Optimization of scheduling in endoscopy

Evaluation of various scheduling rules and schedules using Discrete Event Simulation

Bob Overbeek

Universiteit Twente
de ondernemende universiteit

Optimization of scheduling in endoscopy

Evaluation of various scheduling rules and schedules using Discrete Event Simulation

Bob Overbeek
May 2008

Master's Thesis
Industrial Engineering and Management

University of Twente
Department of Operational Methods for Production and Logistics

Supervisors:

Dr. Ir. E.W. Hans, University of Twente, Enschede, The Netherlands
MASc. P.T. VanBerkel, University of Twente, Enschede, The Netherlands
Ir. P.E. Joustra, Academic Medical Center, Amsterdam, The Netherlands

Preface

You are about to read my Master's thesis about the evaluation of various scheduling rules and schedules in an endoscopy department of a learning hospital. I did my research in the Academic Medical Center (AMC) located in Amsterdam. The department in which I worked is called Quality and Process Innovation (KPI in Dutch) and it handles various optimization projects all over the AMC. I was working on the optimization project in the AMC endoscopy department. This study aims to reduce the access times for this endoscopy department. I developed a simulation model to measure the impact of various planning rules and to test several schedules. To everyone who is interested in my report: enjoy reading it!

There are a lot of people to thank. First of all, my supervisors: Erwin Hans, Peter VanBerkel and Paul Joustra. Thanks for all critical but always constructive comments on my work and report. I would like to thank Veerle Struben and Henk Greuter, the project managers of the endoscopy project, for all their useful help and cooperation during my research. Jesse de Wit, thank you too for numerous useful discussions about this subject since you started your thesis in December on a complementary project. Furthermore, I thanks all direct colleges in the AMC who made my time in the AMC to what is has become. I also thank all doctors and staff from the endoscopy department for all their help. An finally, last but not least, I thank my parents who have always supported me, not only during my Master's thesis. Thank you all!

Bob Overbeek
Enschede, May 2008

Summary

Introduction

The endoscopy department of the Academic Medical Center (AMC) in Amsterdam has a long access time in the specialism Hepato-Gastroenterology (HGE). This access time is up to twelve weeks, which is unacceptable according to the AMC. This case study aims to improve the scheduling of patients in this department with the objective to reduce access time of the endoscopy department.

Context

The patients of this department need various consultations, with two possible urgency labels: urgent (treated within one day) and elective. The department uses an individual-block, variable interval planning rule. This means the length of the appointment depends on the consultation type and the schedule consists of blocks dedicated to certain treatments. Patients are scheduled directly after their request for appointment.

The department has seven treatment rooms, with various restrictions. It employs ten physicians and six residents. Of the various consultation types, three have an access time of more than twelve weeks.

Method

To analyse various scheduling rules and schedules we develop a simulation model of the department. Is uses probability distributions for the arrivals of patients and the closures of rooms that occur occasionally. The model includes the following experimental factors:

- Number of occasional closures
- Usage of time slots
- Rejection policy
- Capacity of the schedule

The main outputs of the model are:

- access time
- double bookings (urgent patients that are booked alongside another patient, resulting in overtime)

Based on various scenario evaluations, we propose a set of scheduling rules and a schedule. All these scenarios use the expected arrival of appointment requests caused by a newly proposed and adopted rejection policy and the extra demand expected by a planned medical examination.

Results and conclusions

The simulation model shows that access times can be reduced without the application of extra doctors or equipment. Through a better balanced schedule 95% of all patients are treated within the standard times that we defined. This schedule reduces the amount of urgency blocks by 50%. To prevent the amount of overtime to increase, we implement week urgency.

This week urgency patients account for 20% of the urgent patients, but can be treated within one week. This reduces the overtime work by 3.3%. The number of occasional closures has a major impact on access times. The elimination of these closures can reduce the access time by up to 74.0%. By creating a schedule that consists of back up doctors for every shift, we can reduce these occasional closures to 2% of all shifts.

The schedule and scheduling rules that we propose are a 134.5 hour schedule (instead of 135.75 currently), dedicated block usage, blocks for week urgency as well as day urgency and 2% occasional closures. This results in 95% of all patients are treated within their standard times and a reduction of the overtime work by 35.8% to 126.85 hours per year.

Samenvatting

Aanleiding

De endoscopie afdeling van het Academisch Medisch Centrum (AMC) te Amsterdam kampt met lange toegangstijden bij het specialisme Maag-Darm-Lever (MDL). Deze toegangstijden lopen op tot twaalf weken, hetgeen voor het AMC als onacceptabel wordt bestempeld. Deze studie focust op de verbetering van de patiëntenplanning op deze afdeling en heeft als doel de toegangstijden te verkorten.

Context

De patiënten die behandeld worden op deze afdeling zijn van verschillende behandelingstypes. Er bestaan twee verschillende urgentie indicaties: urgent (binnen één dag geholpen) of electief. Het percentage urgente patiënten bedraagt 19%. De afdeling gebruikt een individueel-blok, variabel interval planningsregel. Dit betekent dat de lengte van de afspraak afhangt van het behandeltype en het rooster bestaat uit blokken die toegewezen zijn aan bepaalde behandelingen. Patiënten worden direct ingepland nadat ze om een afspraak gevraagd hebben.

De afdeling bestaat uit zeven behandelkamers, die elk bepaalde restricties hebben. Er werken tien stafartsen en zes arts-assistenten. Van de verschillende behandelingstypes zijn er drie die een toegangstijd hebben van meer dan twaalf weken.

Methode

Om verschillende planningsregels en roosters te analyseren ontwikkelen we een simulatiemodel van de afdeling. Dit model gebruikt kansverdelingen voor de aankomsten van patiënten en de sluitingen van kamers die incidenteel voorkomen. Het model bezit de volgende experimentele factoren:

- Aantal incidentele sluitingen
- Gebruik van blokken
- Vraagsturing (aantal aanvragen voor afspraken)
- Capaciteit van het rooster

De belangrijkste uitkomsten van het model zijn:

- Toegangstijden
- Dubbelplanningen (urgente patiënten die naast een andere patient worden gepland, wat resulteert in overwerk)

We evalueren verschillende scenario's en doen een voorstel voor een aantal planningsregels en een rooster. Alle scenario's gebruiken de verwachte vraag die anders is dan de huidige. Dit komt door een onlangs voorgestelde en ingevoerde vraagsturing en een extra vraag die verwacht wordt vanwege een bevolkingsonderzoek.

Resultaten en conclusies

The simulatiemodel laat zien dat de toegangstijden verkort kunnen worden zonder de inzet van extra artsen of apparatuur. Door middel van een beter gebalanceerd rooster kan 95% van alle patiënten behandeld worden binnen de gestelde normen. Dit rooster bestaat uit 50% minder spoedblokken dan het originele rooster. Om te voorkomen dat de hoeveelheid overwerk toeneemt, introduceren we weekspoed. Twintig procent van alle urgente patiënten hoeft niet op dezelfde dag geholpen te worden, maar kan ook binnen een week behandeld worden. De introductie van weekspoed vermindert de hoeveelheid overwerk met 3,3%. Het aantal incidentele sluitingen heeft veel impact op de toegangstijden. Het elimineren van deze sluitingen kan de toegangstijden tot 74,0% verkorten. Door het maken van een rooster waarbij voor elke shift een back-up arts is gepland, kunnen we de incidentele sluitingen terugbrengen tot 2% van alle shifts.

Het rooster en de planningsregels die we voorstellen zijn een rooster met een capaciteit van 134,5 uur (in plaats van 135,75 in het huidige rooster), gebruik van specifieke blokken, blokken voor zowel weekspoed als dagspoed en 2% incidentele sluitingen. Dit resulteert in het feit dat 95% van alle patiënten behandelt wordt binnen de gestelde normen en dat de hoeveelheid overwerk afneemt met 35,8% naar 126,85 uur per jaar.

Table of contents

1. Introduction

The Endoscopy department of the Academic Medical Center (AMC) has a long access time in the specialism Hepato-Gastroenterology (HGE). This access time is up to twelve weeks, which is unacceptable according to the AMC. This case study focuses on the scheduling of patients in this department with the objective to reduce access time of the endoscopy department.

The remainder of this chapter is organized as follows. In Section 1.1 we give a short context description after which we define the main problem in Section 1.2. Given this problem definition we state an objective for this study in Section 1.3. Based on this objective, we formulate some research questions in Section 1.4.

1.1 Context description

Academic Medical Center
The AMC one out of two university hospitals situated in the city of Amsterdam in the Netherlands. It is connected to the University of Amsterdam for research and educational purposes. With more than 6000 FTEs and more than 1000 beds it is one of the largest hospitals in the Netherlands, as well as one of the most prominent medical centers.

Team Innovation and Process management
The Team Innovation and Process management (TIP) is a department within the AMC that consists of about twenty people of different backgrounds, like logistics, medical, service and quality management. TIP comes up with improvements for departments of the AMC and advices those departments and the Board of Directors. The objective of TIP is to support departments and help them improve the logistics and service in healthcare processes. The activities are on a project basis and are in close cooperation with the concerning department. TIP handles projects that result in direct and visible improvements in healthcare as well as projects that create conditions for optimization of healthcare processes. Since the beginning of 2008 TIP has become an official department known as Quality and Process Innovation (KPI in Dutch).

Endoscopy
Endoscopy is a minimally invasive medical procedure used to look inside certain organs by inserting a tube into the body. The word endoscopy literally means "looking inside". An endoscope (Figure 1) is used to execute this procedure. The endoscope consists of a camera at the end of the tube, as well as a light and a lens. Besides that there exists an extra channel in the tube to insert medical instruments, for example to take a sample of the organ tissue. The part of the endoscopy department that is studied in this case study is specialized in stomach, colon and liver procedures.

Figure 1.1: An Endoscope

The studied part of endoscopy department in the AMC handles about 9300 patients each year. The department consists of seven different treatment rooms and has 10 doctors and 6 residents (*in Dutch: arts-assistenten*). The total number of FTEs of nurses is 19,4.

Access time
The access time of a department is the time between the patient's request for an appointment and the moment the patient arrives at the clinic. The access time can be measured in various ways. Section 4.2.2 discusses the method we use. Access time is not the same as waiting time, since waiting time is defined herein as the time a patient spends in the waiting room.

1.2 Problem definition

The main problem is:

The endoscopy-HGE department of the AMC has long access times to such a degree that patients have to wait up to twelve weeks and the department has to forward patients to other hospitals.

1.3 Objective

The objective of this study is to reduce the access times for the endoscopy-HGE department of the AMC by improving the department's scheduling of patients.

1.4 Research Questions

Given the objective of this study, we formulate the following research questions:

1. *What are the current processes, planning and control and performances in the endoscopy-HGE department?*
   To get insight in the current situation, Chapter 2 describes the current processes in the endoscopy department. This includes patient flows within the department, as well as the way in which patients are planned in this situation. In addition we present data that represent the current access times, production data and production requests.

2. *What are the relevant theoretical methods about modeling in healthcare?*
   In Chapter 3 we present the relevant modeling methods that are available for this study.

3. *Which modeling method and input parameters should be used to build a model?*
   We build a simulation model which we discuss in Chapter 4. In this chapter we discuss the method, define the input parameters and the assumptions made, as well as the output data and the validation process.

4. *What scenarios are appropriate to evaluate and what are the quantitative outcomes?*
   In Chapter 5 we use the simulation model to perform measurements on various scenarios and the current situation. We determine which alternative scenarios to evaluate, compare all outcomes of simulation and finally come to a conclusion on which scenario is the best for the endoscopy department.

5. *Which changes or investments are necessary to implement the best scenario(s)?*
   It is possible that for implementation of some of the scenarios investments are necessary, like investments in new techniques or ICT systems. In the remainder of Chapter 5 we compare the investment costs of the scenarios with each other and with the quantitative outcomes.

In Chapter 6 we present the final conclusions and recommendations for the endoscopy-HGE department.

2. Context Analysis

This chapter describes the current situation in the endoscopy-HGE department of the AMC. Section 2.1 discusses the processes in this department, while Section 2.2 presents the performances in the current situation.

2.1 Processes in the current situation

Every year, more than 9000 requests for appointments arrive at the endoscopy-HGE department. A triage is performed on their urgency and type, patients are scheduled by the counter personnel. After a certain time on the waiting list patients arrive at the AMC for their treatment. They check in at the counter and enter the waiting room. A nurse accompanies them to the treatment room where after some preparations the treatment is performed. In case of sedation the patient is brought to the recovery room after the endoscopy, if not he leaves the department. The remainder of this section gives a more detailed description of these processes. Appendix A presents the flow charts of the complete process.

Consultation types
Patients that contact the department can be new patients or review patients. In scheduling there is no difference between these types since both have the same amount of scheduled consultation time. There are various ways patients are referred to the department. Patients can be referred by (1) medical specialists from the AMC, (2) general practitioners or (3) other hospitals. Presently, requests by general practitioners are forwarded to other hospitals because of the long access times, in order to cope with the demand. The main groups of endoscopies executed at the department are Endoscopic Retrograde CholangioPancreatography (ERCP), which uses the support of X-ray technology, Endoscopic UltraSound (EUS), which uses the support of ultrasound technology and more general endoscopies like Colonoscopy (Colo), Sigmoidoscopy (Sigmo), Oesophago-Gastro-Duodenoscopy (OGD) and Proctoscopy (Procto).

Urgency
The endoscopy department distinguishes three types of urgency:
- Emergency patients
- Urgent patients
- Elective patients

As the endoscopy department does not state clear definitions of these types of urgency, we define them as follows: Emergency patients are patients that have to be treated within 24 hours, while we define urgent and elective as the same: patients that need an appointment somewhere in the future. The planning system is not able to register the number of emergency and urgent patients, so it is difficult to determine that exact amount. We assume that patients that had an access time of zero days (in 2006) are emergency patients. Some patients are treated directly on the day they arrive, but are registered in the planning system afterwards. In the data this gives a negative access time number, since the appointment date is earlier than the registration date. We assume that patients of which the appointment date is one day earlier than the contact date are also emergency patients for which the registration was done after they had been treated. This results in a 19% percentage emergency rate in 2006. Doctors indicate that this number is a good approximation, as they consider an emergency rate of about 20%. Emergency patients are the only patients that can be double-booked, which means they can be planned in a time slot where already another patient is planned, eventually resulting in some overtime in that treatment room.

The distribution of the emergencies over the weekdays (Figure 2.1) shows that most of the emergencies are on Fridays. This can be explained by the fact that a lot of patients are that urgent that they need to be treated before the weekend.

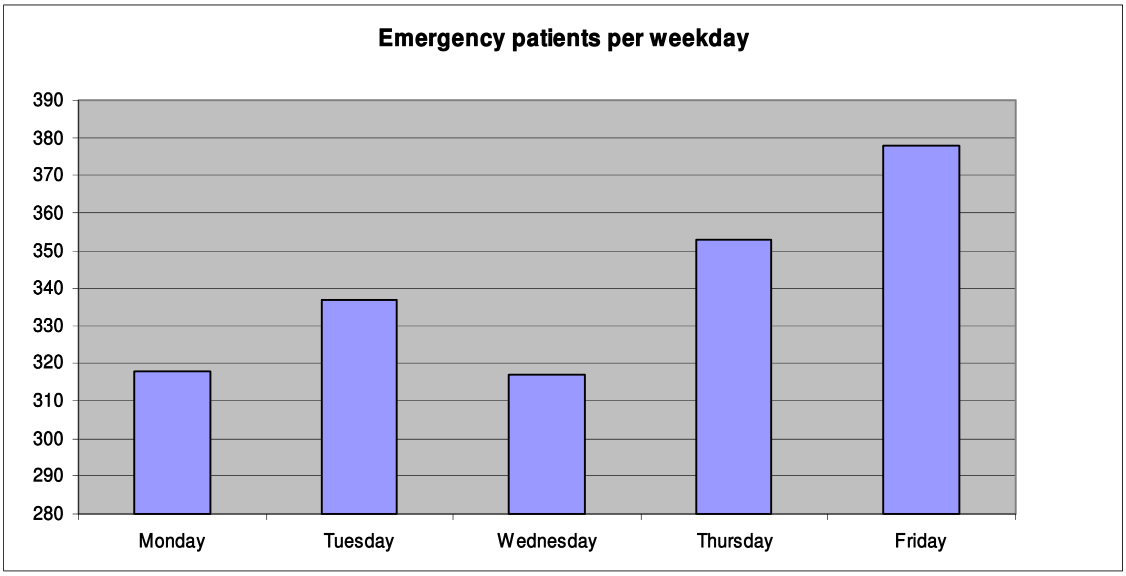


Figure 2.1 Emergency arrivals per weekday (n = 1703, 2006)

Appointment scheduling
Scheduling of patients is done at the reception desk. An incoming request for an appointment this is planned at once. There are various time slots dedicated to specific consultations. There are special time slots for emergency patients. A time slot is an appointment interval that has a pre-defined length, which is used for a certain treatment. All time slots remain dedicated until one week in advance, after which these slots are unblocked so they can be used for other treatments. This is not the case for emergency blocks. Appendix B shows the current planning schedule. This schedule is based on the availability of doctors. Appointments are planned with an individual-block and variable interval rule *(Cayirli & Veral, 2003)*. This means there are never two patients planned on the same appointment time in the same room, and the length of appointment depends on the type of consultation (Figure 2.2).

