create a scaled Euclidian distance matrix. We then use an agglomerative clustering algorithm with wards linkage to determine a suitable K number of clusters, based on visual inspection of a dendogram. Such a dendogram provides us with information on the loss of information when reducing the number of clusters. For traumatology, such a dendogram is visualized in Figure 4.3. Two key factors when determining K are that we preserve enough demand for each cluster and that we minimize surgical procedure distance within each cluster. We determine on a range of K clusters for that we will later examine performance. We then re-cluster the dataset again with a K-means algorithm. The result is a K number of clusters with aggregated demand and homogenous duration and variability. We call these clusters surgical procedure types, compared to the individual surgical procedures we used as input.

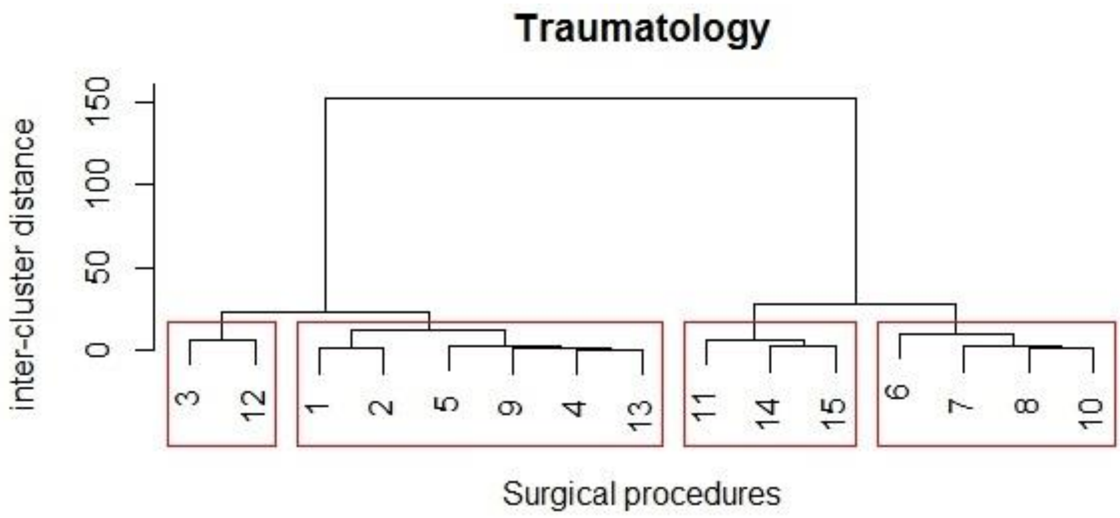

Figure 4.3: A dendogram of the merging process within sub-specialty traumatology. Surgical procedures (x-axis) are plotted against the distance between clusters (y-axis). Each horizontal line represents a merge. The boxes present an example of 4 surgical procedure types.

To determine the duration of a surgical procedure type, we fit a distribution based on all individual procedures that comprise a procedure type. In section 0, we determined that monthly demand for sub-specialties fluctuates. From an economic point of view, we would like to prevent idle time by adjusting capacity to such demand fluctuations, similar as to what happens in a non-cyclic schedule. We reason that any fixed capacity will ultimately lead to situations where capacity is either lacking or underutilized. We therefore use a quantile function to describe demand spread for each surgical procedure type. The result is a distribution of monthly observed demand over 4 quartiles, each with a 25% probability. The quartiles can be visualized with a box-and whiskers plot, shown in Figure 4.4. Rather than only using the average monthly demand, we can now separate demand into a fixed and variable component. The fixed component equals the lower observed quartile of monthly demand. The variable component fluctuates between 0, the second (that equals the median), third and fourth quartile of monthly demand, and is adjusted when management perceives that waiting lists are either shrinking or growing.

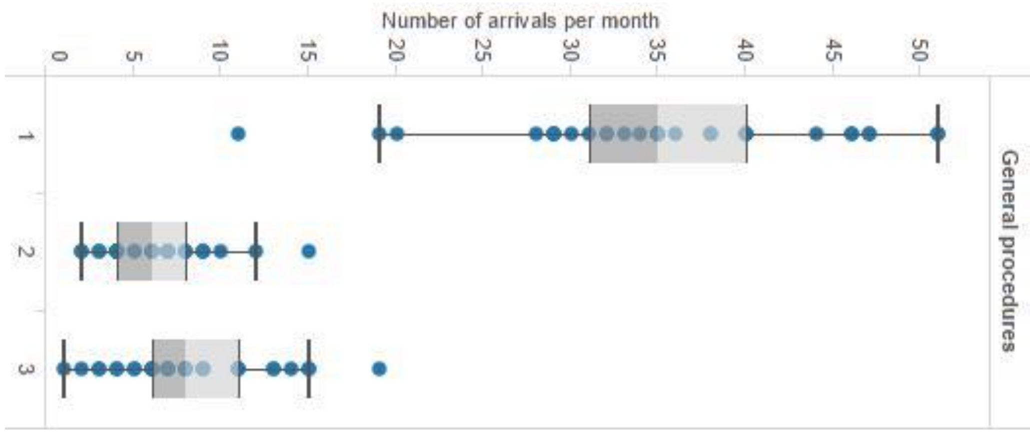

Figure 4.4: Quartile box-and whiskers plot of monthly demand of the (general) first 3 surgical procedure types (Source: SAP, data taken between 03-13 and 10-15).

A detailed motivation of the decisions used in clustering is given in Appendix C, cluster results are given in Appendix D. An unfortunate collateral effect of clustering surgical procedures is that we cannot manipulate required changeover times by sequencing specific individual procedures. However, we predict that the beneficial effects of clustering outweigh those of manipulating total required changeover time. To incorporate changeovers in our model, we clusters all changeovers and define them to be a surgical procedure type. In subsection 4.3.3 we will model the special requirements related to changeovers.

4.3.2 Queuing theory

Our initial aim is to determine how much capacity we should reserve such that semi-urgent patients have a certain probability of receiving access within a week time. We fitted a Poisson distribution to describe weekly semi-urgent arrival, but are still uncertain about how much capacity we should reserve to achieve our access time goal. Reserving too much capacity will reduce access time but also result that capacity is frequently left idle. Reserving too little capacity will result in frequent “knock-off” effects as is currently the case, but also higher OR utilization and more frequent overtime. Reserving capacity is a trade-off between these scenarios that will both occur to some extent.

Weekly required capacity is determined by 2 components: the number of patients per week and the duration of their surgical procedures. To simplify our problem, we temporarily ignore procedure duration uncertainty and assume that each procedure has a deterministic duration of one time unit. The only question remaining is then how many time slots of one time-unit we require on weekly basis, with consideration to both independent new arrivals and “knock-off” effects from earlier weeks. Using the deterministic queuing theory model presented by *Kortbeek et al.* [2] we may determine the number of slots per week required to accommodate semi-urgent patients within their allowed access time of 1 week in 90% of the times. In other words, we may determine the number of semi-urgent patients we need to reserve time for per week.

We used Minitab to identify distributions that might describe semi-urgent arrival demand. We removed outliers outside a two sided confidence interval of 95%. However, poor registration and unnatural confounding influences prevent us from finding surgical procedure-, or even sub-specialty specific arrival distributions. We can only describe the weekly arrival of the complete set of semi-urgent patients by a Poisson process with rate λ_w. We use that distribution and may later distribute the total determined capacity amongst surgical sub-specializations based on historical distribution.

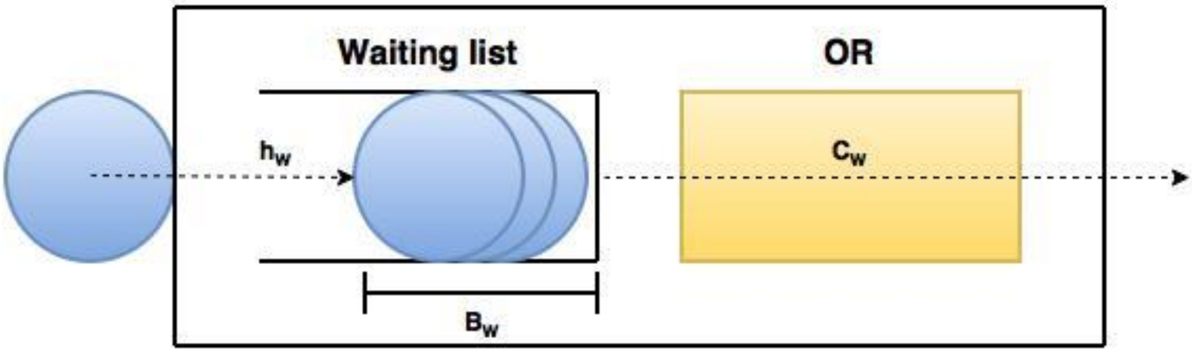

Figure 4.5: Visualization of Queuing model with blue dots representing arriving patients.

We first need introduce some mathematical notations before we can effectively describe the queuing theory model in technical terms. These notations are also summarized in Table 4.1.

Symbol	Definition
w	Current week
W	Cycle length in weeks
λ_w	Average arrival rate per week
B_w	Backlog at the beginning of week w
c_w	Capacity at week w
A_w	Number of arrivals in week w
ES_w	Number of empty patient slots in week w
AT_w	Access time of a patient arriving in week w

Table 4.1: Symbols used in discrete time queuing slot (Source: Kortbeek et al.)

On arrival, patients join a single queue that is served on first-come first served base by a single server. We assume that patients cannot be scheduled in the same week as they arrive. This may seem a limitation, but has no real significance due to the stationary nature of the model. Weekly, the batch of patients that is removed from the queue can be described by:

$$(B_w - c_w)^{+} \tag{4.1}$$

In which B_w is the backlog of patients at week w, c_w the capacity, or number of slots, at week w and $(x)^+ = \max(x, 0)$. In words, the batch of patients removed equals at most the weekly number of slots. This process is visualized Figure 4.5. We introduce the number of weekly arrivals A^w and can now define the backlog in the next week as:

$$B_{w+1} = (B_w - c_w)^{+} + A_w \tag{4.2}$$

The probability of transitioning from a backlog i in week w to i' in week $w+1$ can then be described by:

$$P(B_{w+1} = i' \,|B_w = i) = \begin{cases} P(A_w = i) & if\ B_w \leq 0 \\ P(A_w = i' - i + c_w) & if\ B_w \geq 0 \end{cases} \tag{4.3}$$

Kortbeek et al. presents an exact method based on probability generating functions to determine the stationary backlog probability distribution, but finds that it is only computationally feasible for small problems. For larger problems, he approximates the stationary vector by determining the eigenvector of a large but finite stochastic matrix with backlog probabilities. This not alternative methods makes the model suitable for larger instances such as this one. With the stationary number of patients waiting in queue known, other important statistics can be determined such as expected access time and expected number of empty slots. The expected weekly number of slots is given by:

$$\mathbb{E}(ES_w) = \sum_{b=0}^{e_w-1} (c_w - b)\, ¶(B_w = b) \tag{4.4}$$

We may also identify the expected access time for semi-urgent patients, which is our performance indicator of interest. The probability of having an access time 0 weeks is 0, the probability of an access times larger than y weeks is given by:

Where the number of patients that arrived in week w and is scheduled within y weeks is represented by s. If the backlog in week w exceeds the capacity, then s is given by $\sum_{i=0}^{y} c_{w+i} + i - b$. If not, then s is given by $\sum_{i=0}^{y} c_{w+i}$. Hence:

$$s = min\left\{\sum\nolimits_{i=1}^{y} c_{w+i}\ ,\sum\nolimits_{i=0}^{y} c_{w+i} - b\right\} \tag{4.5}$$

In other words, the conditional probability of an access time of at least y weeks is 1 when backlog b is larger than available capacity till week y. If capacity exceeds backlog, all semi-urgent patients until s can be scheduled within y weeks. The conditional access time is then given by:

$$\mathbb{E}(AT_w \mid B_w = b) = \sum_{y=0}^{\infty} P(AT_w > y \mid B_w = b) \tag{4.6}$$

Which can be simplified into:

$$\mathbb{E}(AT_w) = \sum_{b=0}^{\infty} \mathbb{E}(AT_w \mid B_w = b) *\ P(\,B_w = b) \tag{4.7}$$

And thus:

$$\mathbb{E}(AT) = \sum_{w=1}^{W} \frac{\mathbb{E}(AT_w) *\ \mathbb{E}(A_w)}{\sum_{q=1}^{W} \mathbb{E}(A_q)} \tag{4.8}$$

We can determine the minimal number of weekly slots required to provide 90% of patients with an access time of up to one week. We can now regard semi-urgent patients as another "elective" surgical procedure type and the weekly determined number of slots as its demand. Figure 2.1 visualizes the distribution of this demand amongst the sub-specialties based on historical observations. For each distribution, we fit a 3 parameter log-normal distribution to describe surgical duration using Minitab. Details are provided in Appendix C.

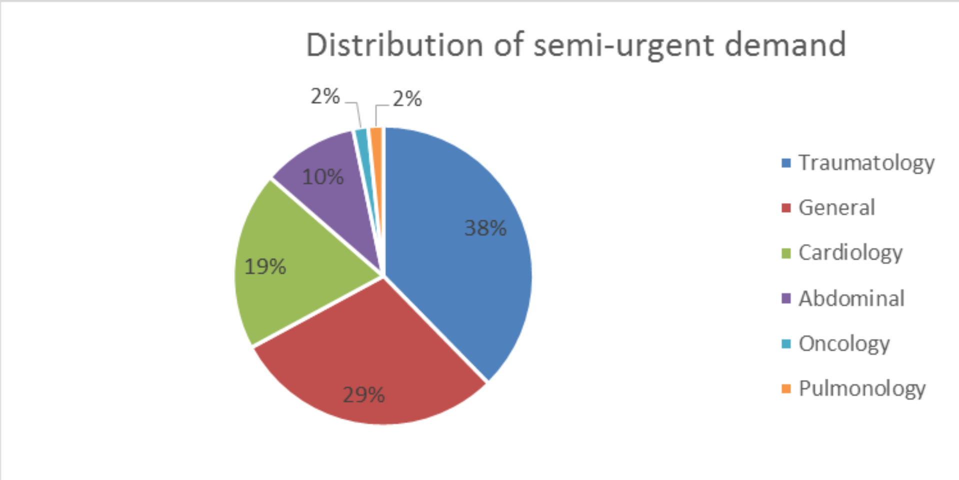

Figure 4.6: Distribution of semi-urgent demand amongst sub-specialties (Source: SAP, data taken between 03-13 and 10-15).

We can convert the weekly demand to four week demand to match the planning cycle. We will ensure for special constraints in the mathematical programming model presented in subsection 4.3.3 to ensure that the weekly required capacity is still met when constructing the SPTS. This queuing model can easily be generalized to accommodate semi-urgent patients with any access time by converting it back to its original form that determines daily required capacity. We can then apply the model to check any SPTS for access-time feasibility. Since our semi-urgent patients have an maximum allowed access time of exactly one week, the current model will suffice as long as we ensure that our SPTS has enough semi-urgent slots per week.

4.3.3 Mathematical programming

In this subsection, we introduce a mathematical programming approach to allocate capacity to surgical procedure types and construct a SPTS.

Base model

We propose a mathematical programming approach to assign surgical procedure types to specific days and ORs in a SPTS. The underlying objective is to minimize the required OR capacity to cover demand and ensuring timely access for semi-urgent procedure types. We first need to introduce some mathematical notations before we can effectively describe the model in technical terms. These notations are also summarized in Table 1.1.

