Outpatient appointment scheduling

An evaluation of alternative appointment systems to reduce waiting times and underutilization in an ENT outpatient clinic

Jochem Westeneng

Erasmus MC
University Medical Center Rotterdam

University of Twente
Enschede - The Netherlands

Outpatient appointment scheduling

An evaluation of alternative appointment systems to reduce waiting times and underutilization in an ENT outpatient clinic

Jochem Westeneng

July 2007

Master's thesis

Industrial Engineering and Management

University of Twente, Enschede, The Netherlands

School of Management and Governance

Department of Operational Methods for Production and Logistics

Supervisors:

Dr. Ir. E.W. Hans, University of Twente, Enschede, The Netherlands

Dr. Ir. J.M.J. Schutten, University of Twente, Enschede, The Netherlands

Ir. J.M. van Oostrum, Erasmus University Medical Center, Rotterdam, The Netherlands

Summary

Introduction

The Ear, Nose, and Throat (ENT) outpatient clinic of Erasmus University Medical Center (Erasmus MC) in Rotterdam, the Netherlands (further referred to as '*the outpatient clinic*') is confronted with long waiting times for patients, overtime for doctors and nurses during clinic sessions, and peak workloads for its counter personnel. Previous studies (Huang, 1994) show that long waiting times are an important dissatisfaction for patients. According to doctors and personnel, overtime and peak workloads are potential threats for the quality of care and the quality of labor. This case study focuses on outpatient scheduling, by evaluating different appointment scheduling systems as a means to reduce patients' waiting times, to improve the outpatient clinic's utilization, to reduce doctor's overtime and idle time, and to focus on the counter personnel's peak workload as well.

Context

In the current situation, the schedulers of the outpatient clinic immediately schedule a patient upon his request for an appointment, using two consultation types: *new* and *review*. The lengths of appointment intervals (time slots) are fixed in advance, and *dedicated* to either one of the consultation types. Patients are scheduled on a first-call-first-appointment basis, up to three months in advance (i.e. a scheduling horizon of three months). The schedulers do not adjust the appointment time of a new patient for the attendance of a *medical student*, who treats new patients in advance of their consultation with a *resident* or *medical specialist*, thereby causing idle time for the doctor.

This current appointment system (i.e. the combination of control parameters and mechanisms that determines the way of scheduling patients) results for the outpatient clinic in an average *internal waiting time* for patients of 19 minutes. The internal waiting time includes all the time patients have to wait after their scheduled appointment time has been passed. About 45% of the clinic sessions last at least 30 minutes longer than scheduled (*overtime*). Counter personnel face peak workloads on Tuesday, Wednesday, and Friday mornings.

Methods

To improve the current appointment system, we evaluate six experimental factors:

- Method of decision-making, which involves the possibility to schedule patients in batches at the same time (static scheduling), or immediately when the patient requests an appointment (dynamic scheduling, as in the current situation).
- Length of the scheduling horizon.

- Usage of dedicated time slots, either dedicated to consultation types in advance (as in the current situation), or *pile-up scheduling* (i.e. scheduling the consultations connected together, irrespective of the consultation types).
- Sequence of patients in a clinic session.
- Number of patients to arrive at the first appointment time of a clinic session.
- Adjustment of appointment times for the attendance of medical students.

We evaluate 45 alternative appointment systems, constructed as combinations of the values of the experimental factors. We thereby introduce a new classification of consultation types that enables to forecast a patient's consultation type better in advance. A discrete-event simulation model evaluates the performances of these alternative appointment systems, using theoretical probability distributions for the arrival rate of patients, punctualities, and consultation durations, as input parameters (among others).

Results and conclusions

Following the simulation results, an efficient frontier for the trade-off between internal waiting time, overtime, and doctor's utilization shows three alternative appointment systems to perform efficient. A *Data Envelopment Analysis* of the results shows that two of these appointment systems perform clearly better than the others, when the access time (i.e. the number of days a patient has to wait for an appointment) is taken into consideration as well. However, the results are very sensitive to an increasing number of appointment requests.

We conclude that an appointment system with dedicated time slots and only one patient scheduled for the first appointment time, combined with a sequencing rule that assigns patients with low variance of consultation duration to the beginning of a clinic session, is able to achieve a 55 % reduction of average internal waiting time, and an average overtime that falls from 19 % to 11 % of the clinic session duration. This appointment system involves a relatively small number of changes to the current appointment system. However, the utilization rate for this appointment system is 89.5 %, whereas the configuration for the current situation has a utilization rate of 93.3 %. To treat the same number of patients per year with this lower utilization rate, the capacity has to be increased with two clinic sessions per week. Nevertheless, the advantages of this appointment system, as stated above, outweigh the decrease of utilization rate. The *Data Envelopment Analysis* supports this conclusion. Therefore we recommend the implementation of this appointment system to the outpatient clinic.

The length of the scheduling horizon, and the correction of appointment times for the presence of medical students, have little effect on the performance of the appointment systems. The effects of the other experimental factors depend heavily on the combination of experimental factor values. An appointment system with dynamic decision-making will not further increase the pressure on the workload for the counter and scheduling personnel, whereas static scheduling involves an extra call-back to each patient for the counter personnel, thereby raising the workload. A better balancing of clinic sessions over the week can improve the workload.

Samenvatting

Aanleiding

Bij de polikliniek van de afdeling Keel-, Neus- en Oorheelkunde (KNO) van het Erasmus Universitair Medisch Centrum (Erasmus MC) in Rotterdam (in deze samenvatting verder aangeduid als '*de polikliniek*') bestaan lange wachttijden voor patiënten, uitloop van spreekuren en piekbelastingen in de werklast voor het baliepersoneel. Eerdere onderzoeken (o.a. Huang, 1994) laten zien dat lange wachttijden tot grote ontevredenheid onder patiënten leiden. Volgens artsen en polipersoneel kunnen uitloop en piekbelastingen de kwaliteit van zorg en arbeid in gevaar brengen. Bij de casus in dit onderzoek bekijken we de afspraakplanning van de polikliniek, door middel van een evaluatie van alternatieve afspraaksystemen. Het doel is om gelijktijdig de wachttijden te verkorten, de benutting van de polikliniek te verbeteren, de uitloop en leegstand van spreekuren terug te dringen en de piekbelasting voor het baliepersoneel te verminderen.

Context

Als een patiënt in de huidige situatie een wil afspraak maken, dan krijgt hij meteen te horen op welke dag en tijd hij terecht kan op de polikliniek. De lengte van afspraakintervallen staat van te voren vast in een blauwdruk van *spreekuurplaatsen*, die tevens zijn toegewezen aan een van de consulttypen *nieuw* of *controle*. Patiënten worden ingepland in volgorde van de aankomst van afspraakverzoeken, tot drie maanden in de toekomst (d.w.z. een planningshorizon van drie maanden). *Co-assistenten* zien de nieuwe patiënten voorafgaand aan het consult met een *arts-assistent* of *medisch specialist*, waardoor het voorkomt dat de arts moet wachten op de co-assistent. Desondanks worden afspraaktijden voor nieuwe patiënten niet aangepast op de aanwezigheid van co-assistenten.

Een afspraaksysteem is de combinatie van bestuurparameters en –mechanismen die de manier van inplannen van patiënten bepalen. Met het huidige afspraaksysteem hebben patiënten van de polikliniek gemiddeld 19 minuten *interne wachttijd*. Onder interne wachttijd verstaan we de tijd die een patiënt in de wachtkamer moet wachten op de arts nadat de geplande afspraaktijd is verstreken. Ongeveer 45% van de spreekuren loopt minimaal 30 minuten uit. Het baliepersoneel krijgt in de huidige situatie te maken met piekbelastingen op dinsdag-, woensdag- en vrijdagochtenden.

Methode

Ter verbetering van het afspraaksysteem beschouwen we in deze studie de volgende zes experimentele factoren:

- Moment van inplannen. Hierbij onderscheiden we enerzijds het inplannen van patiënten in *batches* op hetzelfde moment (statisch inplannen), en anderzijds het inplannen op het moment dat de patiënt zich aandient (dynamisch, zoals in de huidige situatie).

- Lengte van de planningshorizon.
- Gebruik van vaste spreekuurplaatsen, waarbij we onderscheid maken tussen het vooraf vastleggen van spreekuurplaatsen in een blauwdruk (zoals in de huidige situatie), en een zogenaamd *stapelspreekuur*. In het laatste geval worden consulten aaneengesloten op elkaar gestapeld, onafhankelijk van het consulttype of de duur.
- Volgorde van patiënten in een spreekuur.
- Aantal patiënten dat ingepland wordt voor het eerste afspraaktijdstip van een spreekuur.
- Al dan niet aanpassen van de afspraaktijd van nieuwe patiënten op de aanwezigheid van co-assistenten.

We evalueren 45 alternatieve afspraaksystemen die elk bestaan uit een combinatie van waarden van de bovenstaande experimentele factoren. Daarbij introduceren we nieuwe consulttypen, die het mogelijk maken om de consultduur van een patiënt vooraf beter te voorspellen. Een *discrete-event* simulatiemodel rekent de prestaties van deze alternatieve afspraaksystemen door, waarbij we gebruik maken van theoretische kansverdelingen voor onder andere de volgende inputparameters: aankomstintensiteit van patiënten, punctualiteit van artsen en patiënten, en de duur van consulten.

Resultaten en conclusies

Op basis van de simulatieresultaten zetten we de gemiddelde interne wachttijd van elk alternatief afspraaksysteem uit tegen diens gewogen gemiddelde van uitloop en benutting. Een efficiënte lijn langs die punten laat drie alternatieve afspraaksystemen zien die op deze gebieden het best presteren. Een *Data Envelopment Analysis* bevestigt dat er hiervan twee beter scoren dan de andere alternatieven, wanneer ook de toegangstijd (het aantal dagen dat een patiënt moet wachten tot hij bij een arts terecht kan) in de analyse wordt betrokken. De resultaten blijken echter zeer gevoelig te zijn voor een toename van het aantal afspraakverzoeken per jaar.

We concluderen dat het afspraaksysteem met een vaste blauwdruk van spreekuurplaatsen, waarbij slechts één patiënt wordt ingepland voor het eerste afspraaktijdstip van een spreekuur, gecombineerd met een volgorderegel die ervoor zorgt dat patiënten met een lage variantie in de consultduur in de eerste helft van het spreekuur worden gepland, kan zorgen voor een afname van 55 % van de interne wachttijd. Bovendien kan de gemiddelde uitloop hierbij teruglopen van 19 % tot 11 % van de spreekuurduur. Dit afspraaksysteem behoeft een beperkt aantal aanpassingen ten opzichte van het huidige afspraaksysteem. Echter, de benutting van dit alternatieve afspraaksysteem is slechts 89,5 %, tegen 93,3 % in de huidige situatie. Om desondanks hetzelfde aantal patiënten per jaar te kunnen behandelen, zal de capaciteit van de polikliniek met twee spreekuren per week uitgebreid moeten worden. Desondanks wegen de voordelen van dit alternatieve afspraaksysteem, zoals hierboven weergegeven, op tegen de terugloop van de benutting. De *Data Envelopment Analysis* ondersteunt deze conclusie. Daarom bevelen wij de implementatie van bovengenoemd afspraaksysteem in de polikliniek aan.

De lengte van de planningshorizon en het aanpassen van de afspraaktijd op de aanwezigheid van co-assistenten hebben slecht een beperkt effect op de prestaties van een afspraaksysteem. De invloed van de andere experimentele factoren hangt in hoge mate af van de combinatie waarin ze voorkomen. Een afspraaksysteem waarbij de afspraken dynamisch worden ingepland, zal de piekbelasting van het baliepersoneel niet verder vergroten, terwijl statisch inplannen vergt dat elke patiënt teruggebeld wordt om het definitieve afspraaktijdstip door te geven, en daarmee een grotere belasting voor het baliepersoneel vormt. Wanneer het aantal spreekuren beter gebalanceerd wordt over de week, kan dit de piekbelasting ten goede komen.

Table of contents

Preface

You are about to read my Master's thesis about the evaluation of alternative appointment scheduling systems for the ENT (Ear, Nose, and Throat) outpatient clinic in the Erasmus University Medical Center in Rotterdam, the Netherlands. This study aims at an improvement of the scheduling approach for outpatient appointments, leading to a reduction in waiting times for patients and an improvement of the outpatient clinic's utilization. This report has a number of purposes: it is a scientific study in the field of outpatient scheduling, to which I hope to contribute. This study is an advisory report for the Erasmus MC outpatient clinic as well, about the appointment system to implement. And finally, it is my Master's thesis for Industrial Engineering & Management. Whatever your purpose for reading this report is: I hope you enjoy reading it!

When I asked Erwin Hans more than a year ago about possible assignments for my Master's thesis, he talked for 1.5 hours about all the students he had coached in the Erasmus MC in Rotterdam, and I became more enthusiastic with every story he told. There was no question left: I was to perform my thesis in Rotterdam. The ENT outpatient department I found in September was, and still is, very eager on every improvement project. They welcomed me with open arms. I would like to thank all doctors, assistants and other staff for their cooperation with this project. Especially I thank Anne van Linge, the medical coordinator of the outpatient clinic, Hilger Jansen, the unit manager, and Rob Baatenburg-de Jong, the professor and head of the ENT department, for the opportunity they gave me to perform this study in their outpatient clinic, for the many fruitful discussions we had about this topic, and of course for their confidence in me: they offered me a job to implement this study. I am looking forward to it!

There are more people to thank. First of all I would like to thank my tutors, Erwin Hans, Marco Schutten, Jeroen van Oostrum, Gerhard Wullink, and Tim Nieberg for their useful comments and their patience. Dirk Denef, thanks for your accompany in the hospital since we both started with our thesis in September. You have probably had your beach-weather by now… Thanks as well for the other members of the Clusterbureau and the Health Care Logistics Group. And last but of course not least: my parents and Saskia for their support.

Rotterdam, July 2007
Jochem Westeneng

1. Introduction

The Ear, Nose, and Throat (ENT) outpatient clinic of Erasmus University Medical Center (Erasmus MC) in Rotterdam, the Netherlands is confronted with long waiting times for patients, overtime for doctors and nurses during clinic sessions, and peak workloads for its counter personnel. Research on outpatient clinics shows that waiting times are patients' main dissatisfaction with hospital services (Huang, 1994). According to doctors and personnel, overtime and peak workloads are potential threats for the quality of care and the quality of labor, because they increase stress and time pressure. This case study focuses on outpatient scheduling as a means to solve these problems for the ENT outpatient clinic in Erasmus MC.

1.1 Context description

Rising national health care expenditures and privatization developments put public and economic pressure on hospitals to improve utilization of resources (OECD, 2005; TPG, 2004). Additionally, waiting times in hospitals are of high importance for patients, politicians, and hospital managers. Outpatient clinics are essential services for hospitals. They perform a gatekeeper role and are often a patient's first contact with a hospital.

Erasmus MC is situated in the city of Rotterdam and is the Netherlands' largest university medical center. Table 1.1 denotes the three outpatient clinic locations of the Department of ENT. This study focuses on the *General ENT* and *Head and Neck Surgery and Oncology* clinic sessions of the Central Location, and is referred to as '*the outpatient clinic'* in the remainder of this report.

Figure 1.1 shows a generalized patient flow through the outpatient clinic up to the moment of consultation. The first and the second block in the figure, when read from left to right, indicate contact moments with the outpatient clinic's counter personnel. The last block represents the consultation by a doctor. The other time blocks denote waiting times. The total patient waiting time is divided in (1) access time and (2) internal waiting time. Access time is the time between the patient's

Table 1.1 Erasmus MC ENT outpatient clinics.

Location	Subspecialties (clinic sessions)
Central Location (Dijkzigt)	General Ear, Nose, and Throat Surgery
	Head and Neck Surgery and Oncology
	Speech, Language, and Hearing Centre
Daniël den Hoed Hospital	Head and Neck Oncology
Sophia Children's Hospital	Children's ENT Surgery

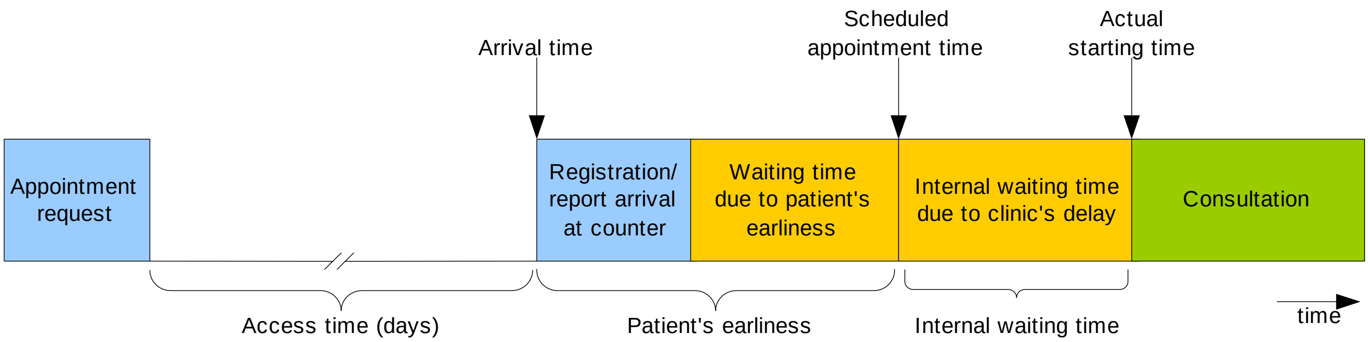


Figure 1.1 The patient flow through the outpatient clinic.

request for an appointment and his arrival at the outpatient clinic. A patient's internal waiting time is the period between the scheduled starting time and the actual starting time of his consultation. Waiting time due to a patient's early arrival is extracted from the internal waiting time, since it is not a consequence of the appointment system (Cayirli and Veral, 2003).