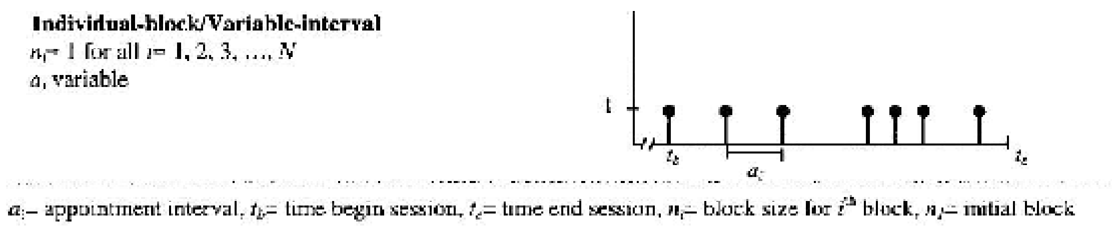


Figure 2.2 Individual-block/Variable-interval planning rule *(Cayirli & Veral, 2003)*

2.2 Performances in the current situation

2.2.1 Performance Indicators

Internal waiting time
The internal waiting time is the time between the appointment time and the time when a patient enters the treatment room. The waiting time due to the earliness of the patient is not considered since this is not influenced by planning. The waiting times are measured for a small group of patients. The mean waiting time of these patients is 8 minutes, but has a high standard deviation. Table 2.1 shows the waiting times per diagnosis type.

Diagnosis	Average waiting time in minutes	95% Confidence Interval
COLO (n=10)	0:02	0:04
ERCP (n=6)	0:08	0:12
EUS (n=13)	0:12	0:12
OESDIL[1] (n=4)	0:15	0:10
HGE (n=17)	0:10	0:07
SIGMO (n=12)	0:04	0:05

Table 2.1 Internal waiting times in 2006

Access times
As earlier indicated the access times for HGE-endoscopies are very high. Figure 2.3 shows the development of the average access times in 2006 (excluding outliers (n = 377)). These access times are still increasing except for ERCP and EUS, as the access times of Colo, HGE and Sigmo are more than doubled in 2007 (Table 2.2). There is a big increase in access time for some treatments in March. We cannot completely explain this, but the upcoming holidays like Easter and others may had some influence on this, since closures can have major impact on access time.

[1] Oesdil is the dilatation of the oesophagus. This treatment is a subgroup of HGE.

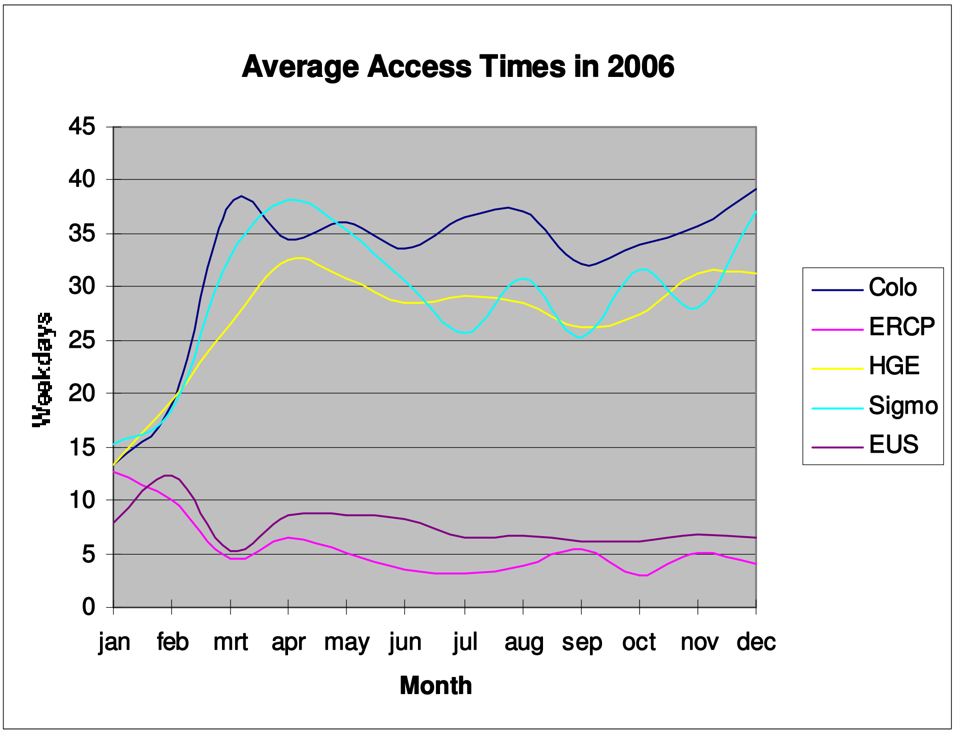


Figure 2.3 Average Access Times in 2006 (n = 7381)

Diagnosis	COLO	ERCP	HGE	Sigmo	EUS
Access Time (days)	77	5	89	89	12

Table 2.2 Access times in days (measured on 22-08-2007)

Consultation times

As the endoscopy department's planning method has variable intervals, the various types of consultations have different planned durations. For every consultation type a fixed duration is considered. Table 2.3 shows the planned length of the consultations as well as the measured length. The planned lengths turn out to be good estimates for the real lengths of the consultations, but leave little slack. However, between two consultations there is some change time: the room has to be cleaned and a new patient should be brought to the room. This time is presented in the fifth column. If we add this to the measured time we clearly see that the planned time is too short.

Consultation times	Planned time	Average Measured time	95% CI	Slack (Planned – average time)	Change time	95% CI
ERCP (n = 28)	1:00	0:58	0:07	0:02	0:14	0:03
EUS (n = 11)	0:45	0:46	0:13	-0:01	0:09	0:07
Colo (n = 21)	1:00	0:56	0:07	0:04	0:21	0:08
Oesdil (n = 2)	0:30	0:40	0:03	-0:10	0:29	0:02
HGE 15 [2] (n = 13)	0:15	0:13	0:01	0:02	0:07	0:05
HGE 30 [3] (n = 14)	0:30	0:25	0:06	0:05	0:13	0:08

Table 2.3 Consultation times in minutes in 2006

2.2.2 Capacity

Daily shifts

This study focuses on the consultations done during normal working days, since during the weekends and evenings only emergency patients are treated. A working day consists of two shifts. We define a shift as a morning or afternoon during which one or more parallel sessions take place. In most rooms the length of the morning shift is 3.5 hours (8:30 AM to 12:15 PM) and the length of the afternoon shift is 2.5 hours (1:15 PM to 4:00 PM). In both shifts a break of 15 minutes is scheduled.

Treatment rooms

The department consist of seven rooms available for consultations. A few rooms are allocated to specific treatments: one room for ERCPs, one for EUS, one for colonoscopy and one for proctoscopy. The three remaining rooms are used for different treatments with one of these rooms indicated as an emergency room. Not all rooms have the same equipment. There is only one room with ultrasound equipment, which is necessary for EUS, and two rooms with X-ray technology, which is necessary for ERCPs.

Besides the treatment rooms there are two waiting rooms and a recovery room. Not all patients have to go to the recovery room, this depends on the need of sedation. The recovery room officially has room for ten beds, though in practice there are placed up to eighteen beds. The recovery room is also used as waiting room for clinical patients and as a pre-assessment room. For pre-assessments there is also another room available.

Personnel

The department employs 10 physicians and 10 residents, of which six are specialized in stomach/colon. In every treatment room one resident or physician works and one or two nurses. On top of that there are three supervisors every shift, which are always physicians. One supervisor is for the ERCP room, one for the EUS room and one for the rest of the rooms. If a physician is planned in either the ERCP or EUS room, a supervisor is not necessary. During every shift one resident is on duty, which means he has to answer the phone and step in as necessary.

[2] HGE 15 is the group of general Hepato-Gastroenterology treatments that is planned for 15 minutes

[3] HGE 30 is the group of general Hepato-Gastroenterology treatments that is planned for 30 minutes

Scopes
For each treatment a new, clean endoscope is used. The total number of scopes available is 54. After each treatment the scope is cleaned in a special washing machine in the disinfection room of the department. Each machine can clean two scopes at a time and a single washing takes 33 minutes. The cleaning is done by specialized personnel that works in this room.

Production and demand
The total demand in 2006 was 9334 treatments, which includes the no-shows. The production that was actually done was 9113. Figure 2.4 shows the distribution of this demand and production over the year.

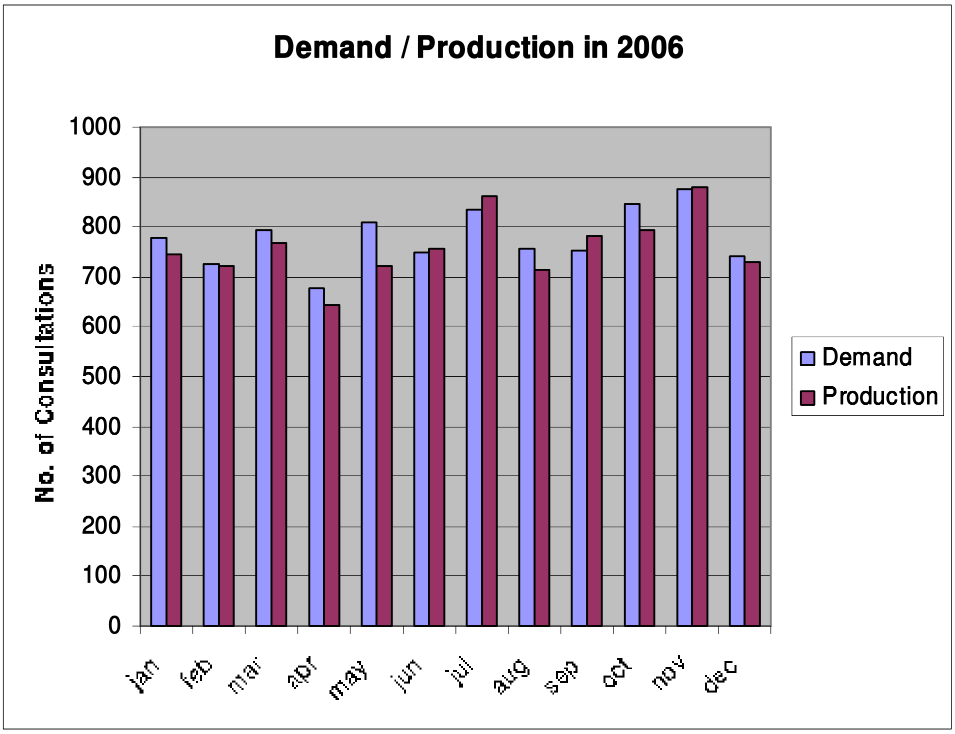


Figure 2.4 Demand and production in 2006 (n = 9334 / 9113)

When we multiply the number of performed treatments by the planned duration of each diagnosis the result is a production time of 5538 hours. The capacity of the current schedule (Appendix B) is 6775 hours, assuming opening of 50 weeks a year. This schedule is corrected by subtracting holidays. Besides the holidays there are seven training days a year, these are always on Wednesdays and three diagnosis rooms are closed then. The number of occasional closures is 117 days, divided over the 7 rooms. After these subtractions the capacity is 5820 hours, resulting in an utilisation rate of 95%. However, if we do not subtract these closures, the utilization rate is only 82%. The real utilisation of the department is somewhat lower because of the production done in overtime. No-shows and cancellation of appointments have a negative impact on the utilization rate.

No-shows
A no-show is a patient that does not show up at his appointment. The average percentage of no-shows over all treatments is 2.11% for this department. The percentage varies somewhat over the various treatments, but is always under 4% (Figure 2.5). Other studies show no-show

probabilities that range from 5 to 30 percent (Cayirli & Veral, 2003). This makes clear that the no-show rate in the endoscopy-HGE department is relatively low.

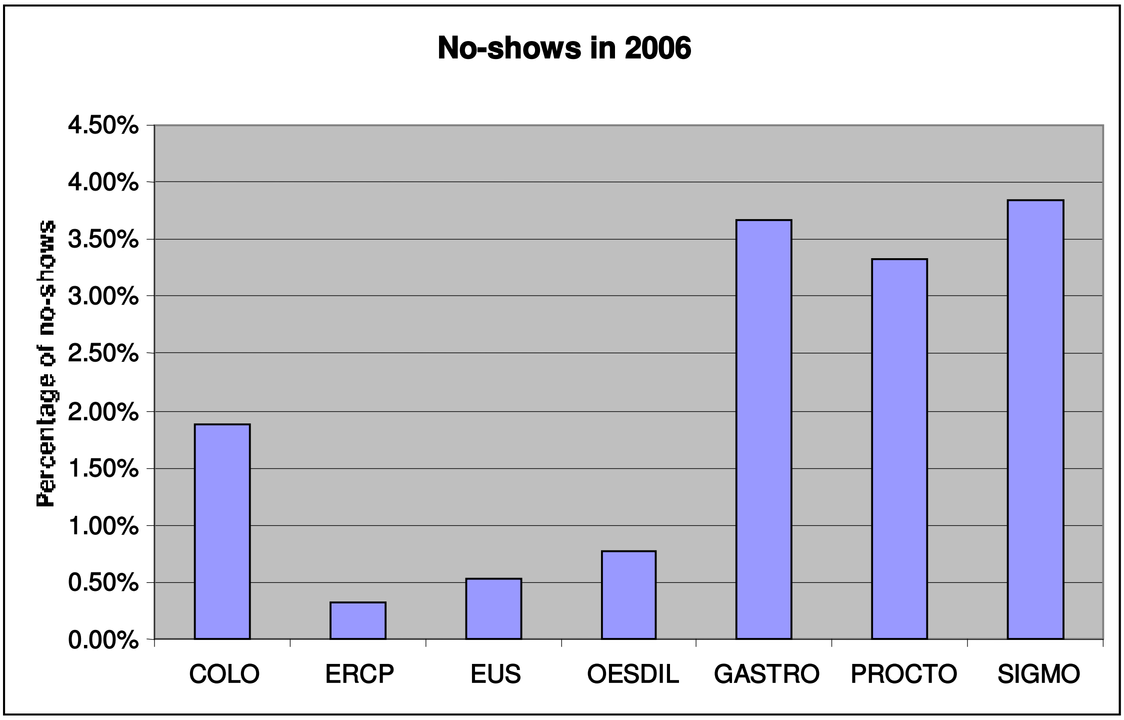


Figure 2.5 Percentage of no-shows per consultation type in 2006

Cancellations and modifications

Besides no-shows, medical departments also face cancellations or modifications of appointments. In contrast to no-shows, which are not announced in advance, a modification is always announced in advance by for example a phone-call. The appointment changes and the original time slot becomes available for other appointments. The amount of modifications is 8% of the total number of appointments and modifications are announced on average 20 days prior to the appointment time. The number of cancellations is not registered yet, the appointment is just removed from the planning.

3. Modeling of the system

3.1 Introduction

There are several methods to study a system. Figure 3.1 presents these methods (Law, 2007, pp. 4).

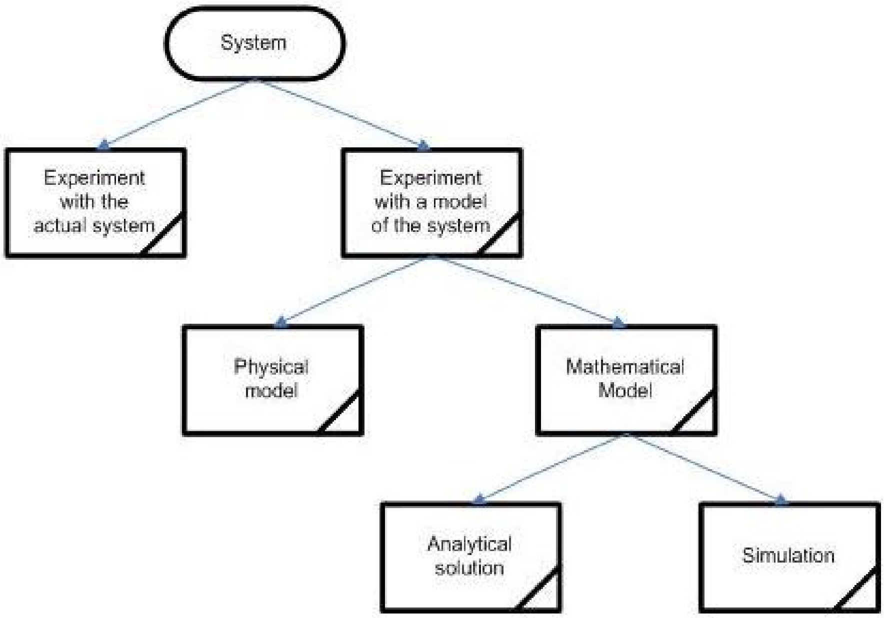


Figure 3.1 Ways to study a system (Law, 2007, pp.4)

The first decision to make is whether to experiment with the actual system or with a model of the system. We choose to analyse our system with a model of the system. This has many advantages, which we mention below:

- *Cost*: the costs of developing a computer simulation model are much less than implementing changes in the real system

- *Time*: with a simulation model we can evaluate certain interventions in compressed time, which means we can simulate several years within a few hours.