Indices	
$\boldsymbol{r}$	Operating Room capacity type of set $\boldsymbol{R}$
$\boldsymbol{s}$	Sub-specialty $\boldsymbol{s}$ of set $\boldsymbol{S}$
$\boldsymbol{i}$	Surgical procedure type $\boldsymbol{i}$ of set $\boldsymbol{I}$
$\boldsymbol{w}$	Week $\boldsymbol{w}$ of set W
$\boldsymbol{t}$	Day $\boldsymbol{t}$ of set $\boldsymbol{T}$
$\boldsymbol{j}$	Operating room $\boldsymbol{j}$ of set $\boldsymbol{J}$
Variables	
$\boldsymbol{V_{rwtj}}$	Use OR $\boldsymbol{j}$ of capacity type $\boldsymbol{r}$ on day $\boldsymbol{t}$ of week $\boldsymbol{w}$
$\boldsymbol{W_{rwtjs}}$	Assign sub-specialty $\boldsymbol{s}$ to OR $\boldsymbol{j}$ of capacity type $\boldsymbol{r}$ on day $\boldsymbol{t}$ of week $\boldsymbol{w}$

$\boldsymbol{Z_{rwtji}}$	Number of procedure type $\boldsymbol{i}$ in OR $\boldsymbol{j}$ of capacity type $\boldsymbol{r}$ on day $\boldsymbol{t}$ of week $\boldsymbol{w}$
Parameters	
$\boldsymbol{d_r}$	Operating room capacity type $\boldsymbol{r}$ duration in minutes
$\boldsymbol{s_i}$	Minimum demand for surgical procedure type $\boldsymbol{i}$
$\boldsymbol{M_i}$	Maximum number of surgical procedure type $\boldsymbol{i}$ that fits in largest capacity type
$\boldsymbol{A_{wtj}}$	Availability of operating room $\boldsymbol{j}$ on day $\boldsymbol{t}$ of week $\boldsymbol{w}$
$\boldsymbol{B_{wts}}$	Sub-specialty $\boldsymbol{s}$ is available on day $\boldsymbol{t}$ of week $\boldsymbol{w}$
$\boldsymbol{D_{si}}$	Ability of sub-specialty $\boldsymbol{s}$ to perform surgical procedure type $\boldsymbol{i}$

Table 4.2: Queuing model identifiers

We introduce the set W of weeks $(w = 1, \dots, W)$, set T of days t $(t = 1, \dots, T)$. Together they describe all the different dates in the cycle period. We also introduce set J of operating rooms j $(j = 1, \dots, J)$ and capacity type r of set R $(r = 1, \dots, R)$. The capacity of each operating room can be described in minutes by parameter d_r. If we define binary variable $\boldsymbol{V_{rwtj}}$ as the use of operating room j of capacity type r on day t of week w, we can describe our objective of minimizing the total required periodic capacity as:

$$minimize \sum_{r=1}^{R} \sum_{w=1}^{W} \sum_{t=1}^{T} \sum_{j=1}^{J} V_{rwtj} * d_r$$

We now introduce I as the set with surgical procedure types i $(i = 1, \dots, I)$ that should be scheduled in the SPTS and parameter $\boldsymbol{s_i}$ to denote the minimal periodic demand for each procedure type. Parameter M_i describes the maximum number of surgical procedure of type i that fits in an operating room with the largest capacity. We can then describe the number of scheduled procedure type i in OR j of capacity type r on day t of week w by variable Z_{rwtji}. Constraint (4.9) ensures that an OR is considered "used" when procedures are scheduled using the big-M method:

$$Z_{rwtji} \leq M_i * V_{rwtj} \quad \forall r, w, t, j, i \tag{4.9}$$

And we introduce another constraint to make sure that periodic demand is met:

$$\sum_{r=1}^{R} \sum_{w=1}^{W} \sum_{t=1}^{T} \sum_{j=1}^{J} Z_{rwtji} \geq s_i \quad \forall i \tag{4.10}$$

We assumed an unbounded number of ORs in section 4.2 but will model some "virtual" OR availability limits to smooth ORDS assignment throughout the planning cycle. We introduce binary parameter A_{rwtj} to denote the availability of operating room j on day t of week w. To ensure that ORs are only used when they are available we introduce constraint:

$$V_{rwtj} \leq A_{rwtj} \quad \forall\, r, w, t, j \tag{4.11}$$

We would like to ensure that the probability of overtime is limited to some safety factor α, that is chosen by management. Let $f_{rwtj}(Z)$ describe the probability distribution function of the sum of all surgical procedure types in OR j of capacity type r on day t of week w, where Z denotes the set of stochastic variables. We can then introduce the probabilistic constraint:

$$P\left(f_{rwtj}(z) \leq d_r\right) \geq 1 - \alpha \quad \forall\, r, w, t, j \tag{4.12}$$

We introduce binary variableW_{rwtjs} to denote whether sub-specialty s is assigned to OR j of capacity type r on day t of week w. Let binary parameter D_{si} denote whether procedure type i belongs to sub-specialty s. Using the big-M method we can force to assign sub-specialties to OR j of capacity type r on day t of week w if we schedule one of their procedure types:

$$Z_{rwtji} \leq \sum_{s=1}^{S} D_{si} * M_i * W_{rwtjs} \quad \forall r, w, t, j, i \tag{4.13}$$

And we can ensure that only one sub-specialty is assigned to an OR:

$$\sum_{s=1}^{S} W_{rwtjs} \leq 1 \quad \forall\, r, w, t, j \tag{4.14}$$

Let binary parameter B_{wts} denote whether a sub-specialty is available on a specific date, constraint that sub-specialties only perform procedures when one of their surgeons is available:

$$\sum_{r=1}^{R} \sum_{j=1}^{J} W_{rwtjs} \leq B_{wts} \quad \forall\, w, t, s \tag{4.15}$$

Finally, we add a special constraint to ensure that the weekly number of "semi-urgent" surgical procedure types at least equals the output from our queuing model. If W denotes the size of the set of weeks in our planning period, we can recalculate back to the weekly required capacity. Constraint ensures that semi-urgent capacity in our SPTS is still robust:

$$\sum_{r=1}^{R} \sum_{w=1}^{W} \sum_{t=1}^{T} Z_{rwtji} \geq \frac{s_i}{W} \quad \forall\, w, i_{semi-urgent} \tag{4.16}$$

Summarizing, the base model can be formulated as a generic mixed integer programming model:

Minimize:

$$\sum_{r=1}^{R} \sum_{w=1}^{W} \sum_{t=1}^{T} \sum_{j=1}^{J} V_{rwtj} * d_r$$

Subject to:

$$Z_{rwtji} \leq M_i * V_{rwtj} \quad \forall r, w, t, j, i$$

$$\sum_{r=1}^{R} \sum_{w=1}^{W} \sum_{t=1}^{T} \sum_{j=1}^{J} Z_{rwtji} \geq s_i \quad \forall i$$

$$V_{rwtj} \leq A_{rwtj} \quad \forall\, r, w, t, j$$

$$P\left(f_{rwtj}(z) \leq d_r\right) \geq 1 - \alpha \quad \forall\, r, w, t, j$$

$$Z_{rwtji} \leq \sum_{s=1}^{S} D_{si} * M_i * W_{rwtjs} \quad \forall r, w, t, j, i$$

$$\sum_{s=1}^{S} W_{rwtjs} \leq 1 \quad \forall\, r, w, t, j$$

$$\sum_{r=1}^{R} \sum_{j=1}^{J} W_{rwtjs} \leq B_{wts} \quad \forall\, w, t, s$$

$$\sum_{r=1}^{R} \sum_{w=1}^{W} \sum_{t=1}^{T} Z_{rwtji} \geq \frac{S_i}{W} \quad \forall\, w, i_{semi-urgent}$$

$$V_{rwtj} = \{0,1\}, W_{rwtjs} = \{0,1\}, Z_{rwtji} = integer$$

Unfortunately the solution space of this model becomes computationally infeasible for larger problem instances. An additional issue is that probabilistic constraints often have to be solved using heuristic approximations [43]. We therefore propose a column generation approach aimed at reducing the solution space. In the remainder of this section we therefore discuss:

- The column generation algorithm
- The column generation primal model
- The column generation pricing model
- The technical implementation of the column generation model

Column generation

To reduce the computational burden, we propose to decompose the base model into two smaller models. In this way, our problem is comparable to a standard cutting stock problem as described by *Gilmore et al.* [44] and *Bisschop et al.* [45]. Instead of calculating the complete solution space explicitly, such problems are solved by implicitly expanding the model solution space with only improving solutions that are generated by a submodel. Implicit refers to the fact that we do not calculate the entire solution space prior to running the selection model (also referred to as an explicit solution space). This technique is also called delayed column generation, since each solution is essentially a column of the underlying linear program. It is successfully exploited by *van Oostrum et al.* [27], from who we in turn derive our mathematical model with some additional input from *Bosch et al.* [34].

We apply a column generation approach to our problem and identify a primal- and a pricing model, that are respectively the main- and sub model described earlier. In the primal model, we select that combination of ORDs that minimize the required demand and assign these to specific dates and ORs in a SPTS. We follow *van Oostrum et al.* and define ORDs to be an OR-day that is completely scheduled with surgical procedure types [3]. New ORDs are iteratively generated in the pricing model and added to the implicit set of ORDs until newly generated ORDs offer no improvement anymore.

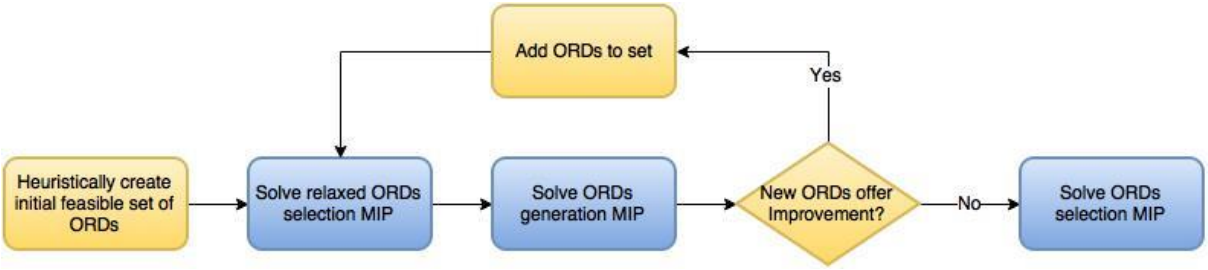

Figure 4.7: Flowchart of column generation algorithm to construct a SPTS

The choice as to which surgery types are scheduled in the pricing model is derived from their shadowprice of one of the primal model constraints. This constraint explained in detail when discussing the primal model. The pricing model selects those procedure types that maximize the sum of shadowprice and result in a feasible total schedule duration. A result of this sequence is that we require a feasible set of ORDs when running the primal model for the first time. We cannot obtain any ORDs via the pricing model since we have not yet generated shadow prices in the primal model. We solve this issue be initializing a feasible set of ORDs via a heuristic. In section 3.2.3, we determined that LPTF showed good performance and therefor apply this heuristic before we run the primal model.

The total schedule duration in the pricing model is determined by the sum of procedure means and slack, which is dependent on pooled procedure type variance and a chosen probability factor of overtime. The use of slack allows the pricing model generate ORDs that have a limited probability of running into overtime. Schedule duration is constrained by the capacity of that particular ORDs. The new ORDs is then added to the implicit set of the ORDs selection model, which is then solved to obtain new shadow prices. Figure 4.7 depicts the entire process. We implement our model in commercial solver AIMMS. In the remainder of this section we will discuss:

Primal model

In the primal model we want to create a SPTS from generated ORDs. We first need to introduce some mathematical notations before we can effectively describe the model in technical terms. Table 4.3 summarizes these notations.

Indices	
r	Operating Room capacity type of set R
u	Operating Room Day types of set U
s	Sub-specialty of set S
i	Surgical procedure type of set I
w	Week w of set W
t	Day t of set T
j	Operating room j of set J
Variables	
X_{ru}	Number of selected ORDs u of capacity type r
Y_{ruwtj}	Assign ORDs u of capacity type r to operating room j on day t of week w

Parameters	
$\boldsymbol{d_r}$	Operating room capacity type $\boldsymbol{r}$ duration in minutes
$\boldsymbol{a_{rui}}$	Number of surgical procedures type $\boldsymbol{i}$ in ORD $\boldsymbol{u}$ capacity type $\boldsymbol{r}$
$\boldsymbol{s_i}$	Minimum demand for surgical procedure type $\boldsymbol{i}$
$\boldsymbol{m_r}$	Number of available operating rooms of capacity type $\boldsymbol{r}$
$\boldsymbol{A_{wtj}}$	Availability of operating room $\boldsymbol{j}$ on day $\boldsymbol{t}$ of week $\boldsymbol{w}$
$\boldsymbol{B_{rus}}$	Sub-specialty $\boldsymbol{s}$ performs the surgical schedule of ORDs $\boldsymbol{u}$ of capacity type $\boldsymbol{r}$
$\boldsymbol{D_{wts}}$	Availability of sub-specialty $\boldsymbol{s}$ on day $\boldsymbol{t}$ of week $\boldsymbol{w}$

Table 4.3: Primal model identifiers

Let set U with ORDs u $(u = 1, \dots, U)$. Since not all ORDs have an equal size, we also introduce the set R with capacity types r $(r = 1, \dots R)$, sorted in descending order of capacity size. To denote the selection of optimal ORDs we introduce variable X_{ru} that denotes the integer number of selected ORDs u with capacity type r. Parameter d_r denotes the capacity of type r in minutes, m_r denotes the maximum number of ORDs u that we allow to be scheduled in the SPTS. The value of m_r is a management decision. The total number of capacity we use in our SPTS is determined by the sum of all ORDs of all capacity types:

$$\sum_{r=1}^{R} \sum_{u=1}^{U} X_{ru} * d_r$$

We re-introduce I as the set with surgical procedure types i $(i = 1, \dots, I)$ that should be scheduled in the SPTS. Each surgical procedure type i has an expected duration μ_i, variance θ_i and minimal cyclic demand of s_i. For any ORD u of capacity type r, we denote the number of surgical procedure types i scheduled with the parameter a_{rui}. Our goal is to select a set with ORDs that meets our demand requirement s_i and semi-urgent access time target. Preferably, we would meet these requirements with as few ORDs as possible, so we could view this as a minimization problem. This objective function can be mathematically described as:

$$minimize \sum_{r=1}^{R} \sum_{u=1}^{U} X_{ru} * d_r$$

We add a constraint to ensure that enough ORDs are selected to cover demand of each surgical procedure type i:

$$\sum_{r=1}^{R} \sum_{u=1}^{U} X_{ru} * a_{rui} \geq s_i \quad \forall\, i \tag{4.17}$$

We also constrain the maximum number of ORDs to be selected for each capacity type r:

$$\sum_{r=1}^{R} \sum_{u=1}^{U} X_{ru} \leq m_r \quad \forall\, r \tag{4.18}$$

So far, we have mainly introduced variables and parameters that refer to the selection of ORDs in the cycle. However, we also need to assign these ORDs to specific dates and ORs. Let W be the set of weeks w $(w = 1, \dots, W)$ in the planning horizon and T the set with weekdays

t ($t = 1, \ldots, T$) in each week. We also introduce set J of actual ORs ($j = 1, \ldots, J$). We now have everything available to define binary ORDs assignment variable Y_{ruwtj}, that assigns ORDs u of capacity type r to operating room j on day t of week w. We add a constraint to force the assignment of all selected ORDs:

$$\sum_{W}\sum_{T}\sum_{J} Y_{ruwtj} = X_{ru} \; \forall\, r, u \tag{4.19}$$

Since the number of available ORs may vary per day, we introduce binary parameter A_{wtj} that denotes the availability of operating room j on day t of week w. We add the constraint that ORDs may only be assigned to actual available ORs:

$$\sum_{R}\sum_{U} Y_{ruwjt} \leq A_{jt} \;\; \forall\, w, t, j \tag{4.20}$$

Let S be the set with sub-specialties ($s = 1, \ldots, S$). We now define two binary parameters. Let B_{rus} denote whether sub-specialty s performs the surgeries in ORD u of capacity type r and D_{wts} denote the availability of sub-specialty s on day t of week w. We may then add the constraint that sub-specialties may only be scheduled when they are available:

$$\sum_{R}\sum_{U}\sum_{J} B_{rus} * Y_{ruwtj} \leq D_{wts} \; \forall\, w, t, s \tag{4.21}$$