Bailey (1952) already announced that an appointment system is a trade-off between doctors' and patients' waiting times. Although the Erasmus MC outpatient clinic's average internal waiting times are long, doctors frequently have idle time. Patients who do not show up or who are late for their appointments cause idle time for doctors, leading to temporary underutilization of the outpatient clinic's capacity. Gaps in the appointment schedules also cause underutilization of the doctor's time.

1.2 Problem definition

Given the problem context, we formulate the problem as follows:

The patients of Erasmus MC's central ENT outpatient clinic often have long access times and long internal waiting times for their appointments, while the doctors face idle time and overtime.

1.3 Objectives

The aim of this study is to improve the appointment system for the Erasmus MC ENT outpatient clinic such that patients' internal waiting times and access times are shortened, the personnel's workload is balanced, and the outpatient clinic's utilization rate is increased at the same time.

We choose for outpatient scheduling in this project because it involves the construction of an appointment system to schedule appointments such that waiting times and overtime are minimized. Outpatient scheduling is a form of resource scheduling under uncertainty (Cayirli and Veral, 2003). Additionally, it can positively affect an outpatient clinic's access time and peak workload.

We construct a discrete-event computer simulation model to analyze alternative scheduling methods and to determine the most appropriate appointment scheduling system for the Erasmus MC ENT outpatient clinic. The performance indicators for the alternative appointment systems are listed below:

- Patients' access time
- Patients' internal waiting time
- Doctors' idle time
- Doctors' overtime
- Doctors' utilization rate
- Counter personnel's peak workload

1.4 Research questions and approach

We formulate the following research questions:

1. What is the performance of the current situation, with respect to the access times, capacity utilization, internal waiting times, idle time, overtime, and counter personnel's workload?
2. Which input parameters and control parameters and mechanisms for appointment systems are described in the literature, and which ones should be incorporated in the evaluation of alternative appointment systems for the Erasmus MC ENT outpatient clinic?
3. Which modeling technique is most suitable to evaluate these appointment systems?
4. Which alternative appointment systems are feasible to evaluate for the outpatient clinic?
5. What are values, averages, and/or variances of the following input parameters?
    - Yearly production
    - Percentage of cancelled appointments
    - Percentage of not-attending patients (no-shows)
    - Return rate of patients
    - Patients' and doctors' punctualities
    - Arrival rate of emergency patients
    - Durations of consultations

    (Chapter 2 introduces and explains these input parameters)
6. What probability distributions are suitable to model these input parameters?
7. What is the modeled performance of the alternative appointment systems?
8. Which alternative appointment systems are efficient on at least one of the performance indicators?
9. Which appointment system is best suitable for implementation in the outpatient clinic?

For a better understanding of the current situation, we analyze the processes and current method of appointment scheduling in the outpatient clinic. Time measurements, observations, and data mining in the hospital's management information system show the performance of the outpatient clinic in the current situation. Chapter 2 describes the context analysis and the performance of the current situation (research question 1).

Many authors contributed to the field of outpatient scheduling. Chapter 3 describes the most important input parameters and control parameters and mechanisms for appointment systems that we found in literature. We compare the characteristics of Erasmus MC's outpatient clinic to the models and case studies described in the literature and decide on the input parameters and control parameters and mechanisms that we include in the evaluation of alternative appointment systems (question 2).

We construct a model of the outpatient clinic to evaluate the appointment systems. Chapter 4 introduces the discrete-event simulation modeling technique, which enables to compare the alternatives easily, fast and realistic, without adjustments to the real world (question 3). The same chapter describes the construction of alternative appointment system configurations, i.e., the combination of control parameters and mechanisms forming complete appointment systems (question 4).

Chapter 5 describes the time measurements we perform to gather the input data for the model. The tables and graphs in this chapter show the values, frequencies, averages, and variances of these data (question 5). We analyze the data and fit theoretical probability distributions on the data to model the system properly (question 6).

Chapter 6 presents the results. We score the alternative appointment systems on each performance indicator separately. Additionally, we analyze the contribution of all control parameters and mechanisms, or experimental factors (question 7). A Data Envelopment Analysis (DEA) shows which alternative appointment systems are efficient on a combination of performance indicators (question 8). Finally, we recommend the outpatient clinic to implement one particular appointment system (question 9).

2. Context analysis

This chapter describes and analyzes the context of appointment scheduling in the Erasmus MC outpatient clinic of the ENT department. First, Section 2.1 describes the process flow of patients through the outpatient clinic. Section 2.2 shows the performance of the current situation, which includes tables and figures about the outpatient clinic's capacity, production, waiting times, and overtime.

2.1 Process flow

Figures 2.1 and 2.2 depict the detailed process flow of patients through the outpatient clinic. The processes are grouped: appointment-scheduling processes, processes at the counter, paramedical processes and doctors' processes. Paramedical processes concern the actions performed on the patient by paramedical personnel before or after the doctors' processes. This study focuses on the effect that changes in appointment scheduling processes have on the performance of the doctors' processes, and patients' waiting times.

2.1.1 Appointment-scheduling processes

Patients request appointments at the counter in the outpatient clinic, by phone or by email. Upon an appointment request, the scheduler performs a number of tasks:

The scheduler

- performs a triage on the patient's urgency,
- determines the patient's consultation type,
- determines which doctor(s) is/are able to treat this patient,
- determines an appropriate time slot code for this appointment,
- and schedules the patient.

With respect to urgency, the outpatient clinic distinguishes three categories:

- Elective patients (who may have a preferred appointment date and time, and are scheduled for an appointment sometime within the scheduling horizon),
- Urgent patients (who need treatment within a few days),
- Emergencies (who have to be treated by a doctor as soon as possible, and arrive to the outpatient clinic without appointment).

A patient's consultation type indicates the phase in the clinical pathway of a patient's treatment. Currently, the outpatient clinic uses two consultation types: *new* and *review*. A new patient either visits the outpatient clinic for the first time, or visits the outpatient clinic with a new medical complaint. A review patient has visited the outpatient clinic before with the current medical complaint.

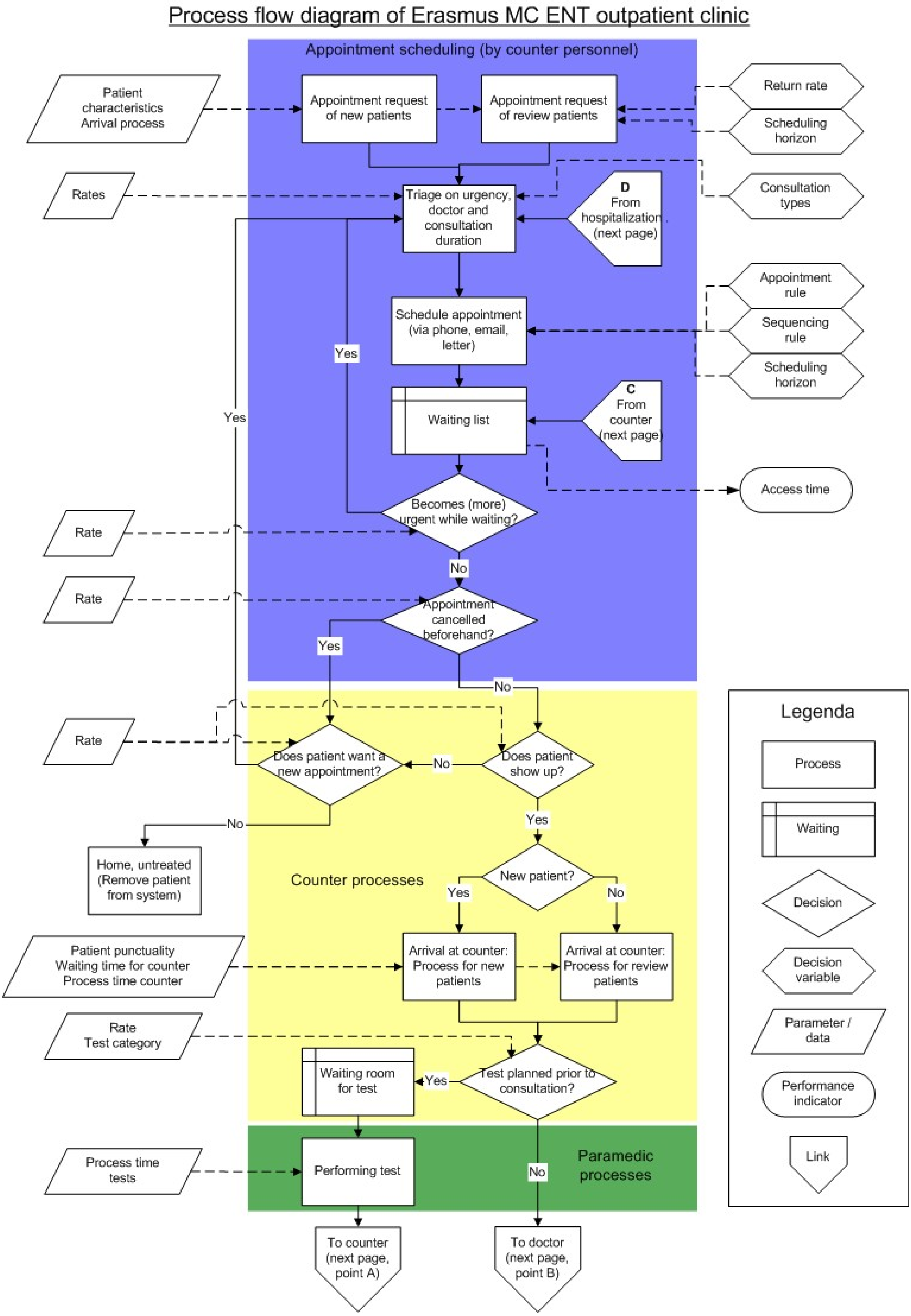


Figure 2.1 Erasmus MC process flow diagram (part 1/2)

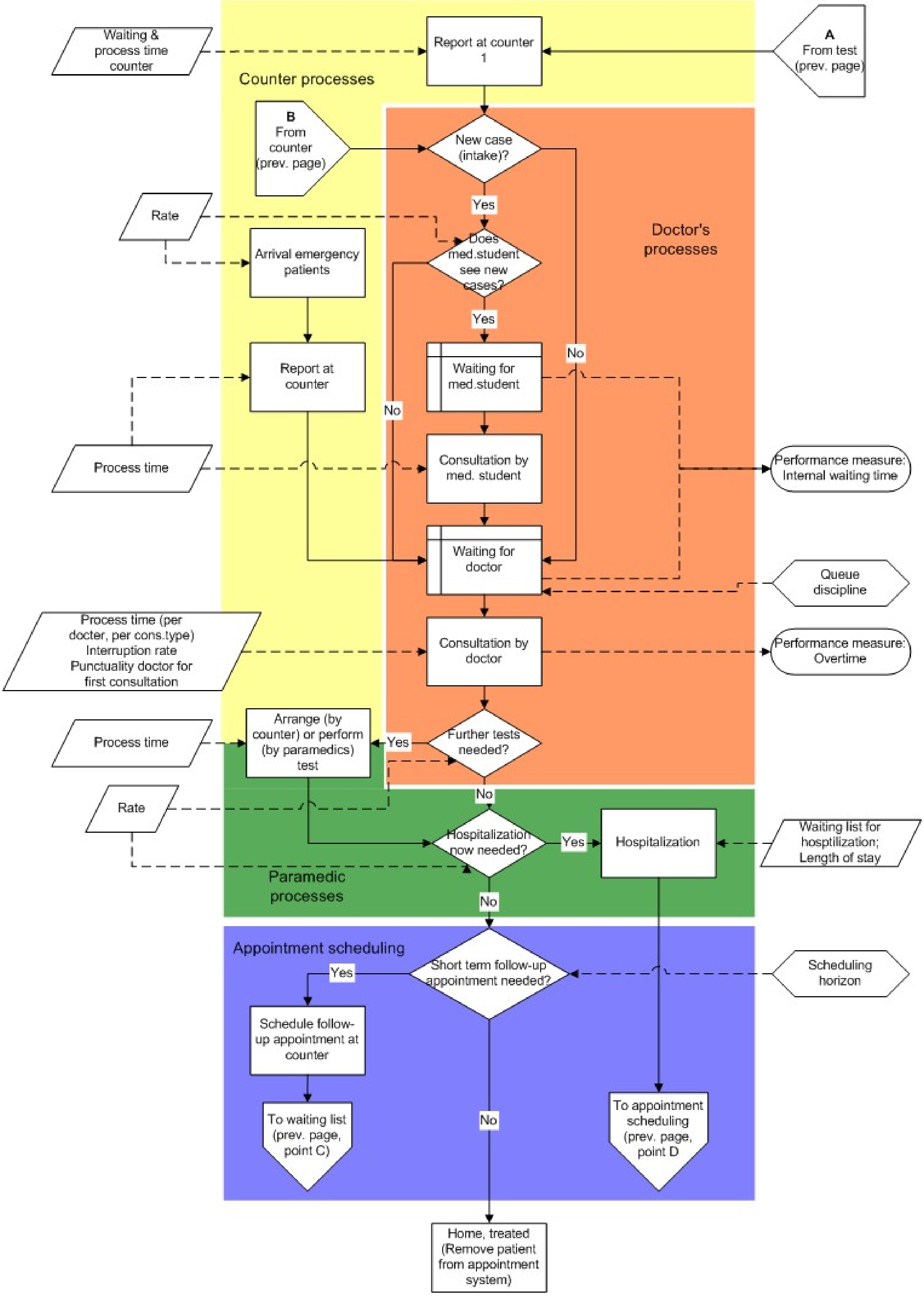


Figure 2.2 Erasmus MC process flow diagram (part 2/2)

The outpatient clinic employs about 23 doctors, of which 15 medical specialists and 8 resident doctors. Medical specialists are specialized in one subspecialty of ENT, such as otology, rhinology, or head and neck oncology. They restrict themselves to treat only patients in their own field. Resident doctors (in Dutch: *arts-assistenten, AIOS* or *ANIOS*) do have a medical degree (M.D.), but are not specialized (yet). Resident doctors treat most new patients who are referred directly to the Erasmus MC ENT outpatient clinic by their general practitioner (i.e. second-echelon patients). Review patients usually return to the doctor who treated them on their previous visit to the outpatient clinic.

In addition to the specialists and residents, medical students (in Dutch: *co-assistenten*) attend some of the clinic sessions of residents or specialists. Medical students consult the new patients in a clinic session separately from a resident or specialist, but they are not allowed to diagnose a patient or to perform medical actions themselves. Medical students always work under supervision of a resident or specialist. The attendance of medical students is not included in the appointment schedules in the current situation, resulting in long internal waiting times for review patients and idle time for residents and specialists.

A time slot is an appointment interval with a pre-defined length that is designated to patients with a certain consultation type or urgency class. The length and consultation types corresponding to the time slots are already specified when a new, empty appointment schedule is created. In other words: the outpatient clinic uses *dedicated* time slots, contrary to *pile-up scheduling*, where no time slots are specified in advance. Schedulers currently use over twenty different time slot codes for appointments, which can be grouped in:

- Regular time slots for *new* patients, with urgency *elective* or *urgent*;
- Regular time slots for *review* patients, with urgency *elective* or *urgent*;
- Overflow time slots for *urgent* patients only, either with consultation type *new* or *review*.

The length of a time slot varies form five minutes (overflow time slot) to one hour (new patients in an oncology specialist's clinic session).

When a patient is scheduled, he receives an appointment date and time. If an audiometric or other diagnostic test is scheduled prior to the consultation with the doctor, the appointment time is adjusted accordingly. This consultation with the doctor is scheduled to start at the beginning of a 5-minute interval and lasts for one or more 5-minute intervals. A doctor treats a number of patients successively during a clinic session, although idle periods during clinic sessions occur. We define a *shift* as a morning or afternoon of a working day, in which one or more parallel clinic sessions take place. Appointments are scheduled up to a certain number of shifts in the future, i.e. the rolling scheduling horizon. This scheduling horizon is currently about 3 months (140 shifts, equivalent to 70 working days). Appointments are scheduled with an *individual-block, variable-interval* appointment rule.

'Individual block' means there are no two patients scheduled for the same appointment time for the same doctor. The time difference between the patient's request for an appointment and the consultation is the access time. Cancelled clinic sessions and 'gaps' in the appointment schedule, in which patients could have been scheduled otherwise, cause longer access times for patients.

2.1.2 Counter and paramedic processes

Upon arrival in the outpatient clinic, all patients have to report their arrival at the counter. For new patients, the counter personnel create an empty medical status record and name labels and verify his address and insurance information. This takes some minutes and is in some cases a cause of a doctor's idle time, who is not informed about the arrival of a patient before the new status record has been created. If necessary, paramedic personnel perform diagnostic tests prior the consultation with the doctor. It is very seldom that these tests cause delays in the doctors' schedules. Therefore, we do not take the sub-department Audiometric and tests into account in the remainder of this study.

2.1.3 Doctor's processes

The sequence of calling patients from the waiting room (the queue discipline) generally follows the appointment schedule (first appointment first serve: FAFS). Patients that arrive late are skipped initially and are served at the first possible moment after their arrival, although individual doctors may deviate from this policy. The doctor on duty (an alternating task among residents) sees the emergency patients, who arrive without an appointment. The doctor on duty has thirty minutes of reserved time at the end of his clinic session to treat the emergencies.