- *Disruptions in physical system*: implementing changes in the real system causes a lot of disruptions. Simulation prevents the need to implement more changes than necessary.

- *Control of experimental conditions*: In simulation it is easy to change the experimental factors, while in the real system this is much more difficult.

Then again there are two options: build a physical model or build a mathematical model. Physical models are not the type of models that are usually of interest in operations research and systems analysis (Law, 2007, pp. 4), which concerns our case too. Therefore we decide to build a mathematical model of the system.

Now there are two options left: the analytical solution and simulation. In this chapter we study the possibilities of an analytical solution with the use of queuing theory. We choose queuing theory as we intend to calculate waiting times influenced by some stochastic processes. Queuing theory matches this the best. Section 3.2 gives a short introduction to queuing theory and presents some interesting queuing models. Readers that have some background in queuing theory can skip this section. Section 3.3 tries to fit a queuing model to our actual system while Section 3.4 presents the conclusion of our theoretical survey.

3.2 An introduction to queuing theory

Queuing theory provides methods to calculate waiting times for a process with infinite duration. This is only possible for processes where the demand does not exceed the capacity.
A queuing model consists of three main components: The arrival process (A), the service process (B) and the number of servers (c). The general notation we use for queuing models is A/B/c, which is the notation of Kendall. The arrival process is in many cases a Poisson process. We denote this as M (Markov). Other possibilities of arrival processes are a deterministic arrival process (D) or an arrival process with a general distribution (GI). There exist three different service processes, which have the following characteristics: (1) Exponential service times (denoted as M), (2) Deterministic service times (D) and (3) Service times can follow any distribution (G).

Many variations exist in queuing theory. We do not intend to do an exhaustive survey of queuing theory, but we mention five groups of queuing models that are potentially interesting for our case study: (1) Standard queuing models, (2) Finite source models, (3) Batching systems, (4) Priority Queuing and (5) Polling systems.

3.2.1 Standard queuing models

The most basic queuing model is an M/M/1 queue. Due to the memoryless (Markov) arrival and service processes it is an easy model to make calculations. When one or more of these processes are not Markov, it becomes more complex to do calculations, depending on what distributions are used.

3.2.2 Finite source models

In many applications of queuing theory the assumption is made that there exists an infinite buffer for waiting jobs. However, in practice we can think of some systems which don't have an infinite buffer. We call these systems finite source models. We distinguish two types of finite source models: Models that draw arrivals from a finite pool of jobs and models that have a limited buffer for waiting jobs. The so called machine repair model is an example of a system that draws arrivals from a finite pool of jobs. (Winston, 2004, pp. 1099)

3.2.3 Batching systems

In a batching system jobs are serviced in batches. Jobs are collected in a batch after which the complete batch is serviced. This means that the waiting time of a job consists of two parts: (1) the waiting time to form the batch and (2) the queuing time of the batch when the batch is formed. A batch always consists of a predefined number of jobs. Some authors that have contributed to this subject are Chao (1998) and Economou (2001).

3.2.4 Priority Queuing

A priority queuing system exists of various waiting lines that contain jobs of a certain priority. Jobs of the highest priority are served first. A distinction can be made between systems with (preemptive) and without interruption (non-preemptive) (Takács, 1964). A system with interruptions stops servicing the current job if a job with a higher priority arrives, while a system without interruptions first finishes the current job and then starts with the job with the highest priority. One of the uses of a priority system is to model machine breakdowns. A breakdown then is a job with the highest priority, which postpones all other jobs with lower priorities. This type of models are also known as vacation models. For further reading on this topic we refer to Cobham (1954), Holley (1954), White & Christie (1958), Thiruvengadam (1963), D'Antone (2000) and Alfa (2002).

3.2.5 Polling systems

A typical polling system consists of a number of queues, attended by a single server in a fixed order (Van Vuuren & Winands, 2007). Figure 3.2 shows a graphical representation of such a system. There exist two traditional service disciplines for polling systems: exhaustive policies and gated policies. An exhaustive policy means a queue must be empty before the server starts servicing another queue, while in case of a gated policy only the customers that are in the queue when the server arrives are serviced. A more sophisticated service discipline is the *k*-limited policy (Ozawa, 1998; Van Vuuren & Winands, 2007). This policy implies that a server works until it services *k* customers or the queue is empty. This can be illustrated by a bus stop. Customers arrive at the bus stop from where a bus with a certain capacity transports them to the service location. Other interesting literature on this topic is written by Borst *et al.* (1993) and their references.

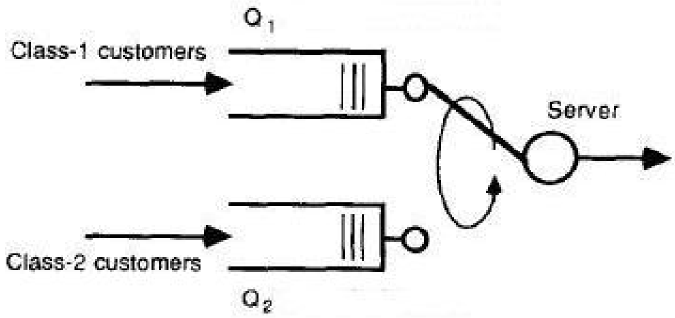


Figure 3.2 Graphical representation of a 2-queue polling system (Ozawa, 1990)

Besides the models we describe above, there is a lot of other interesting literature available about applications of queuing theory. One other interesting article can be that of Ivaneshkin (2007), which describes a system with a variable number of servers.

3.3 Queuing theory applied to the actual system

3.3.1 Basic queuing models

The system we study consists of various types of treatments. We decide to model each treatment as a separate queue, to simplify the problem. The most basic queue in our case is the ERCP queue, as the number of time slots available for ERCPs is five for every working day, except for Wednesday, when the treatment room is closed. We use deterministic service times as the number of patients that can be treated every day is fixed. We can model this queue as an M/D/1 queue or as an M/D/5 queue. Intuitively we model this queue as an M/D/5 queue, as the first few jobs in an M/D/1 queue have extra waiting time. These jobs have to wait for each other, while in an M/D/5 system the first few jobs are serviced right away, which resembles the real system the most. Initially we assume no closures at all. Every time slot available is modeled as one server, while we consider one day as a time unit. A job (= patient) takes one time unit to be served. Franx (2001) gives a solution for the M/D/c waiting time distribution. When we compare the queuing model with a simulation model (and an M/D/1 queue with adapted service times) for a similar situation they show comparable outcomes for most utilization rates (Table 3.1 and Figure 3.3).

Utilization rate	M/D/1	M/D/5	Simulation**
0.708*	0.242	0.14	0.14
0.8	0.4	0.29	0.27
0.9	0.9	0.78	0.78
0.95	1.9	1.77	1.77
0.99	9.9	9.77	7.92

Table 3.1 (Mean) Access times in days

* Utilization rate of the real system

** 30 simulation runs of 41400 patient arrivals (13800 warm-up)

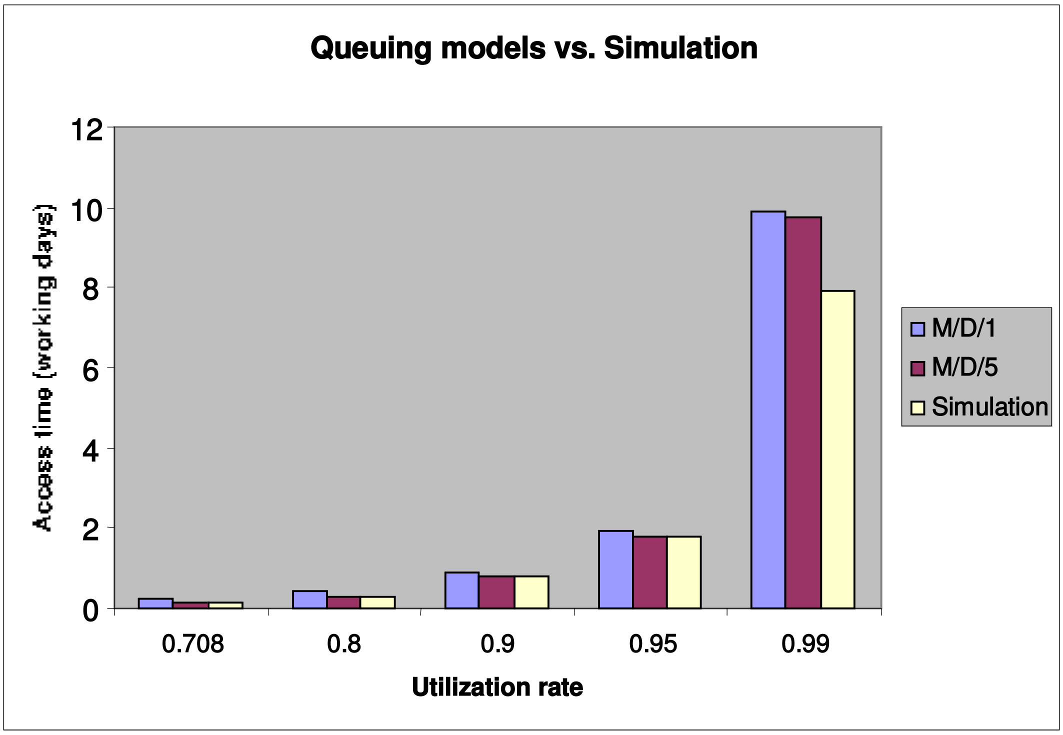


Figure 3.3 (Mean) Access time in days for various utilization rates

As becomes clear, until a utilization rate of 0.95, an M/D/5 queuing model is a realistic reproduction of the (simulated) reality in this case. An M/D/1 model, which we also plotted, turns out to be a good approximation of both other models. However, the closures on Wednesdays, holidays and occasional closures we do not take into account. A standard queuing model is thus not an applicable model to analyze our actual system.

3.3.2 Finite source models

The machine repair model, an example of finite source models, does not model machine breakdowns or vacations in a normal queue as the name could presume. So a finite source model is not applicable to analyze our system because of this incapability and the fact that the source of job arrivals is virtually infinite in the system we study.

3.3.3 Batching models

Batching models could possibly model an M/D/1 queue in such a way that always 5 jobs are served at once. This could simplify the calculations compared to an M/D/5 queue, which we defined in Section 3.3.1. However, a batching model waits until a certain number of patients is collected in a batch and then serve them all at once. This is not equivalent to the actual system, which does not wait until there are enough patients to fill a complete day. This implies that batching models are not applicable for analysis in our case study.

3.3.4 Priority queuing

As we mentioned in Section 3.2.4, priority queuing can be used to model server vacations and breakdowns. To make sure that when the department is closed all servers are down, we model the system as an M/D/1 queue in this case, instead of an M/D/5 queue. This should not be a major problem, as we saw in Section 3.3.1 an M/D/1 queue is a good approximation for an M/D/5 queue, especially for higher utilization rates. We model a vacation as a job with priority 1 and service time 1, while a normal job has priority 2 and service time 1/5. The time unit still is one day. The total number of closures consists of 52 Wednesdays, seven holidays and a number of occasional closures, which is variable. The service is non-preemptive. The formulas to calculate waiting times for priority models with deterministic service times are (D'Antone, 2000):

$$W_i = \frac{W_0}{(1-\sigma_i)(1-\sigma_{i-1})}$$

with

$$W_0 = 1/2\sum_{i=1}^{N} \lambda_i s_i^2$$

In this equation N is the number of priority classes, λ_i represents the arrival rates of jobs with priority i, s_i represents the service processes $1/\mu_i$ of jobs with priority i and $\sigma_i = \sum_{k=1}^{i} \rho_k$ is the utilization of the server by customers of priority 1-i. We are interested in W_2. We can change the number of occasional closures and see what the impact of it is. Again we compare these theoretical outcomes with results from a simulation model that represents this situation. Table 3.2 and Figure 3.4 show the results.

Number of (occasional) closures (days per year)	Without closures	6	8	**10 ***	12	14	16
Priority queuing model (1)	0.242	6.21	7.84	**10.41**	15.07	26.14	85.41
Sim. with fixed Wednesdays (2)	0.14	4.03	4.91	**6.46**	8.72	13.62	34.46
Sim. with variable placed Wednesdays (3)	0.14	6.41	7.89	**9.90**	13.53	23.27	51.03
Sim. with variable placed and variable number of Wednesdays (4)	0.14	6.39	8.14	**10.26**	13.86	25.44	57.44
Utilization rate (ρ_2)	0.708	0.944	0.953	**0.963**	0.974	0.984	0.995

Table 3.2 Average access times in days for priority model and simulations

* Number of closures in the real system

The length of the simulation runs is 72 years with 15 years warm-up time, the number of runs is 10

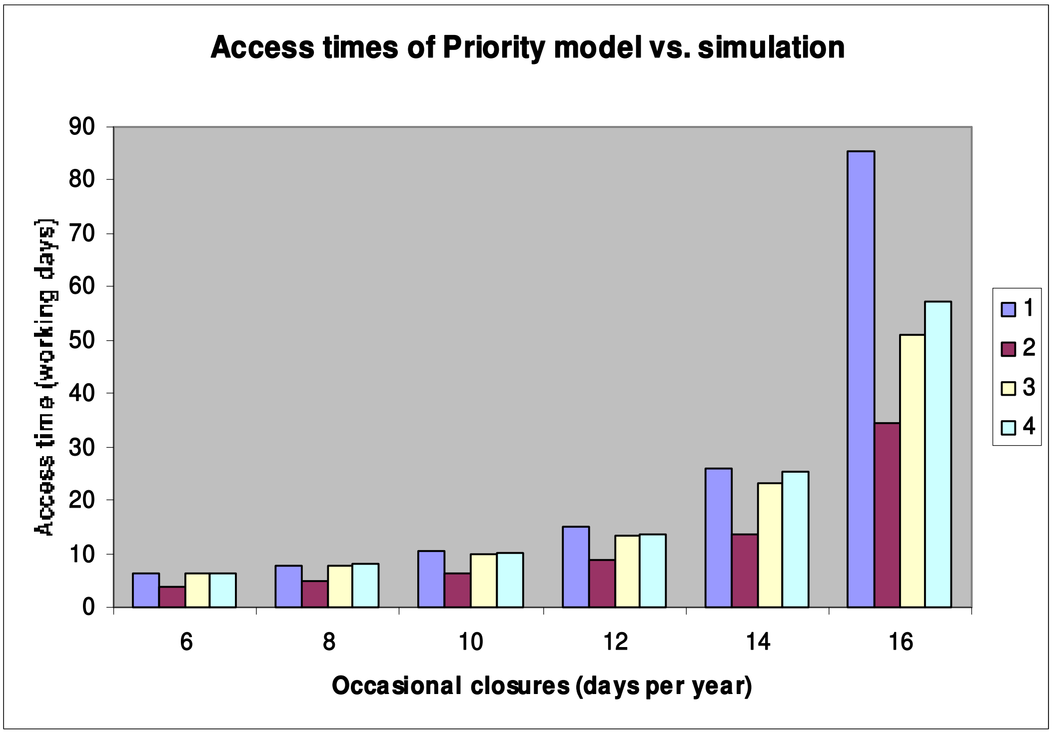


Figure 3.4 Average access times in days for priority model and simulations

The real system is closed every Wednesday. This involves that this number of closures is fixed as well as evenly distributed over the time. As randomness is the major cause of waiting times in queuing systems (Vastola, 1996), this results in lower access times than the theoretical queue shows (2). However, when we distribute the closures randomly over time, but still keep the number fixed, the access times are already higher due to the increase of randomness (3). When we finally also take the total number of closures as a Poisson distributed variable, which again increases randomness, the simulation shows output values that approach the theoretical values (4).

We can conclude from these outcomes that this model has some shortcomings for modelling the actual system, such as the possibility to fix the number of closures over time and their distribution. This difference between the queuing model and reality turns out to have a major impact on access times.

3.3.5 Polling systems

To get closer to reality we can model the system as a k-limited policy model. We use a 2-queue polling system in which one queue uses an exhaustive policy (arrivals of closures) and the other queue a k-limited policy (patient arrivals). In this case $k = 5$, which means that the server collects a maximum of 5 jobs every time it visits the patient queue. After that it visits the "closures"-queue until that queue is empty, after which it starts all over again. This system resembles the priority queuing system, except for the k-limited policy. This means all patients scheduled on a day are served before the closure is "served". As we concluded for the priority queue, that model was not suitable for modeling the real system due to the randomness of the closures. This k-limited polling model has the same shortcomings, so it will not be a realistic reflection of the reality either. Therefore we do not calculate the waiting times for this system.