For semi-urgent surgical procedure types, we need an extra constraint to ensure that each week we have at least as many patients planned as the weekly number of slots we determined in subsection 4.3.2. Therefore, we add constraint:

$$\sum_{R}\sum_{U}\sum_{T}\sum_{J}\sum_{I_{su}} Y_{ruwtj} * a_{iru} \geq \sum_{I_{su}} \frac{s_i}{W} \;\; \forall\, w \tag{4.22}$$

Summarizing, we may describe the primal model as:

Minimize:

$$\sum_{r=1}^{R}\sum_{u=1}^{U} X_{ru} * d_r$$

Subject to:

$$\sum_{r=1}^{R}\sum_{u=1}^{U} X_{ru} * a_{rui} \geq s_i \;\; \forall\, i$$

$$\sum_{r=1}^{R}\sum_{u=1}^{U} X_{ru} \leq m_r \;\;\; \forall\, r$$

$$\sum_{W}\sum_{T}\sum_{J} Y_{ruwtj} = X_{ru} \; \forall\, r, u$$

$$\sum_R \sum_U Y_{ruwjt} \leq A_{rjt} \; \forall \, r, w, t, j$$

$$\sum_R \sum_U \sum_J B_{rus} * Y_{ruwtj} \leq D_{wts} \; \forall \, w, t, s$$

$$\sum_R \sum_U \sum_T \sum_J \sum_{I_{su}} Y_{ruwtj} * a_{iru} * B_{rus} \geq \sum_{I_{su}} \frac{s_i}{W} \quad \forall \, w, s$$

$$X_u = int, \qquad Y_{ruwtj} = \{0,1\}$$

Pricing model

The pricing model is solved subsequently for each capacity type and each sub-specialty. Any ORDs that offer an improvement to the primal set are added. The primal model is then solved with the updated ORDs set. This process is visualized in Figure 4.7. We introduce some additional identifiers in the pricing model, which are listed in Table 4.4. New ORDs are deemed an improvement to the set if they violate the reduced cost criterion:

$$\sum_{i=1}^{I} \lambda_i * Z_i^r > d_r - \bar{\pi}_r + \delta \tag{4.23}$$

In which λ_i denotes the shadow price of the primal surgery type i demand and π_r the shadow price of the maximum allowed ORDs of capacity type r constraint. We define a small margin of error δ to allow for numerical inaccuracies in the shadow price. Variable Z_i^r denotes the number of scheduled surgical procedure types i in the new ORDs. It represents the parameter a_{rui} from the primal model. The index r may be dropped from this variable since the pricing model generates an ORD for each capacity type separately.

Indices	
$\boldsymbol{n}$	Breakpoint n of set $\boldsymbol{N}$
Parameters	
$\boldsymbol{d^r}$	Operating room capacity type r usage
$\boldsymbol{\lambda_i}$	Shadow price of surgical demand $\boldsymbol{s_i}$ of primal model
$\boldsymbol{\mu_i}$	Expected duration of surgical procedure $\boldsymbol{i}$
$\boldsymbol{\sigma_i}$	Square root of variance of surgical procedure type $\boldsymbol{i}$
$\boldsymbol{\alpha}$	Risk of overtime
Variables	
$\boldsymbol{Z_i^r}$	Number of surgical procedures of type $\boldsymbol{i}$ planned in an ORDs with capacity $\boldsymbol{r}$
$\boldsymbol{\rho_n}$	

Table 4.4: Additional identifiers of the pricing model

Our goal in the pricing model is to find a column of Z_i^r that violates the reduced cost criterion, preferably by as much as possible. Let binary parameter C_{si} denote the ability of sub-specialty s to perform surgical procedure type i. The pricing objective may then be mathematically denotes as:

$$maximize \sum_{i=1}^{I} C_{si} * \lambda_i * Z_i^r$$

We require a constraint to ensure that the total schedule duration does not exceed the OR capacity. The total required duration is determined by the sum of individual expected durations and their slack. Slack is in turn determined by the safety factor α and surgical procedure type variance. To reduce the size of summated slack, we would like to implement the portfolio effect by pooling variance of multiple scheduled procedures. In section 4.3.1 we determined that surgical procedure type durations are described by both Gaussian and 3 parametric log-normal distributions. These distributions require a different approach to incorporate the portfolio effect. Unfortunately, we cannot simply implement a non-linear probabilistic constraint (4.12) into our linear program. We therefore discretize the values for expected duration μ_i and variance σ_i^2 of surgical procedure type i, and will discuss approximations of the portfolio effect for both distributions in our model.

For Gaussian variables, we assume the sum of multiple Gaussian variables i to be also Gaussian distributed with $\mu_{sum} = \sum_{i=1}^{I} \mu_i$ and variance $\sigma_{Sum} = \sum_{i=1}^{I} \sigma_i^2$. It can then be shown that the total duration is then determined by the square root of the pooled variance, that is again non-linear. In Appendix B, we describe a linear piecewise approximation between intervals of the square root function in the ORDs capacity constraint as demonstrated by *Bosch et al.* [34]. We introduce a set N of breakpoints $(n = 0,1, \dots N)$ that separate each linear interval. We introduce the breakpoint value x_n and breakpoint function value y_n. It can be shown that each interval can then be described as a weighted sum of the breakpoints, where the sum equals 1. Let parameter ρ_n denote the weights for each breakpoint. We can then formulate the following constraints to incorporate the portfolio effect for Gaussian distributed duration in the ORDs capacity constraint as:

$$\sum_{i=1}^{I} (Z_i * \mu_i) + (1 - \alpha) * \sum_{n \in N} \rho_n * \gamma_n \leq d^r \tag{4.24}$$

$$\sum_{n \in N}^{I} \rho_n \chi_n = \sum_{i \in I} Y_i \, \sigma_i^2 \tag{4.25}$$

$$\sum_{n \in N} \rho_n = 1 \tag{4.26}$$

Due to their open form, we cannot (power) summate 3-parameter log-normal variables like we did with Gaussian variables, which makes pooling variance difficult. We therefore propose the approximation of the portfolio effect by *van Oostrum et al.* [3]. We describe this approximation in detail in Appendix B. This approximation first requires us to convert our 3-parameter log-normal to 2-parameter log-normal distributed values. We achieve this by subtracting the third parameter that describes the shift of the log-normal function, from all the observations in the empirical dataset. We then refit a 2-parameter to the adjusted dataset. We introduce prediction bound n_i^a that denotes the upper bound value for which the duration of a single surgical procedure type i is smaller than with probability a. This prediction bound is a shorthand notation and can again be decomposed into the expected duration and slack. To determine the total duration of multiple procedure types, we summate the prediction bounds and subtract an approximated value of the portfolio effect. The size of the portfolio effect is determined by the difference between the summated prediction bounds and the prediction bound of a Schwartz-Yeh-Ho approximation of the power sum of all the scheduled procedures i [46]. We can then model the portfolio effect as a function g that only depends on the number of surgical procedures i:

$$\sum_{i=1}^{I}(Z_i * n_i^a) - g(\sum_{i=1}^{I} Z_i) \leq d^r \tag{4.27}$$

To implement this function in a commercial solver, we introduce set E $(e = 0,1 \dots, E)$ of power summated duration log-normal distributed surgical procedure types in an ORDs. We introduce the binary counter variable F_{ei} that denotes whether there is an e^{th} surgical procedure type i scheduled. Let g_{ei} represent the portfolio effect when e surgical procedures of type i with log-normal distributed durations are scheduled. We can then formulate the following constraints to incorporate the portfolio effect for log-normal distributed duration in the ORDs capacity constraint as :

$$\sum_{i=1}^{I}(Z_i * n_i^a) - \sum_{i=1}^{I}(g_{ei}) \leq d^r \tag{4.28}$$

$$Z_i = \sum_{e=0}^{E} F_{ei} \quad \forall\, i \tag{4.29}$$

$$F_{ei} \geq F_{e+1,i} \quad \forall\, e, i \tag{4.30}$$

We can combine the Gaussian and log-normal constraints to a single constraint by adding parameter g_i and l_i, that indicate the distribution of each surgical procedure type i. We formulate the set of constraints as:

$$\sum_{i=1}^{I}(Z_i * l_i * n_i^a) - \sum_{i=1}^{I}(g_{ei} * l_i) + \sum_{i=1}^{I}(Z_i * g_i * \mu_i) + (1-\alpha) * \sum_{n \in N} \rho_n * \gamma_n \leq d^r \tag{4.31}$$

$$\sum_{n \in N}^{I} \rho_n \chi_n = \sum_{i \in I} Y_i * g_i * \sigma_i^2$$

$$\sum_{n \epsilon N} \rho_n = 1$$

$$Z_i = \sum_{e=0}^{E} F_{ei} \quad \forall\, i$$

$$F_{ei} \geq F_{e+1,i} \quad \forall\, e, i$$

Recall that we consider change-overs to be a surgical procedure type. In the primal model, its periodic demand s_i is zero and therefore its shadow price is also zero. To ensure that an appropriate amount of changeovers are scheduled we add another constraint:

$$Z_i = \sum_{i=1}^{I_{procedures}} (Z_i) - 1 \ \forall \ i_{Changeover} \tag{4.32}$$

Apart from the ORDs capacity constraints, we require a constraint to prevent overzealous pooling of semi-urgent procedure types. This might would result in infeasibility for constraint (4.22) in the primal model. We re-introduce M_i as the maximal number of procedure types i in an ORDs, and set the value for semi-urgent procedures equal to the weekly demand. We can formulate this constraint as:

$$\sum_{i \in I_{SU}} Z_i^r \leq M_i \tag{4.33}$$

The pricing model can then be summarized as:

Maximize:

$$\sum_{i \in I_s} \lambda_i * Z_i^r$$

Subject to:

$$\sum_{i=1}^{I} (Z_i * l_i * n_i^a) - \sum_{i=1}^{I} (g_{ei} * l_i) + \sum_{i=1}^{I} (Z_i * g_i * \mu_i) + (1 - \alpha) * \sum_{n \in N} \rho_n * \gamma_n \leq d^r$$

$$\sum_{n \in N}^{I} \rho_n \chi_n = \sum_{i \in I} Y_i * g_i * \sigma_i^2$$

$$\sum_{n \epsilon N} \rho_n = 1$$

$$Z_i = \sum_{e=0}^{E} F_{ei} \ \ \forall \ i$$

$$F_{ei} \geq F_{e+1,i} \quad \forall \ e, i$$

$$\sum_{i \in I_{SU}} Z_i^r \leq M_i$$

$$Z_i \in \mathbb{N}, \forall i \ \in \ I_s$$

Technical implementation

We use the commercial solver AIMMS 4.15 to model the mathematical programs. For end-users like the planners at the admission office, a graphical user interface is added. It does not only show an overview of the SPTS, but also information about which surgical procedure types an ORDs contains and estimated procedure starting and duration times. A Gantt chart provides additional visual context.

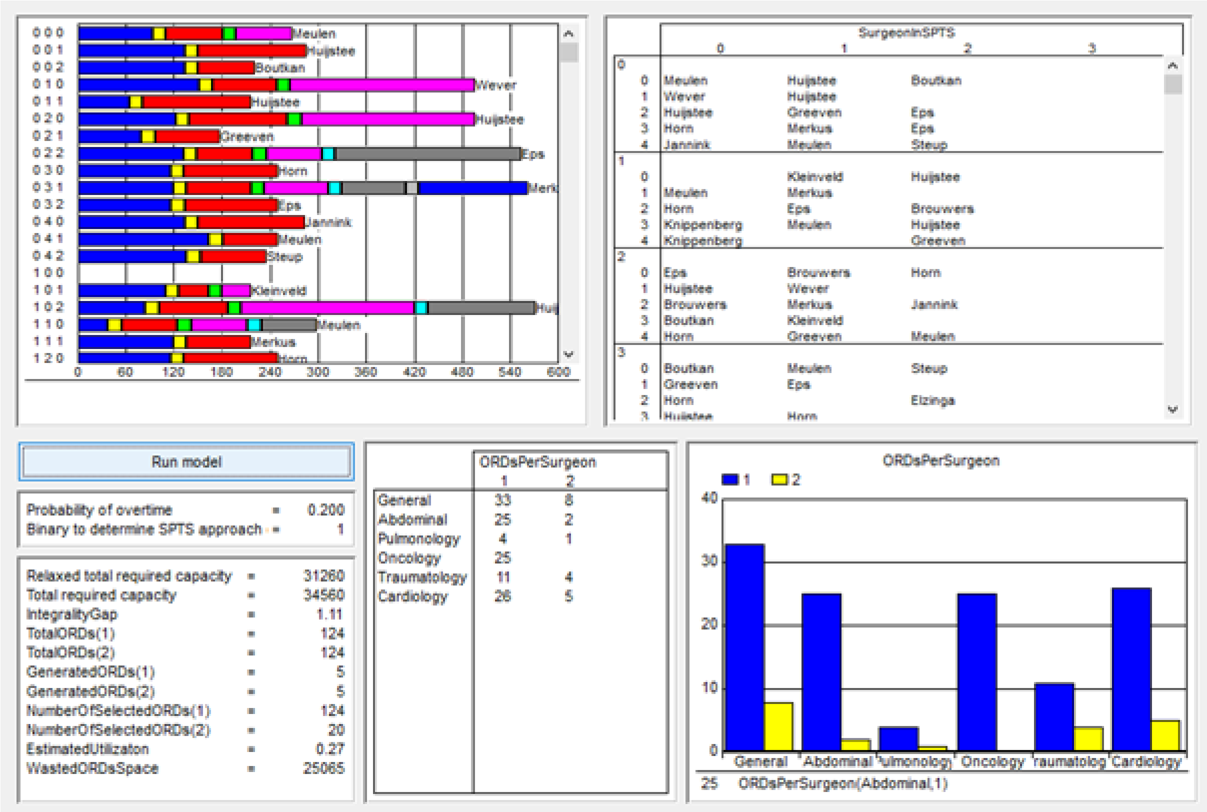

Figure 4.8: Graphical user interface of SPTS mathematical programming model (Source: AIMMS).

Input parameters such as surgical demand, duration, OR availability, probability of overtime and OR capacity can be adjusted and loaded through an user friendly excel file. Most of these parameters rarely need to be adjusted. We chose for an optimality tolerance of 1% for the primal model.

Validation

We will perform a simulation to determine the validity of our SPTS. Our main interests are whether our estimated utilization and overtime probability are valid

We first introduce the default scenario. This scenario is determined to closely mimic the current situation at Haga in order to make a fair comparison in performance. The current planning horizon and opening hours of the OR department are chosen as input. We choose a 30% probability of overtime and determine that 11 weekly semi-urgent slots are an appropriate choice, based on interviews with professionals. We choose the median, or third whisker, as demand input. The default scenario is:

- Median monthly demand (third whisker)
- Planning horizon of 4 weeks
- 11 semi-urgent slots per week/ 44 semi-urgent slots per cycle
- ORDs capacity of 8 hours
- Probability of overtime $\alpha = 0.3$

Indicator	Model	Simulation

	Estimate	Confidence interval (95%)	Average
Utilization	79%	[77.2%-78.5%]	77.9%
Overtime	30%	[19.9%-22.7%]	21.0%

Table 4.5: Average monthly utilization and overtime frequency model estimates and simulation results (N = 30 runs)

We performed a simulation to test whether our model estimates for utilization and overtime probability are valid. Our main interest is whether these performance estimates hold when we schedule surgical cases rather than the surgical procedure types planned used the SPTS. We therefore scheduled randomized surgical procedures, based on their relative demand within a surgical procedure type, into the default scenario SPTS. We performed thirty runs, equivalent to 30 months. We deem this number to be large enough to create a reliable confidence interval. Table 4.5 contains the confidence intervals and averages. Our model estimates utilization based on the ratio mean duration sum of scheduled procedure types and ORDs capacity. Changeovers are considered as idle time. In the simulation, we draw semi-random surgical procedures, based on their relative number of historical observations, for each scheduled surgical procedure type. For surgical duration, we draw a random values from the surgical procedure distributions which we used as input in subsection 4.3.1. We consider overtime when the sum of surgical procedure durations exceeds 8 hours. We measured an 21.0% overtime frequency and a 14.2% overtime frequency of overtimes that exceed half an hour. The average duration of overtime is 17.7 minutes. We see that for both performance indicators in Table 4.5: Average monthly utilization and overtime frequency model estimates and simulation results (N = 30 runs)Table 4.5, the model estimated values lie outside of the simulation confidence interval results. Despite this, the utilization simulation average is close to our model estimate, and can be seen as a rough approximation. Even though our model estimate of overtime frequency is off by some margin from the confidence interval, we can reason that a worst case scenario overtime frequency of up to 30% holds.