We define the gross consultation or service time as all the time during which a patient claims a doctor's attention, or at least prevents him from seeing the next patient (Bailey, 1952). The gross consultation time consists of (1) preparation time, in which the doctor reads the patient's status record and test results and/or prepares material requirements for treatment, (2) net consultation time, which we define as all time the patient is in the consultation room, and (3) doctor's time for administration and cleaning when the patient has left the room. Medical students see most new patients prior to the patient's consultation with a resident or medical specialist. After most consultations, doctors need a few minutes to make calls, to talk with the outpatient clinic's assistants or to explain something to a medical student. We define these necessary interruptions, which cannot be related to any one patient directly, as *desirable inter-consultation time* (DICT). Some previous studies use the term *interruption level* for the DICT. A doctor has idle time if he is prevented from consulting in the time between two consultations, because there are no patients waiting to be seen (Cayirli and Veral, 2003). Doctors' idle time can be seen as *undesirable* inter-consultation time. Overtime is defined as the positive time difference between the scheduled completion time of the clinic session and the actual end of the doctor's administration time for the last patient. Figure 2.3 shows the coherence of these definitions.

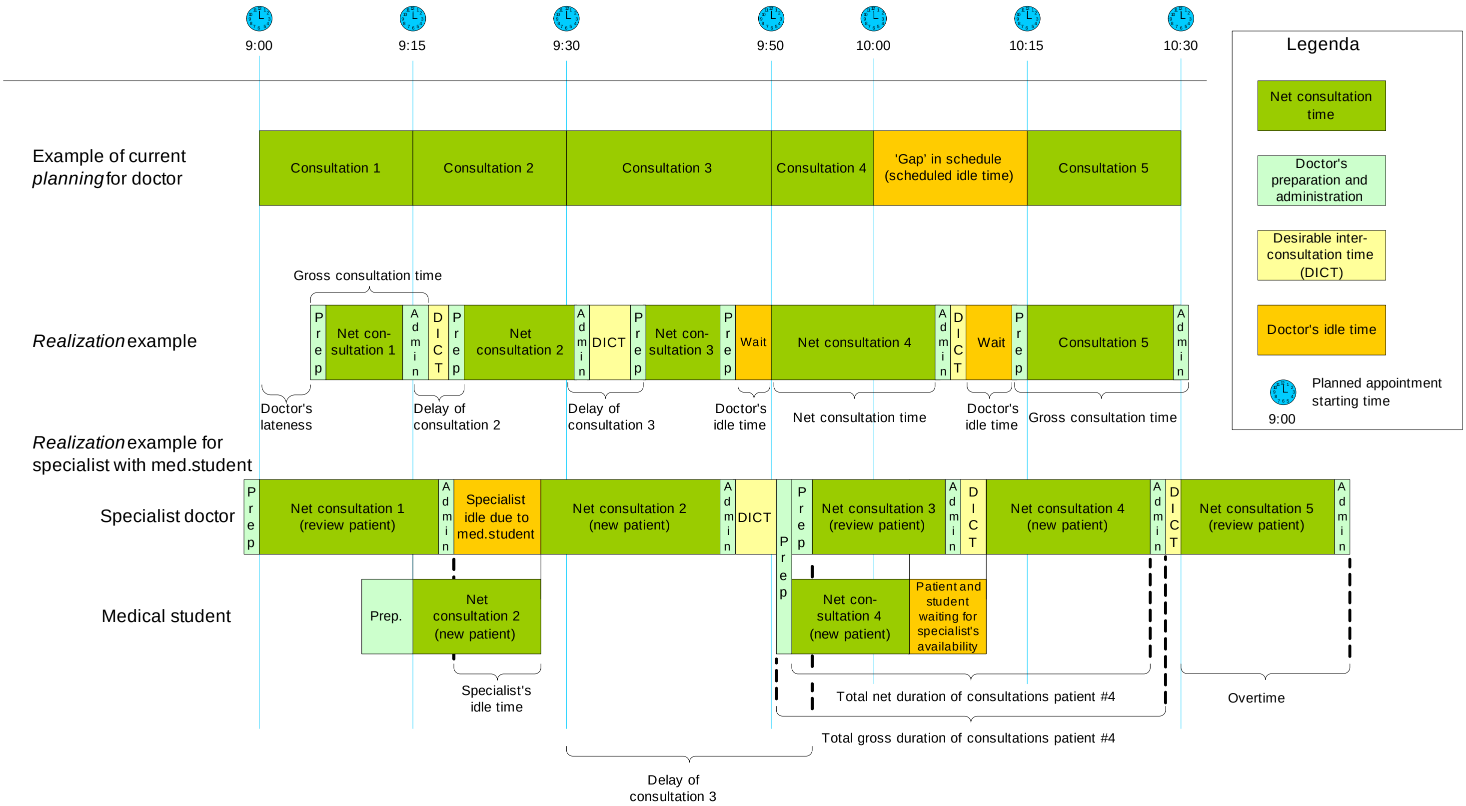


Figure 2.3 Consultation time definitions.

After a patient's consultation, a number of actions are possible, depending on the medical outcome of the consultation. These actions are listed below.

- The patient is cured. No review appointment is needed.
- The patient needs a follow-up appointment at the ENT outpatient clinic. The doctor gives an approximation for the new appointment date. If this date falls within the scheduling horizon, the patient requests a review appointment immediately at the counter. If this date is beyond the scheduling horizon, the patient has to call or (e-)mail later to schedule an appointment.
- The patient needs to be hospitalized. A nurse takes care of the necessary arrangements with the inpatient department.
- The patient needs (additional) diagnostic tests, either within or outside the ENT outpatient clinic. Nurses or counter employees schedule these tests, and paramedics perform some of them immediately.

Combinations of the above actions are possible. Hospitalization and tests are beyond the scope of this study.

2.2 Performance of the current situation

For a quantitative insight of the performance of the current situation, we measure the performance indicators that we stated in Section 1.3. Chapter 6 compares the performance of alternative appointment systems to the performance of the current situation.

2.2.1 Access time

We measure the access time in accordance with CBO (2004): the number of days until the third empty appointment slot. CBO (2004) ignore the first and second empty slots, because these slots have a large probability of being ‘occasionally empty’, for example if a patient has cancelled his appointment, and are therefore not representative for the access time (CBO, 2004). We do not measure the realized access time of scheduled patients, because review patients may request an appointment ‘in advance’ and do not want their access time to be as short as possible. The access time is measured weekly.

Table 2.1 shows the averages results of three measurements in April 2007 in the outpatient clinic. For residents and non-oncology medical specialists (i.e. specialists with another subspecialty than oncology) there is no difference in average access time between new and review patients. Oncologists, however, have extra capacity to ensure their new patients a short access time, which is needed for medical reasons.

2.2.2 Outpatient clinic’s capacity

The outpatient clinic schedules its clinic sessions in two shifts per day: mornings and afternoons. On average, there are 51 scheduled clinic sessions per week (27 morning sessions and 24 afternoon sessions) divided over five working days per week. Hence, on average there are about 5 parallel clinic sessions per shift. The length of a normal shift is 4 hours and 20 minutes for mornings (from 8:10 AM to 12:30 PM) and 2.5 hours for afternoons (1:30 to 4:00 PM), but exceptions occur for individual doctors.

Table 2.1 Average access time in working days.

Average access time in working days	New patients	Review patients
Residents	20.2	20.5
Non-oncology medical specialists	34.7	35.4
Oncology medical specialists	5.5	18.8

Hence, the capacity is 260 minutes for a morning session and 150 minutes for an afternoon session. This is equivalent to a maximum of 640 consultations per week. In 240 effective working days per year (46 weeks), the outpatient clinic is able to treat a maximum of 30,720 patients per year. However, this number is not reached for several reasons, as we observed:

- The above number of clinic sessions is based on the peak season capacity (October – April). There is less capacity during the summer period, due to vacancies of doctors and personnel, and fewer occurrences of ENT related diseases during the summer.
- A number of patients cancel their appointments shortly before the consultation date, which makes it difficult to schedule other patients in their appointment slots.
- Some patients do not show up for their appointment.
- Patients give up their appointment request when the access time is too long. Because there are several other hospitals in the area that treat second-echelon patients (i.e. patients who are referred directly to the Erasmus MC ENT outpatient clinic by their general practitioner), there is an alternative for patients in case they believe the access time is too long.

We take these factors into account in this study.

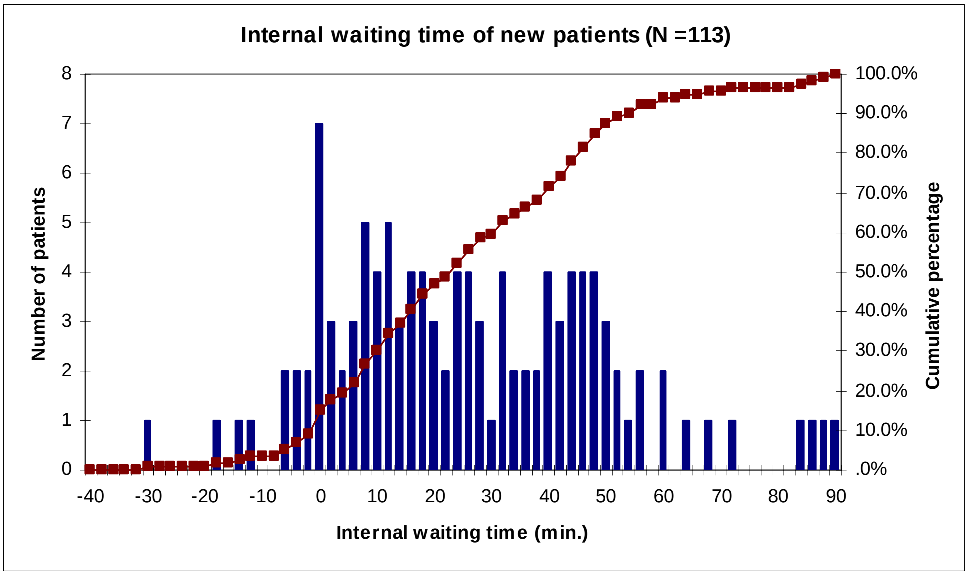
One of the objectives of this study is to improve the utilization rate of the outpatient clinic. However, we do not measure this by dividing the total realized treatment time by the capacity per year. Instead, we calculate the utilization per clinic session when we evaluate alternative appointment systems.

As one doctor told me, the doctors generally value the utilization of their time during the planned duration of a clinic session twice as important as the avoidance of overtime at the end of a clinic session. In other words, a clinic session A with 15 minutes of idle time during the planned duration and no overtime, is valued to perform equally well with regard to utilization as another clinic session B without idle time during the planned duration and 30 minutes of overtime.

2.2.3 Internal waiting time

Previous studies report that outpatient consultation durations have a coefficient of variation of 0.35 to 0.85 (Cayirli and Veral, 2003). This variability causes internal waiting times if appointments are scheduled too tightly. Other causes of internal waiting time for the Erasmus MC outpatient clinic are scheduled appointments with consultation times that do not reflect the expected duration, unexpected disturbances in consultations, urgent consultations, which cannot be scheduled in advance, and patients who arrive late.

Our time measurements of the internal waiting time in the outpatient clinic show the following results (Figure 2.4). We distinguish new patients and review patients, because new patients may have additional internal waiting time when a medical student sees them first.



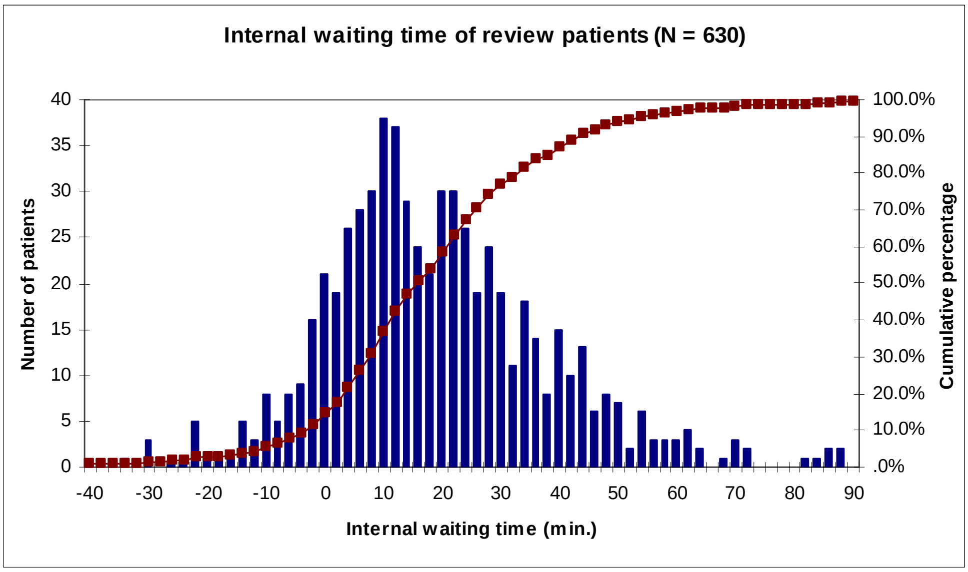


Figure 2.4 Internal waiting time of new and review patients in the current situation.

Table 2.2 Internal waiting time of patients in current situation

Internal waiting time of patients	New	Review
N	113	630
Mean (minutes)	25.6	18.0
Median (minutes)	23	16
Standard deviation (minutes)	23.6	18.3
Coefficient of variation	0.9	1.0
Percentage of patients with negative waiting time	10.6 %	12.2 %
Lower bound of 95% confidence interval of mean	21.2	16.6
Upper bound of 95% confidence interval of mean	29.9	19.4

From the frequencies in Figure 2.4 we calculate that a considerable amount of patients (25.7 %) waits more than 30 minutes from their scheduled appointment time until the actual start of the consultation. Patients with a negative internal waiting time arrived early and were treated before the scheduled appointment time.

Table 2.2 describes the basic statistics of the internal waiting time in the current situation. As we expected, new patients have a significantly longer internal waiting time. A considerable variation exists for both patient groups, since the internal waiting time varies from –30 minutes to 1.5 hours. Time measurements show that the internal waiting time increases towards the end of a clinic session. During the course of a clinic session, the variation in consultation durations and punctuality of patients accumulates, which results on average in longer internal waiting times.

2.2.4 Overtime

Although the overall capacity of the outpatient clinic is large enough to treat the yearly production of 22,000 to 23,000 patients, a large number of clinic sessions have overtime. Overtime has very different causes: emergency arrivals during the clinic session, a sequence of elective consultations with a longer duration than expected, the absence of a medical status record or materials, and late arriving patients and doctors are all causes of overtime we observed. However, there is no correlation between overtime and, for example, the amount of DICT incurred in a clinic session, or the punctuality of the doctor.

Of the 73 clinic sessions that we observed, only 16 sessions (22 %) had an overtime of at most 10 minutes. 33 sessions (45 %) had more than 30 minutes overtime. 12 sessions (16 %) had more than an hour overtime (Figure 2.5). These numbers show that overtime is a serious problem of the outpatient clinic. Doctors arrive late for other activities in the hospital, whereas counter personnel and paramedics are forced to stay longer at the outpatient clinic, which they dislike.

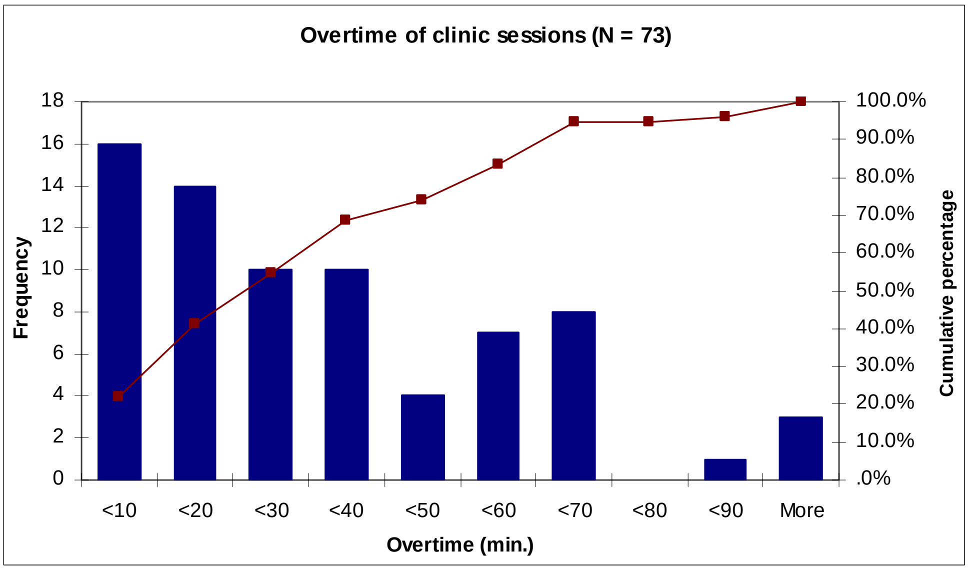


Figure 2.5 Overtime of clinic sessions

2.2.5 Counter personnel workload

Because all patients report their arrival to the counter, and the counter personnel schedules review appointments as well, the counter personnel's workload is related to the number of patients that visits the outpatient clinic. The counter personnel face peak workloads during some moments of a week, especially on Tuesday, Wednesday, and Friday mornings. Figure 2.6 and Figure 2.7 show these peak loads in number of clinic sessions per shift in the week, and the average number of patients treated per hour, respectively.

Other causes of peak workload for the personnel are the appointment requests by telephone. Many patients call on Mondays and Tuesdays, which results in an overload of the telephone lines and a peak workload for the counter personnel. At moments of peak workload for the counter personnel, patients have to queue for the counter and sometimes arrive late in the consultation room.