Another reason for that is that hardly any exact results for this type of model has been obtained to this very day, but also their derivations give little hope for extensions to more realistic systems (Van Vuuren & Winands, 2007).

Since we are not able to satisfactory model the waiting line for the ERCP queue, which is the most simplified one, we assume that modeling more complex queues with queuing theory gives even more complications. We could model these queues as a k-limited polling system with k as a random variable, but we do not see this as a useful expansion to our study due to the reasons we mentioned earlier.

3.4 Conclusion

As we discovered by studying queuing theory it is very hard to model the actual system with queuing models. The system has many complications that cannot be modeled with the use of existing theory. We were able to estimate the access times for some simple models, but as soon as the system gets more complicated queuing theory does not seem applicable any more. Therefore we decide to use the other mathematical method for studying the real system, which is discrete event simulation.

4. Experiment Design

This chapter describes the experimental design of our study. Section 4.1 formulates the problem formally, to clarify the problem in headlines, while Section 4.2 describes the problem approach. This chapter also discusses the simulation model we build, in which Section 4.3 describes the model itself while Section 4.4 discusses the verification and validation of the model.

4.1 Formal problem definition

To get a better insight into the general problem, we write it down in a formal problem definition.

In the system studied, a doctor sees patients during shifts ($s \in S$). A shift has a duration d_s and takes place in treatment room $r \in R$. Treatment rooms r have some restrictions on the types of patients that can be treated. Each patient p is of type t. All patients of type t have a fixed average scheduling time of l_t. Within shifts, specific time intervals are dedicated to one or more patient types t. We call these time intervals blocks b_{st}. In each block a number of n_{tb} patients is scheduled. When a block is not completely filled with planned patients there is idle time I_{st}. The difference between the patient's contact date c_p and the patient's appointment date a_p we define as the access time of the patient AT_p.

The aim of our model is to minimize the access times AT_p for all patients.

4.2 Problem approach

4.2.1 Experimental factors and responses

This section distinguishes two types of parameters: (experimental) factors and responses (Law, 2007) in which experimental factors are input parameters or structural assumptions while responses are output performance measures. Factors can be either quantitative, which assume numerical values, or qualitative, which represent structural assumptions.

As experimental factors we define the following:

- *Capacity of the schedule:* the capacity of the current schedule is fixed on 135.75 hours a week. This capacity can be changed in order to analyse scenarios. This is a quantitative factor.

- *Number of (occasional) closures:* Occasional closures occur when doctors or nurses are sick, personnel requests a day off or doctors that go to conferences. By making clear that these occasional closures influence the outcomes of the system a lot, the number of these closures will become more controllable. There are also several training days during the year when part of the department is closed. This is a quantitative experimental factor.

- *Usage of time slots:* Time slots can be dedicated, which means they are dedicated to certain treatments in advance, or not dedicated. The schedule can exist of for example the current schedule, an optimized dedicated schedule or a schedule with no dedicated slots at all. Usage of time slots is a qualitative factor.

- *Rejection policy:* The arrival of requests for appointments is an uncontrollable factor, but the number of accepted requests can be influenced. For example, patients from outside Amsterdam can be forwarded to other regional hospitals. This is a quantitative factor.

The scenarios we evaluate are combinations of these experimental factors, which we define as configurations. Besides the experimental factors there exist some responses from the system. These responses are:

- *Access time:* We calculate the access times by subtracting the date of referral from the date of appointment. We do this for all patient types so we get access times for every patient category. The fact that not every patient requests the first available time slot causes some noise in those data. The next subsection discusses how we measure access time.

- *Utilization rate:* The department's utilization rate we calculate by the availability of each room after a run divided by the initial availability. In that way we can determine the utilization rate for each separate treatment room.

- *Double bookings:* The simulation model registers the number of double-bookings, which occur due to too little time for urgent treatments. Those double-bookings result in overtime work.

4.2.2 Measurement of access time

There are various ways in which we can measure access time. These methods can be split up in two groups: prospective and retrospective methods. The CBO (2004) defines access time as the time until the third available appointment slot. The reason that not the time until the first appointment slot is measured, is that there exist many fluctuations in this period due to cancellations of appointments. This is a prospective way of measuring, which means the access time can be measured in present time. A retrospective way of measuring access time is to analyze a dataset of a certain period and measure what the actual access time of all patients has been. A disadvantage of this retrospective method is that it is difficult to filter out the patients' preferences, for example when patients request an appointment to take place in three weeks. The advantage of taking the third available appointment slot is that these preferences do not influence this access time measurement. However, the reason why it is specifically the *third* available slot is quite subjective: why not take the second or the fifth slot? To reduce fluctuations we prefer to take the average of a reasonable amount of historical data.

Since the simulation model generates data for a longer time interval, we decide to measure the access time retrospective. The historical dataset we have is of the same format, so we are better able to compare the former situation to new scenarios when we measure access time retrospective. However, some patients have a preferred delay since they only want an appointment in a few weeks. Most of these patients are review patients. To reduce noise we leave out patients that have a requested delay of more than a five days. We keep this calibrated delay during the first few days since this is not only influenced by the patient's preferences but also by the preparations that still have to be done.

4.3 Model description

4.3.1 Discrete event simulation

So far we substantiated why simulation is the most applicable method in this case study. However, within simulation we define various types of simulation models:

Static vs. Dynamic simulation models: a static model is a representation of a system in which time plays no role, while dynamic models represent systems that evolve over time.

Deterministic vs. stochastic simulation models: Deterministic models do not contain any probabilistic or random variables, so once the input variables are set, the output is also determined. Stochastic models however are models that consist of at least one random variable or probability function, so the output is not determined in advance.

Continuous vs. discrete simulation models: a continuous model consists of variables that change continuously, while a discrete model includes variables that change only at certain points in time.

Given these definitions, our system resembles a model with dynamic, stochastic and discrete features. Such a model is called a discrete-event simulation model. We construct our model using MedModel version 7.0.

4.3.2 Scope of the model

The scope of the model refers to what is included in the model. In our case we are mainly interested in the access times of patients. The access time of a patient is already known after the appointment is scheduled. Since the process after scheduling the patient does not influence the access time, our model only includes the scheduling process, from the moment of invoice until the patient gets an appointment time. This corresponds with the planning processes in the flowchart in Figure 4.1. For a complete flowchart see Appendix A.

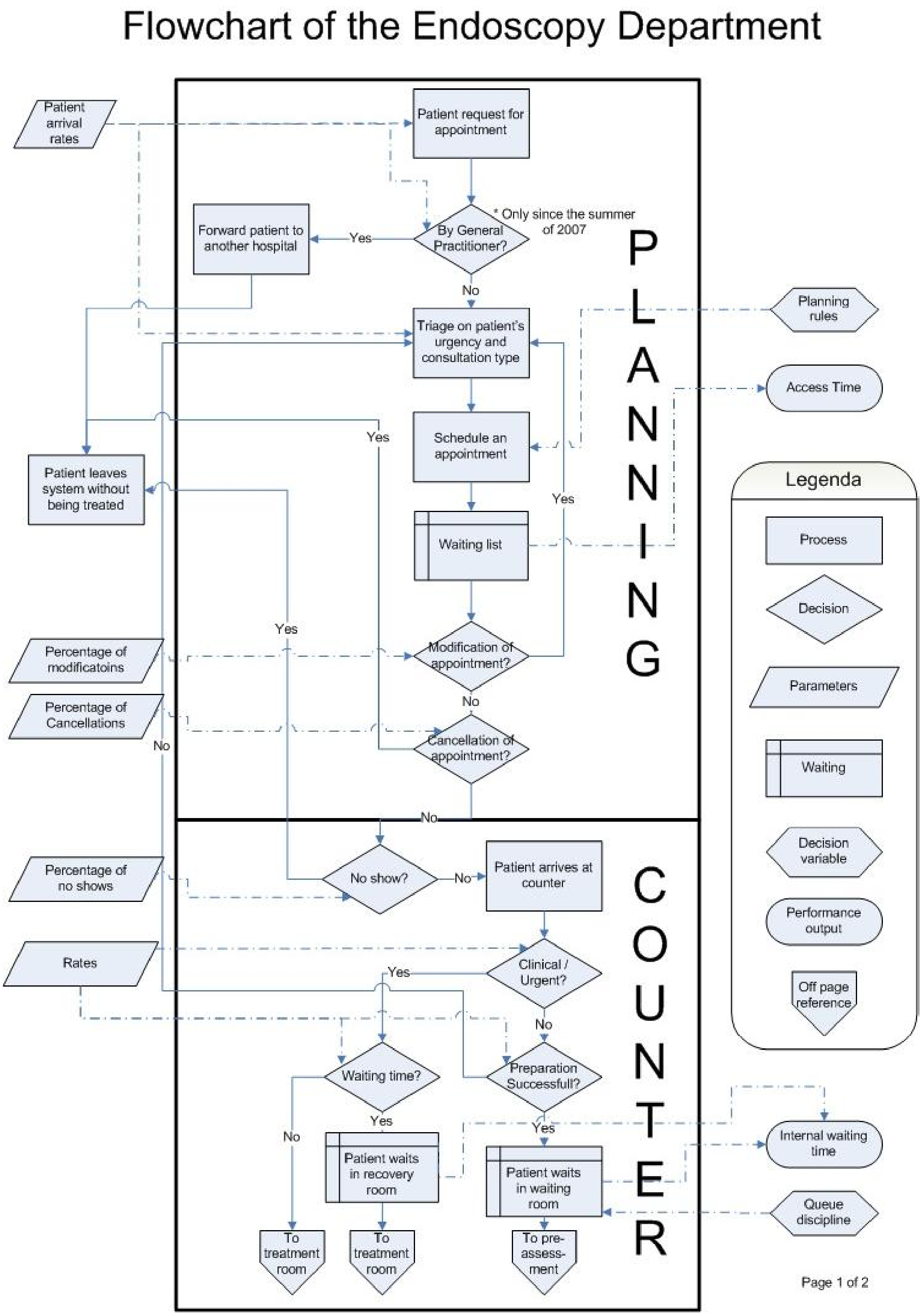


Figure 4.1 Part of the flowchart of the endoscopy department (remainder in Appendix A)

4.3.3 Level of detail

Due to the complexity of the system, it is impossible to model an exact representation of the real system. In order to keep the model manageable, we have to make some assumptions. These assumptions simplify the model and influence the level of detail. This section describes the assumptions we make.

- The number of appointment requests is on average every day the same for elective patients. For these patients we assume no weekly or seasonal patterns. This reflects the reality.

- The number of appointment requests for all urgent patients follows a weekly pattern as shown in Figure 2.1. Again we assume no seasonal fluctuations. This reflects the reality.

- Patients do not always take the first available appointment slot. This is on the one hand caused by the preparations that are not finished yet, and on the other hand a preferred delay of the appointment by the patients. We calibrate this delay of appointments with the use of our simulation model.

- Patients do not have to be seen by a specific doctor, except for the research patients that have specific blocks. We assume that all doctors can do every treatment. This does not completely reflect reality, but we do not consider it as a major shortcoming since it is already partly considered by the research blocks as well as the delay some arriving patients get.

- When an urgent patient cannot be treated within a day this patient can be double booked. All urgent patients should be treated within the day of referral. When there is no empty time slot on that day and the patient is double booked, this results in overtime work.

- Patients are scheduled conform a First-Come-First-Served rule. The only exception on this rule is the delay a patient has due to his preferred day of treatment.

- Patients may change their scheduled appointments prior to their appointment day. We assume the moment of these modifications follows a uniform distribution over the patient's access time.

- New and follow-up patients are equal with respect to consultation time and appointment type. We assume that since all those patients need the same treatment, this does not influence consultation time.

4.3.4 Input Data

We construct a simulation model. This section describes the input parameters we use. The input parameters we use are listed below:

Demand

As the number of requests for appointments we use the demand data from 2006. This is split up in data for different consultation types. In the model we implement these as Poisson arrival processes with the average arrival value per day from 2006. The arrival rate of urgent treatments we implement as a mean per weekday.

Schedule
The master schedule we initially use is the schedule from 2006. This schedule defines various dedicated blocks for certain treatments and for urgent treatments. The horizon of this schedule is one week, which is repeated over and over again by the simulation model.

Number of closures
Each year treatment rooms are closed for several days. These closures include holidays, training days and occasional closures. There are ten holidays each year, of which some of them are during the weekends. On top of that seven training days are randomly drawn each year, always on a Wednesday. Then, some occasional closures occur, which are also randomly drawn over the year. We take the number of occasional closures in 2006 as input for the model.

Modifications and no-shows
Patients change their appointment time or do not show up at all. These modifications of appointments and no-shows we implement in our model. The numbers of these modifications and no-shows are obtained from data from 2006.

Method of scheduling
The whole model is based on specific scheduling rules. The rules we implement are obtained from reality by interviewing people that perform the scheduling.

Consultation lengths
The planned durations for all consultation types are fixed. These durations are used as input for the simulation model. Since we consider access time instead of internal waiting time of the patient we do not study the real consultation durations as these do not have impact on the access time.

4.3.5 Common random numbers

Common random numbers is a variance reduction method that reduces variance between scenarios. The idea is to compare alternative scenarios under similar experimental conditions. The method uses the same random numbers stream for every scenario. We use these common random numbers for the number of occasional closures, since this turns out to have major impact on the access time due to the theory in Chapter 3. Besides that we use CRN for the arrival processes of patients.

To use CRN it the various scenarios should be synchronized. This means that a specific random number used for a specific purpose one configuration should be used for exactly the same purpose in all other configurations (Law 2007, pp. 582). To make sure this is the case we generate a common arrivals array that we use for all scenarios with the same arrival rates.

4.3.6 Patient creation routine

In our model one source creates all patients. Every patient type has it's own arrival process, which is for all types a Poisson process. At the arrival, patients get attributes that determine their characteristics. We use the following attributes:

- Patient number
- Patient type
- Service duration
- Urgency
- Modification of appointment
- No-show

Every patient has it's own unique patient number, which makes it possible to recognise individual patients. Besides that every patient is of a pre-defined type, with a corresponding service duration. The urgency attribute indicates if the patient needs urgent treatment (1) or elective treatment (0). Finally the attributes for modification and no-show indicate whether or not a patient is going to change the appointment time and on what day, and if the patient is going to show up (0) or not (1).

4.3.7 Scheduling routine

The scheduling routine exists of two parts. First the availability schedule is created and secondly the patients are scheduled to that schedule.

At the beginning of each run our model creates an availability schedule. For this creation of the schedule we use one basic schedule, the master schedule of one week length. We repeat this schedule for a number of times sufficient for the run time. Then, we introduce the closures. These closures consist of ten holidays, which are at pre-defined times, based on the randomly drawn weekday of January 1st and the date of Easter. After that we draw the training days. These training days occur seven times a year on Wednesdays, during which three pre-defined treatment rooms are closed. Finally, we draw the occasional closures. The number of these occasional closures varies per treatment room and these numbers are based on data from 2006. These occasional closures cannot be on the same day as earlier drawn closures.

After the availability schedule is created, patients are planned to this schedule. The model starts searching in the first row of the schedule where a patient can be scheduled. A row represents a block b_{st}, which are sorted chronologically. The algorithm then walks through the following steps:

1. Check if the number of the selected day is larger than the contact day of the patient plus its preferred delay. If not: increase row

2. Check if the patient can be treated in the selected room. If not: increase row

3. Check if the urgency of the patient resembles the urgency label of the block. If not: increase row

4. For dedicated slot use: Check if the block type t resembles the type of patient. If not: increase row

5. Check if the available time in the block is enough to plan the patient. If not: increase row.

When all these checks have a positive outcome, the patient is planned. All data are written to a patient planning array.

An additional rule for urgent patients is that if the patient cannot be planned within the day of arrival it can be double-booked. The search starts again at the first block of the current day and uses the same rules, except for rule 5. This now changes in:

5. Check if the available time in the block or its double-bookings time is enough to plan the patient. If not: increase row.

4.4 Model verification and validation

A system only reaches a stable situation when the utilization rate is less than 100%. Since this utilization rate is over hundred percent for some consultation types, we can only analyze the system for the three consultation types that do have an utilization rate of less than 100%: ERCP, EUS and Oesdil.

4.4.1 Warm-up period

The system we analyze is a non-terminating system. A non-terminating system is a system for which there is no natural event E to specify the length of a run (Law, 2007). Such a system is developing towards a steady state on the long run. In order to reach this steady state, we determine the so called "warm-up period". By leaving out the data from within the warm-up period we are able to make better calculations on the remaining output data with respect to averages et cetera.

The method we intend to use to determine the warm-up period is Welch's graphical procedure (Welch, 1981, 1983). With this procedure we plot the output data in a diagram and determine from which point a stable situation is reached. To get a smoother diagram we use the technique of moving averages.