We believe that discrepancies between model estimates and the simulation results are caused by inter-cluster duration variability within surgical procedure types, as also reported by *van Oostrum et al.* [22]. We should take our model estimates with a grain of salt, with utilization being a "rough" estimate and overtime frequency a worst case scenario result.

4.3.4 Flexible SPTS approach

In this subsection, we discuss a flexible approach to applying our SPTS. This approach consist of a fixed component and a variable number of monthly add-on ORDs. This approach presents two advantages. The fixed component allows for easy scheduling for both surgeon planners, staff planners and admission office schedulers to cover the majority of monthly demand. The relative small variations in monthly demand are accounted for with the variable component, which ensures an overall close match of demand and capacity. We expect that this will prevent ORDs waste.

Fixed SPTS component

The fixed component of the SPTS model need only be updated once a year when a new case-mix is determined. When the case-mix changes, new elective surgical procedure types should be clustered as presented in subsection 4.3.1, with according new demand functions. It is also recommended to update the number of slots determined in subsection 4.3.2, based on a new Poisson distribution that describes semi-urgent arrivals. Haga should decide on a probability that they want to risk overtime, which is required as input for the SPTS model. Haga should also decide on a number of semi-urgent slots, based on their access time target. We assume semi-urgent demand to be adequately described by the same Poisson distribution throughout the year, which results in a fixed number of slots as input for the SPTS model.

The seasonal variability of elective demand described in section 0 is also represented through the quartile distribution. For demand of the fixed component of the SPTS, we decide to take the second whisker of the quartile function. This accounts for the lowest 25% of observed elective monthly demand. Figure 4.9 depicts the position of the second whisker in a box-plot. With this relative low expected monthly demand, we can reasonably assume that little to no OR capacity is wasted if Haga keeps a sufficient waiting list buffer of one month. If Haga has decided on a probability of overtime, we can use the model presented in section 4.3.3 to construct the fixed component of the SPTS.

The model consists of a set of ORDs assigned to specific dates, which contain surgical procedure types. Each surgical procedure type in turn consists of a set of surgical procedures. Staff- and surgeon planners can simply assign surgeons to the sub-specialty ORDs assigned in the SPTS. If a certain assignment turns out to be infeasible for the surgeon, the constraint input in the excel file for the SPTS model can be adjusted and the model re-solved. The surgical case schedulers at the admission office can use the graphical user interface to determine which surgical procedure types should be scheduled in which ORDs. Using an included excel tool, they can simply process the waiting list in FCFS order and assign each surgical case to the first available ORDs. Surgical duration estimates and slack are determined by the model, and estimations for surgical starting time can be read from the graphical user interface. We do recommend that patients arrive early by some margin to avoid OR idle time when a surgical schedule advances quicker than expected.

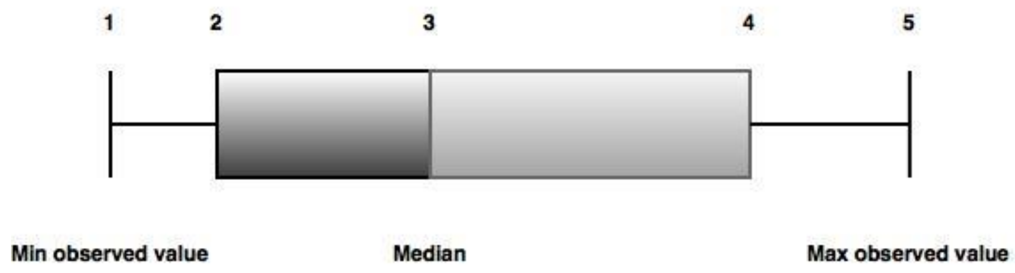

Figure 4.9: Whisker numbers of a box-and whisker plot. The shape of the box is determined by the spread of observations.

Variable SPTS component

With the variable component, we likely require additional capacity to cover monthly demand. Since monthly demand will vary, we propose the add a flexible number of extra ORDs. The exact number of these ORDs should be determined monthly, based on the waiting list arrivals of the previous month. Our model includes an excel tool which can be used to convert waiting list demand into surgical procedure type demand. This demand can then be used as input for the SPTS model to create add-on ORDs. Since the differences in demand between the second, third and fourth whiskers are generally small, we can expect a relatively low number of add-on ORDs. These ORDs can then be added on top of the fixed component based on surgeon and OR availability. Patients can be scheduled into the SPTS in the same way as with the fixed component. Table 4.6 depicts how the demand input would be for the mathematical program for both components.

Approach	Component	Demand input per whisker			
		2	**3**	**4**	**5**
Static	Fixed	2	3	3	3
Flexible	Fixed	2	2	2	2
	Variable	-	(3-2)	(4-2)	(5-2)

Table 4.6: Top four whiskers demand input for the SPTS components.

4.4 Summary

In this chapter, we presented a clustering method based on *van Oostrum et al.* [1] to form elective surgical procedure types within sub-specialty domains with reduce demand uncertainty and included variable levels demand in our models. We presented a queuing theory method based on *Kortbeek et al.* [2] to determine weekly capacity for semi-urgent patients with the underlying condition that arrival follows a Poisson process. We also presented a column generation approach based on *van Oostrum et al.* [3]to create a SPTS with inclusion of semi-urgent patients and changeovers. In this model, we fill ORDs with patients and assign these to specific dates and ORs. The result is a SPTS in which each ORDs can be filled with a specific number of surgical procedures from that sub-specialty and assigned to a surgeon. This model incorporates slack and the portfolio effect by *Hans et al.* [4] to manage overtime probability. Furthermore, we include both Gaussian and log-normal distributed durations in our model while most literature only incorporates one. An underlying condition is that each surgeon can perform all surgical procedures within his sub-specialty. However, the model can also be applied to instances without this condition by determining surgeon specific demand. The result would be a SPTS with ORDs assigned to surgeons instead of sub-specialties. Lastly, we showed how this model can be implemented in a flexible way that deals with elective patient demand fluctuations.

5 Results

In this chapter, we perform a quantitative analysis of the solution approach introduced in chapter 4. In section 5.1, we describe the performance indicators by which we measure outcome. In section 5.2, we introduce the factors which we use in our experimental approach. In section 5.3, we present and discuss the results of our experiments. We summarize this chapter in section 5.4.

5.1 Performance indicators

We first describe indicators that provide insight in the performance of our solution approach. We introduce the probability of access time within a week for semi-urgent patients as a performance indicator to quantify our robustness for arrival uncertainty. To quantify the economic performance of our solution, we introduce utilization estimates and monthly required capacity as performance indicators. We approximate utilization in our SPTS model by determining the ratio between total scheduled procedure duration and ORD capacity. The total required capacity in a planning horizon is the result of selected ORDs and their capacity.

To gain some more insight in the computational performance of our mathematical programming model, we measure integrality gap and model runtime. Our SPTS calculates the required capacity in a planning horizon to cover demand. This capacity will likely not be optimal since the model is an integer programming model and our problem is non-polynomial hard, which results in an enormous computational burden. Such problems are circumvented by relaxing the model to a linear program, which is then solved in conjunction with a solution rounding strategy. The integrality gap can be seen as an approximation ratio for the approximation algorithm that attempts to find the best integer solution. It is determined by the ratio of the best found integer solution and the optimal (fractional) solution, which is at least 1. The runtime provides us with information about the size of the solution space and the efficiency of our model. To summarize, the solution related performance indicators are:

- Semi-urgent access time
- Utilization
- Required monthly capacity

And the model related performance indicators are:

- Integrality Gap
- SPTS model runtime

5.2 Experiment approach

We will introduce the default scenario and a number of internal- and external experimental factors. In each experiment in subsection 5.2.2, we adjust the value of a single factor and compare it against a default scenario.

5.2.1 Default scenario

We first introduce the default scenario. This scenario is determined to closely mimic the current situation at Haga in order to make a fair comparison in performance. The current planning horizon and opening hours of the OR department are chosen as input. Based on interviews with, we choose a 30% probability of overtime and 11 weekly semi-urgent slots. For elective demand, we choose the most likely demand encountered at Haga during four weeks between 3-13 and 10-15, the median demand (or third whisker). The default scenario is:

- Median monthly demand (third whisker)
- Planning horizon of 4 weeks
- 11 semi-urgent slots per week/ 44 semi-urgent slots per cycle
- ORDs capacity of 8 hours
- Probability of overtime $\alpha = 0.3$
- Third whisker demand input

5.2.2 Experimental factors

In this subsection, we introduce several internal experimental factors. To avoid misinterpretation of effects, we will only perform experiments adjusting one experimental factor at the time. This means that if one factor is adjusted for testing, all other factors will be fixed at the default scenario setting described in subsection 5.2.1.

Number of semi-urgent slots

The number of weekly reserved semi-urgent slots influences the probability of a one week access time, but also utilization and the required number of ORDs. We conduct experiments to with a range of slots quantify this influence. Table 5.1 provides an overview of the input values.

OR opening hours

Currently, the ORs are opened for 8 hours per day. Management is considering to increase opening hours by an extra hour. We conduct experiments with a range of daily opening hours to determine the effects on utilization and the number of required ORDs. Table 5.1 provides an overview of the input values. For ORDs combinations with five and three or eight and four hours, we need to elaborate. We assume that physical ORs are opened eight hours per day. Therefore, we can assign two four hour ORDs, ore one five hour and one three hour ORDs, to the same OR on the same day. The exception is the combination with 4.5 and nine hour ORDs. In this option, we assume that physical ORs are opened nine hours a day. That would mean that two 4.5 hour ORDs can be assigned to the same OR at the same day. In all scenario's, we still only allow one surgical sub-specialty per ORD. In reality, it could mean that in the morning, one sub-specialty is scheduled in a four hour ORDs, and in the afternoon another specialty in another ORDs.

Overtime probability

We adjust the safety factor α for the probability of overtime. We perform various experiments with adjustments of α to gain more insight in this trade-off. Table 5.1 provides an overview of the input values.

Planning horizon

The current MSS planning horizon for all specialties at Haga is four weeks. We conduct an experiment to determine how performance compares against a cycle of two weeks. Other cycle lengths would be ungainly to implement since it would not match properly with other specialty planning cycles.

Relaxation of constraints

Currently, only surgical procedure types of one sub-specialty are allowed in an ORDs to simplify the planning of surgeons and staff. This assumption is enforced by constraint (4.20) in the primal model. We perform two experiments with semi-urgent and all procedure types in which we relax to constraint to allowing multiple sub-specialties, to examine the potential benefits of such a challenging scheduling approach.

SPTS schedule approach & variable demand

We test how both the flexible and static SPTS approaches described in subsection Flexible SPTS approach 4.3.4 perform with variable amounts of demand to determine the best approach. For each whisker, we will compare the average utilization of both components of the flexible approach against utilization of the static approach. We will test this for only eight hour ORDs, and for a combination of eight and four hour ORDs. Table 4.6 provides the whisker input demand for the static approach and two components of the flexible approach.

Expected Surgical duration increase

We test how utilization and the number of required ORDs in the default scenario are influenced by an increase of 20% in the expected duration of surgical procedures. Such a scenario may occur because of new clinical techniques or due to an increase in unexperienced residents or surgeons.

Factor	**Values**
Number of slots	43-50
OR hours	4, 7, 8, 8.5, 9, 10, 8+4, 5+3, 9+4.5
Probability of overtime	20%, 30%, 40%
Planning horizon	2 weeks, 4 weeks
Surgical duration increase	20%
Allow semi-urgent procedures in any specialty ORDs	Relaxation
Primal model constraint (4.20)	Relaxation
Elective demand variability 8 hour ORDs	2-4 whisker demand
Elective demand variability 8+4 hour ORDs	2-4 whisker demand

Table 5.1: Values of experimental factors

5.3 Results

In this section, we present the performance results of our experiments. We first discuss the performance results of the default scenario, then the results of the internal- and external factor experiments.

5.3.1 Default scenario

In this subsection, we describe the results of our default scenario described in subsection 5.2.1. Table 5.3 contains the results for the default scenario. We can observe a utilization of 79% with a 30% probability of overtime and with a 85% probability of access within a week for semi-urgent patients. With eight hour surgical time, we monthly require 60 ORDs. This is a slight increase compared to the current number of 55 ORDs. However, in the current situation we also observe an increase of the waiting list, hence the current capacity is not a fair comparison. Figure 5.1 provides a graphical example of how surgical procedure types are scheduled in an OR in the default scenario.

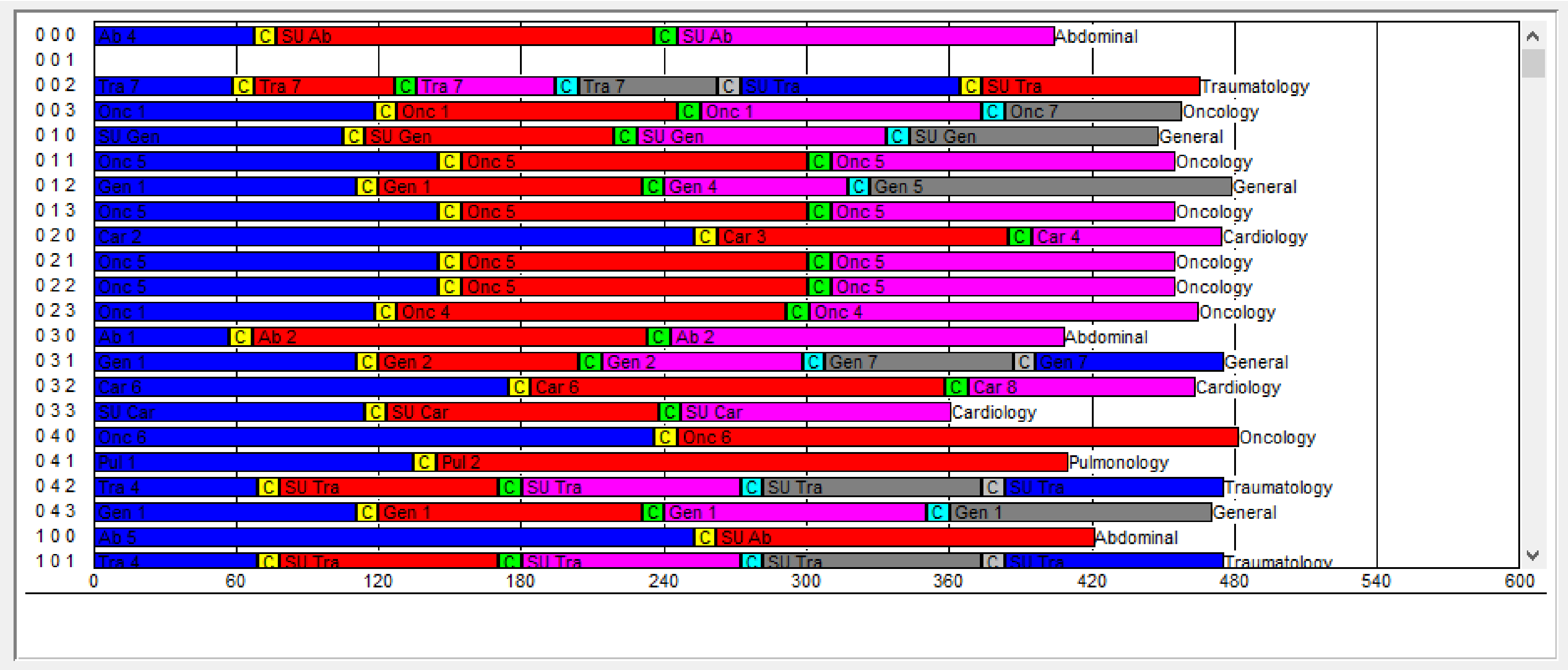

Figure 5.1: Gantt chart of a part of the default SPTS. Visible are the scheduled surgical procedure types and changeovers per each week, day and OR. Each bar consists of an ORDs of the specified sub-specialty on the right side of the bar. The x-axis depicts the duration in minutes. The y-axis depicts the week-number, day-number and OR-number (AIMMS).