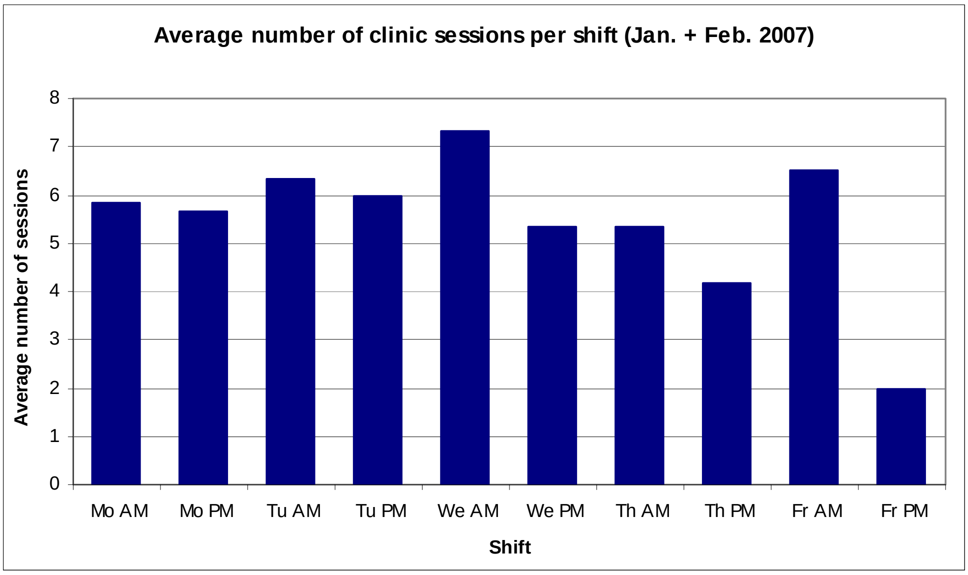


Figure 2.6 Average number of clinic sessions per shift in the week.

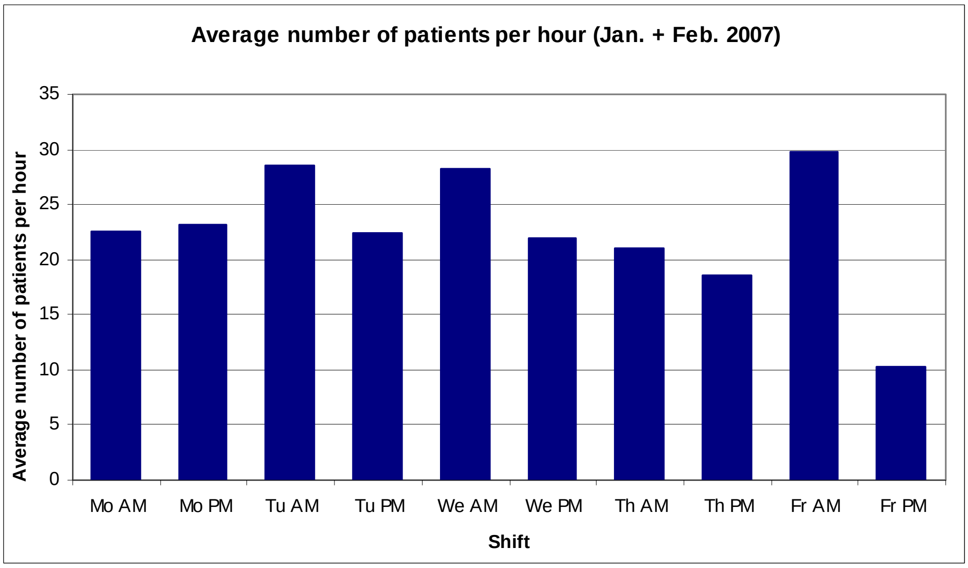


Figure 2.7 Average number of patients per hour.

3. Outpatient scheduling in the literature

3.1 General formal problem definition

Outpatient scheduling involves making patients' appointments for an outpatient clinic or general practitioner, and can be compared to scheduling an appointment with a hairdresser or bank office. Appointment scheduling is an adjusted form of a queuing process, where appointment times regulate the arrival process of patients to the system. However, in the application to an outpatient clinic, the steady state of a queuing process is never reached due to the limited number of patients per clinic session (Bailey, 1954). The aim of a general appointment system is to balance patients' waiting time, doctors' idle time, and doctor's overtime. We add the doctor's utilization rate and the counter personnel's workload to the objectives. Input parameters, or environmental factors, of the arrival processes and of the doctors' processes can be added to an appointment system model for realistic conditions. Some examples of such input parameters are no-show and walk-in rates, and punctuality.

In the formal problem definition of the general case of an appointment system (without additional input parameters), a doctor s sees patients during well-defined shifts $j=1,2,3,\ldots$ (clinic sessions) with a duration of $T \cdot q$ minutes (T intervals of q minutes). The appointment of patient p has a planned duration of y_{ps} successive intervals of q minutes. For interval t a number of n_{jts} patients are scheduled, which is called a block. All patients in a block are scheduled for the same appointment time, but are treated successively. If the sum of all appointment intervals for a doctor in a shift exceeds the shift duration $T \cdot q$, there is planned overtime.

3.2 Literature review

The objective of this literature review is to analyze the input parameters and appointment system control parameters and mechanisms used in previous studies, and to evaluate their applicability to the Erasmus MC outpatient clinic. Additionally, we work the other way around: we verify whether certain input parameters and control parameters and mechanisms we use, are described in the literature.

Cayirli and Veral (2003) present a comprehensive overview of the appointment-scheduling literature. Tables 3.1 and 3.2 show a summary of the most important articles, according to the literature review by Cayirli and Veral. We select studies that have commonality with our case. ENT and university outpatient clinics may for example differ from outpatient clinics in regional hospitals by the variability of consultation time and the range of doctors' subspecialties. Table 3.1 and the leftmost three columns and the rightmost two columns of Table 3.2 are directly retrieved from Cayirli and Veral (2003) and comprise methodology, control parameters and mechanisms, input parameters, and performance measures.

Table 3.1 Input parameters from the literature.

INPUT PARAMETERS	Service time distribution	Patient punctuality (mean, st.dev)	No-shows (p = no-show probability)	Walk-ins (regular and emergency)	Doctors' lateness	Doctors' interruption level
Erasmus MC ENT clinic	*Gamma*	*N(-13, 17)*	*p = 0.05*	*Emergency only*	*Late N(5,15) minutes*	*yes (DICT)*
Bailey (1952)	Gamma	Punctual	p = 0	None	Punctual	None
Blanco White & Pike (1964)	Gamma	Gamma, mu=0	p = 0, 0.09 and 0.19	None	0, 5, 10, 15 or 20 min.	None
Cayirli, Veral & Rosen (2004)	Lognormal	N (-15, 25)	p = 0 and 0.15	0 to 15%, also regular	Punctual	None
Cayirli, Veral & Rosen (2006)	Lognormal	N(0,25) and N(-15,25)	p = 0 and 0.15	0 to 15%, also regular	Punctual	None
Chen & Robinson (2005)	Randomly	Unpunctual, mu=0	p = 0	None	Punctual	None
Clague et al. (1997)	Randomly	Punctual	p = 0, .2, .3	None	Punctual	None
Denton & Gupta (2003)	Uniform, Gamma an Normal	Punctual	p = 0	None	Punctual	None
Fetter & Thompson (1966)	Empirically collected	Late allowed to max. 5 min.	p=[0.04-0.22] with mean 0.14	7 to 58% with mean 38%	0, 30 or 60 min	None
Fries & Marathe (1981)	Negative. Exponential	Punctual	p = 0	None	Punctual	None
Harper & Gamlin (2003)	Not specified	Unpunctual (mean 8.3 min early, SD=14.7 min)	p > 0 (not specified)	Urgent	Unpunctual	None
Ho, Lau & Li (1995)	Uniform, exponential	Punctual	p=0, 0.10, 0.20	None	Punctual	None
Hutzschenreuter (2004)	Triangular, Gamma	Unpunctual, (-10, 10)	p=0.10	None	Punctual	None
Kaandorp & Koole (2007)	Exponential	Punctual	p = 0, 0.1, 0.25, 0.5	None	Punctual	None
Klassen & Rohleder (1996)	Lognormal	Punctual	p = 0.05	Max 2 emergencies per session	Punctual	None
Klassen & Rohleder (2004)	Lognormal	Punctual	p = 0.05	10 % of patients	Punctual	None
Lehancy, Clarke & Paul (1999)	Not specified	Punctual	p = 0	None	Punctual	yes
Liu & Liu (1998)	Uniform, exponential, Weibull	Punctual	p = 0, 0.10, 0.20	None	Uniform over [0,6] min. late	None
Robinson & Chen (2003)	Generalized Lambda	Punctual	p = 0	None	Punctual	none
Rohleder & Klassen (2000)	Lognormal	Punctual	p = 0.05	Max. 2 emergencies per session	Punctual	None
Vanden Bosch, Dietz & Simeoni (1999)	Erlang	Punctual	p = 0	None	Punctual	None
Vissers & Wijngaard (1979)	General	In system earliness	Included by adjusting service times	Included by adjusting service times times	In system earliness	None
Welch & Bailey (1952)	Gamma	Punctual	p = 0	None	Punctual	None

Table 3.2 Control parameters and mechanisms from the literature.

CONTROL PARAMETERS ANS MECHANISMS	Methodology	Size Number of doctors (S); Number of patients per session (N) and Duration of session (T)	Appointment rule BW=Bailey-Welch; V=Variable; F=Fixed; I=Individual; N=Block; A=Interval	Sequencing rule	Patient classification	Adjustments on basis of patient classification	Scope	Stages	Queue discipline	Performance measurement pw=patients' waiting; di = doctors; idle time; do=doctors' overtime
Erasmus MC ENT clinic	*Simulation*	*S=10; T=150, 260 min, N varies*	*IN/VA & BW/VA*	*FCFA*	*New/return/tel.*	*Interval and sequencing*	*Rolling planning horizon*	*Four stages*	*FAFS*	*time: pw, di, do; utilization, workload*
Bailey (1952)	Simulation	S=1; N=10, 15, 20, 25; T=125 min.	BW		None	N/A	One session	Single stage	FCFS	time: pw, di, queue length, do
Blanco White & Pike (1964)	Simulation	S=1; N=10, 20, 30, 40, 50, 60; T=150 min	For punctual: BW, for unpunctual: VN/FA		Punctual/unpunctual	Appointment system	One session	Single stage	FCFS	time: pw, di and % patients within 30min.
Cayirli, Veral & Rosen (2004)	Simulation	S=1; N=10	VN/VA		New/return	Sequence and appointment interval	One session	Single stage	FAFS	time: pw, di, do, "fairness" of AS
Cayirli, Veral & Rosen (2006)	Simulation	S=1; N=10,20, T=210	VN/VA and solve for sequencing	Various	New/return	Sequence and appointment interval	One session	Single stage	FAFS, adj for late and walk-ins	time: pw, di, do
Chen & Robinson (2005)	Analytical	S=1; N=2	VN/VA		None	N/A	One session	Single stage	FAFS	time: pw, di
Clague et al. (1997)	Simulation	S=3; N=36-45	FN/F&VA		New/return	Interval	One session	Single stage	Chose shortest queue	time: pw, di
Denton & Gupta (2003)	Analytical	S=1; N=3, 5, 7	IN/VA		Mean service	Interval	One session	Single stage	FCFS	costs: pw, di, do
Fetter & Thompson (1966)	Simulation	S=3; N=26	IN/FA		Elective/walk-ins	Service times and sequencing	One session	Single stage	FCFS, walk-ins to first available	time: pw, di and #patients seen per session
Fries & Marathe (1981)	Analytical	S=1; N=24	VN/FA		None	N/A	Multiple sessions	Single stage	FCFS	time: pw, di, do
Harper & Gamlin (2003)	Simulation	S=22, N and T not specified	V&FN/V&FA	Various	5 classes	Block size & interval length	Ten sessions (1wk), 40 runs	Two to seven stages (varies per pt)	FCFS	time: pw
Ho, Lau & Li (1995)	Simulation	S=1; N=10, 20, 30	IN/VA		None	N/A	One session	Single stage	FCFS	time: pw, di
Hutzschenreuter (2004)	Simulation	S=1; N=6, 18; T=180	FN/FA	Various	Mean and SD of service time	Sequencing and intervals	One session, 300 runs	Single stage	FCFS	time: pw and doctor's utilization
Kaandorp & Koole (2007)	Analytical	S=1, N=8 to 20, T = 240	VN/VA		None	N/A	One session	Single stage	FCFS	time: pw, di, do
Klassen & Rohleder (1996)	Simulation	S=1; T=210min; N=19, 20, 21 (depends on urgent calls received)	IN/FA; 2 slots open for urgent walk-ins	LVBEG	Low/high variance in consultation time	Sequence	One session	Single stage	FCFS for regular	time: pw, di; mean and max completion times, % of urgent pt served
Klassen & Rohleder (2004)	Simulation	S=1, T=420min; N=48	IN/FA and BW	LVBEG	Low/high variance in consultation time	Sequence	10 day rolling horizon	Single stage	FCFS for regular	time: pw, di, do, server utilization, access time
Lehancy, Clarke & Paul (1999)	'Soft-simulation'	S=3; N=11	IN/VI	First short proc, times	None	N/A	One session	Multi-stage	FCFS	time: pw and other
Liu & Liu (1998)	Simulation	S=2, 3, 5; N = 46	VN/FA		None	N/A	One session	Single stage	FCFS	costs: p flow time, di
Robinson & Chen (2003)	Analytical	S=1; N=3, 5, 8, 12, 16	IN/VA	Pre-defined	None	N/A	One session	Single stage	FCFS	costs: pw, di
Rohleder & Klassen (2000)	Simulation	S=1; T=210min; N=19, 20, 21 (depends on urgent calls received)	IN/FA; 2 slots open for urgent walk-ins		Low/high variance in consultation time	Sequence	One session	Single stage	FCFS for regular	time: pw, di; mean and max completion times, % of urgent pt served, mean max waiting time, % pt waits < 10min, % pt who receive slot requested
Vanden Bosch, Dietz & Simeoni (1999)	Analytical	S=1, N and T vary	VN/FA		None	N/A	One session	Single stage	FCFS	cost: pw, do
Vissers & Wijngaard (1979)	Simulation	S=1; N=10, 20, 30 ,40, 50, 60	IN/FA and FN/FA		None	N/A	One session	Single stage	FCFS	time: pw, di
Welch & Bailey (1952)	Simulation	S=1, N=10, 15, 20, 25; T=125 min.	BW		None	N/A	One session	Single stage	FCFS	time: pw, di, queue length, do

For a detailed description of these factors, we refer to Cayirli and Veral (2003). We add the other columns with control parameters and mechanisms to Table 3.2 for a more sophisticated comparison with our case. The following sections briefly describe those input parameters and control parameters and mechanisms.

3.3 Input parameters

We define input parameters as the factors that serve as input for the modeled outpatient clinic and its appointment system, and that are not controllable by the outpatient clinic's management of medical staff. Table 3.1 lists the input parameter values for all reviewed articles. The following subsections correspond to the columns in this table.

3.3.1 Service times

The service time distribution describes the behavior of the consultation time. We assume a doctor always uses as much time for a patient as medically required, regardless of work pressure, waiting times or overtime. Therefore, the management cannot control the service times. All reviewed studies draw the service time of a consultation with a doctor from a continuous probability distribution, but the distributions they use vary widely. Some of the distributions we found in the literature are Gamma (Bailey, 1952; Vanden Bosch and Dietz, 2000, and others), lognormal (Cayirli, Veral, and Rosen, 2006; Rohleder and Klassen, 2000, and others), Weibull (Liu and Liu, 1998), uniform (Ho, Lau and Li, 1995, and others) and (negative) exponential (Fries and Marathe, 1981; Liu and Liu, 1998). Some of them are fitted to empirical data (Bailey, 1952; Blanco White and Pike, 1964; Fetter and Thompson, 1966; Robinson and Chen, 2003).

3.3.2 Patients' arrival processes

Four input parameters describe the arrival process of patients: their punctuality with regard to the scheduled appointment time, the percentage of patients that cancels the appointment in advance, the percentage of patients that do not show up (no-show rate), and the number or percentage of patients that arrive without appointment (walk-ins), such as emergencies. The outpatient clinic cannot control the arrival process of patients easily, although this process has a significant impact on the waiting time, idle time and overtime of both patients and doctors (Blanco White and Pike, 1964). Therefore it is remarkable that only a few studies include punctuality, a no-show rate and walk-ins in their analyses. To our knowledge, there are no studies that include appointment cancellations.

With regard to patients' punctualities, Blanco White and Pike (1964) were the first to measure the arrival times and fitted a Gamma distribution. Most recent studies do allow for early and late arrivals and show a mean arrival time of zero to fifteen minutes prior to the scheduled appointment time (for example: Cayirli, Veral, and Rosen, 2004 and 2006, and Hutzschenreuter, 2004). Vissers and Wijngaard (1979) calculate a *system earliness factor* for their simulation model, in which they include the mean punctuality of patients and doctors.

Some patients cancel their appointment before the execution date. These patients may request another appointment. If a patient cancels his appointment, the outpatient clinic faces a 'gap' in the appointment schedule. This results in scheduled idle time, if the schedulers are unable to schedule another patient for this appointment interval. To our knowledge, there are no studies that include appointment cancellations.

Patients who do not cancelled their appointment in advance, and who do not show up, are classified as a *no-show*. The no-show rate of patients is included as a fixed value in the majority of studies, and varies from 0% to 20%. Fetter and Thompson (1966) use a stochastic no-show probability from a uniform distribution.

Unscheduled patients who arrive during the course of the day are generally called *walk-ins*, and include emergencies. Rohleder and Klassen (2000) allow for two emergency arrivals per clinic session in their appointment schedule, by leaving two appointment slots free. Cayirli, Veral, and Rosen (2006), and Fetter and Thompson (1966) add a certain percentage to the number of arriving patients, who they classify as *walk-ins*. Vissers and Wijngaard (1979) adjust the mean and variance of the service time to correct for *no-shows* and *walk-ins*.

3.3.3 Punctuality of doctors

When the doctor arrives late for the first appointment, the internal waiting times of all patients in the clinic session are affected by this delay. The percentage of doctors that do not arrive in time for the first appointment of their clinic session, is described by the input parameter *doctors' lateness*. Most studies assume that doctors arrive punctually, although the simulation study of Blanco White and Pike (1964) allows doctors to arrive late 0 to 20 minutes, in steps of 5 minutes. Fetter and Thompson (1966) assume doctors to arrive 0, 30 or 60 minutes late. Liu and Liu (1998) model the doctors' lateness uniformly on a 0 to 6 minutes interval.