In our case the diagram is not reaching a stable situation, due to seasonal fluctuations in the capacity of the department and a utilization rate over hundred percent for some treatments. We eliminate the seasonal influences of holidays and we plot the access time of colonoscopies over a run of one year. This graph (Appendix C) shows an increasing access time, as we expected. Although it is not completely clear, it seems to have a stable increase after two to three months. To also analyze a stable system and get a clearer warm-up period we decrease the number of arrivals of the treatments with an utilization rate of more than a hundred percent to generate a utilization rate of 99%. Again we plot the access time of these treatments over a run of a year.

Appendix C shows the graphs that result from this graphical procedure. Despite some variance in the output, we can conclude from these graphs that the warm-up period is three months. This becomes clear from the graph of HGE 30, which shows the longest warm-up period.

4.4.2 Run length and number of replications

To get statistical significant outcomes we determine the run length and the number of replications. These two numbers are dependent of each other. The longer one run, the less runs we have to perform for statistical relevant outcomes and vice versa. We define the run length as one year. Based on this run length we can determine the number of runs we have to perform. We calculate this number of runs with the replication-deletion method (Law, 2007), which uses the following formula:

$$n_r^*(\gamma) = \min\left\{ i \geq n : \frac{t_{i-1,1-\alpha/2}\sqrt{S^2(n)/i}}{\left|\overline{X}(n)\right|} \leq \gamma' \right\}$$

with:

$$\gamma' = \frac{\gamma}{1+\gamma}$$

The variables represent the following:

n_r: the number of runs necessary to get statistical significant outcomes
i: the counter
t: value of the student-distribution with i-1 degrees of freedom and confidence interval of 1-α.
$S^2(n)$: the variance of the sample
X(n): the mean value of the sample
γ: the preferred relative error

For alpha we take 0,05, while we define the relative error (gamma) as 5%. This results in a number of 19 independent replications to get statistically significant outcomes. However, we run the model with 30 replications, which corresponds to a relative error of 3,0% for the most differentiating treatment. We choose to run 30 replications because of the fact we discovered some variations in the relative error as we ran some other scenarios.

4.4.3 Verification and validation

In the verification and validation phase we check if our model is working correctly and if it is representing the real system. This is the most time consuming part of simulation. As Neelamkavil (1987) states for validation: "True validation is a philosophical impossibility and all we can do is either invalidate or fail to invalidate." We distinguish three phases: verification, qualitative validation and quantitative validation.

Verification
Verification concerns the debugging and checking the simulation model for inconsistencies. We verify whether the model is working correctly and performs all intended actions in the right way. We perform many test runs to do so.

Qualitative validation
In the qualitative validation we use expert opinions to further validate our model. We show our assumptions, the planning rules in our model and the flow charts to the planners and the head of the department. These people indicate that the model is an applicable model to simulate the actual system, assuming we use the correct input data.

Quantitative validation
The final validation we perform is quantitative validation. We compare the outcomes of our model to the historical output data of the actual system. First we calibrate the model with output data of 2006, after which we validate it with data from 2005. Since we consider access time as the most important response from the system, we validate our model on access time.

There are various validation methods that can be used for quantitative validation, like the inspection approach, the confidence interval approach and time series approaches (Law, 2007). Most of these approaches consider two datasets containing many data that can be compared to each other. Since we only have one dataset (from the simulation) and some values from the real system we use the basic inspection approach to validate our model.

The basic inspection method uses sample means to compare the model with the real system. We take the average access times of 2005 and 2006 and compare them to sample means from our simulation model with corresponding inputs. Since we have only one dataset from the two years from the real system, these data are sensitive for variation.

Our approach measures the percentage that the average simulated value differs from the value from the real system. First we use the performance data of 2006 to calibrate the simulation model. For the actual validation we use the performance data of 2005. The actual error is about 15% for 2005, while it is about 5% for 2006. This amount is reasonable due to the stochastic nature of the dataset from the real system. Figures 4.2 – 4.7 show histograms of the access times from the real system and the simulation. The left side represents the data from the real system, the right side the data from the simulation.

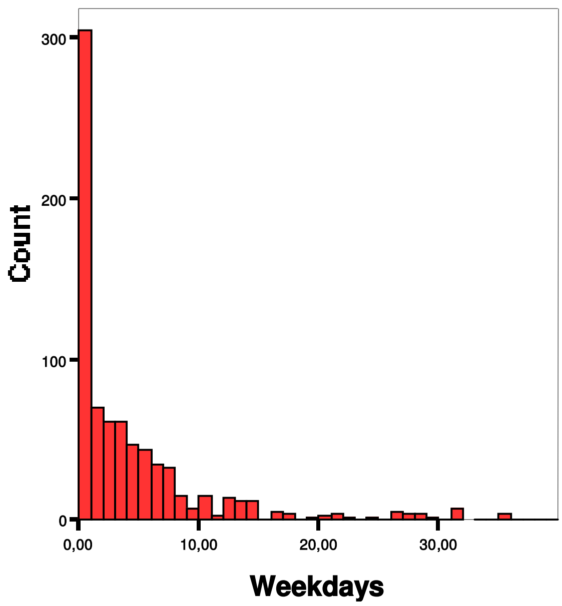


Figure 4.2 Distribution of access time for ERCP in 2006

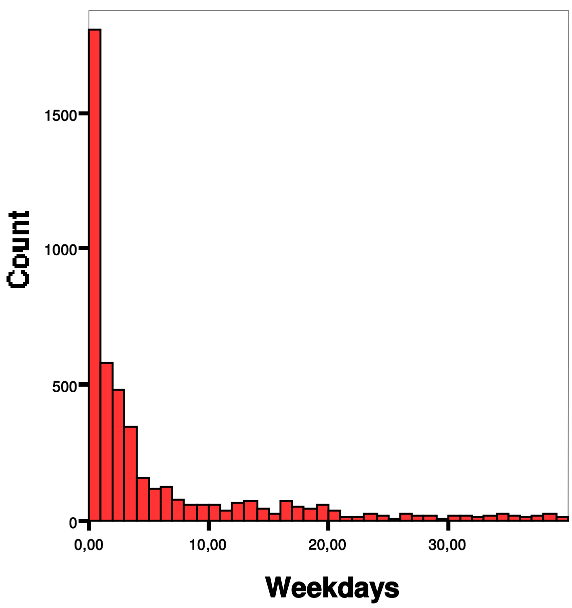


Figure 4.3 Simulated distribution of access time for ERCP

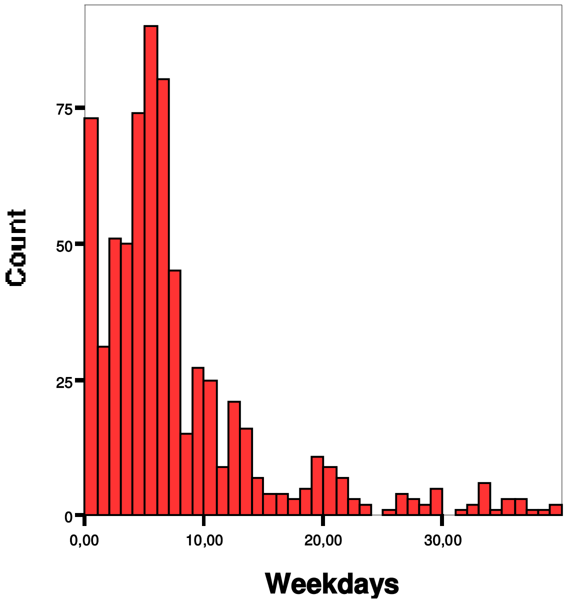


Figure 4.4 Distribution of access time for EUS in 2006

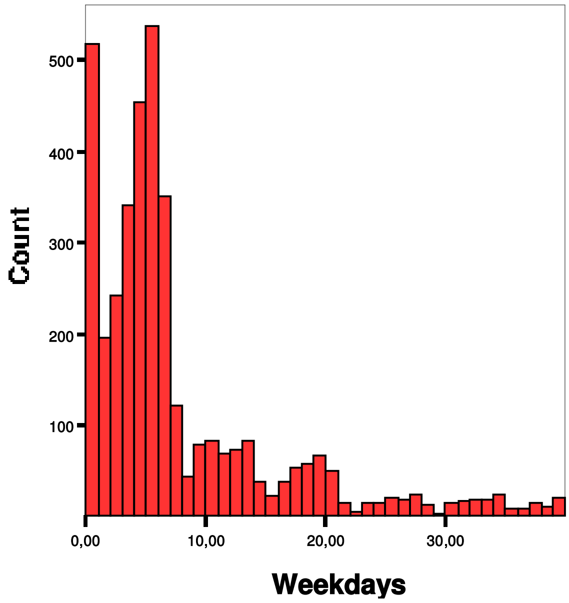


Figure 4.5 Simulated distribution of access time for EUS

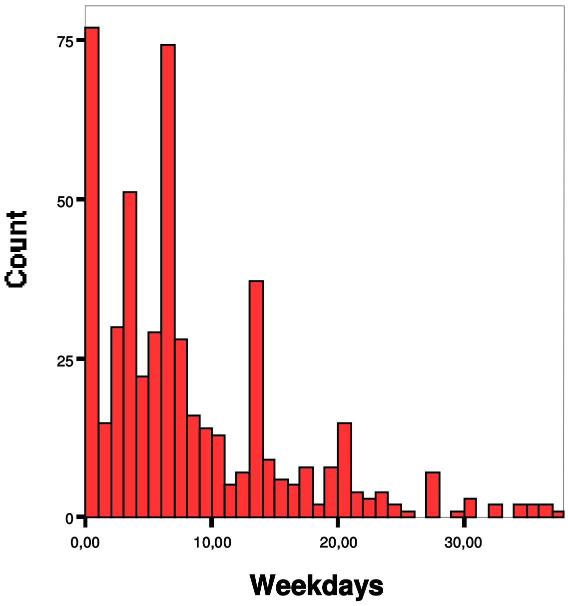


Figure 4.6 Distribution of access time for Oesdil in 2006

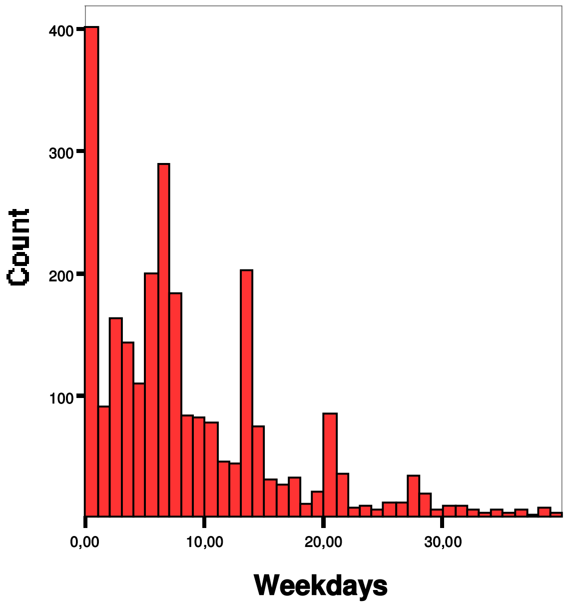


Figure 4.7 Simulated distribution of access time for Oesdil

5. Evaluation of various scenarios

This chapter presents the scenarios we evaluate using the simulation model described in the previous chapter. First we define the scenarios and consultation types we use (Section 5.1). Section 5.2 presents the results from various simulation runs while Section 5.3 discusses the elimination of backlog.

5.1 Definition of scenarios

5.1.1 The scenarios

Chapter 4 describes the experimental factors included in the model. These factors are the basis of the scenarios we define. Initially we generate the scenarios that are extremes of these factors. To compare these scenarios we define a null-scenario: a scenario that eliminates all restrictions. The Table 5.1 shows an overview of the various scenarios.

	Rooms	Blocks	Day urgency	Week urgency	HGE15	Closures	Hours
NULL					√	0	135.75
A	√	√	√			120	135.75
B	√	√	√			0	135.75
C	√		√			120	135.75
D			√			120	135.75
E	√	√				120	135.75
F	√	√	√	√	√	2%	134.5 – 159.5

Table 5.1 Theoretical scenarios

Now we will explain the experimental factors we use. These are in the header row of Table 5.1.

Rooms: All rooms have restrictions with respect to the patient types it can handle. Aspects like the size of the room, the medical equipment it houses or the level of hygiene it needs influence this.

Blocks: The schedule that we use as input for the model consists of dedicated blocks for specific patient types. These (types of) patients can only be scheduled in these blocks. If we do not take these dedicated blocks into account, all patient types can be scheduled in all blocks, as long as other restrictions are satisfied.

Day Urgency: The schedule consists of some dedicated blocks for day-urgency patients. These patients should be treated on the same day of their arrival at the department. Therefore in these day-urgency blocks no patients are planned until the day of their existence. However, if we eliminate this restriction day-urgency blocks are treated like normal blocks.

Week-Urgency: Some urgency patients do not have to be treated on the day of their arrival, but can be treated within one week. Therefore we introduce week-urgency blocks. These blocks are dedicated to week-urgency patients, but can also be filled with day-urgency patients if still empty on that day.

HGE15: Currently, HGE15 treatments are scheduled for 15 minutes in room 213, but for 30 minutes in all other rooms. This scheduling rule is based on historical developments, like some years ago there where not enough nurses available to schedule two nurses in every room (only in room 213). Since there are now enough nurses to do so we can schedule all HGE15 treatments for 15 minutes. We concluded this from consultation with some staff members and the head of the department. If the cell of HGE15 is checked (√) in Table 5.1, all HGE15 consultations are scheduled for 15 minutes.

Closures: The number presented in the closures-column of Table 5.1 is the number of days one room is occasionally closed within one year, one shift is counted 0.5. We use occasional closures as a variable since we can influence this by making clear the impact of those closures to the doctors or to schedule a back-up doctor for every shift.

Hours: The number of hours the master schedule of one week consists of. This includes all (dedicated) elective blocks as well as all urgency blocks.

Scenario A is the simulated situation of the year 2006. We use this scenario as a comparison to other scenarios. We can define the Null scenario as a lower bound and the Scenario A as an upper bound for the new scenarios.

We also define four theoretical scenarios. The first theoretical scenario (B) eliminates the occasional closures, while the second one (C) eliminates all blocks, so the system does not have dedicated time slots any more. In the scenario C we keep two rooms dedicated to ERCP/Oesdil and EUS, since these treatments have a standard of one week, contrary to other treatments which have standards of three weeks. Scenario D eliminates the room restrictions and the block restrictions, while scenario E does not take the urgency blocks into account.

Scenario F is the scenario we use for further analysis of various master schedules.

Since these theoretical scenarios are not suitable to implement in the real system and still have increasing access times, we define some scenarios that are more applicable to implement. In these scenarios we include some new circumstances, based on consultations with staff members. These are:

- *Demand reduction:* routine treatments from outside the region can be forwarded to other hospitals in other regions. This action will come into effect right away.
- *Extra demand:* Due to a medical examination of the population which is starting one month after our investigation, there will be an expected increase in colonoscopies. We already take this into account in our calculations.
- *Extra research block:* Due to a lot of patients with the same disease that have to be seen by a few specific doctors, it is necessary to implement one extra research block. This will take one shift a week and is included in all our scenarios and is named IBD.
- *Elimination of Proctoscopies:* Proctoscopies, which are now done once every two weeks, are in the future done in another department. We eliminate the arrivals of these patients from our simulation model.

Since we concluded from the theoretical scenarios that the elective capacity is not sufficient to achieve our goals, this capacity is increased in our new scenarios. Therefore the time slots for urgent patients will decrease. To make sure the amount of overtime work does not increase (too much), we introduce a new urgency type: week urgency. To pool all urgency for the week reduces variance and makes it easier to plan these emergencies. This implies a higher utilization rate of these urgency blocks and therefore less overtime work. Doctors from the department indicate that about 20% of the patients that are now seen as urgency of the day can also be seen within one week. Therefore we use the amount of 5 hours of week-urgency blocks each week.

Furthermore we still use dedicated blocks, but we pool the blocks for HGE15/HGE30 and COLO treatments, since these treatments have to satisfy the same standards and their restrictions are the same except for one room restriction.

The new schedules we use are based on an ILP solution (De Wit et al., 2008). Due to the constraint in the ILP that always a back-up doctor should be available for every shift, we can reduce allowed occasional closures to 2% of all days. This includes two conferences each year during which the department is closed.

To analyze the impact of week urgency and the HGE15 rule, we define some extra scenarios. Table 5.2 shows these scenarios.

	Rooms	Blocks	Day urgency	Week urgency	HGE15	Closures	Hours
G	√	√	√	√		2%	158.5
H	√	√	√			2%	158.5
I	√	√	√	√		2%	159.5
J	√	√	√	√	√	2%	159.5

Table 5.2 Additional scenarios for analysis

5.1.2 Consultation types

To compare the various scenarios we define a standard for the various consultation types. Since we focus on access time, the standards only takes in account the access time. The standards resulted from a discussion with doctors, nurses and other employees of the department. Table 5.3 shows an overview of all consultation types we use.