5.3.2 Experimental results

In this section, we describe the performance results of the experiments described in subsection 0. For each performance indicator, we will discuss the influence of the experiments. Table 5.3 provides an overview of the results.

Number of semi-urgent slots

We ran the queuing model described in subsection 4.3.2 in Matlab to determine the performance influence of the number of semi-urgent slots. Weekly semi-urgent can be described by a Poisson distribution with rate $\lambda_w = 10.5$. To ensure a stable model, we require a capacity that at least equals demand. This means that in a 4 week cycle, we require at least 43 slots. We determine the influence of the amount of slots on access time. Table 5.2 depicts the results.

Number of slots	1 week access	2 week access	3 week access
43	82%	99%	100%
44	85%	99%	100%
45	87%	100%	100%
46	89%	100%	100%
47	91%	100%	100%
48	92%	100%	100%
49	93%	100%	100%
50	94%	100%	100%

Table 5.2: Access time probability per monthly number of semi-urgent slots, in which probabilities are rounded up to the nearest integer.

We can obtain several insights when we inspect the results. First, we can observe that the probability of access within 2 weeks is already nearly 100% with the minimum required number of slots for a stable queuing system. Second, we can notice the decrease in marginal benefit when adding extra slots. This can be explained by the nature of the Poisson distributed arrival process in which the duration between two consecutive arrivals is exponentially distributed. If we are persistent in guaranteeing a 90% probability of access within a week, it will be relatively more costly than if we settle for a 82% or 85% access time. Management should therefore decide how much OR time they want to invest for a particular access time.

Utilization

We determine the effects of internal experimental factors on utilization, which are visualized in Table 5.3. The effects of OR opening hours on utilization are interesting to note. A half hour increase in opening hours actually results in an equal utilization, but seven, nine or ten hour days result in A 1% increase compared to the default eight hours. Four hour ORDs lower the estimated utilization to 77%. We might argue that the ratio of surgical procedure type duration to ORDs capacity influences the number of scheduling options for our model.

Experiment	Total ORDs I	Generated ORDs I	Selected ORDs I	Total ORDs II	Generated ORDs II	Selected ORDs II	Objective	Objective (relaxed)	Integrality gap	Run time	Estimated utilization
Default scenario	105	32	60	0	0	0	32400	27035	20%	48.1	79%
Opening hours: 4	156	9	139	0	0	0	33360	30320	10%	20.6	77%
Opening hours: 7	123	29	70	0	0	0	29400	26670	10%	33.5	80%
Opening hours: 8.5	107	31	62	0	0	0	31620	27160	16%	41.6	79%
Opening hours: 9	137	43	75	0	0	0	32400	27035	20%	47.2	80%
Opening hours: 10	129	41	69	0	0	0	32400	26607	22%	94.4	80%
Opening hours: 8 and 4 hours	101	15	42	144	7	32	28800	25957	11%	46.6	82%
Opening hours: 5 and 3 hours	166	8	67	247	1	44	28020	25640	9%	47.7	85%
Opening hours: 9 and 4.5 hours	207	13	31	206	7	48	29700	26294	13%	93.8	81%
Overtime probability: 20%	79	25	72	0	0	0	34560	29239	18%	41.8	72%
Overtime Probability: 40%	74	40	60	0	0	0	28800	24182	19%	60.6	89%
Relaxation SU-ORDs constraint	67	30	58	0	0	0	29580	27168	9%	42.4	80%
Relaxation ORDs constraint	319	282	54	0	0	0	27540	26358	4%	381.2	82%
Semi-urgent slots: 43	107	53	59	0	0	0	31860	26728	19%	76.2	79%
Semi-urgent slots: 45	108	35	60	0	0	0	32400	26968	20%	69.9	79%
Planning horizon: 2 weeks	81	44	34	0	0	0	36720	30762	19%	35.5	76%
Whisker 2, 8 hours	50	20	51	0	0	0	24480	20157	21%	17.5	78%
Static SPTS whisker 3, 8 hours	63	22	65	0	0	0	31200	27113	15%	24.2	79%
Static SPTS whisker 4, 8 hours	86	24	85	0	0	0	40800	36614	11%	28.7	80%
Flexible SPTS whisker 3, 8 hours	89	56	70	0	0	0	33600	27536	-	-	80%
Flexible SPTS whisker 4, 8 hours	97	59	90	0	0	0	43200	36984	-	-	80%
Whisker 2, 8+4 hours	48	18	29	82	4	35	22320	19510	4%	28.2	78%
Static SPTS whisker 3, 8+4 hours	57	16	37	115	7	46	28800	25957	11%	43.1	82%
Static SPTS whisker 4, 8+4 hours	78	16	47	160	6	64	37920	35422	7%	57.9	83%
Flexible SPTS whisker 3, 8+4 hours	74	41	36	108	13	54	30240	26270	-	-	81%
Flexible SPTS whisker 4, 8+4 hours	84	37	56	144	7	54	39840	35738	-	-	81%
Procedure duration increase: 20%	93	37	72	0	0	0	34560	29565	17%	35.3	72%

Table 5.3: Performance results of default scenario and experiments.

As the ratio becomes smaller, our model has less maneuverability to schedule procedure types. For openings hours larger than four, we might argue that after some threshold the number of hours is no bottleneck in terms of utilization performance. We do see that shorter ORDs can contribute to efficient allocation when combined with longer ORDs. The opening hour combinations show an increased estimated utilization, with the combination of three and five hours resulting in 85%. It seems that this combination caters very well to efficient scheduling options. Other options with multiple capacity types also show an increase in utilization.

When we review the influence of the probability of overtime on utilization, we may notice a classic trade-off. Decreasing the probability of overtime also decreases utilization, and vice versa. A larger probability of overtime decreases the duration of procedure types, which makes them easier to schedule. While the resulting utilization of 89% is very high, we then also have to contend with a 40% probability of overtime. Similar, reducing the risk of overtime to 20% also reduces our utilization to 72%.

A utilization increase is gained by dropping the constraint that prevents that semi-urgent schedules are scheduled in ORDs of other sub-specializations. This results in a utilization of 80%, a minor increase. It can be argued that an increased set of options gives our model more maneuverability to schedule procedure types economically. This effect is increased when relaxing the entire constraint, which allows multiple sub-specialties in the same ORD. This results in a utilization of 82%. The setback is a higher challenge in scheduling surgeons and staff to accommodate this maneuverability. We observe that a planning horizon of two weeks performs 3% worse in terms of utilization to a four week horizon. We can attribute that to the reduced number of planning options reduces when scheduling in only two weeks, and that the portfolio benefits are therefore less.

When we compare our approaches in constructing the SPTS against demand variability, we may observe small variations in utilization performance. For only eight hour ORDs, we can observe that the static SPTS is slightly outperformed by the flexible SPTS at the third whisker., and that utilization is equal or the fourth whisker. This is contrary to what one could expect, since the static approach should have larger demand for a single SPTS, and thus more scheduling options. It seems that the higher utilization is caused by the variable component of the flexible approach, that has a 81% utilization. We cannot provide a solid explanation for this behavior. For the combination of eight and four hour ORDs, we do see the effects that we would expect. The largest difference between the static and flexible approach is 2% at the fourth whisker.

The surgical procedure type duration increase has a negative effect on utilization in that it decreases to 72%. It is reasonable to assume that there is less room to maneuver surgical procedures as their durations increase, similar to the effect when decreasing the number of opening hours.

Monthly required OR capacity

Interesting are the effects of opening hours on total required capacity. Shortening the opening hours to seven decreases the required capacity, while increasing it to eight and a half hours increases the required capacity. When only operating with one number of opening hours, a seven hour ORDs shows the best results with 29400 minutes of required OR capacity per month. If Haga is committed to only performing surgery within office hours, this may actually increase the burden on OR capacity. Nine and ten hour ORDs have no impact on the required capacity, and a four hour ORD increases the required capacity. Shorter ORDs can probably not be filled economically, which leads to an increase in ORDs and hence an increase in capacity. When looking at the combinations of OR opening hours, we see the same performance increase as with utilization. The five and three hour combination again performs best, with only 28020 minutes of required capacity per month, followed by the eight and four hour combination with 28800 minutes per month.

We can spot a sizeable decrease to 27540 minutes per month when we relax the sub-specialty constraint, but this again costs us in terms of staff planning challenges, and a similar but smaller effect when we relax the constraint only for semi-urgent procedures. A reduction in required capacity to 28800 minutes per month can be observed when we increase the allowed probability of overtime. However, this improvement is deceiving since the number of overtimes will likely increase. Capacity will probably still be required, but more frequent after office hours. While a two week and four week horizon perform equally in terms of utilization, we may now notice that a two week horizon underperforms in terms of capacity requirement. We may attribute this to some extent to rounding errors in demand. Demand for a two week horizon was derived from the monthly demand, and fractional demand was rounded upwards.

We can spot something similar with how the fixed and variable SPTS approaches perform with demand fluctuations, where the fixed SPTSs outperforms our variable SPTS approaches for both opening hour options. For the combination of eight and four hour ORDs, we see that the flexible approach leads to a 5% increase in required capacity. For the option with only eight hour ORDs, the increase is 7.7% for the third whisker and 5.9% for the fourth whisker. It is reasonable to assume that with dividing demand up amongst two SPTSs, scheduling surgical procedure types will be less economic. This effect is reduced when adding four hour ORDs, which increases the number of planning options.

Integrality gap

For most experiments, the integrality gap lies in the range of 10% and 20%, with an average gap of 15%. While this is not negligible, it is still in the acceptable domain. It serves that our model is relatively complex, which is caused by the combination of selecting ORDs and assigning them in a single model. This is emphasized by the fact that the integrality gap is smaller when the sub-specialty constraint is dropped, and the solution space reduced. This feat of combining the selection and assignment of ORDs is only since recently possible due to the advance in computational strength in computer hardware.

Model runtime

A general trend which we may notice is that the runtime increases along with the solution space. When we increase the number of opening hours, procedure demand or overtime probability, the solution space increases too. This is reflected in longer running times. A similar thing can be observed when relaxing the sub-specialty constraint. While one could expect that runtime should decrease when dropping constraints, we also see that the number of scheduling options increases. We might therefore also argue that the particular constraint was not a bottleneck with regards to procedure runtime.

5.4 Summary

In this chapter, we performed experiments and measure model and solution performance.

We performed various experiments with a range of opening hours to determine which options perform best. The highest utilization of 85% is achieved by implementing both five and three hour ORDs, followed by the combination 8 and 4 hour ORDs with 82% utilization. These combinations also require the lowest amount of capacity, with respectively 28020 and 28800 minutes per month. A decent increase of 1% utilization is also shown by only using eight, nine or ten hour ORDs, but with no capacity requirement reduction. Figure 5.2 visualizes the utilization of the five and three hour ORDs against the current results.

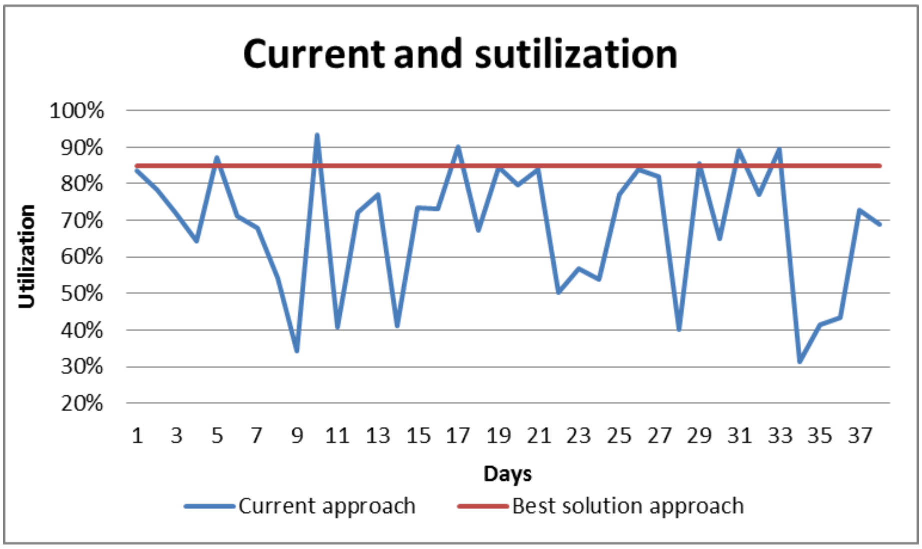

Figure 5.2: Comparison of current and best solution approach utilization (Source current utilization: SAP, data between 06-15 and 11-15).

A minimum of 43 slots is required to stabilize access time for semi-urgent patients, which results in an access time probability of 82%. Access time can be improved by reserving more slots, but the marginal benefit of each slot decreases. Adjusting the number of slots has no influence on utilization, but increase the required amount of capacity. Relaxing the constraint that only one sub-specialty may perform semi-urgent surgery in an ORDs increases utilization but also likely staff planning challenges. These effects are further increased when relaxing the constraint for all surgical procedure types. These effects however small compared the benefits of using multiple ORDs capacity types. There exists a clear trade-off between the willingness to risk overtime and utilization. Utilization is influenced by the amount of slack that is chosen by management to limit the probability of overtime.

We tested a flexible SPTS approach with a fixed and variable component in response to variable demand. The best results for both methods are shown when using eight and four hour ORDs. Then, the flexible approach results in 81% utilization for both whiskers, while the static approach scores 82% for the third, and 83% for the fourth whisker. At both whiskers, the static approach will require 5% less capacity per month than the flexible approach. However, the flexible approach offers some other advantages. For the flexible approach, staff and surgeons for the fixed component will mostly not have to be altered in advance. A relatively low number of additional ORDs is scheduled. A variable set of ORDs can then be added to the existing schedule when required. Compared to the “static” approach, the flexible approach also prevents the risk of idle time when demand is low. Finally, we determined the runtime and integrality gap of our SPTS model to be still in an acceptable range.

6 Conclusion

In this chapter, we present conclusions and recommendations based on our research and results. Section 6.1 discusses the conclusions while section 6.3 provides some discussion about our solution approach. Section 6.3 proposes some recommendations for Haga and future research.

6.1 Conclusions

We conclude that Haga will benefit from using a SPTS. The size of this benefit vary depending on which management choices are made with relation to overtime probability, semi-urgent slots, opening hours and whether a flexible or static SPTS approach is chosen.

The default scenario results in a utilization of 79%, one week access time probability of 85% and overtime probability of 30%. When we add the possibility of multiple OR capacity types, utilization may increase up to 85% and less capacity is required for the same throughput. The best results with a 30% probability of overtime can be achieved when implementing five and three hour ORDs, with a utilization of 85% and capacity requirement of 28020 minutes per month, followed by eight and four hour ORDs with a utilization of 82% and 28800 minutes per month.

Reserving capacity for semi-urgent patients will result in a reduction of last minute changes, elective patient cancellations and improvement of semi-urgent access time. Reserving 43 slots per month results a 82% probability of access within a week. Reserving extra slots results in a higher probability of one week access, but at a marginal higher cost.