3.3.4 Doctors' interruption levels

The doctors' interruption level indicates the percentage of time during a clinic session in which a doctor has to perform other tasks, and therefore cannot see patients. To our knowledge, Lehancy, Clarke, and Paul (1999) are the only authors who take the interruption level of doctors into account. We find this remarkable, since Patel *et al.* (2002) show that on average 41% of the doctor's time during a session is not spent on patients.

3.4 Control parameters and mechanisms

Control parameters and mechanisms are, contrary to with input parameters, subject to management control. This section lists the control parameters and mechanisms that Table 3.2 shows.

3.4.1 Number of parallel doctor's and doctor classification

One of the common characteristics of almost all literature on outpatient scheduling is the limitation of the model to a single doctor with one queue. Exceptions are Fetter and Thompson (1966), who schedule patients for three doctors, and the simulation model of Liu and Liu (1998), which includes two to five doctors. In both cases patients are scheduled in one queue and assigned to the first doctor available when they arrive. This implies that all doctors are able to treat all types of patients. To our knowledge, there are no studies that use a classification of doctors and assign patients of different types to these doctors.

3.4.2 Appointment rule

This subsection uses notation of the general formal problem description, introduced in Section 3.1, to clarify different appointment rules. An appointment rule is the combination of the block sizes (n_{jts} for $t > 1$), the size of the initial block (n_{j1s}) and the appointment intervals ($q \cdot y_{ps}$). Bailey (1952) and Welch and Bailey (1952) concluded in their pioneering studies that for one doctor the following appointment rule performs the best on a trade-off between patients' waiting times and doctor's idle time: an initial block $n_{j1s} = 2$, successive blocks $n_{jts} = 1$ (t > 1) and a fixed appointment interval $q \cdot y_{ps}$ that equals the mean service duration. Ever since, this is called the Bailey-Welch rule. Hutzschenreuter (2004) shows that the Bailey-Welch rule still performs well in case no-shows and punctuality of patients are included in the model. However, very little research has been done on the performance of appointment rules in cases with multiple scheduling shifts and multiple stages.

Both block sizes and appointment intervals can be chosen fixed or variable in an appointment system. Both are fixed in the Bailey-Welch rule, as described above. Fries and Marathe (1981) and Liu and Liu (1998) show examples of variable-block/fixed-interval systems. Denton and Gupta (2003), Ho, Lau, and Li (1995) and Robinson and Chen (2003) used fixed (individual)-block/variable-interval systems for their analyses. Variable-block/variable-interval systems are found in studies by, for example, Cayirli, Veral, and Rosen (2004 and 2006) and Chen and Robinson (2005). See also Figure 3.1.

		Block size	
		Fixed	*Variable*
Appointment interval	*Fixed*	Bailey (1952), Welch and Bailey (1952)	Fries and Marathe (1981), Liu and Liu (1998)
	Variable	Denton and Gupta (2003), Ho, Lau, and Li (1995), Robinson and Chen (2003)	Cayirli, Veral, and Rosen (2004 and 2006), Chen and Robinson (2005)

Figure 3.1 Examples of appointment rules

3.4.3 Sequencing rule

The sequencing rule determines the order of scheduling patients for a clinic session. Klassen and Rohleder (1996) conclude that scheduling patients with a low variance of their service time at the beginning of the clinic session, performed well under most circumstances. Hutzschenreuter (2004) reviews five sequencing rules based on expected service times and shows that best doctor utilization is reached when patients with low expected service times are scheduled towards the beginning of the clinic session, when combined with the Bailey-Welch appointment rule for this group of patients.

3.4.4 Patient classification

Patients differ in terms of service time characteristics, diagnosis, and urgency. If patients are grouped on basis of service time characteristics, the scheduler can adjust the sequencing and/or the appointment interval accordingly. A limited number of studies use a form of patient classification as a scheduling control mechanism (Harper and Gamlin, 2003; Cayirli, Veral, and Rosen, 2004 and 2006; Klassen and Rohleder, 1996; Lau and Lau, 2000; and others). It is also possible to group patients on basis of diagnosis and use this classification while assigning patients to doctors. Doctors may be able to treat only one or more types of diagnoses. Although this is very common in practice, we do not know an example of this type of patient classification in the outpatient scheduling literature.

3.4.5 Scope and scheduling horizon

The scope of a model describes the number of successive shifts that are reviewed in a model. The scheduling horizon determines the number of shifts for which appointment can be scheduled in advance. Most studies have a scope of one shift, which is equal to one clinic session if only one doctor is reviewed. However, Klassen and Rohleder (2004) review a multi-shift environment.

3.4.6 Number of stages

An appointment-scheduling model has multiple stages if it includes, for example, diagnostic tests or consultations with medical students before or after the consultation with a specialist. In that case, the patient 'flows' through a series of process steps. All reviewed articles consider a single-stage process, while we model a multi-stage process.

3.4.7 Queue discipline

The queue discipline determines the order of calling patients from the waiting room by the doctor. Most studies use a first-come-first-serve (FCFS) order, which is based on the arrival times of patients and neglects the original sequence in the appointment schedule. Cayirli, Veral, and Rosen (2004) argue that this sequence does not completely reflect reality in case early and late patient arrivals are allowed, but it is relatively easy to model. Cayirli, Veral, and Rosen (2004) and Chen and Robinson (2005) model the more realistic first-appointment-first-serve (FAFS) queue discipline.

3.5 Conclusion

We believe that all input parameters Section 3.3 mentions, are important for a realistic and valid representation of reality in a model. We measure the service times and interruption levels (the percentage of DICT) explicitly, as well as all parameters that describe the patients' arrival processes and doctors' punctuality. To our knowledge, we are the first to incorporate all those factors in one model.

With respect to the control parameters and mechanisms, to our knowledge there are no previous studies that use a doctor classification applicable to our case, where residents' patients can be scheduled for all residents, whereas all other patients have their own medical specialist. As the literature shows, adjusting the appointment rule and the sequencing rule offers improvement possibilities. Therefore, we include these rules as experimental factors in our model. Sequencing patients based on the expected consultation duration and its variance requires a clear patient classification, in which each patient group has its own service time characteristics. Section 4.2 formulates new consultation types for this purpose. We are particularly interested in the performance of an appointment system over multiple days, for example to evaluate the access time. Chapter 4 includes the length of the scheduling horizon as an experimental factor for our multi-day environment. The outpatient clinic we review has multiple stages, which do not change over time. Therefore, we create a multi-stage model, but we do not vary the number of stages by the experimental factors. The queue discipline FAFS does not change as well.

Additionally to these control parameters and mechanisms, we are interested in the evaluation of the method of decision-making, the usage of time slots versus pile-up scheduling, and the effect of adjusting appointment times for the attendance of medical students. Cayirli and Veral (2003) describe the method of decision-making briefly, but do not advice one of them. To our knowledge, there are no previous studies that evaluate the usage of time slots or the attendance of medical students. Chapter 4 describes the experimental factors we evaluate in detail.

4. Experimental design

To improve the performance of the outpatient clinic, we evaluate alternative appointment systems and compare their performances to the current situation. This chapter describes the evaluation approach and the simulation model, as well as the considered configurations of appointment systems.

4.1 Evaluation approach

To improve the outpatient clinic's appointment system, and to achieve the objectives, we create a model of the current situation. This model represents the real outpatient clinic. The scope (breath) and level (depth) of the model determine respectively which parts of reality are represented, in which level of detail. A validated model can be treated as a legitimate representation of reality.

We formulate experimental factors for alternative appointment systems, based on literature research, interviews, staff meetings and observations. The experimental factors are the methods by which the objectives might be achieved (Robinson, 1994). As the objectives state *what* should be achieved, the experimental factors describe *how* these objectives might be achieved. Experimental factors can be quantitative or qualitative. Quantitative factors assume numerical values, whereas qualitative factors represent the structural decisions (Law and Kelton, 2000).

We formulate the following controllable experimental factors:

- **Method of decision-making**
  The method of decision-making involves the possibility to schedule patients in batches at the same time (static scheduling), or immediately when the patient requests an appointment (dynamic scheduling). This is a quantitative experimental factor, since it specifies the number of shifts between scheduling two batches. This is a positive value for static scheduling, and zero for dynamic scheduling.
- **Length of scheduling horizon**
  The rolling scheduling horizon is the number of shifts (half working days) for which appointment can be scheduled in advance. This is a quantitative experimental factor.
- **Usage of time slots**
  In case the outpatient clinic uses dedicated time slots, appointment intervals are dedicated to certain patient classes in advance. An alternative is pile-up scheduling. The usage of time slots is a qualitative experimental factor.
- **Sequencing rule**
  The sequencing rule determines the order of scheduling patients for a clinic session. This is a qualitative experimental factor.

- **Appointment rule**
  The appointment rule is the combination of the block sizes, the size of the initial block and the appointment intervals. This experimental factor has qualitative and quantitative aspects.
- **Correction for medical students**
  The scheduler can adjust the appointment times of new patients in a clinic session in which a medical student attends. When he asks patients to arrive early, the scheduler corrects for the extra consultation with a medical student. This quantitative experimental factor determines with how many minutes the appointment times of new patients are corrected.

Section 4.4 describes the experimental factors in detail, as well as their possible values in the model. A combination of the values of the experimental factors is a *configuration*.

4.2 New patient classification

In this study we introduce a new form of patient classification. A limited number of studies use a form of patient classification in appointment scheduling (Harper and Gamlin, 2003*;* Cayirli, Veral, and Rosen, 2004 and 2006; Klassen and Rohleder, 1996; Lau and Lau, 2000; and others). However, as we conclude in Section 3.5, these patient classifications are insufficient to forecast the consultation duration, or are not easy to use by the counter personnel. In our opinion, a patient classification on basis of the steps in the care pathway enables the schedulers to forecast a patient's consultation duration properly, and is easy to understand and use as well.

The care pathway is the sequence of healthcare processes a patient undergoes between his first hospital visit and the moment he is cured. We call one such process a step in the care pathway. Examples of these steps are *'intake at the outpatient clinic'*, *'clinical surgery'*, and *'periodical monitoring at the outpatient clinic'*. The exact sequence of these steps varies between patients, but is generally as follows:
Intake (outpatient) → Diagnostics (outpatient) → Surgery → Inpatient care → Follow-up treatment (outpatient) → Periodical monitoring (outpatient).

In dialogue with the residents and medical specialists we created a list of seven new consultation types; the majority is derived from the steps in the care pathway that take place at the outpatient clinic (Table 4.1). The consultation types with the numbers 1 to 5 each represent a step in the care pathway. However, if these consultations take place telephonically we use the sixth consultation type, irrespective of the step in the care pathway, since telephonic consultations significantly vary in duration. The seventh consultation type is included for all consultations that do not coincide with one of the other consultation types. Note that a patient does not necessarily need an appointment for each step in the care pathway (e.g. not all patients need a treatment or surgery).

Table 4.1 Consultation types.

No.	New consultation type	Definition	Corresponding current consultation type
1	Intake: second echelon	New patient, referred by his general practitioner.	New
2	Intake: third or fourth echelon	New patient, referred by another medical specialist.	New
3	Diagnostics	Follow-up, in which the diagnosis is identified and/or a (possible) treatment is explained to the patient.	Review
4	Follow-up treatment	Consultation, which is directly related to a surgery in the nearby past.	Review
5	Periodical monitoring	Follow-up monitoring or treatment on a periodical basis, not directly related to a surgery in the past.	Review
6	Telephonic consultation	Consultation by phone.	Review
7	Other consultation	Any other consultation type.	-

The *intake* consultation types (numbers 1 and 2) correspond to the traditional consultation type *new*. Consultation types 3 to 6 are covered by the traditional type *review*.

4.3 Model description

4.3.1 Discrete event simulation

We use simulation to evaluate and compare alternative appointment systems. Simulation is a means to evaluate and analyze the performance of a system mathematically (Figure 4.1). In this case, the system is the outpatient clinic, whereas the formal problem description in Subsection 3.1 and Appendix B is a mathematical model of this system. We chose for simulation because this mathematical model is too complex to solve analytically in the same level of detail as we intend. As Law and Kelton (2000) state, for a complex system, its mathematical model is complex itself and the model must be studied by means of simulation. Apart from the possibility to evaluate complex systems, simulation has other advantages, which we list below.

- **Risk reduction**
  Simulation provides the possibility to estimate the performance of the system under other conditions than in the current situation, without making changes to the actual system (Law and Kelton, 2000). Using simulation, there are no risks involved in the evaluation of configurations. If all alternative appointment systems would be evaluated in the actual outpatient clinic, significant financial, social and perhaps medical risks would be involved.
- **Flexibility**
  Simulation allows for evaluation of the performance of multiple systems over a long period in compressed time (Law and Kelton, 2000). It is, for example, possible to evaluate the outpatient clinic's performance over a period of two years, by a simulation run of one hour.
- **Enhanced understanding of the situation**
  A computer simulation with a graphical interface and animation effects offers the management insight in the model and can be persuasive in decision-making. Moreover, the complete sequence of processes and its bottlenecks are clear for all stakeholders. A simulation model can act as a motivator to discuss about other problems.

Using simulation, the modeler and the stakeholders must be conscious of the stochastic nature of simulation, which causes the model to produce only estimates of the performance. Multiple or long runs limit the of the stochastic nature. The relatively long development period should be explicitly included in the project planning. If the model is not valid representation of the actual system, the stakeholders must treat the output results with care (Law and Kelton, 2000).

Time plays an important role in our system, since most of our performance measures are time related. The events, such as the transitions between the blocks in Figure 2.1 in Chapter 2, only occur on designated points in time. An event is 'an instantaneous occurrence that may change the state of the system' (Law and Kelton, 2000, pp. 6). This is the case for, for example, the arrival times of patients and doctors. In other words, the system does not change continuously but only at a countable number of points in time. Therefore, discrete-event simulation is an applicable simulation approach.

We construct our simulation model in the simulation package eM-Plant version 7.0 by Tecnomatix.

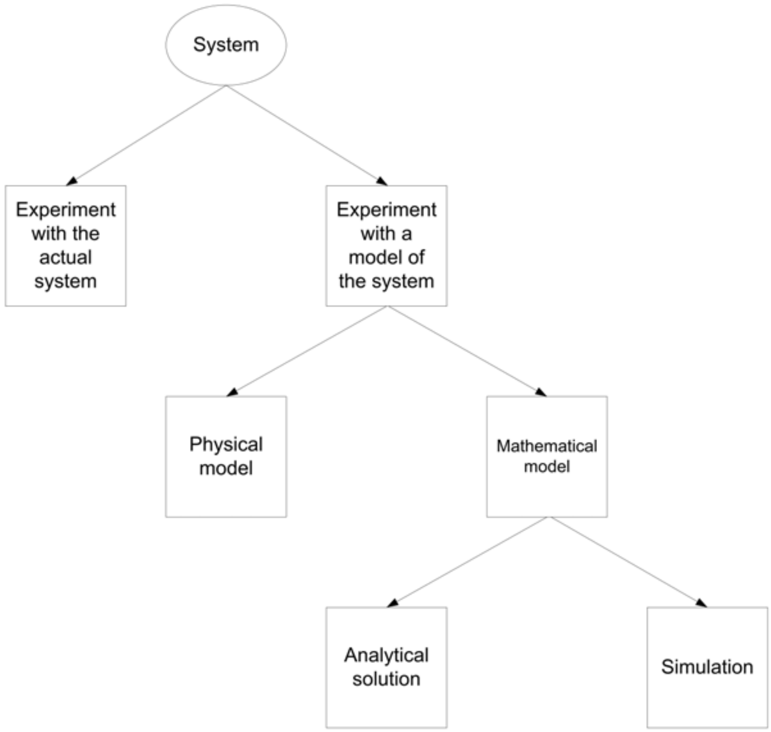


Figure 4.1 Ways to study a system (Law and Kelton, 2000, pp. 4)

4.3.2 Scope of the model

The scope is the range or the breath of the model, and specifies what is included in the model (Robinson, 1994). Recapitulating the process flow diagrams from Chapter 2 (Figure 2.1 and Figure 2.2), our model is restricted to the appointment scheduling processes, the counter processes, and the doctors´ processes (respectively the blue, yellow, and orange blocks in these figures). The process flow diagrams in Figures 4.2 to 4.4 depict the scope of the model. The available capacity and stochastic process times are included for all processes, except for the *Schedule appointment* process. The model includes the doctors' work schedule as well.

Based on interviews with doctors and personnel, and on observations in the outpatient clinic, we assume that the paramedic processes have little impact on the performance indicators of the appointment system, and therefore these processes are beyond the scope of the model. Hospitalization and combined appointments with other outpatient departments are beyond the model scope as well.