Type	No. of arrivals per year *	Standard	Research
ERCP	721	1 week	
EUS	719	1 week	
Colo ***	1157	3 weeks	
HGE 15 ***	1465	3 weeks	
HGE 30 ***	316	3 weeks	
Oesdil	487	1 week	
Fluores	698	n.a.	√
IBD	197	2 weeks**	√
CRC	728	2 weeks	√
FAP	168	n.a.	√
Day urgency	1896	1 day	
Week urgency	198	1 week	

Table 5.3 Overview of used consultation types

* Number of arrivals in the new situation
** Due to an rough estimation of arrivals we do not take these in account in the results
*** Taken together in some scenarios naming them HGE

5.2 Simulation results

This section discusses the simulation results of the various scenarios. Section 5.2.1 discusses the outcomes of the theoretical scenarios defined in Section 5.1. Section 5.2.2 and 5.2.3 subsequently analyze the impact of week urgency and the HGE15 rule. Then Section 5.4 analyzes the impact of extra blocks in the schedule. The final section combines all results and proposes one new planning method and its corresponding schedule.

5.2.1 Theoretical scenarios

Table 5.4 and Table 5.5 give an overview of the simulation results of the theoretical scenarios defined in Section 5.1.

	Access times within standards					
	NULL	A	B	C	D	E
ERCP	100.0%	97.6%	99.5%	93.7%	4.8%	98.4%
EUS	100.0%	93.8%	99.0%	95.1%	3.3%	94.5%
Colo	100.0%	21.0%	67.9%	7.0%	40.1%	100.0%
Oesdil	100.0%	61.0%	77.1%	89.8%	10.0%	95.3%
HGE 15	100.0%	92.6%	100.0%	9.0%	44.2%	100.0%
HGE 30	100.0%	88.6%	100.0%	4.7%	40.8%	100.0%
CRC	100.0%	88.3%	97.3%	2.6%	26.7%	96.6%
Day urgency	100.0%	94.7%	95.5%	94.7%	97.6%	94.9%
Overtime work (hr)	144.7	197.5	158.8	202.3	58.1	360.6

Table 5.4 Simulation results of the theoretical scenarios (1)

	Average access times					
	NULL	A	B	C	D	E
ERCP	5.73	6.82	6.49	7.30	23.33	6.63
EUS	7.30	8.25	7.80	8.25	24.00	8.20
Colo	7.22	69.61	18.07	51.11	24.83	7.79
Oesdil	7.01	16.88	11.01	8.07	22.32	7.45
HGE 15	7.29	13.22	9.46	50.12	23.88	7.46
HGE 30	7.27	15.26	10.24	52.72	24.69	7.59
CRC	7.25	13.35	12.38	51.50	22.78	10.73

Table 5.5 Simulation results of the theoretical scenarios (2)

Initially we expect scenario B and C to perform better with respect to access times than scenario A, because of the elimination of one restriction. For scenario B the results show that the elimination of occasional closures has major impact, like we expected. It even reduces the access time of Colo with 74.0%. However, scenario C does not perform better than A on the percentage of patients that is treated within the standard. We distinguish two causes: (1) the blocks are normally kept dedicated until one week in advance. After that it is allowed to use them for all treatments. This causes that despite the average access time is increasing, still part of the patients is treated within the standard time. (2) The HGE15 treatments are planned for 15 minutes in room 213, but for 30 minutes in all other rooms. Since all HGE blocks are in room 213 in scenario A the majority of these treatments is planned for 15 minutes, while in a scenario C more of these treatments are planned for 30 minutes, which ultimately results in higher access times.

Furthermore, in scenario D we have eliminated the room restriction to create more flexibility in planning. The access times of all patient types are now evenly distributed, however, the percentage of patients treated within the standards is lower than in the current situation. This is caused by the reasons mentioned above, but also implies that there is too little capacity planned for elective patients. The amount of overtime work decreases in this scenario.

To create more elective capacity we define a scenario without urgency blocks (scenario E). As we expect the amount of overtime increases dramatically, but most patients are treated within the standard times. This scenario indicates that we have to look critically to the trade off between elective capacity and urgency blocks.

5.2.2 Week urgency

To reduce overtime work we introduce week urgency. This type of urgency only has impact on the HGE and Colo treatments since all other treatments already have a standard of one week. Therefore we compare the amount of overtime work for these treatments only as well as the amount of overtime work for the whole department. An amount of 20% of all urgent patients for HGE and Colo treatments can be indicated as week urgency, which corresponds to five hours a week. Table 5.6 shows the results from the simulation model.

	No. of double bookings		Overtime work (hours)	
Scenario	G	H	G	H
Colo	14	16	14.0	16.0
HGE 15	52	72	13.0	18.0
HGE 30	27	38	13.5	19.0
Total	93	126	40.5	53.0
Whole department	223	248	152.8	158.1

Table 5.6 Influence of week urgency on double bookings and overtime work

In both scenarios, the day urgency standard is about the same (97.1% vs. 97.2%). The percentage of week urgency patients treated within one week in scenario G is 99.7%. Table 5.6 shows that the overtime work caused by the HGE and Colo treatments decreases with 12.5 hours (23.6%) a year. However, the total overtime work each year only decreases by 5.3 hours (3.3%) a year. We can explain this by the fact that week urgency blocks are filled up earlier with HGE treatments than other day urgencies arrive. This means there are fewer blocks available for day urgencies. Despite the relative small reduction of overtime work, we decide to implement the week urgency, since it still improves the performances.

5.2.3 The HGE 15 rule

Since there are enough nurses available in the current situation to plan two nurses in every room we plan all HGE 15 treatments for 15 minutes in all rooms. This influences the access times and double bookings of all treatments. Table 5.7 shows the simulation results.

	Access times within standards		Mean access time		No. of double bookings	
Scenario	I	J	I	J	I	J
ERCP	97.2%	98.6%	6.82	6.51	81	59
EUS	95.9%	96.8%	7.82	7.69	13	13
Colo	95.0%	100.0%	11.11	8.47	14	12
Oesdil	94.0%	98.6%	8.09	7.50	30	26
HGE 15	94.0%	100.0%	11.51	8.11	36	20
HGE 30	91.9%	100.0%	12.30	8.48	28	12
CRC	99.7%	99.7%	9.81	9.51	0	0
Day urgency	97.1%	97.1%				
Week urgency	99.7%	99.9%				
Overtime work (hours)	143.0	105.3				

Table 5.7 Influence of the HGE 15 rule on access time and double bookings

Table 5.7 shows that the HGE 15 rule has major impact on the average access time as well as on double bookings. The average access time decreases dramatically (28.3%) for the HGE and Colo treatments and also decreases for other treatments. This implies that the percentage of treatments done within the standards is increasing, which Table 5.7 also shows. The amount of overwork time (caused by double bookings) is decreasing enormously too, towards 105.3 hours a year, which implies a decrease of 26.4%.

5.2.4 Capacity of the schedule

Now we have analyzed some additional scheduling rules we have to determine the capacity of the master schedule. Especially the capacity available for elective HGE and Colo treatments should be increased compared to the current situation. We do this by trial and error, and during this process we automatically do some sensitivity analysis. The data in Table 5.8 and 5.9 represent a scenario like scenario I, with varying master schedules.

Hours per week HGE/Colo	33	34	36
HGE 15	92.1%	94.0%	99.1%
HGE 30	89.6%	91.9%	99.1%
Colo	93.1%	95.0%	99.2%

Table 5.8 Access times within standards with varying capacity

Hours per week HGE/Colo	33	34	36
HGE 15	12.78	11.51	9.35
HGE 30	13.88	12.3	9.68
Colo	12.43	11.11	9.41

Table 5.9 Average access times with varying capacity

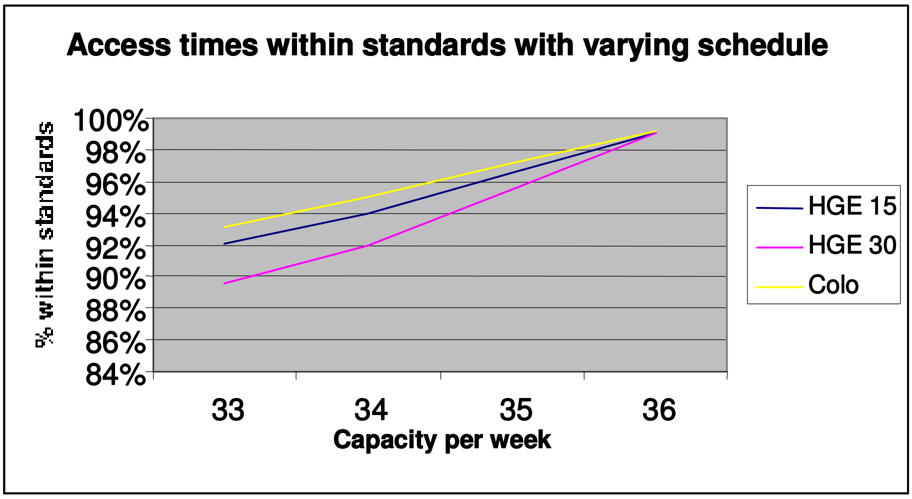


Figure 5.1 Access times within standard with varying capacity

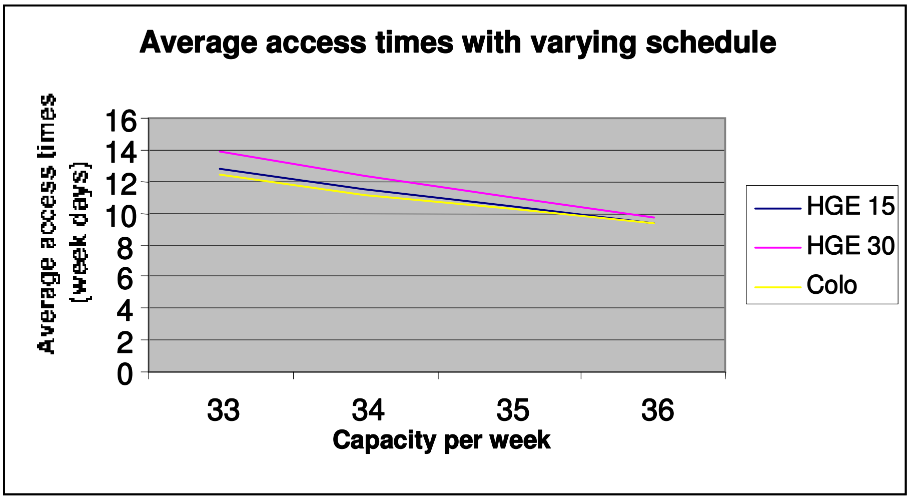


Figure 5.2 Average access times with varying capacity

Table 5.8 and Figure 5.1 show that every extra hour in de schedule results in 2-3 % more patients treated within the standards for percentages over 90%. Table 5.9 and Figure 5.2 show that each extra hour of HGE/Colo reduces the average access time of HGE/Colo with one day on average.

5.2.5 Proposed schedule

From the analysis we did on scheduling rules and the master schedule we propose a master schedule for the department. The scheduling rules in this scenario are presented in Table 5.10 for scenario F.

	Rooms	Blocks	Day urgency	Week urgency	HGE15	Closures	Hours
A	√	√	√			120	175.5
F	√	√	√	√	√	2%	Var.

Table 5.10 Current (A) and proposed (F) scheduling rules

The schedule we propose consists of the following blocks (in hours) compared to the current situation (Table 5.11). This new schedule is determined by solving an ILP problem, which lies beyond the scope of this study (De Wit et al., 2008).

	Current schedule (hours)	Proposed schedule (hours)
ERCP	14	21
EUS	18.75	19
Colo	25	32.5*
HGE	14.5	
Oesdil	4	6
Fluores	8.5	9
CRC	10	18
IBD	-	3
FAP	2.375	2.5
Day urgency	38.625	19
Week Urgency	-	4.5
Total	135.75	134.5

Table 5.11 Capacities of the current and proposed schedule

* Includes Colo and HGE

Table 5.12 shows all results from the current situation (scenario A) and the results of the proposed situation (scenario F with new schedule).

	Access times within standards		Average access time		No. of double bookings	
Scenario	A	F	A	F	A	F
ERCP	97.6%	97.1%	6.82	6.73	141	69
EUS	93.8%	97.1%	8.25	7.76	16	13
Colo	21.0%	95.3%	69.61	11.44	14	19
Oesdil	61.0%	96.9%	16.88	7.68	6	23
HGE15	92.6%	95.9%	13.22	10.97	69	33
HGE30	88.6%	94.6%	15.26	11.93	19	18
CRC	88.3%	99.7%	13.35	10.08	0	0
Day urgency	94.7%	96.6%				
Week urgency	n.a.	99.5%				
Overtime work (hours)	197.51	126.85				

Table 5.12 Simulation results of current situation (A) and proposed situation (F)

The proposed scenario performs better on every aspect. Figures 5.3 and 5.4 also depict this.

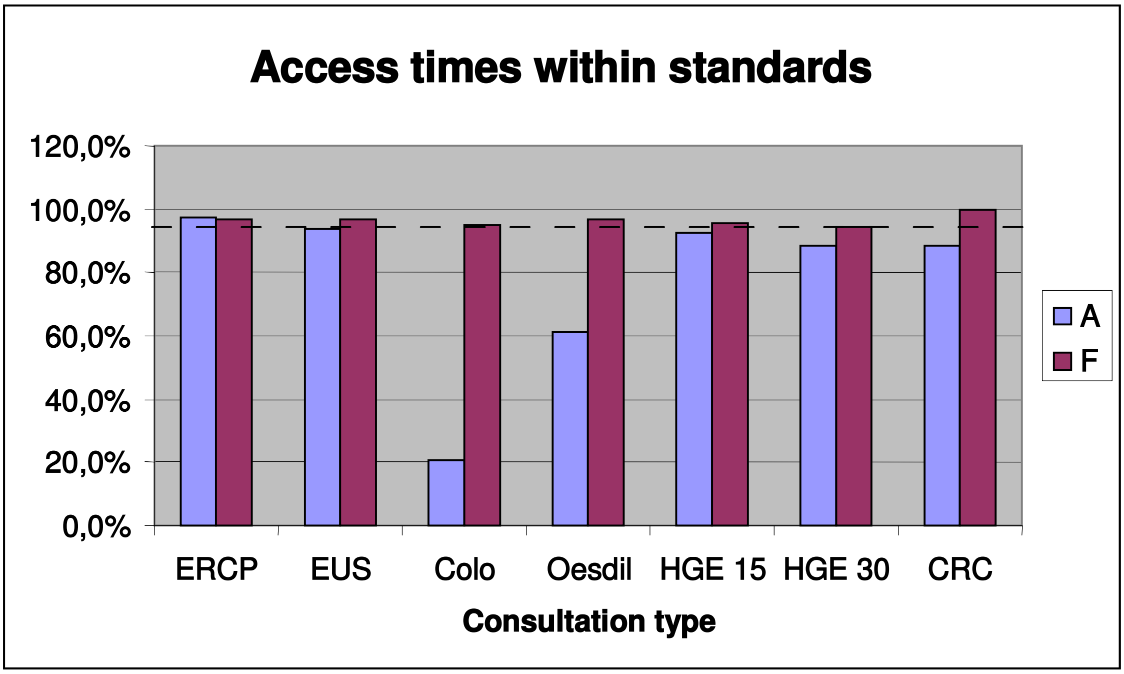


Figure 5.3 Comparison between current situation (A) and proposed situation (F)

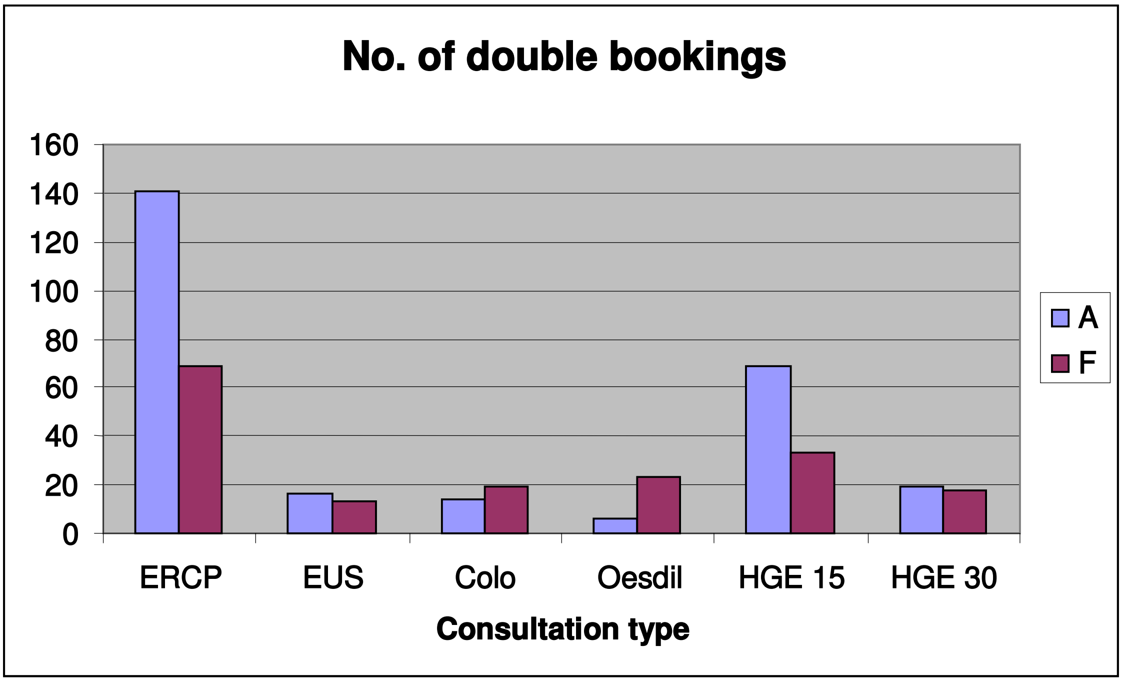


Figure 5.4 Comparison between current situation (A) and proposed situation (F)

Despite 1.25 hours less capacity in the master schedule, the proposed scenario (F) performs better than the scenario of the current situation (A). For all consultation types and urgency types at least 95% of the patients is treated within the standard time, except for the HGE 30 treatments. However, this percentage (94.6%) approaches the 95% closely and since this schedule includes back-up doctors for every shift, we consider this as acceptable.