We argue that Haga will also benefit from using the flexible approach towards using a SPTS. On paper, this approach is slightly outperformed by the static approach in terms of utilization and capacity requirement. However, we anticipate that the flexible approach offers practical benefits in dealing with elective demand fluctuations.

6.2 Discussion

We recall that the objective of our research is "*To develop an OR scheduling approach which manages surgical demand and duration uncertainty for elective and semi-urgent patients".* Based on our performance analysis in chapter 2, we identified twenty core causes of poor performance and demarcated our research to nine:

- Inaccurate estimates for surgical procedures due to biased estimates
- Inaccurate estimates for surgical procedures due to lack of consideration of variability
- Inaccurate estimates for surgical procedures due to conflicting definitions of what the estimate entails
- Inaccurate estimates for changeover times
- Lack of clear standards for "buffer" time
- Lack of consideration of un-arrived "semi-urgent" patients when scheduling
- Lack of a clear planning method for scheduling patients

- Poor OR admission control due to capacity allocation based on the preferences of surgeons instead of demand
- Lack of capacity balancing through periodical capacity allocation variation

Our solution approach incorporates techniques that overcome problems, which is reflected in our model performance. Results show that our model performs well compared to the current situation in terms of utilization, access time- and overtime probability. We anticipate that reserving capacity for semi-urgent patients does not only result in a performance increase in access time, but also a reduction of elective patient cancellations and last-minute changes.

However, we do note that our model performance is a theoretical approximation and susceptible to stronger performance than reality. Whether or not our model is really valid depends on the accuracy of our distribution fits for surgical procedures. In reality, online operational events such as no-show and late starts influence performance even further. Such events are occluded in our models.

We note that it is difficult to determine the optimal distribution fit for surgical procedure. We have the option of 2 parameter, 3 parameter and the method proposed by *Stepaniak et al.* [36] to describe surgical procedure duration. Often, we have to make due with a relative small number of observations. While 3-parameters may describe the procedure best, the log-normal power-sum approximation method requires us to use a 2-parameter distribution. Without, we cannot implement the portfolio effect. Furthermore, we cannot be 100% certain as to the accuracy of the approximation.

Our model selects the minimal amount of ORDs to balance a chosen whisker of demand. It will be interesting to see how demand will be influenced by the new SPTS. If surgeons are allocated more ORDs, their availability for clinical consults will lower. As a result, it is likely that the demand input will be affected by a decrease in clinical consults. This might result in over-allocation of surgical procedure types in the SPTS. While there are currently enough patients on the waiting list, we expect that it is necessary to closely monitor the waiting lists in the initial period to find a proper balance between demand and production.

6.3 Recommendations and future research

In this section, we propose recommendations and suggestions for future research. In subsection 6.3.1, we discuss recommendations based on our research. In subsection 6.3.3, we present some recommendations related to the implementation of our solution approach. In subsection 6.3.3, we propose some future research avenues.

6.3.1 General recommendations

In this subsection, we propose some general recommendations related to core problems we discussed in subsection 2.5.4 .

Performance measurement & data registration

Currently, there is a lack of OR performance measurement. Indicators such as utilization, starting time, overtime and surgical duration estimation accuracy are insufficiently defined and measured. We recommend for management to decide on a set of performance indicators to implement in a dashboard in the new hospital system which Haga recently acquired. Additionally, poor data registration makes it difficult to accurately measure these indicators. A range of timestamps is described subsection 2.1.3 is to measure these performance indicators, but most of these timestamps are rarely used. Is seems that registration of these timestamps is currently too optional of a choice for OR personnel. We recommend that these timestamps are incorporated in a more mandatory way in the OR process.

Strategic capacity dimensioning

At strategic level, the case-mix and production goals are determined in order to estimate the yearly required capacity during weekdays for specialties. Currently, the required utilized capacity is determined by the sum of the surgical procedure products of production and average surgical duration. This capacity is then increased by 15% to compensate for idle time. This increase of 15% accounts for an expected utilization of 87%, which for most specialties is optimistic at best. We recommend to set different and realistic utilization targets for each specialty, rather than one fixed utilization. Utilization targets should be determined based on the case-mix for each specialty and the willingness to risk overtime. Apart from utilization targets, an estimation should also be made for the proportion of surgical procedures that is performed during weekends. These should not be included in the general capacity estimates for weekdays.

Capacity re-allocation

Whenever specialties are not able to perform surgery at a specific date, they are required to timely contact the OR project manager to “return” that ORDs to prevent an idle OR or staff. Reasons to return an ORD could be an event which surgeons want to attend (such as a medical congress). Another reason to return an ORD could be a lack of demand, which means that an OR cannot be fully scheduled with patients. In the current organizational structure, the responsibility for timely returning ORDs lies fully with the specialties. Unfortunately, there seem to be frequent occurrences where ORDs are not timely returned which resulted in a waste of OR resources. We recommend an organizational intervention that puts ORDs utilization more exposed on the agenda at both specialties and the OR project manager. An example of such an intervention could be a monthly meeting between the OR project manager and specialty chiefs of surgery, aimed specifically at production forecasts for the next month, in which specialties can either return or request ORDs accordingly.

OR starting times

We recorded a number of late starts of the OR department in subsection 2.4.2. These are caused by a variety of reasons, of which we will discuss two.

The first reason is that both OR personnel and surgeons are accustomed to waiting for each other. Surgeons may purposefully arrive if they do not expect the OR room to be ready. In turn, this reduces the need for OR personnel to prepare the OR early, since they are accustomed to waiting on surgeons. A clear structure of responsibilities seems to be lacking at OR personnel, which also simulates late preparation of ORs. We propose the appoint team leaders for each OR room, who should ensure timely preparations and increase accountability for both OR personnel and surgeons.

The second reason for late starts is caused by the fact that all ORs start at the same time in the morning. This results in large peak burdens on both the holding and anesthesiologists. With one anesthesiologist per two ORs, the odds are likely that ORs will be waiting on an anesthesiologist in the morning periods. We therefore propose to divide the starting period for ORs over two specific times with at least 15 minutes difference.

6.3.2 Implementation recommendation

As mentioned in section 6.2, we expect that the initial period requires close monitoring to determine a proper balance between production and elective demand. To reduce the current waiting list, we advise to run the initial two months with an SPTS based on the fourth whisker and evaluate after that. After the two initial months, we advise to use the capacity re-allocation meeting recommended in subsection 6.3.1 to estimate the number of variable add-on ORDs required for the next month. The model comes with an excel tool that determines the number of add-on ORDs based the size of the waiting list.

The SPTS comes with a list of procedures that might be assigned at each specific date and OR. Available elective procedures may be scheduled on a first come, first served base. Semi-urgent shots should be kept open for semi-urgent patients. If semi-urgent slots are not filled two days in advance, we advise to schedule an elective patient in the available slot. For each sub-specialty, the added excel tool comes with a list of potential elective procedures that fit in the duration in a semi-urgent slot. The model can be scheduled with available surgical procedures for an entire month such that patients can be notified of their procedure date well in advance. The schedule can be sent to the OR department weekly, after which semi-urgent patients can be scheduled online.

6.3.3 Future research

In this subsection, we present suggestions for future research into extensions of our solution approach.

Extension to other specialties

We have presented an approach to manage demand and duration uncertainty for elective and semi-urgent patients at general surgery. It would be interesting to see if the model can be extended to incorporate multiple specialties. Other specialties might even have a more suitable case-mix for clustering surgical procedures than general surgery has. Additionally, we would then schedule in a larger number of daily ORs. This increase would make it interesting to see if we could then research if emergency patients could be incorporated inside the SPTS without the need of dedicated flex rooms, and whether this offers benefits in performance.

Outpatients and clinical consults

In our solution approach, we focused only on inpatient surgical procedures. It would also be interesting if this model could offer performance benefits if applied to outpatient procedures. similarly, it would be interesting to see if the model could be applied to clinical consults between patients and consults. Clinical consults can have similar challenges in terms of demand and duration uncertainty, and it would be interesting to see if our model would result in performance improvements.

Bed occupation

In our model, pooling of surgical procedures is rewarded with reductions in the total required amount of slack. It is unknown how this pooling might affect bed occupation at the recovery, PACU and general wards. It would be interesting to measure the effects of our model on these wards. Even more interesting would be to incorporate bed leveling in our model. This would have to be closely incorporated with guaranteeing access times for semi-urgent patients.

Accurate estimation of surgical procedure duration

It would be interesting to see if we could combine our historical observed duration with the current practice of surgeons estimating surgical duration. Rather than having a single distribution for surgical procedures, we could determine whether it is feasible to develop multiple categories of duration per procedure. A surgeon could estimate a category for each patient when requesting a surgical procedure. It would be interesting to see if the resulting procedure types are more accurate and economically efficient than the current set.

Bibliography

[1] J. M. van Oostrum, T. Parlevliet, A. P. M. Wagelmans, and G. Kazemier, "A method for clustering surgical cases to allow master surgical scheduling," *INFOR*, vol. 49, no. 4, p. 37, 2008.

[2] N. Kortbeek, M. E. Zonderland, R. J. Boucherie, N. Litvak, and E. W. Hans, "Designing Cyclic Appointment Schedules for Outpatient Clinics with Scheduled and Unscheduled Patient Arrivals," 2010.

[3] J. M. Van Oostrum, M. Van Houdenhoven, J. L. Hurink, and E. Al., "A model for generating master surgical schedules to allow cyclic scheduling in operating room departments," *Dep. Appl. Math. Fac. EEMCS*, 2006.

[4] E. Hans, G. Wullink, M. van Houdenhoven, and G. Kazemier, "Robust surgery loading," *Eur. J. Oper. Res.*, vol. 185, no. 3, pp. 1038–1050, 2008.

[5] H. Heerkens and A. van Winden, *Geen probleem, een aanpak voor alle bedrijfskundige vragen en mysteries*. Van Winden Communicatie, 2012.

[6] E. W. Hans, M. Van Houdenhoven, and P. J. H. Hulshof, "A Framework for Healthcare Planning and Control," *Handb. Healthc. Syst. Sched.*, vol. 168, pp. 303–320, 2012.

[7] N. Zorgautoriteit, "Regeling medisch specialistische zorg," pp. 1–19, 2009.

[8] M. van Houdenhoven, E. W. Hans, J. Klein, G. Wullink, and G. Kazemier, "A norm utilisation for scarce hospital resources: Evidence from operating rooms in a Dutch university hospital.," *J. Med. Syst.*, 2007.

[9] E. W. Hans and P. T. Vanberkel, "Operating theatre planning and scheduling," *Handb. Healthc. Syst. Sched.*, no. 2011, pp. 105–130, 2012.

[10] Plexus, "Een methodiek voor optimalisatie van de capaciteit van OK-complexen," no. 0332, 2012.

[11] F. Milewski, "Operating Room Utilization and Perioperative Process Flow," pp. 1–15.

[12] A. Testi and E. Tànfani, "Tactical and operational decisions for operating room planning: Efficiency and welfare implications," *Health Care Manag. Sci.*, vol. 12, no. 4, pp. 363–373, 2009.

[13] H. O. H. S. Chmitz, N. K. K. Wak, and P. J. K. Uzdrall, "The determination of surgical suite capacity and an evaluation of patient scheduling policies," *Reserche Oper.*, vol. 12, no. 1, pp. 3–14, 1978.

[14] F. Dexter, R. D. Traub, and A. Macario, "How to release allocated operating room time to increase efficiency: predicting which surgical service will have the most underutilized operating room time.," *Anesth. Analg.*, vol. 96, no. 2, pp. 507–12– table of contents, 2003.

[15] F. Dexter, A. Macario, and R. D. Traub, "Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Use of bin packing algorithms and fuzzy constraints in operating room management.," *Anesthesiology*, vol. 91, no. 5, pp. 1491–1500, 1999.

[16] J. Zhou and F. Dexter, "Method to assist in the scheduling of add-on surgical cases--upper prediction bounds for surgical case durations based on the log-normal distribution.," *Anesthesiology*, vol. 89, no. 5, pp. 1228–1232, 1998.

[17] O. M.a, T. C.b, and C. L.c, "Structural estimation of the newsvendor model: An application to reserving operating room time," *Manage. Sci.*, vol. 54, no. 1, pp. 41–55, 2008.

[18] F. Dexter, R. D. Traub, and P. Lebowitz, "Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations.," *Anesth. Analg.*, vol. 92, no. 4, pp. 943–946, 2001.

[19] B. Cardoen, E. Demeulemeester, and J. Belien, "Operating room planning and scheduling: A literature review," *Eur. J. Oper. Res.*, vol. 201, no. 3, pp. 921–932, 2010.

[20] 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," vol. 1, no. 2, pp. 129–175, 2012.

[21] B. Cardoen, E. Demeulemeester, and J. Beliën, "Operating room planning and scheduling : A literature

review," *Eur. J. Oper. Res.*, vol. 201, no. 3, pp. 921–932, 2010.

[22] M. Van Houdenhoven, J. M. van Oostrum, E. W. Hans, G. Wullink, and G. Kazemier, "Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling.," *Anesth. Analg.*, vol. 105, no. 3, pp. 707–714, 2007.

[23] H. M. Markowitz, "Portfolio Selection: Efficient Diversification of Investments." 1959.

[24] M. Lamiri, X. Xie, A. Dolgui, and F. Grimaud, "A stochastic model for operating room planning with elective and emergency demand for surgery," *Eur. J. Oper. Res.*, 2008.

[25] J. Bowers and G. Mould, "Managing uncertainty in orthopaedic trauma theatres," *Eur. J. Oper. Res.*, vol. 154, no. 3, pp. 599–608, 2004.

[26] D. Gupta, "Surgical Suites ' Operations Management," vol. 16, no. 6, pp. 689–700, 2007.

[27] J. M. Van Oostrum, M. Van Houdenhoven, J. L. Hurink, E. W. Hans, G. Wullink, and G. Kazemier, "A master surgical scheduling approach for cyclic scheduling in operating room departments," *OR Spectr.*, vol. 30, no. 2, pp. 355–374, 2008.

[28] M. E. Zonderland, R. J. Boucherie, N. Litvak, and C. L. A. M. Vleggeert-Lankamp, "Planning and scheduling of semi-urgent surgeries," *Health Care Manag. Sci.*, vol. 13, no. 3, pp. 256–267, 2010.

[29] I. Adan, J. Bekkers, N. Dellaert, J. Jeunet, and J. Vissers, "Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 290–308, 2011.

[30] N. Dellaert and J. Jeunet, "Hospital admission planning to optimize major resources utilization under uncertainty," *BETA*, vol. 319, no. July, 2010.

[31] J. M. H. Vissers, I. J. B. F. Adan, and N. P. Dellaert, "Developing a platform for comparison of hospital admission systems: An illustration," *Eur. J. Oper. Res.*, vol. 180, no. 3, pp. 1290–1301, 2007.

[32] J. Beliën and E. Demeulemeester, "Building cyclic master surgery schedules with leveled resulting bed occupancy," *Eur. J. Oper. Res.*, vol. 176, no. 2, pp. 1185–1204, 2007.

[33] E. Schmidt, M. Dada, J. Ward, and D. Adams, "Using cyclic planning to manage capacity at ALCOA," *Interfaces (Providence).*, vol. 31, no. 3, pp. 16–27, 2001.

[34] M. B. Erwin, W. H. Mark, and V. H. Johann, "Improve OR-Schedule to Reduce Number of Required Beds," 2012.

[35] D. P. Strum, J. H. May, and L. G. Vargas, "Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models.," *Anesthesiology*, vol. 92, no. 4, pp. 1160–7, 2000.

[36] de V. G. Stepaniak PS, Heij C, Mannaerts GH, de Quelerij M, "Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study.," *Anesth Analg.*, vol. 109, no. 4, 2009.

[37] F. Dexter, A. Macario, R. D. Traub, M. Hopwood, and D. A. Lubarsky, "An Operating Room Scheduling Strategy to Maximize the Use of Operating Room Block Time," *Anesth. Analg.*, vol. 89, no. 1, pp. 7–20, 1999.