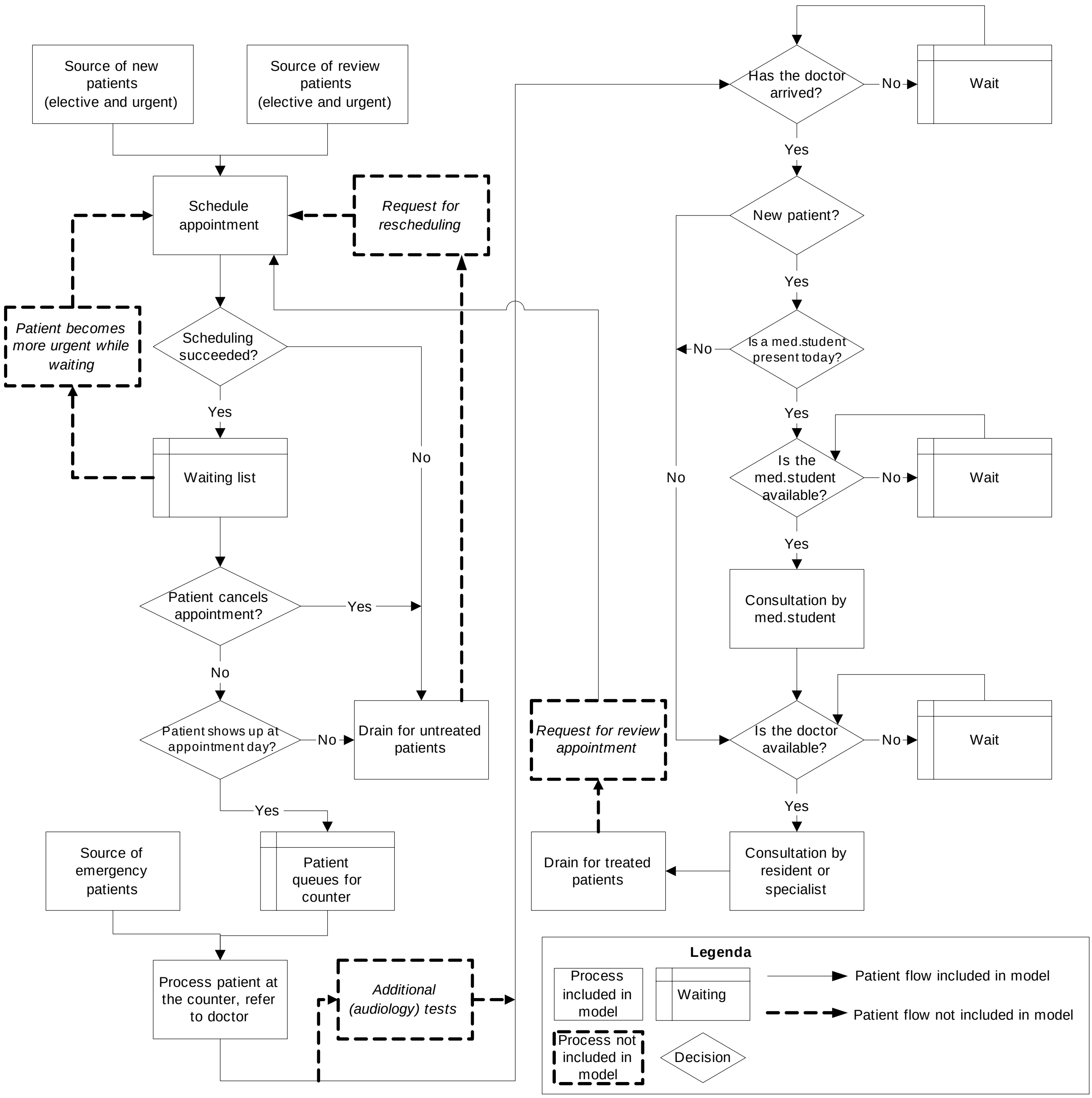


Figure 4.2 Patient's process flow of the simulation model.

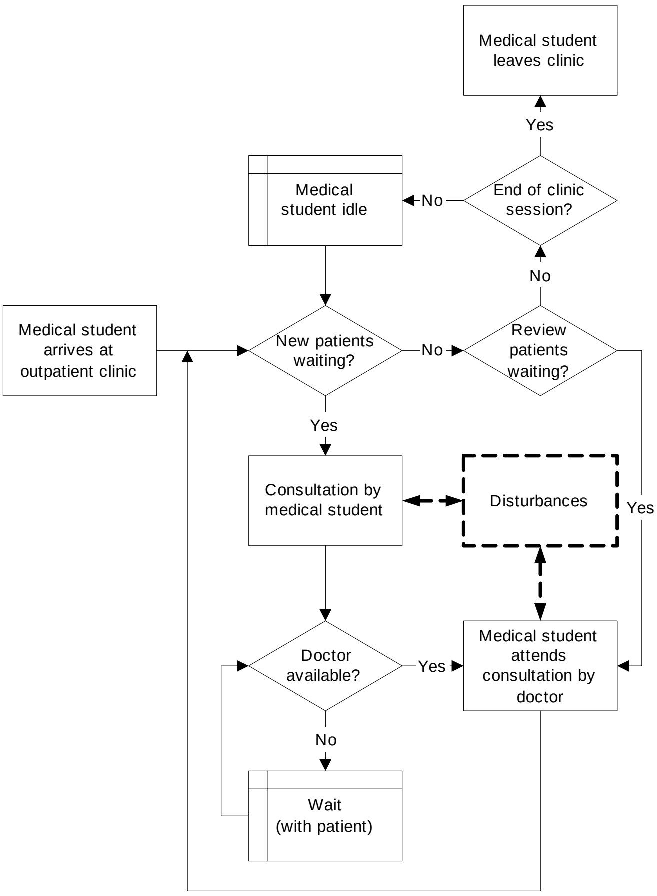


Figure 4.3 Doctor's process flow diagram of simulation model. For legend: see Figure 4.2.

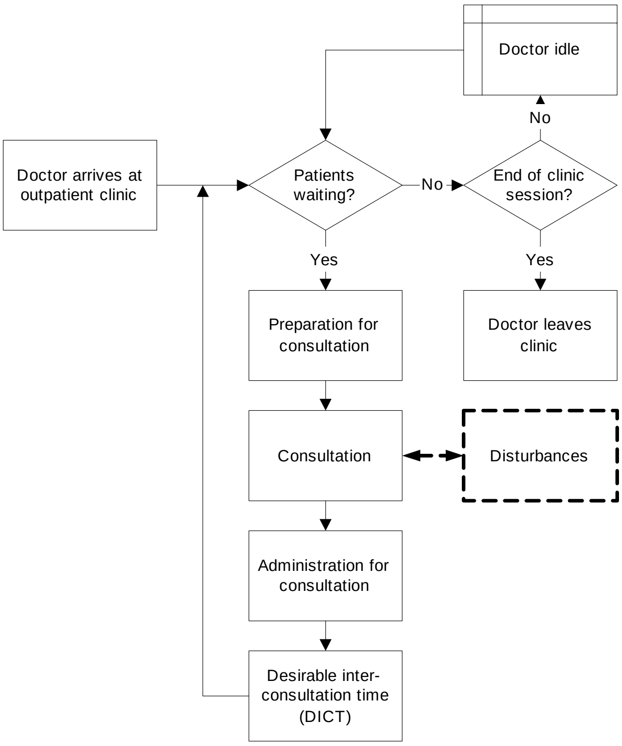


Figure 4.4 Medical student's process flow of the simulation model. For legend: see Figure 4.2.

Incurring these assumptions, the model comprises of three to four stages, depending on the attendance of medical students:

- Schedule appointment,
- Process patient at counter,
- (Consultation by medical student),
- Consultation by resident or specialist.

4.3.3 Level and assumptions of the model

The model's level is the amount of detail to be modeled, or the model's depth (Robinson, 1994). Because it is practically impossible to model all details of the real system, we assume the following:

- The modeled arrival rate of patients' appointment requests does not change over time. In other words: the modeled outpatient clinic receives on average the same amount of appointment requests during all days of the week. This does not reflect reality, since the outpatient clinic receives more requests on Mondays. However, we do not consider this as a serious shortcoming, since the total number of patients remains unchanged and the appointments spread out over the week to fill the capacity. We do not model seasonal demand effects, but model a continuous duration of a peak season instead.

- Fluctuating access times are often used as measures for the (in)balance between the production and the demand (CBO, 2004). Since both production and demand are no experimental factors in the model and they are estimated on basis of real data, we do not evaluate this imbalance.

- In case of dynamic scheduling, a patient's appointment is scheduled on a first-come-first-serve basis. This means that a patient is scheduled directly by the scheduler once he requests an appointment. Since the capacity of the scheduler to schedule patients is not experienced as a bottleneck, the model does not incur processing time to schedule a patient.

- Patients prefer certain days and appointment times when they request an appointment. However, patients are less selective when access times are long. We simplify the appointment scheduling process by neglecting the day and time preferences of patients. Instead, we always schedule a patient at the earliest possibility. As a result, the utilization rate outputs of the model should be treated as upper bounds to the utilization rate, especially for short access times. We do take duration, slot code, urgency and doctor constraints into account while scheduling a patient. Subsection 4.3.5 describes the scheduling routine in detail.

- If it is impossible to schedule a patient within the scheduling horizon for his own doctor, scheduling fails. In reality, the scheduler asks the doctor for his medical opinion whether or not to find a creative solution for this patient (e.g. to classify the patient as an *emergency* patient, or to treat the patient after office hours), or the patient decides to request an appointment in another hospital. In our model, we cannot include these creative solutions. Therefore, we delete this group of patients and adjust the arrival rates for the percentage of *failed* patients.

- Each scheduled patient cancels his appointment with a certain probability. We assume the moment of cancellation follows a uniform distribution over the patient's access time.

- Patients who cancel their appointments and patient who do not show up, may request a new appointment at another day. However, the model deletes these patients. We adjust the arrival rates for the percentage of cancelled appointments and no-shows.

- The processing times of patients at the counter differ between new and review patients (see subsection 2.1.2). This aspect is included in the model.

- Doctors and patients arrive early, punctually or late for their appointments. We explicitly model the arrival processes of patients and doctors, using theoretical probability distributions. A clinic session starts when the doctor and at least one patient have arrived. Consultations are allowed to start before the scheduled consultation time when the patient has arrived and the doctor is available. Therefore, negative internal waiting times may occur. However, in the model a clinic session never starts more than ten minutes before the scheduled appointment time of the *first* patient. This reflects reality as far as possible.

- We include the presence of medical students in the doctors' work schedule in the model for three random clinic sessions per shift. However, we model the medical students to show up with a probability of 80%. There are no data available about the presence of medical students in reality. We assume that our observations on this aspect reflect reality.

- If a medical student attends the clinic session of a resident or specialist, the student treats all new patients of that clinic session. We assume the expected net consultation duration for this consultation is 1.2 times the expected net consultation duration for the respective patient. However, we do not take preparation times, administration times and DICT into account for medical students, since they tend to prepare their consultations during the doctor's consultations with review patients. Medical students do not have administrative tasks.

- After a medical student's consultation, he follows the patient to the doctor and attends the doctor's consultation with the same patient. Meanwhile, the medical student is unavailable for other patients.

- During the presence of a doctor at the outpatient clinic, his status is either *busy* (consulting, preparation, administration or DICT) or *idle*. In reality, doctors always find a useful activity for their idle time, which makes the term *idle* perhaps inappropriate. However, the objective is to minimize this idle time, irrespective of the usefulness of the activities performed in it.

4.3.4 Patient creation routine

The subsections 4.3.4 and 4.3.5 are added for the interested reader. They focus on the creation of patients in the model, and on the details of the *Schedule appointment* process of Figure 4.2, respectively. These subsections may contain too much detail for some readers. We would like to refer these readers to subsection 4.3.6.

The model's three sources create new, review, and emergency patients respectively. The sources create patients following a negative exponential distribution with a mean inter-arrival time of 25.0 minutes for new patients, 5.7 minutes for review patients, and 171.9 minutes for emergency patients. Upon the creation of a patient, the model assigns the following attributes to the patient:

- Unique patient ID number
- Urgency
- Consultation type (following the new consultation types in Table 4.1)
- Treating specialist

The assigned values of these attributes are drawn based on historical and/or measured ratios. For new and review patients, the assigned urgency is either *elective* or *urgent*. All emergency patients receive urgency value *emergency* and consultation type *other consultation*. The value for 'treating specialist' determines the doctor for which the patient requests an appointment. This value can be 'resident' as well, indicating that the patient requests an appointment with one of the resident doctors, instead of an appointment with a medical specialist.

4.3.5 Scheduling routine

The appointment-scheduling process is one of the core processes of the simulation model. We describe the scheduling routine step by step. This model routine corresponds as much as possible to the current scheduling routine in the real outpatient clinic. Particular steps in this routine differ for alternative appointment systems. Below we describe the basic routine, and the Section 4.4 adds the changes to be made for other configurations.

At the start of each day, the model creates new empty appointment schedules for the first day after the scheduling horizon, while the appointment schedules of the past day are deleted. The model creates empty schedules for the complete scheduling horizon at the start of the simulation run. For the configuration representing the current situation, each appointment schedule consists of dedicated slots with codes *new*, *review*, and *overflow*. For alternative appointment schedules with dedicated slots, the time slots have codes *low* and *high*. The sequencing rule determines the definition of the codes *low* and *high* (e.g. short and long expected consultation duration, respectively). Section 4.4 explains the alternative sequencing rules in detail. For a dedicated time slot, the appointment interval, time slot code, and appointment starting time are defined in advance.

For alternative appointment schedules with pile-up scheduling, each empty appointment schedule is split in two parts. The sequencing rule and the expected number of patients in each 'sequencing class' determine the size of these parts. We call such a part of a schedule a *partial appointment schedule*. The model assigns a 'slot code' *low* or *high* to a partial schedule. With the sequencing rule 'First patients with long expected service time', the first partial appointment schedule receives the slot code *High* and is reserved for patients with a larger than average expected gross consultation time. Section 4.4 explains about the usage of slots and the sequencing rule in detail.

The routine to schedule a patient for his own doctor is as follows:

1. Determine the preferred slot code, based on the patient's consultation type.
2. Determine the expected gross consultation time G_p.
3. Determine the preferred appointment interval.
4. A. For dedicated slot usage:
    - Sort the empty slots of the doctor's appointment schedule in chronological order.
    - Search the first empty slot with the preferred slot code up to 10 working days in the future.
    - Check if the appointment interval of the empty slot differs at most 5 minutes (for the configuration that reflects the current situation: 10 minutes) from the preferred interval. If this is not the case, search for the next empty slot within the time frame.
    - If a slot is found: schedule the patient in the slot, and stop this routine. Else: go to step 5.

   B. For pile-up scheduling:
    - Sort the partial clinic sessions with the preferred slot code that have available time in chronological order.

- Search for the first partial clinic session up to 10 working days in the future that has at least the preferred appointment interval of available time.
- If such a partial clinic session is found: schedule the patient at the beginning of the available time, reduce the amount of available time with the appointment interval of this patient, and stop this routine. Else: go to step 5.

5. If step 4 was unsuccessful, change the search criteria according to the following scheme:
   A. For patients whose own doctor is a resident:
      - Repeat step 4, but search for empty slots further down the scheduling horizon.
      - If this is still unsuccessful: repeat step 4 with the enlarged time frame, but search for another slot code. Change the slot code according to the scheme in Table 4.2.

Table 4.2 Time slot code changing schema for scheduling routine

Preferred slot code (as determined in step 1)		New slot code
New	→	Review
Review	→	Overflow
Overflow	→	Review
High	→	Low
Low	→	High

      - If this is still unsuccessful: repeat step 4 and allow combining two successive slots of any type to one larger slot. This is only possible in configurations with dedicated slot usage.
      - If this is still unsuccessful: repeat step 4 and search for empty slots further down the scheduling horizon, allow the appointment interval to differ 10 minutes from the preferred interval, and allow to combine successive slots of any type (in configurations with dedicated slot usage).
      - If this is still unsuccessful: scheduling has failed. The patient is removed from the system, and the model increases the 'failed patients count' by one.

   B. For patients whose own doctor is a specialist:
      - Repeat step 4, but search for another slot code according to the scheme in Table 4.2.
      - If this is still unsuccessful: repeat step 4 and allow combining two successive slots to one larger slot. This is only possible in configurations with dedicated slot usage.
      - If this is still unsuccessful: repeat step 4 and search for empty slots further down the scheduling horizon.
      - If this is still unsuccessful: repeat step 4 and search for empty slots further down the scheduling horizon, allow the appointment interval to differ 10 minutes from the

preferred interval, and allow to combine successive slots of any type (in configurations with dedicated slot usage).

- If this is still unsuccessful: scheduling has failed. The patient is removed from the system, and the model increases the 'failed patients count' by one.

4.3.6 Warm-up period

At the start of a simulation run, the model is 'empty', which means there are no patients on the waiting list and there are no appointments scheduled yet. Since this is not a representative situation, we have to 'warm up the model' until the system reaches a steady state (Robinson, 1994). The output data from the warm-up period should be deleted. Although a steady state of an outpatient clinic is never reached during a clinic session (Bailey, 1952), the model can reach a steady state over day averages. Robinson (1994) describes that closely inspecting throughput, work-in-progress and queue size reports suffice to determine the warm-up period.

We evaluate the number of patients on the waiting list (Figure 4.5), the daily throughput of new and review patients (Figure 4.6), and the return rate (Figure 4.7) for the configuration representing the current situation. We conclude from all three figures together that a warm-up period of 150 working days is sufficient. Evaluating these figures for other configurations results in a comparable warm-up period.

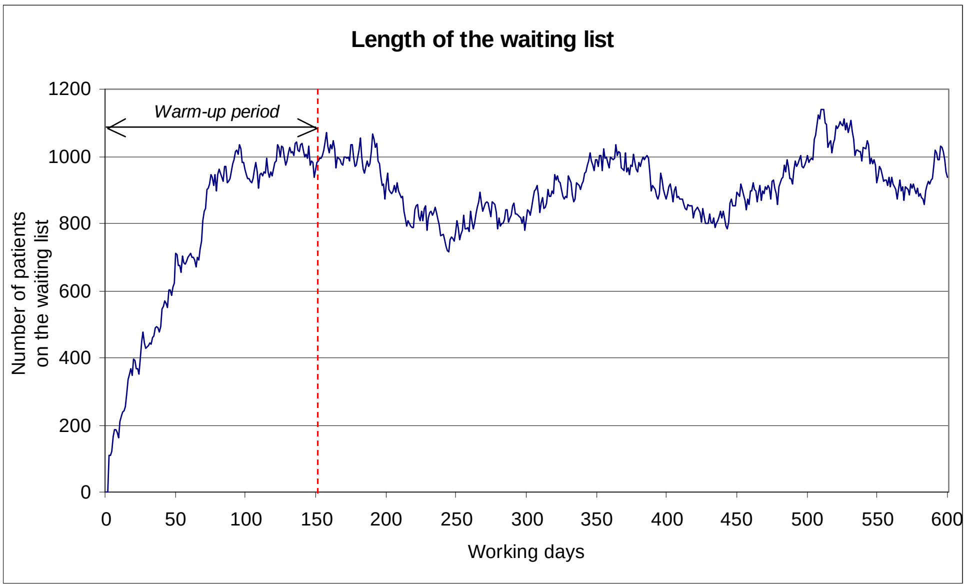


Figure 4.5 Length of the waiting list with warm-up period for the configuration for the current situation.