The total overtime work is decreasing with 35.8% to 126.85 hours per year. For most treatments the number of double bookings is decreasing, but still for some this number is increasing. We can explain this by the fact that the urgency blocks in the new schedule are differently distributed than the old ones. This can cause some fluctuations in double bookings due to room restrictions.

5.3 Backlog

The implementation of these new planning rules and master schedule should not cause too many problems and no investments have to be done. However, there still is a backlog of HGE and Colo patients that has to be handled. Therefore we decide to temporarily create some extra capacity of nine hours a week. There are two options for the implementation of the extra capacity. We can start right away with this capacity or we can introduce it when the new schedule is implemented, in three months (the size of the backlog). Figures 5.5 and 5.6 depict what impact this extra capacity has on the backlog and access time for HGE and Colo treatments.

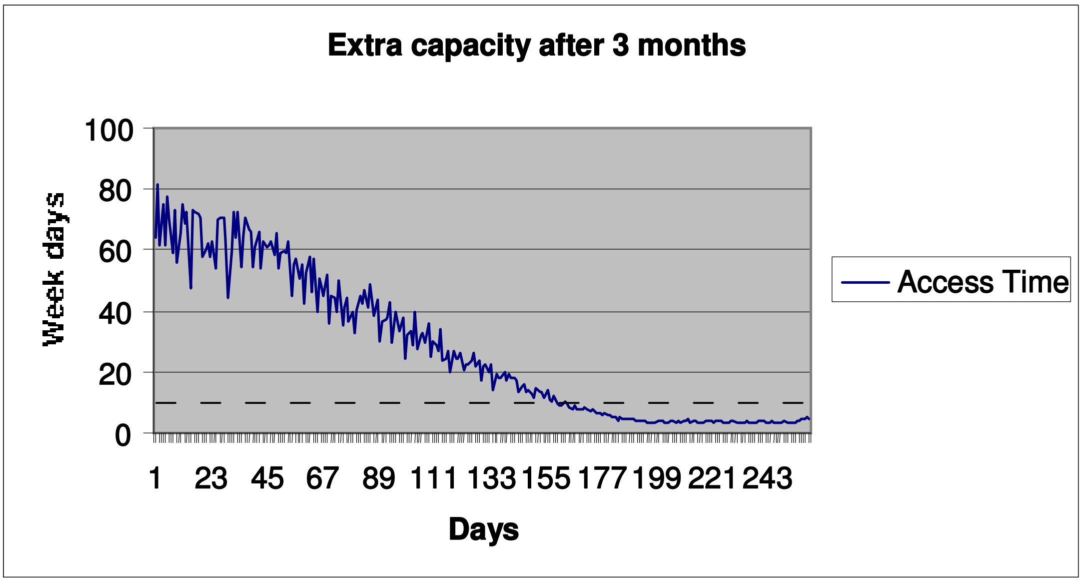


Figure 5.5 Backlog reduction if extra capacity is implemented after 3 months

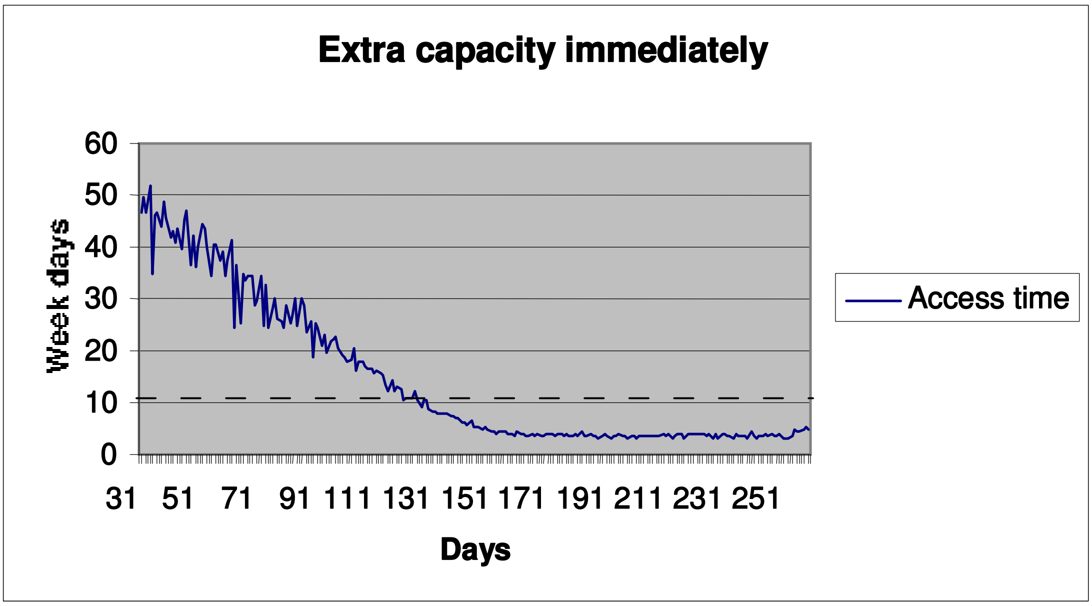


Figure 5.6 Backlog reduction if extra capacity is implemented directly

Since Table 5.10 shows that in a stable situation of Scenario F the access time for HGE and Colo patients is between ten and eleven days, the access time has to reach this level to represent a stable situation. Figure 5.5 shows that it takes about 160 days to get to that stable situation if we implement the extra capacity after three months. Figure 5.6 shows that if we implement these extra hours directly, it still takes about 135 days to get to the stable situation. Therefore we prefer to implement this extra capacity at the same time the new master schedule takes in effect, to prevent unnecessary difficulties with varying schedules besides the varying capacity.

6. Conclusion and recommendations

6.1 Conclusions

We refer back to the aim of this study which we defined in Chapter 1:

The objective of this study is to reduce the access times for the endoscopy-HGE department of the AMC by improving the department's scheduling of patients.

The simulation model of the endoscopy department shows that there are various ways to reduce the access times without increasing overtime or increase the capacity of the schedule. Both theory and the model show that occasional closures have a negative impact on access times. To reduce these occasional closures we initiate a schedule were every block has a back up doctor, so if the originally scheduled doctor is unavailable, the back up doctor can replace him. With this back up system we assume we can decrease the number of occasional closures to two percent.

Furthermore, we increase the capacity available for HGE and Colo treatments. This capacity we create by decreasing the capacity for urgent patients. To prevent the double bookings to increase we introduce week urgency. Week urgency consists of 20% of all urgent patients of HGE and Colo. This reduces the total overtime with 3.3%.

Another change we make to the scheduling rules is the scheduling of the HGE 15 treatments. We now plan all of them for 15 minutes, which reduces access times and overtime work with 28.3% and 26.4% respectively.

We combine all these scheduling rules with a new master schedule to come to a proposed schedule. This schedule has 1.25 hours (0.9%) less capacity than the current schedule, but results in better access times and less overtime work. Of all patients, 95% is treated within the standards, the access times all decrease compared to the current situation. The overtime work also decreases with 35.8 % to 126.85 hours per year.

6.2 Recommendations

We recommend the endoscopy department of the AMC the following:

- We recommend to implement the proposed schedule. This schedule balances the capacities for all consultation types and urgency types. This schedule results in that 95% of all patients are treated within their standard times.

- We recommend the department to reduce the occasional closures by using back-up doctors. The schedule we propose consists of back-up doctors, who can replace the originally planned doctors when they are unable to attend their clinic sessions. This has a positive impact on the access times of patients.

- Since there are enough nurses to plan two nurses in each room, we recommend planning all HGE 15 treatments for 15 minutes in every room. This again has positive impact on the access times.

- To get the access times for HGE and Colo treatments more equal to each other, we recommend clustering the blocks for these treatments. This results in an equal access time for both groups and a decrease of the variance, which influences access time in a positive way.

- With the introduction of week urgency we reduce the number of double bookings. Week urgency can be introduced for HGE and Colo treatments, and since it reduces overtime work we recommend to do so.

Recommendations for further research:

- We assume that the system with back-up doctors will decrease the number of occasional closures. However, this implies a culture change among the doctors. The results of this study show the impact the reduction of occasional closures have on the access time, which is an argument for implementing it, but we recommend to study a way for successful implementation of this system.

- We assume a constant capacity, but the impact of closures like holidays and conferences on access times is significant. We recommend to further research the possibility of using flexible capacity to prepare for these holidays.

- Many patients that visit the AMC need more than one diagnostic or treatment. This study focuses only on the appointment system for the endoscopy department. However, it is interesting to further investigate the possibilities for a multiple appointment system.

References

Alfa, A.S. (2003). Vacation Models in Discrete Time. *Queueing Systems* 44, 5-30.

D'Antone, I. (2000). Pipeline length optimization in a digital asynchronous trigger controller. *Nuclear Instruments and Methods in Physics Research* A 451, 439-448.

Borst, S.C., Boxma, O.J. and Combé, M.B. (1993). An M/G/1 queue with customer collection. *Stochastic Models,* 9:3, 341-371.

Cayirly, T. and Veral, E. (2003). Outpatient scheduling in healthcare: A review of literature. *Production and Operations Management*, Vol. 12, No. 4, 519-549.

Chao, X. and Zheng, S. (1998). A result on networks of queues with customer coalescence and state-dependent signalling. *Journal of Applied Probability* 35, 151-164.

Cobham, A. (1954). Priority assignment in waiting line problems. *Journal of the Operations Research Society of America*, Vol. 2, No. 1, pp. 70-76.

Economou, A. (2001). An Alternative Model for Queueing Systems with Single Arrivals, Batch Services and Customer Coalescence *Queueing Systems* 40, 407–432.

Franx, G.J. (2001). A simple solution for the M/D/c waiting time distribution. *Operations Research Letters* 29, 221-229.

Holley, J.L. (1954). Waiting line subject to priorities. *Journal of the Operations Research Society of America,* Vol. 2, No. 3, pp. 341-343.

Ivaneshkin, A.I. (2007). Optimizing a multiserver queuing system with a variable number of servers. *Cybernetics and Systems Analysis,* Vol. 43, No. 4, 542-548.

Neelamkavil (1987), Francis Computer Simulation and Modeling, John Wiley & Sons.

Ozawa, T. (1990). Alternating Service Queues with Mixed Exhaustive and *K*-limited Services. *Performance Evaluation*, Vol.11, No. 3, pp.165-175.

Takács, L. (1964). Priority Queues. *Operations Research,* Vol. 12, No. 1, pp. 63-74.

Thiruvengadam, K. (1963). Queuing with breakdowns. *Operations Research*, Vol. 11, No. 1, pp. 62-71.

Vastola, K. (1996). Performance Modeling and Analysis of Computer Communication Networks, *http://networks.ecse.rpi.edu/~vastola/*

Vuuren, M. van. and Winands, E.M.M. (2007). Iterative approximation of *k*-limited polling systems. *Queueing Systems* 55, 161-178.

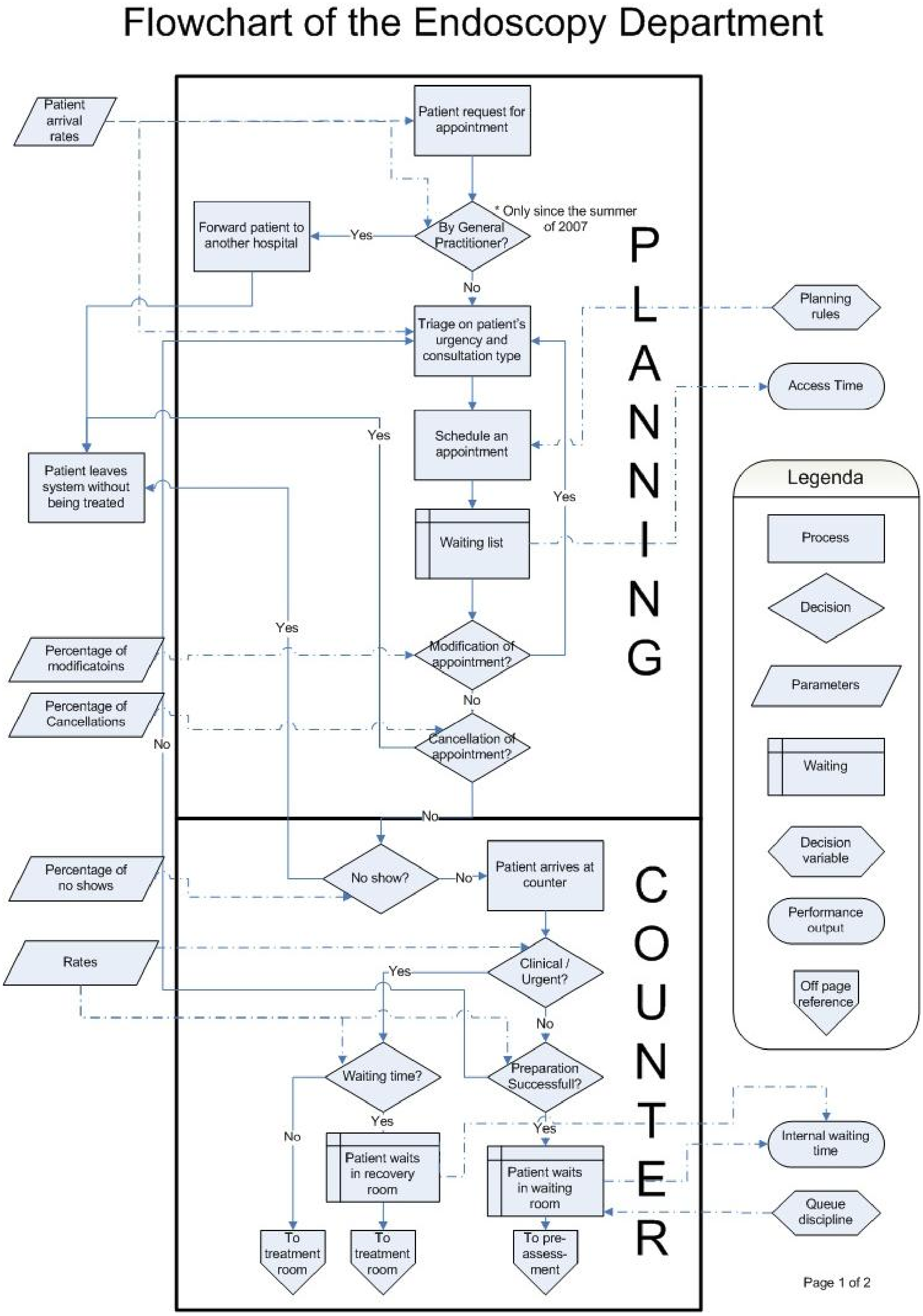
Welch, P.D. (1981). On the problem of the initial transient in steady-state simulation, IBM Watson Research Center, Yorktown Heights, New York.