[38] D. Min and Y. Yih, "An elective surgery scheduling problem considering patient priority," *Comput. Oper. Res.*, vol. 37, no. 6, pp. 1091–1099, 2010.

[39] B. Cardoen and E. Dmeulemeester, "A DECISION SUPPORT SYSTEM FOR SURGERY SEQUENCING AT UZ LEUVEN'S DAY-CARE DEPARTMENT.," *Int. J. Inf. Technol. Decis. Mak.*, vol. 10, no. 3, pp. 435–450, 2011.

[40] I. J. Van Oostrum, "Wachttijdverkorting voor spoedpatiënten door schuiven met de planning van electieve operaties Samenvatting," pp. 147–154.

[41] N. K. Kwak, P. J. Kuzdrall, and H. H. Schmitz, "THE GPSS SIMULATION OF SCHEDULING POLICIES FOR SURGICAL PATIENTS," *Manage. Sci.*, vol. 22, no. 9, pp. 982–989, 1976.

[42] B. Denton, J. Viapiano, and A. Vogl, “Optimization of surgery sequencing and scheduling decisions under uncertainty,” *Health Care Manag. Sci.*, vol. 10, no. 1, pp. 13–24, 2007.

[43] D. Reich, “Discrete Optimization: A linear programming approach for linear programs with probabilistic constraints,” *Eur. J. Oper. Res.*, vol. 230, no. 3, pp. 487–494, 2013.

[44] P. C. Gilmore and R. E. Gomory, “A linear programming approach to the cutting stock problem.” 1963.

[45] J. Bisschop and M. Roelofs, “AIMMS Modeling Guide,” 2016.

[46] C. L. Ho, “Calculating the mean and variance of power sums with two log-normal components,” *IEEE Trans. Veh. Technol.*, vol. 44, no. 4, 1995.

[47] D. Dilts, J. Khamalah, and A. N. N. Plotkin, “Using Cluster Analysis for Medical Resource Decision Making,” 1995.

[48] L. Fenton, “The sum of log-normal prob- ability distributions in scatter transmission sys- tems.,” *IRE Trans. Commun. Syst.*, vol. 8, no. 1, 1960.

[49] S. Schwartz and Y. S. Yeh, “On the distribution function and moments of power sums with log-normal compo- nents.,” *Bell Syst. Tech. J*, vol. 6, 1982.

[50] P. Pirinen, “STATISTICAL POWER SUM ANALYSIS FOR NONIDENTICALLY DISTRIBUTED CORRELATED LOGNORMAL SIGNALS,” no. 1.

Appendix A – List of figures

Appendix B – List of tables

Appendix C - Clustering

In this section, we discuss clustering as an identification method for elective patient groups and we will argument our selection of choices. Literature about clustering is very extensive, but we recommend *Dilts et al*. for a comprehensive review on available options on application of clustering in a medical oriented environment [47].

What to cluster?

We want to cluster similar surgical procedures in terms of some variable together as surgical procedure types. If surgical procedure A and B are similar, we cluster them together and pool their demand. This way, we pool the demand uncertainty of both procedures. We create clusters for each sub-specialty.

Variables

One of the key factors in clustering performance is the choice of variables. We would like to restrict our variables to those that measure our domains of interest. Variables could be urgency, average demand, surgeon and expected duration. Variables are all standardized to ensure that some variables will not dominate the clustering decision. We decided to cluster on expected duration and standard deviation to create homogenous aggregated groups which can be scheduled as a dummy surgery instead of individual surgical procedures. Expected duration is a suitable variable for obvious reasons, and using standard deviation will improve benefits from the portfolio effect.

Data preparation

We assume all surgical procedure duration to be 3 parameter log-normal distributed [35,13]. However, as also explained in Appendix B, we will approximate the power sum of 2-parameter log-normal distributed variables in a later stage of the solution approach. We therefore "transform" our 3-parameter distribution into a 3-parameter distribution. For a normal 2 parameter log-normal distribution, the lower bound of the distribution function nears zero and the x-axis. The third parameter in a 3 parameter distribution, or threshold, determines the shift from zero compared with a 2-parameter log-normal distribution. A 3-parameter log-normal distribution with the third parameter equaling zero is essentially a 2-parameter distribution. We therefore can remove the third parameter by finding the smallest observed duration in the underlying dataset of our distribution, and by subtracting that from each observation.

Clustering algorithm

Clustering algorithms can be categorized in either Hierarchical or Partitioning algorithms. Partitioning algorithms iteratively attempt to form a fixed amount of K clusters from K chosen starting points. K-mean clustering is the most common partitioning algorithm. It starts with K-chosen centroids and iteratively adds or switches data points to clusters to minimize the sum of squared distances from data point to cluster centroid. Hierarchical algorithms may work either agglomerative or divisive. Agglomerative is a bottom-up approach where each observation starts as an individual clusters, which iteratively merges into new clusters until 1 large cluster is left. Divisive is a top-down approach that starts with 1 large cluster, which iteratively partitions until resulting clusters cannot be further partitioned. An advantage of hierarchical algorithms is that a suitable number of clusters does not has to be determined prior to clustering, and can be determined through a dendogram. A disadvantage is the lack of continuous cluster optimization since each observation can only be partitioned from, or merged with a cluster once, whereas partitioning algorithms continuously try to improve their clusters. Partitioning algorithms tend to perform poorly with non-convex data distributions. We will use both agglomerative and k-means algorithms, the reasoning behind this decision will be explained later when discussing the number of clusters.

Measure of distance

An important factor for clustering performance is how to "determine" distances between observations. Euclidian distance is commonly used when using multiple variables and is our preferred method. Other methods such as Manhattan distance are mostly suitable for very specific situations which do not apply for our data-set.

Method of linkage

Another important factor for clustering performance is how to quantify dissimilarities between clusters, the so-method of linkage. We will shortly describe 5 methods of dissimilarity quantification, namely single-, complete-, average-, centroid- and Ward's linkage. In single linkage, the minimum of all the possible differences between two clusters is determined and used when updating the matrix of dissimilarities. In complete linkage, the maximum dissimilarity is used in the matrix. Since the remaining dissimilarities within the cluster are less than the level which is used to update the dissimilarity matrix, complete linkage usually leads to more homogenous clusters compared single linkage. Average linkage is a compromise between these 2 quantification methods where for each observation in the first cluster, the average dissimilarity compared to all the points in the second cluster is used to update the matrix. In centroid linkage, the distance between the centroids of the two clusters are used as dissimilarity. Ward's linkage determines the sum of squared deviations between all the points in a cluster and its centroid. It compares the change in dissimilarity when merging clusters and creating 1 new centroid, compared to the unmerged situation with 2 centroids. We use Ward's linkage since the more homogenous our clusters are in terms of standard deviation, the stronger the portfolio effect will be. Ward's linkage tends to result in very homogenous clusters compared to other methods of linkage [1].

Number of clusters

With each new cluster, the variance within a cluster increases. The number of clusters is a trade-off between having a suitable number of clusters for later use and inter-cluster variance. Agglomerative clustering will result in one large final cluster. To obtain a preferred and usable number of clusters, we have to determine where to "stop" the agglomerative process. One method is simply through visual observation. The clustering process is depicted Figure C.0.1 in a dendogram, where each linkage is graphed against new cluster variance. Also common is the elbow method, which aims on the percentage of total variance, which is ratio of inter cluster and total variance, against the number of clusters. In the elbow method, one tries to identify the number of clusters where percentage of total variance "jumps", which looks a bit like a bend (or elbow) when graphed. Many other methods of determining the "right" set of clusters exist.

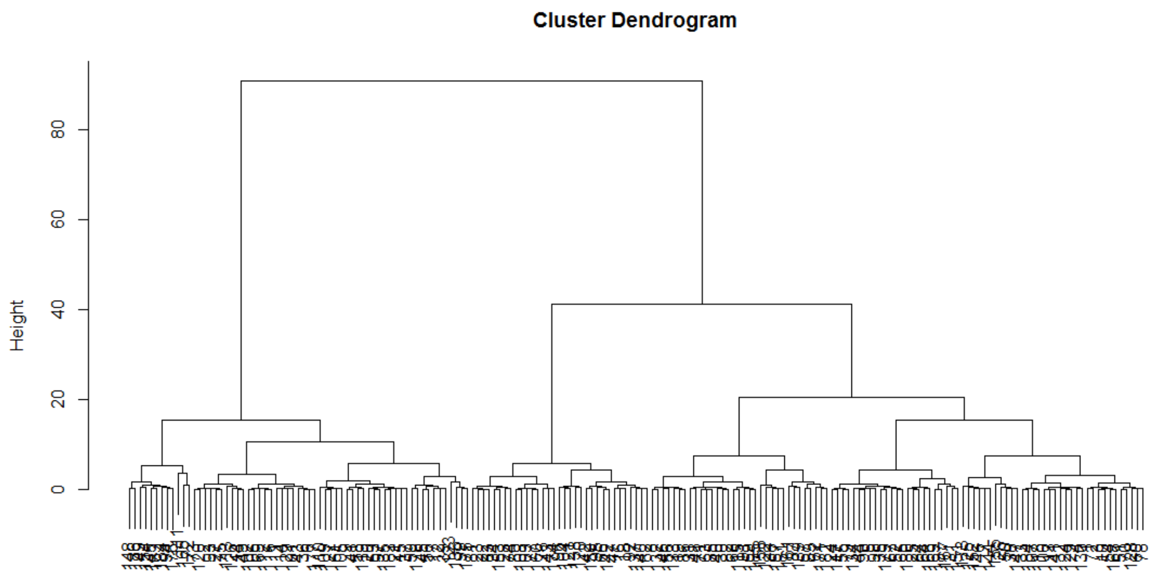

Figure C.0.1: Dendogram of agglomerative clustering process. Surgical procedures are portrayed on the x-axis, inter-cluster distance on the y-axis. Horizontal lines refer to merged clusters, vertical lines to the distance between those clusters. (Source: R)

Validation

Often, cluster results are checked through some other variable. In our case, we would like to determine if cluster variance for individual surgeons is not too large compared to cluster mean.

Software

We will use R to model our patient group identification which is a popular programming language used by statisticians and data miners. There are many suitable languages and software available, but we favor R due to its free availability and extensive features. We use Minitab to identify distributions that may describe procedure type duration.

Fitting a distribution to describe duration

We now have generated clusters that contain a collection surgical procedures. However, we steed need to determine the duration distribution for these clusters. Recall that for each surgical procedure, we subtracted the minimum observed value of duration from the complete dataset to transform from a 3 parameter to a 2 parameter lognormal distribution. We combine the complete adjusted dataset and perform a goodness-of-fit test for both a Gaussian and 2 parameter log-normal distribution. We determine the best fit by checking for the highest probability value and through visual inspection of the cumulative distribution plots. These plots are visualized in Figure C.0.2.

We can convert 3-paramteric lognormal variables to 2-paramater lognormal variables using the method described by *Stepaniak et al.* [36]. We can transform our procedure durations to a 2-parameter log-normal distribution prior to clustering. Since the 3-parameter log-normal distribution is simply the usual 2-parameter distribution with a location shift, we propose a method to drop the threshold parameter. Let X be a random variable with a 3-parameter distribution with μ, σ and γ, then:

$$Y = X - \gamma \tag{C.1}$$

Where Y is log-normal distributed with 2 parameters. For each surgical procedure, we therefore find the smallest observed value of surgical duration in the dataset and assume that to be threshold parameter γ. We subtract γ from all individual observations in the dataset and refit the adjusted dataset as a 2-parameter log-normal variable. We can also ensure that we will include γ in any later calculations with the 2-parameter log-normal variables.

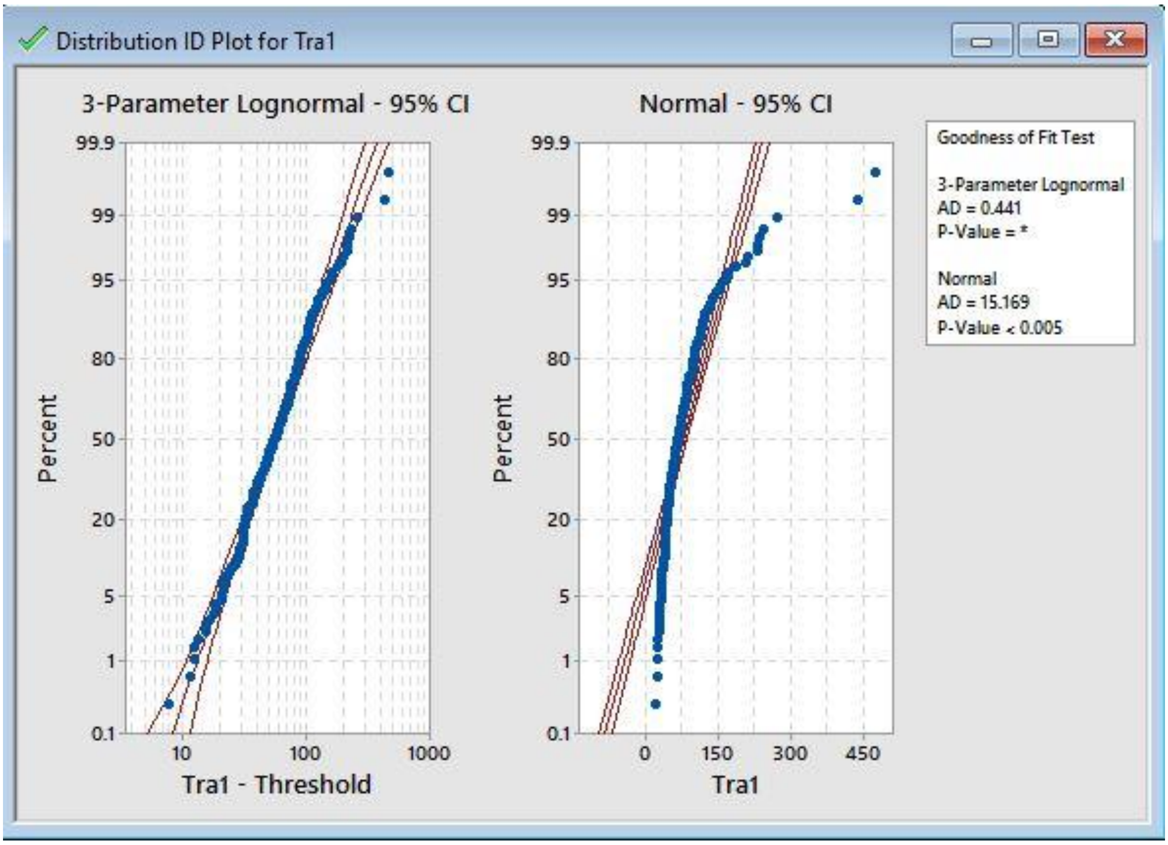

Figure C.0.2: Cumulative distribution plot of a 3-parameter log-normal distributed variable and Gaussian distributed variable, in which the variable seems to have a snuf fit for 3-parameter log-normal distribution. (Source: Minitab).

If the goodness-of-fit out for the log-normal outperforms the Guassian fit, we can determine the cluster to be log-normal distributed with a threshold comprised of the average of all observed minimum values. However, if the Gaussian fit outperforms the log-normal fit, we cannot directly determine it to be Gaussian. We then perform another goodness-of-fit on the unadjusted dataset of that cluster. If this then turns out to be log-normal again, we determine it to be 2 parameter log-normal with threshold 0. If the Gaussian has the highest probability of fit, we assume it to be Gaussian distributed.