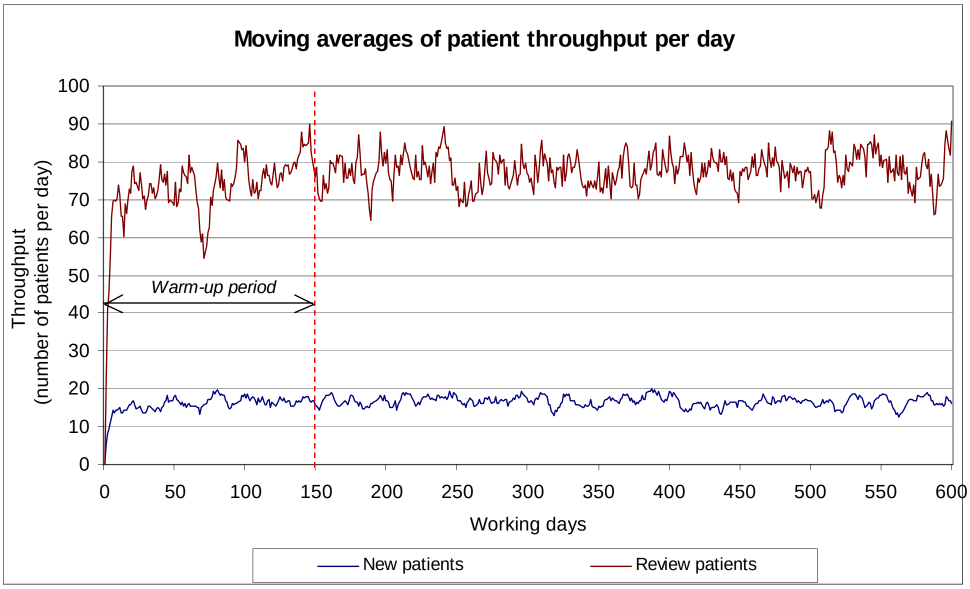


Figure 4.6 Moving averages of patient throughput with warm-up period for the configuration for the current situation.

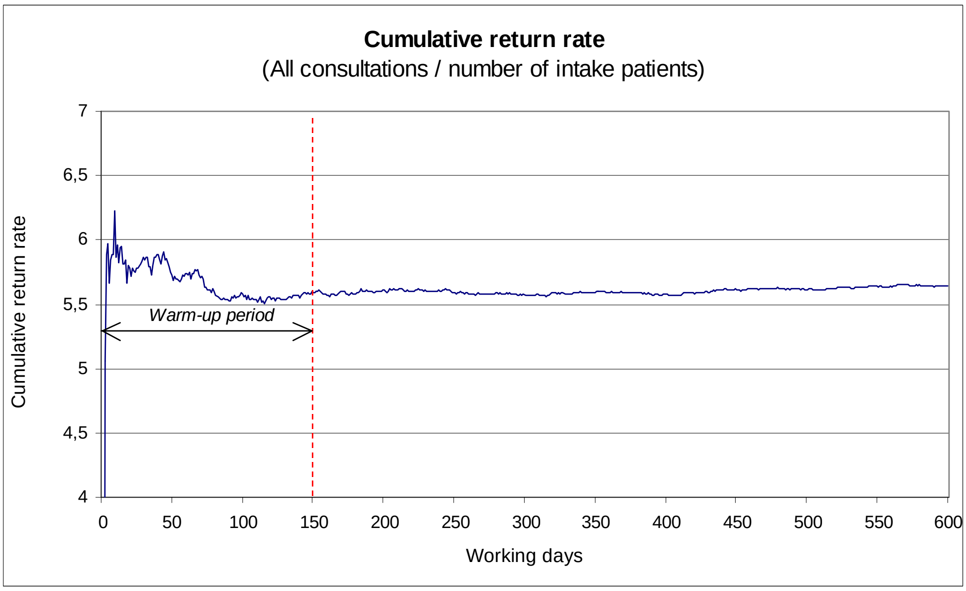


Figure 4.7 Cumulative return rate with warm-up period for the configuration for the current situation.

4.3.7 Run length

Since appointment scheduling with a rolling planning horizon can theoretically continue infinitely, we use a non-terminating simulation. The performance indicator *average internal waiting time* is independent between any two days. We use this performance indictor to determine the run length using the sequential approach (Law and Kelton, 2000). The reliability of the model results increases as the run length or the number of runs increase. However, executing too much or too long runs is a time consuming task. Since the warm-up period is considerably long, we prefer to execute one long run instead of multiple shorter runs for each configuration.

The purpose of the sequential approach is to find the run length for which the 95 % confidence interval for the expected internal waiting time has a relative error of less than 7.5 %.

- The average internal waiting time after $n = m - l$ working days is $\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ with X_i the observed average internal waiting time at day i, and m the run length with a warm-up period of l days.
- $\delta(n,\alpha) = t_{n-1.1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$ with the t-distribution with n-1 degrees of freedom, an approximate $100(1-\alpha)$ percent confidence interval, and $S^2(n)$ the sample variance.
- $\gamma' = \frac{\gamma}{(1+\gamma)}$ with preferred relative error γ.

The sequential approach procedure is a follows:

1. Run the model with an arbitrary run length of n_0 days, and set $n = n_0$.
2. Calculate $\overline{X}(n)$ and $\delta(n,\alpha)$ from $X_1, X_2, \ldots, X_n$.
3. If $\frac{\delta(n,\alpha)}{\left|\overline{X}(n)\right|} \leq \gamma'$ then for a run length of n the conditions are satisfied. If this is not the case, increment n by 1 and return to step 2.

We chose a relative error of 7.5 % and a two-tailed confidence interval of 95 %, which results in a run length of 455 working days plus the warm-up period of 150 days for the configuration representing the current situation. Since 480 working days represent a period of two years, we decide to execute all runs with a length of 150 + 480 = 630 working days.

4.3.8 Verification and validation

Verification of the model is a 'micro' check of the correctness of the model (Robinson, 1994). We verify the correctness of the methods and procedures during the model coding. A scheduler of the outpatient clinic successfully checks the steps of the model's scheduling routine for correctness. Before running the simulations, we perform many test runs to debug the model. The scheduling routine and the doctor arrivals are especially vulnerable for small but far-reaching errors. Testing and debugging is a very time-consuming activity.

The validation is the 'macro' check of a model and comprises the testing of the overall accuracy of the model and its ability to meet the objectives (Robinson, 1994). We validate the model in two ways: qualitatively and quantitatively. For the qualitative validation we show the model, its flow diagram, and its main outputs to a medical specialist, and discuss the applicability of the model to the outpatient clinic's appointment system. According to the medical specialist, the model's flow diagram is a good representation of the major processes of the outpatient clinic and its appointment scheduling. Using correct input data, the model is valid and applicable to represent the outpatient clinic and its appointment system, according to the medical specialist.

The quantitative validation comprises a comparison of the model's output data with the real system. A performance indicator that is influenced by the scheduling routine as well as by the doctors' processes is the internal waiting time. Therefore, we use the internal waiting time output to validate the model. Figure 4.8 depicts the combined cumulative frequencies of the internal waiting times from the real outpatient clinic and from the model's configuration that represents the current situation.

As Figure 4.8 shows, the internal waiting times differ between the model and the real system. An explanation for this difference is the habit of doctors to reduce the consultation time if many patients are waiting in the waiting room, and to increase the consultation time if the internal waiting time is short. This habit prevents the occurrence of very long and very short internal waiting times: if the internal waiting time is short at a certain moment, the doctor increases the consultation time because he experiences little time pressure, resulting in a longer internal waiting time for the next patient, and vice versa. The doctors recognize this phenomenon in their own clinic sessions, and they admit to have this habit. Figure 4.9 supports this explanation and the existence of this habit. This figure shows that the average net consultation duration in the real system is longer if there are fewer patients in the waiting room. The model does not adjust the consultation times for the internal waiting time or the number of waiting patients, resulting in more outliers and a lower average internal waiting time.

Concluding, the verification and qualitative validation show that the model is a valid representation of reality. The quantitative validation shows a shortcoming of the model. However, we have an explanation for the optimistic representation of the internal waiting time.

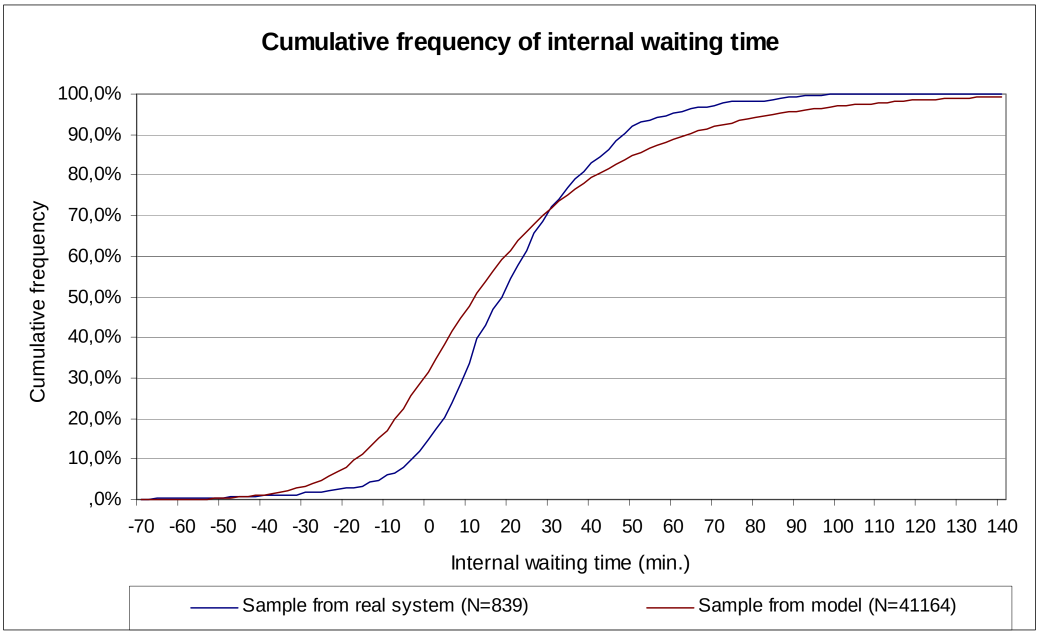


Figure 4.8 Cumulative frequency of internal waiting time.

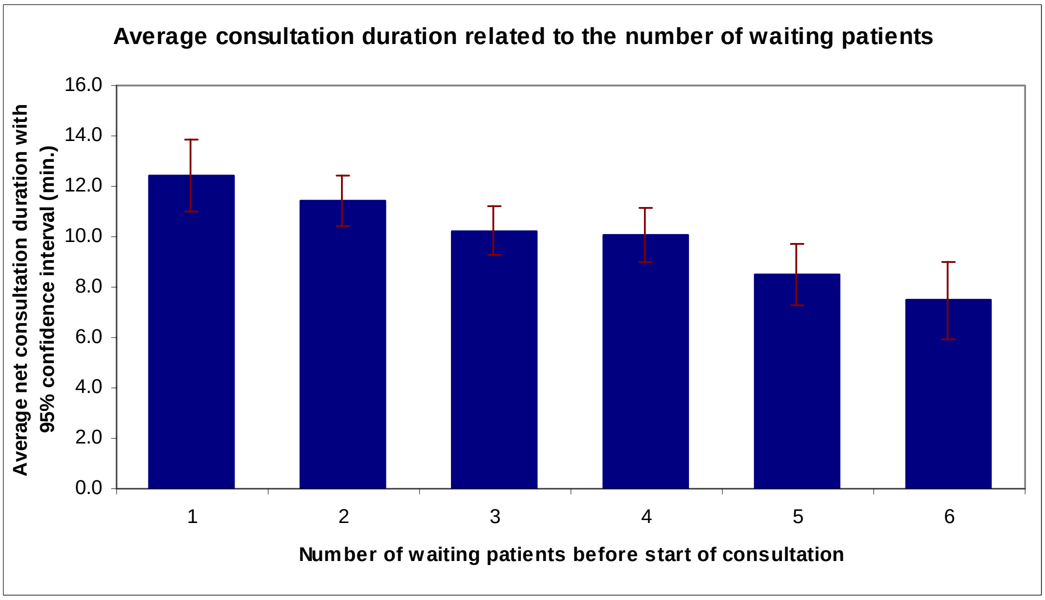


Figure 4.9 Average consultation duration related to the number of waiting patients in the real system (N = 769).

4.4 Experimental factors

This section describes the experimental factors and their values in detail. For each experimental factor we evaluate two to six optional values, resulting in a theoretical maximum of 288 alternative appointment system configurations (Figure 4.10). Below we describe the possible values. Section 4.5 selects configurations of these values, resulting in an initial evaluation of 45 configurations.

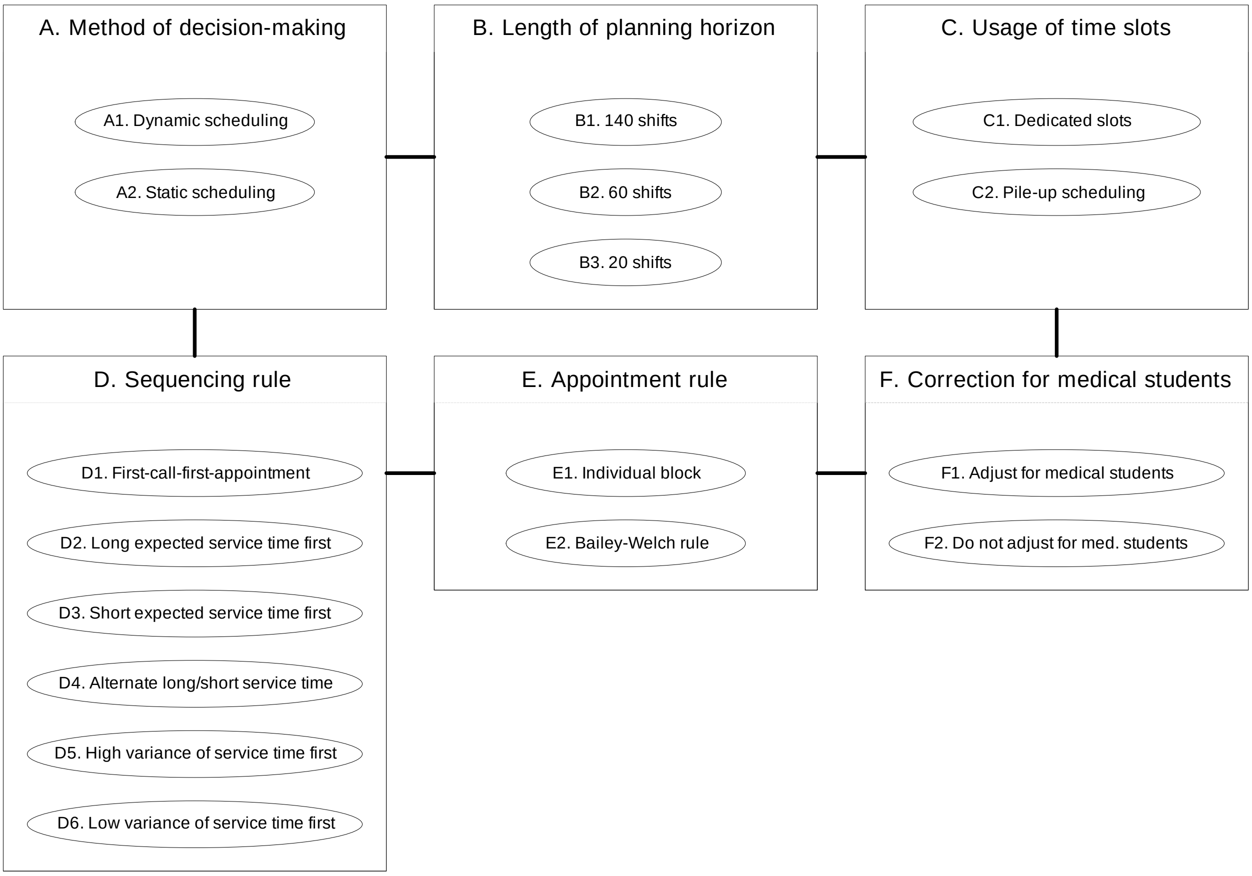


Figure 4.10 The six experimental factors (rectangles) with two to six possible values each (ellipses).

4.4.1 Method of decision-making

The method of decision-making involves the possibility to schedule patients in batches at the same time. We evaluate two methods of decision-making:

A1. **Dynamic scheduling** is what Klassen and Rohleder (2004) define as ‘scheduling patients throughout the day without knowledge of the type and number of clients that will call for an appointment later’. The scheduler schedules an appointment immediately upon the patient’s request. This may involve a good patient service level, since the patient immediately receives an appointment date and time. On the other hand, the appointment schedule is always sub-optimal because the scheduler cannot take later requests for appointments for the same scheduling horizon into account.

A2. **Static scheduling** is ‘scheduling a finite number of patients simultaneously’ (Wang, 1999), using all appointment requests that arrived after the last moment of scheduling. With static scheduling a scheduler batches appointment requests. The advantage of static scheduling is the possibility to improve (parts of) appointment schedules with sequencing rules (see Subsection 4.4.4) and/or with local search heuristics (see Section 4.7). The disadvantage is that the scheduler has to contact the patients again once he has planned their appointments.