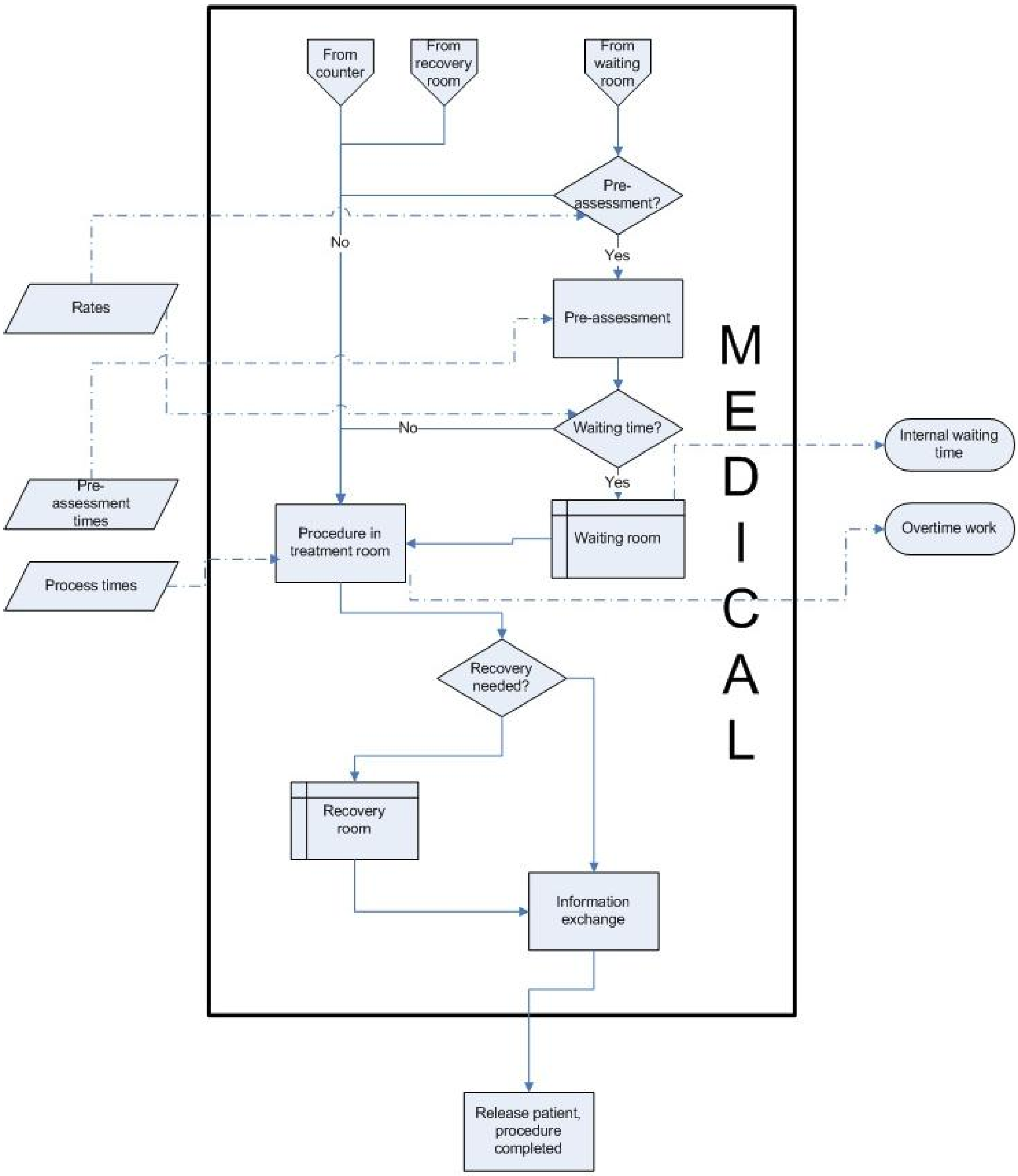
Welch, P.D. (1983). The statistical analysis of simulation results, *The Computer Performance Modeling Handbook,* S.S. Lavenberg, Ed., Academic Press, New York.

White, H. and Christie, L.S. (1958). Queuing with preemptive priorities or with breakdown. *Operations Research,* Vol. 6, No. 1, pp. 79-95.

Wit, J. de, (2008). Looking inside the endoscopy department, Master's thesis, unpublished.

Appendix A: Flowchart of the endoscopy department




From counter
From recovery room
From waiting room
Pre-assessment?
No
Yes
Rates
Pre-assessment
Waiting time?
No
Yes
Pre-assessment times
Procedure in treatment room
Waiting room
Process times
Internal waiting time
Overtime work
MEDICAL
Recovery needed?
Recovery room
Information exchange
Release patient, procedure completed

Page 2 of 2

Appendix B: Standard schedule in 2007

Room No.	Diagnosis Type	Equipment	Number of doctor available	Planned duration (min)	Amount	Part of day	Monday	Tuesday	Wednesday	Thursday	Friday
K213	OGD/AN OESDIL SIGMO COLO PEG IN/ UIT VSM/VSZ/button	X-Ray technology Mobile towers	10	15/30 30 15 60 30/30 30		morning	HGE General 2 x emerg. oes dil HGE General	HGE General 2 x emerg. oes dil HGE General	HGE General 2 x emerg. oes dil HGE General	HGE General 2 x emerg. oes dil HGE General	HGE General 2 x emerg. oes dil HGE General
						afternoon	1 x colo HGE General Urgent	1 x colo HGE General Urgent	1 x colo HGE General Urgent		1 x colo HGE General Urgent
K132	OGD/AN SIGMO COLO PEG IN/ UIT VSM/VSZ/button	Mobile towers	10	15/30 30 60 60/30		morning	2 x colo Emergencies		2 x colo Emergencies		2 x emerg. colo Emergencies
						afternoon	Emergencies	Emergencies	Emergencies	Emergencies	Emergencies
K219	OGD/AN SIGMO COLO PEG VSM/VSZ/button	Mobile towers	10	15/30 30 60 60/30		morning	7 x OGD	3 x colo	4 x colo	3 x emerg. colo 1 x emergency	7 x OGD
						afternoon	3 x colo	3 x colo	2 x emerg. colo emergencies	emerg. Colo	
K120	COLO	Mobile towers	10	60	max 18	morning		3 x emerg. colo	3 x colo	3 x colo	
						afternoon	2 x colo 1 x emerg. colo	2 x colo 1 x emerg. colo		3 x colo	
K218	PROCTO	Special 'bed'	10	15	max 14	morning				procto 1x per 2 weeks 10 x reg / 4 x emerg.	
						afternoon					
K124	ERCP OESDIL	X-ray technology Mobile towers	7	60 30	max 11 max 20	morning	2 x oes dil 3 x ercp	2 x oes dil 2 x ercp		2 x oes dil 3 x ercp	2 x oes dil 3 x ercp
						afternoon	2x emerg. ercp	1x ercp 2x emerg. ercp		1x ercp 1x emerg. ercp	1x ercp 1x emerg. ercp
K127	EUS Plex blok	Ultrasound equip. Mobile towers	6	45	24 / 25	morning	4 x EUS	4 x EUS	4 x EUS	4 x EUS	
						afternoon	3 x EUS	3 x EUS		3 x EUS	

Appendix C: Welch's graphical procedure

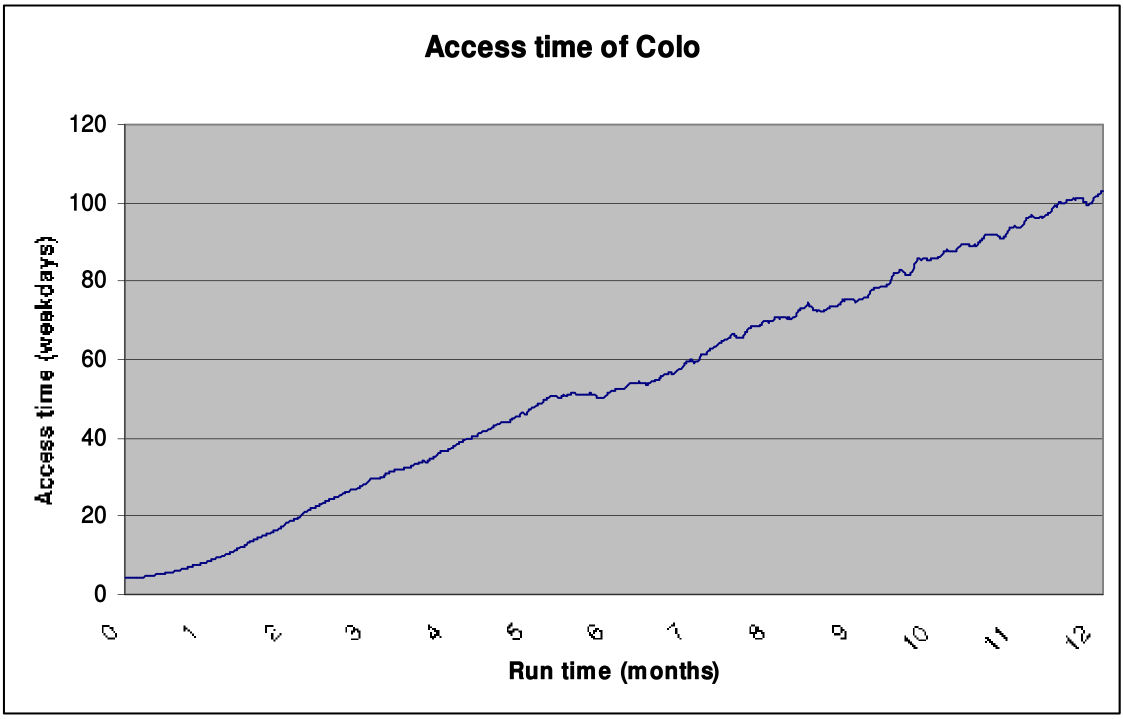


Figure C.1 Access time of Colo with a utilization rate of 146%

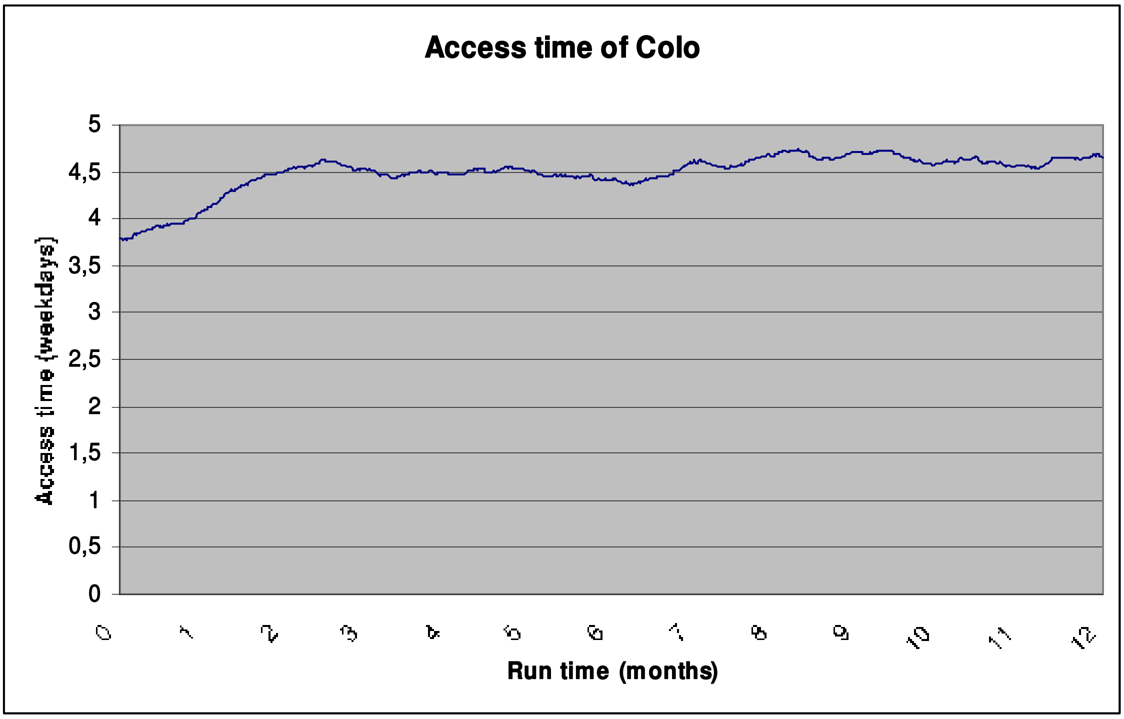


Figure C.2 Access time of Colo with a utilization rate of 99%

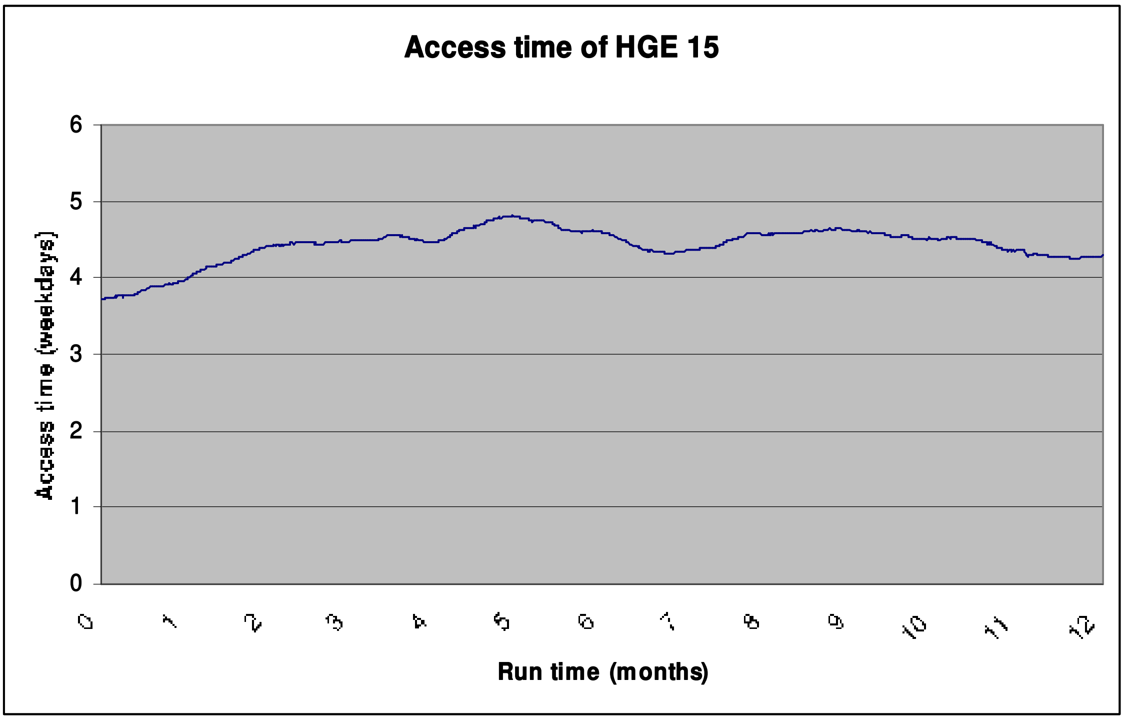


Figure C.3 Access time of HGE 15 with a utilization rate of 99%

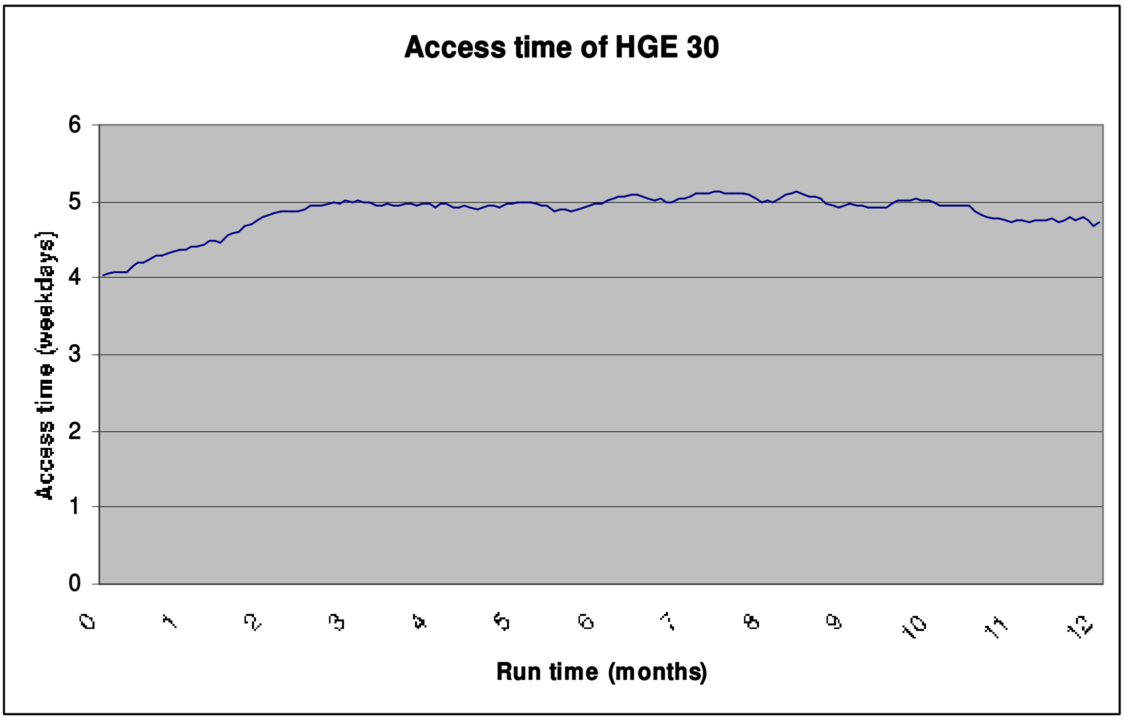


Figure C.4 Access time of HGE 30 with a utilization rate of 99%