Appendix D – Cluster results

					Whiskers				
Procedure type	Distributior	Mu	Sigma	Treshold	0	1	2	3	4
General 1	LogNormal	4.55573	0.28643	0	8	19	22	26	36
General 2	LogNormal	4.28906	0.27811	0	4	12	15	20	24
General 3	LogNormal	4.74463	0.63025	0	0	1	1	3	6
General 4	LogNormal	4.12528	0.41726	0	2	11	14	17	24
General 5	LogNormal	4.83671	0.36164	0	0	0	1	3	5
General 6	LogNormal	4.66472	0.34462	0	0	0	1	1	2
General 7	LogNormal	4.162	0.39707	0	1	4	5	6	10
Abdominal 1	LogNormal	3.83954	0.40323	0	1	6	8	11	17
Abdominal 2	Normal	166.2252	63.14174	0	1	7	9	12	19
Abdominal 3	Normal	200.7	78.83713	0	0	1	1	1	4
Abdominal 4	LogNormal	4.07025	0.2726	0	0	1	2	4	8
Abdominal 5	LogNormal	5.40591	0.24142	0	0	2	3	4	6
Abdominal 6	LogNormal	5.27873	0.4525	0	0	0	1	2	4
Abdominal 7	LogNormal	4.75474	0.37461	0	0	0	0	1	4
Abdominal 8	Normal	289.8333	100.6586	0	0	0	0	1	2
Pulmonology 1	LogNormal	4.71764	0.35721	0	0	2	2	4	9
Pulmonology 2	LogNormal	5.31996	0.5027	0	0	0	1	1	5
Pulmonology 3	LogNormal	5.31779	0.24207	0	0	1	2	4	8
Pulmonology 4	LogNormal	4.70821	0.36263	0	0	0	0	1	5
Oncology 1	LogNormal	4.57293	0.3813	0	1	10	13	14	24
Oncology 2	LogNormal	3.9583	0.2224	0	0	0	0	1	3
Oncology 3	LogNormal	4.79211	0.3557	0	0	2	3	11	18
Oncology 4	LogNormal	4.9768	0.23485	0	0	1	2	3	4
Oncology 5	LogNormal	4.80815	0.32881	0	1	9	17	23	29
Oncology 6	LogNormal	5.33483	0.24832	0	0	1	2	3	9
Oncology 7	LogNormal	4.16278	0.29117	0	0	0	0	1	3
Traumatology 1	LogNormal	4.21005	0.49814	0	0	3	5	6	15
Traumatology 2	LogNormal	3.83625	0.25478	0	0	0	1	1	3
Traumatology 3	LogNormal	4.09643	0.45312	0	0	1	2	3	5
Traumatology 4	LogNormal	3.97516	0.49342	0	0	1	1	2	3
Traumatology 5	LogNormal	4.62648	0.33806	0	0	0	1	2	5
Traumatology 6	LogNormal	4.54054	0.31791	0	0	0	1	2	4
Traumatology 7	LogNormal	3.879	0.36364	0	0	0	1	1	3
Traumatology 8	LogNormal	4.26134	0.24501	0	0	0	0	0	2
Cardiology 1	LogNormal	4.35265	0.27077	0	0	0	0	1	3
Cardiology 2	LogNormal	5.4231	0.20968	0	0	3	3	5	8
Cardiology 3	LogNormal	4.59599	0.411	0	1	3	4	6	9
Cardiology 4	LogNormal	4.23718	0.27541	1	0	3	4	6	14
Cardiology 5	LogNormal	4.23718	0.27541	2	0	1	2	3	8
Cardiology 6	LogNormal	4.96492	0.37628	3	0	2	3	4	9
Cardiology 7	LogNormal	5.57401	0.23193	4	0	0	2	2	5
Cardiology 8	LogNormal	4.40464	0.29898	5	0	0	1	2	3
Cardiology 9	LogNormal	5.63692	0.41641	6	0	0	0	0	1
Semi-urgent General	LogNormal	4.39933	0.48343	5.7101	13	13	13	13	13
Semi-urgent Abdominal	LogNormal	4.86044	0.39863	12.28005	5	5	5	5	5
Semi-urgent Cardiology	LogNormal	4.29778	0.83816	24.76976	9	9	9	9	9
Semi-urgent Traumatology	LogNormal	4.25795	0.50672	2.03805	17	17	17	17	17
Changeover	LogNormal	2.1	0.3	0	0	0	0	0	0

Table D.1: Cluster results. Surgical procedure type, distribution type, parameters and demand per whisker.

Appendix E – Portfolio effect

In this appendix, we discuss the implementation of the portfolio effect for cases with Gaussian and log-normal distributed variables.

Let the function $f(V)$ describe the joint probability distribution function of the sum of all surgical procedure types scheduled in an ORD. We can describe the probability that total schedule duration in any ORD exceeds capacity d_r as probabilistic constraint:

$$P\left[f\left(Z_{wtji}\right) \leq d_r\right] \geq \alpha \quad \forall\, w, t, j \tag{E.1}$$

However, probabilistic constraints are non-linear and therefore infeasible for a commercial linear program solver. We propose linearization of probabilistic constraints by determinisation of the stochastic variables. We will describe a different method for Gaussian and log-normal variables.

Gaussian distributed variables

We introduce a Gaussian distributed random variable for surgical procedure type duration $f(V) \sim N(\mu, \sigma^2)$ with parameters mean μ and sigma variance σ^2. If we want to reserve capacity when scheduling such a procedure, we have to reserve time up to some prediction bound. We may decide on reserving the expected duration μ and some safety margin of the variance σ^2, which we denote as slack. We introduce the safety factor α which denotes the probability of exceeding prediction bound X. We would like to determine our prediction bound such that the probability of exceeding it is α, which can be decided by management. In standard form, we can write this as:

$$\Phi\left(\frac{X-\mu}{\sigma}\right) \geq a \tag{E.2}$$

Which we can also write as:

$$\mu + \Phi^{-1}(1-\alpha) * \sqrt{\sigma^2} \leq X \tag{E.3}$$

In which $\Phi^{-1}(1-\alpha)$ refers to the corresponding value from the standard normal table. When planning multiple surgical procedures in an ORD with capacity d_r, the total required duration is determined by the sum of individual expected durations and slack. To reduce the size of summated slack, we would like to employ the portfolio effect by pooling variance of multiple scheduled procedures. The portfolio effect relies on the mathematical rule that the sum of multiple Gaussian distributed independent variables i is in turn a Gaussian joint-distributed variable [23], with mean $\mu_{sum} = \sum_{i=1}^{I} \mu_i$ and variance $\sigma_{Sum} = \sum_{i=1}^{I} \sigma_i^2$. Using this rule, we can rewrite constraint (E.3) for an instance with multiple procedures i in standard form as:

$$\sum_{i=1}^{I} (Z_{wtji} * \mu_i) + \Phi^{-1}(1-\alpha) * \sqrt{\sum_{i=1}^{I} Z_{wtji} * \sigma_i^2} \leq d_r \quad \forall\, w, t, \tag{E.4}$$

Note however that total duration is determined by a square root function, which is non-linear. We therefore propose a linear approximation of the square root function as demonstrated by *Bosch et al.* [34] and described in the AIMMS commercial solver manual by *Bisschop et al.* [45].

Bosch et al. shows that the square root function can be approximated as a set of linear functions. Each linear function approximates some part of the square root function within an interval. The intervals are in turn determined by a set $N = \{0,1,\dots,m)$ breakpoints. Within each interval $[x_n, x_{n+1}]$ the linear function is a tangent line of the square root function and described by:

$$h_n(x) = a_n + b_n * x \qquad \textit{(E.5)}$$

In which b_n is the first derivative of the square root function and a_n the result of equalizing the square root function against the linear approximation.

$$b_n = \frac{d\sqrt{t_n}}{dx} = \frac{1}{2} * \sqrt{\frac{1}{t_n}} \qquad \textit{(E.6)}$$

$$\sqrt{t_n} = a_n + b_n * t_n \quad \rightarrow \quad a_n = \frac{1}{2} * \sqrt{t_n} \qquad \textit{(E.7)}$$

And thus:

$$h_n(x) = a_n + b_n * x = \frac{1}{2} * \sqrt{t_n} + \frac{x}{2} * \sqrt{\frac{1}{t_n}} \qquad \textit{(E.8)}$$

We define y_n to be function value of the linear approximation at breakpoint n, so that $h_n(x_n) = y_n$. For any linear line, any point between two breakpoints can be described as a weighted sum of those breakpoints. We apply this principle and introduce non-negative weights p_n and the conditions $\sum_{n=1}^{N} p_n = 1$, and the condition that two adjacent breakpoint weights are non-zero. We can then rewrite constraint (E.3) as:

$$\sum_{i=1}^{I} (Z_{wtji} * \mu_i) + \sum_{n=1}^{N} (p_n * y_n) \le d_r \qquad \forall\, w, t, j \qquad \textit{(E.9)}$$

With additional constraints:

$$\sum_{n=1}^{N} (p_n * x_n) = \sum_{i=1}^{I} (Z_{wtji} * \sigma_i^2) \qquad \textit{(E.10)}$$

$$\sum_{n=1}^{N} p_n = 1 \qquad \textit{(0.11)}$$

More breakpoints results in a more accurate approximation of the square root function but also a larger computational burden. We would like to determine the minimal number of breakpoints that is required to approximate the square root function by some maximum error of difference. We define the error δ_n at breakpoint $n \in N$. Knowing that the largest error between approximation and the squared root lies at the breakpoints, we can define:

$$\delta_0 = h_1(x_0) - \sqrt{x_0} \qquad \textit{(E.12)}$$

$$\delta_n = h_n(x_n) - \sqrt{x_n} \tag{E.13}$$

$$\delta^{max} = \max(\delta_n) \tag{E.14}$$

We can minimize δ_{max} by equalizing the breakpoints errors. Mathematically, we can notate this concept through the following set of equations:

$$h_n(x_n) = h_{n+1}(x_n) \tag{E.15}$$

$$\delta_n = \delta_{n+1} \tag{E.16}$$

$$x_n < t_{n+1} < x_{n+1} \tag{E.17}$$

If the approximation errors at all breakpoints are equal, it follows that we can describe the maximum error in terms of the number of breakpoints m:

$$\delta^{max} = \delta_m = h_m(x_m) - \sqrt{x_m} \tag{E.18}$$

In which δ^{max} is determined by m and the maximum value of the interval of. If we determine on a value of x_m, we can determine the minimal number of breakpoints required to approximate within a chosen maximum error.

Log-normal distributed variables

We introduce a log-normal distributed random variable for surgical procedure type duration $f(V) = \ln(Z)$ with $Z \sim N(\mu, \sigma^2)$ and parameters mean $\bar{\mu}$ and variance $\bar{\sigma}^2$. Again, we seek a prediction bound X which we can use when scheduling a surgical procedure. For a single variable, the prediction bound is given by:

$$e^{\bar{\mu} + \Phi^{-1}(1-\alpha) * \sqrt{\sigma^2}} \leq X \tag{E.19}$$

Unfortunately, working with the (power) sum of log-normal distributed variables is far less convenient than with Gaussian variables. We cannot directly determine a prediction bound for the joint-distributed variable due to its open form. However, we can approximate the power sum of log-normal variables with the assumption that the sum of log-normal distributed variables is also log-normal distributed. Several methods to approximate this sum are described in literature, in varying levels of complexity. We will briefly compare the two most popular methods.

The Fenton-Wilkinson method (which is also used by *van Oostrum et al.)* matches the first two moments of the power sum of log-normal variables to estimate a single lognormal variable [48]. The Schwartz-Yeh method involves the exact computation of a 2 power sum log-normal joint- distributed variable, which can be extended to a larger sum of variables through a recursive approach [49]. A quantitative analysis by *Pirinen et al.* finds that the Fenton-Wilkinson approach performs well in approximating the tail region (CDF(0.9-0.999)) of the approximated log-normal variable but poor in the body region (CDF(0-0.9)) [50]. The Schwartz-Yeh method performs well in the body-region, but less so in the tail region. Tail regions are of interest in areas such as signal processing, but our upper bounds for surgical procedure duration all lie in the body region. We therefore prefer the Schwartz-yeh method to approximate power sum log-normal distributions for surgical duration.

We propose an adaptation of the Schwartz-Yeh method presented by *Ho et al.* [46]. *Ho et al.* circumvents the computational burden of exact calculation and by using trapezoidal rule to approximate complicated integrals. *Pirinen et al.* finds strong performance in the body-region of the approximated sum of log-normal variables with this method. While summing log-normal variables with non-identical parameters is possible, *Pirinen et al.* reports for that accurate approximation is limited to sums of log-normal variables with identical parameters. The method can also be extend to correlating variables, though we assume all our variables to be independent. The expressions for calculating mean and standard deviation are presented in constraints (E.20-E.28):

$$m_z = m_y + G_1 \tag{E.20}$$

$$\sigma_z^2 = \sigma_{y_1}^2 - G_1^2 - 2\sigma_{y_1}^2 * (I_2 + I_0) + G_2 \tag{E.21}$$

$$G_1 = E[\ln(1 + e^w)] = A_0 + I_1 \tag{E.22}$$

$$G_2 = E[ln^2(1 + e^w)] = I_3 + 2 * I_4 + 2 * \sigma_w^2 * I_0 + m_w * A_0 \tag{E.23}$$

$$G_3 = E[(w - m_w)] * \ln(1 + e^w) = \sigma_w^2 * (I_2 + I_0) \tag{E.24}$$

$$I_4 = \sigma_w^2 * [f_w(0) * \ln(2) - I_5] + m_w * I_6 \tag{E.25}$$

$$A_0 = m_w * I_0 + \frac{\sigma_w}{\sqrt{2\pi}} * e^{-m_w^2 / 2*\sigma_w^2} \tag{E.26}$$

$$I_i = \int_0^1 (h_i(v) * v^{-1})\, dv \tag{E.27}$$

Formulas for recursive approximation with i variables.

$h_i(v) =$	$i = 0$	$(\sqrt{2\pi})^{-1} * e^{\frac{-(\ln(v) + m_w/\sigma_w)^2}{2}}$
	$i = 1$	$[f_w(\ln(v)) + f_w(-\ln(\mathrm{v})] * \ln(1 + \mathrm{v})$
	$i = 2$	$[f_w(\ln(v)) + f_w(-\ln(\mathrm{v})] * (1 + \mathrm{v}^{-1})^{-1}$
	$i = 3$	$[f_w(\ln(v)) + f_w(-\ln(\mathrm{v})] * \ln^2(1 + \mathrm{v})$
	$i = 4$	$-f_w(-\ln(v)) * \ln(v) * \ln(1 + v)$
	$i = 5$	$f_w(-\ln(v)) * (1 + \mathrm{v}^{-1})^{-1}$
	$i = 6$	$f_w(-\ln(v)) * \ln(1 + v)$

(E.28)

$$f_w(w) = \frac{e^{-\frac{(w - m_w)^2}{2*\sigma_w^2}}}{\sqrt{2 * \pi * \sigma_w^2}}$$

To incorporate the portfolio effect in the power sum of log-normal variables, we implement an approximation presented by *van Oostrum et al.* [1]. We sum the prediction bounds X_i of individual surgical procedure types and subtract time that corresponds with the portfolio effect. Portfolio time is determined by the difference between the prediction bound of individual variables and the power sum variable. Since we cannot implement a stochastic variable in a linear constraint, we use the deterministic values mean and variance, and implement a function with explicit joint log-normal distributed prediction bounds. Mathematically, we can then describe constraint (E.19) as:

$$\sum_{i=1}^{I} Z_{wtji} * \mu_i - g(Z_{iwtj}) \quad \forall\, w, t, j, i \tag{E.29}$$

Combining Gaussian and log-normal portfolio effects

We can incorporate the linear approximations of Gaussian and log-normal distributed surgical procedure types fairly easily in a single ORD. Log-normal variables can be scheduled directly. If the remaining capacity of an ORD is scheduled with Gaussian distributed variables, we can assume that the this capacity is indeed described by a linearized approximation of the root function. This means first that slack and the portfolio effect are determined separately for log-normal and Gaussian functions. Second, for log-normal distributed duration surgical procedure types we also calculate the portfolio effect per individual type. This choice is based on reported inaccurate approximation of non-identical joint distributions by *Pirinen et al.*