Dynamic scheduling and *static scheduling* are both a form of *operational offline planning.* Houdenhoven *et al.* (2006) define *operational offline planning* as “the in-advance day-to-day control of expected activities”. The arrival of appointment requests is expected, and the appointment schedule is always created in advance of the day of execution of the schedule. This is irrespective of the method of decision-making. The definition of Houdenhoven *et al.* (2006) for *operational online planning* (“all control mechanisms that deal with monitoring the process and reacting to unforeseen or unanticipated events”) does apply to the arrival of emergency patients at the outpatient clinic, who have to be treated by the resident on duty.

4.4.2 Length of scheduling horizon

The rolling scheduling horizon is the number of shifts for which appointments can be scheduled in advance. As we described in Chapter 2, the scheduling horizon is currently approximately 140 shifts (70 working days). With a shorter scheduling horizon, there are fewer patients on the waiting list. Hence, the number of cancellations by patients is less. For a cancelled appointment a ‘gap’ in the appointment schedule emerges, which has to be filled by the scheduler. For cancelled clinic sessions, the scheduler has to reschedule all patients who already had an appointment for that clinic session. In both cases, the amount of work for the scheduler increases. The disadvantage of a shorter scheduling horizon is the smaller number of scheduling possibilities for appointment requests. This may lead to situations where it is impossible to schedule a patient.

One of the objectives for this study is to balance the work pressure for the schedulers. Therefore, we evaluate configurations with a shorter scheduling horizon and include the number of times a scheduler has to contact a patient, in the scheduling objective function (Appendix B).

Because most appointment scheduling studies have a scope of one clinic session (see Chapter 3), there are no specific research results available on the optimal length of the scheduling horizon for outpatient scheduling. However, based on some empirical cases, CBO (2006) proposes to reduce the scheduling horizon to 6 weeks (60 shifts) once the access time for new patients has been reduced to one week, for the reasons stated above. As a third option, we include an even shorter scheduling horizon to the configurations, to test whether we can reduce the work pressure further.

Thus, we evaluate the following three options for the scheduling horizon:

B1. 140 shifts (70 working days), as in the current situation.

B2. 60 shifts (30 working days), as CBO (2006) proposes.

B3. 20 shifts (10 working days), as an extra improvement possibility.

4.4.3 Usage of time slots

In the current situation, the scheduler designs a grid of appointment intervals for all clinic sessions before he schedules the first patient to this clinic session. All intervals are dedicated for one of the three consultation types currently in use (*new*, *review* and *overflow*; see Chapter 2). The number of time slots for each consultation type and the length of the intervals are based on historical data of patients' visits, and on production targets about the number of new patients between the ENT department and higher management. These numbers and the length of the intervals are predetermined and are not adjusted according to individual patients' requests. At first glance this method of time slot usage is suboptimal, especially if we extend the number of consultation types to seven (see Section 4.2), since for individual clinic sessions the number of consultations of one type may exceed the number of dedicated slots for that consultation type.

An alternative to dedicated slot usage is so-called pile-up scheduling. With pile-up scheduling, one uses a grid of intervals with a length of five minutes that are not dedicated for consultation types beforehand. Instead, the scheduler can schedule a patient for one or more successive intervals, starting at the earliest free interval of the clinic session. The amount of time a patient is scheduled for, is decided by the expected gross consultation time. Since intervals lengths occur in multiples of five minutes, we round the expected gross consultation time to the nearest integer multiple of five minutes. We base the expected value of the consultation times on empirical data, as Chapter 5 explains. Disadvantages of pile-up scheduling are difficulties to meet the production targets, and combination appointments with other departments. We do not take the latter disadvantage into account, since we only model one the outpatient clinic of one medical department.

From the previous studies, most articles advocate (implicitly) a form of pile-up scheduling (e.g. Kaandorp and Koole, 2007; Klassen and Rohleder, 1996) or assume a homogeneous group of patients (e.g. Bailey, 1952). However, most outpatient clinics in the Netherlands use an appointment system with dedicated slots.

We evaluate the following usages of time slots.

C1. Dedicated slots for consultation types.

C2. Pile-up scheduling: round expected gross consultation time to the nearest integer multiple of five minutes.

4.4.4 Sequencing rule

Different studies found good results on patients' waiting times and doctors' idle times and overtime with different sequencing of patients, as indicated below. The reason those studies found different best sequencing rules lies in their respective modeling and outpatient clinic environments, and their assumptions to reality. Since we use another model as well, we evaluate a relatively large number of sequencing rules. For configurations with dedicated slots, a sequencing rule is only used to create the fixed grid of intervals. A sequencing rule is used for each batch in case of pile-up scheduling.
We evaluate the following sequencing rules:

D1. First-call-first-appointment

D2. First patients with long expected service time

D3. First patients with short expected service time (performed best for Hutzschenreuter, 2004).

D4. Alternate patients with short and long expected service time (performed best for Cayirli, Veral, and Rosen, 2006).

D5. First patients with high variance of service time

D6. First patients with low variance of service time (performed best for Klassen and Rohleder, 2004 in a multi-period environment).

The results of Hutzschenreuter (2004) are comparable to those of Klassen and Rohleder (2004) because she varied the coefficient of variation (CV) of the service time and thereby indirectly the standard deviation, resulting in the same sequencing rule outperforming the others. Although we use fixed CVs, we review both sequencing rules D3 and D6.

4.4.5 Appointment rule

As we described in Chapter 3, the schedulers use an individual block appointment rule in the current situation. Hutzschenreuter (2004) found that the individual block rule performed best when the doctors' utilization was valued relatively low with respect to patients' waiting times in the objective function, whereas the classic Bailey-Welch rule performed best when the patients' internal waiting times were valued relatively high. Since this trade-off presumably depends on other factors incorporated in her model as well, we evaluate both the individual block and the Bailey-Welch rule.

Thus, we evaluate the following appointment rules:

E1. Individual block

E2. Bailey-Welch rule

4.4.6 Correction for medical students

Medical students see some new patients in the outpatient clinic. An example in Figure 2.3 in Chapter 2 shows that the specialist's idle time and patients' internal waiting times can be affected by the attendance of medical students. Therefore, it would be useful to correct new patients' appointment times for clinic sessions with a medical student, for example by asking these patients to arrive 30 minutes early. However, medical students have a non-zero no-show rate as well. If a medical student does not show up, these patients have to incur longer waiting times and are only treated by the resident or specialist. To evaluate these effects, we evaluate configurations with and without correction of appointment times for new patients. As far as we know, there are no previous studies that take this aspect into account.

- F1. Correct appointment time of new patients with 30 minutes for attendance of medical students.
- F2. Do not correct appointment time for attendance of medical students.

4.5 Appointment system configurations

We combine the values for the experimental factors described above to complete appointment scheduling configurations. The current scheduling method, as described in Chapter 2, can be summarized as the A1-B1-C1-D1-E1-F2 configuration. Initially, we only evaluate the values of the experimental factors in the current scheduling method, and combinations with alternative values for each factor that are most likely to perform well. If these alternative values do not perform better than the current value, it is unlikely that other values for the same experimental factor perform better. For the *sequencing rule* factor (D), we initially evaluate two of the three alternatives that performed well in previous studies.

Initially, we evaluate:

- A1 and A2.
- B1 and B2.
- C1 and C2.
- D1, D3 and D6.
- E1 and E2.
- F1. We evaluate F2 only for the top-3 performing configurations.

In configurations that include the combination C1-D1, we use the dedicated time slot layout that is currently in use in the outpatient clinic, which assumes an individual block rule (E1). Therefore, C1-D1-E2 combinations do not exist. This results in an initial evaluation of 45 scheduling configurations, including the current scheduling configuration.

4.6 Scheduling objective function

For configurations containing static scheduling (A2), we calculate the Scheduling Objective value *SO* (see formula below) upon scheduling a patient in the model. The *SO* objective function minimizes the doctor's expected overtime and idle time (and thereby maximizes expected utilization), the patient's expected internal waiting time and the counter's peak load that we expect to incur. If there are at least two scheduling possibilities for a patient, we schedule the patient for the position in the appointment schedule that results in the lowest *SO* value. Calculating the *SO* after scheduling each patient, we know its incremental contribution to the *SO* value.

$$SO = \min\left\{\gamma_{PO}\sum_{s=1}^{S} PO_s + \gamma_{PI}\sum_{s=1}^{S} PI_s + \gamma_{PIW}\sum_{p=1}^{P} PIW_p + \gamma_{PPL} PPL\right\}$$

In this formula, PO_s and PI_s are the planned overtime and planned idle time respectively for the doctors, PIW_p is the planned internal waiting time for patients and *PPL* is the planned peak workload for the counter, when all patients that have been scheduled are taken into account. The γ-factors are weighing factors. For the derivation of *SO* we refer to Appendix B.

4.7 Local search for new patients with static scheduling

We use a local search heuristic in the model to improve (partial) schedules in configurations with static scheduling (A2). Patients who are under treatment of a specialist doctor cannot be treated by another doctor, because of the highly specialized medical care of their own specialist. However, resident doctors treat the patients with medical diagnoses that require less specialized medical care. Therefore, we can move these patients between schedules of different resident doctors, in order to obtain an improved *SO*.

After a batch of appointments has been scheduled, we perform the following 2-opt algorithm.

Step 1. Select two resident doctors' appointment schedules *x* and *y* that will be executed during the same shift *j* in the future, each containing at least one new patient that has just been scheduled in this batch. If no two such appointment schedules exist: stop.

Step 2. Take $Q = \{q_1,\ldots, q_{Nx}\}$ and $R = \{r_1,\ldots, r_{Ny}\}$ as the set of new patients that have just been scheduled for appointment schedules *x* and *y* respectively.

Step 4. For all possible combinations of an element from *Q* and an element from *R*, determine the incremental change of the *SO* function when these elements would be exchanged between *Q* and *R*.

Step 5. If the incremental change of the *SO* value results in a better *SO* value: perform the corresponding swap, calculate the new *SO* and repeat step 4. Otherwise: go to step 6.

Step 6. Repeat step 1 to 5 for all possible combinations of appointment schedules *x* and *y*, that satisfy the constraints in step 1.

Since the number of patients that are scheduled in the same batch for the same shift j for different appointment schedules will be limited, it is possible to enumerate all possible swaps in step 3. For example, suppose 150 patients request an appointment on a certain day (on average, this amount is less than 100 in the outpatient clinic, see Chapter 5). In the exceptional case where 40% of them are new patients, of which 50% can be scheduled for the next day, the above algorithm concerns 30 patients. This corresponds to a theoretical maximum of 225 combinations that have to be reviewed, every time step 3 is performed. In the outpatient clinic, at most four resident doctors have a clinic session during the same shift. On an average day, it concerns about 10 patients for two to four resident doctors.

5. Input data gathering and analysis

5.1 Time measurements

We performed time measurements at the outpatient clinic to collect data for the input parameters of the appointment system. These parameters are the necessary input data for the simulation model. Although many previous studies include quantitative data about outpatient clinics, it is important to collect specific data about the case at hand to build a reliable model.

The measurements we perform are listed below, as well as their relevance.

- **Arrival times of patients**
  Hutzschenreuter (2004) concluded recently from measurements in a university hospital outpatient clinic in Amsterdam that patients arrive on average 10 minutes early. Although that situation is comparable to ours, we perform arrival time measurements ourselves. We also want to measure the internal waiting times and the peak workloads for the counter personnel, and for those calculations we use the arrival times.
- **Preparation times**
  All previous studies that measured consultation times neglect the doctor's preparation time. Using only gross consultation times for scheduling purposes is suitable from a doctor's perspective. However, a patient's internal waiting time ends when his net consultation time starts, instead of when his gross consultation time starts. Therefore, we measure the preparation times explicitly.
- **Net consultation times**
  As Chapter 3 shows, the service time distributions used in previous studies vary widely. We perform our own measurements here as well.
- **Post-consultation administration process times**
  To complete the gross consultation time measurements, we also measure the administration process times.
- **Desirable inter-consultation time (DICT) durations**
  It is generally assumed in the literature that doctors use all non-idle time to treat patients. Since this is not the case in practice, we want to know what percentage of time during a clinic session is not spent directly on treating patients.
- **Doctors' idle time and pauses in-between consultations**
  Other activities of doctors during clinic sessions are waiting for patients, and pausing. Since these two activities cannot be distinguished clearly from each other, we measure them as one activity. We assume doctors will not take a break when there are patients waiting to be seen, which is generally true in the outpatient clinic. Under this assumption, pausing and idle time are the same.

For the measurements during the doctors' processes we use a customized time measurement application on four laptop computers that are operated by the doctors themselves. During a period of five weeks we measure the consultation times of 985 patients in 109 clinic sessions of 21 doctors. Using these measurements and the original appointment schedules, we calculate the punctuality of patients and doctors, the inter-arrival times of emergency patients, internal waiting times, no-show rates, and overtime. For other factors we use the Erasmus MC Management Information System "Business Objects" and estimates.

5.2 Appointment scheduling processes

The Sections 5.2 to 5.5 follow the sequence of processes of Figures 2.1 and 2.2 in Chapter 2.

5.2.1 Production

Based on the Erasmus MC Management Information System "Business Objects", the outpatient clinic performed 22,228 consultations in 2006. Table 5.1 shows the relative numbers of patients, grouped by the current patient classes. The second column shows the correspondence to the new classification in consultation types, introduced in Section 4.2.

5.2.2 Percentage of canceled and rescheduled appointments and no-shows

Based on the same production figures, 29,177 appointments were scheduled in 2006. Of this total, 4,332 patients (14.8 %) cancelled their appointment prior to the day of consultation. 1,246 appointments (4.3 %) were cancelled and rescheduled by the outpatient clinic because of unexpected absence of a doctor. Of the remaining appointments, 5.8% of the patients did not show up.

5.2.3 Return rate

CBO (2004) calculates the return rate of patients for a certain period as follows:

$$\text{Return rate} = \frac{\text{All consultations}}{\text{Number of intakes}}$$

We use the same definition. Based on the production figures of 2006, the return rate of the outpatient clinic is 5.38. This means that patients visit the outpatient clinic on average 5.38 times.

Table 5.1 Relative shares of consultation types in 2006 production.

Current patient classification	New classification: consultation types	Percentage of 2006 production
New patients	Intakes	18.6 %
Review patients	Diagnostics, Follow-ups, and Periodical treatments	71.7 %
Telephonic consultations	Telephonic consultations	2.4 %
Other consultations	Other consultations	7.3 %

5.3 Counter processes

5.3.1 Patients' punctuality

Figures 5.1 and 5.2 show the results of the measurements on arrival times of new and review patients respectively. Statistical analysis does not show significant differences in the punctuality of new and review patients. We assume the punctuality of all patients together follows a normal distribution with a mean of -12.9 minutes and a standard deviation of 17.2 minutes, based on Table 5.2 and the Q-Q plot in Appendix C (Figure C.1). Compared to the VU Medical Center in Amsterdam, in which Hutzschenreuter (2004) performed time studies in an ophthalmic outpatient department, patients arrive on average 2.9 minutes earlier in the Erasmus MC ENT outpatient clinic. According to some of the medical specialists, many patients travel over a considerable distance to the hospital, due to the highly specialized care and the regional function of the oncology clinic sessions. These patients, and elderly people, tend to arrive earlier than other patients.

Table 5.2 Descriptive statistics of patients' punctuality.

Punctuality of patients	New	Review	All
N	113	635	748
Mean (min.)	-19.2	-11.8	-12.9
Median (min.)	-16	-10	-11
Sample deviation (min.)	22.7	16.2	17.2
Coefficient of variation	-1.2	-1.4	-1.3
Percentage of patients on time	87.4%	81.4%	82.1%
95% confidence interval of mean, LB	-20.2	-13.1	-14.0
95% confidence interval of mean, UB	-11.8	-10.6	-11.6

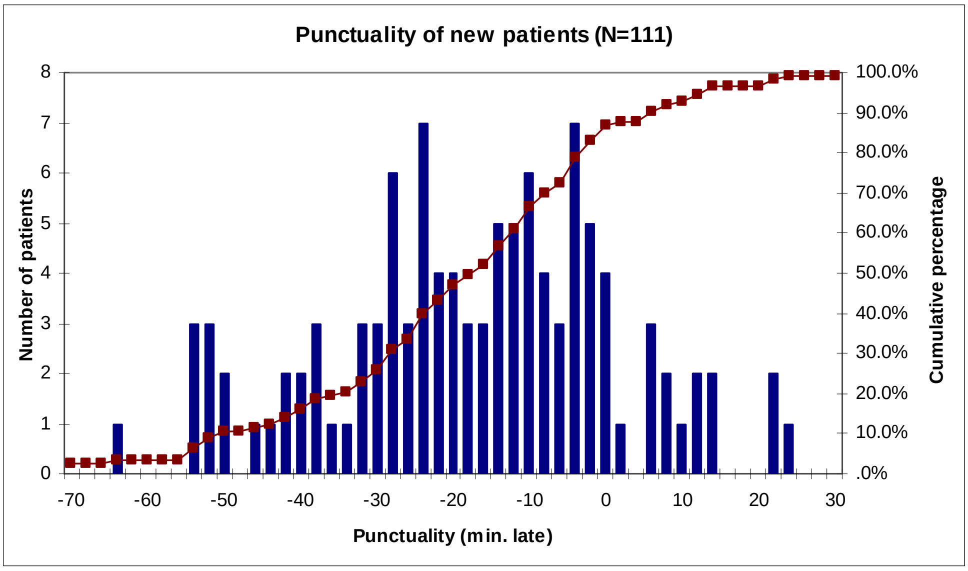


Figure 5.1 Histogram and cumulative frequency for the measured punctuality of new patients (N=111)

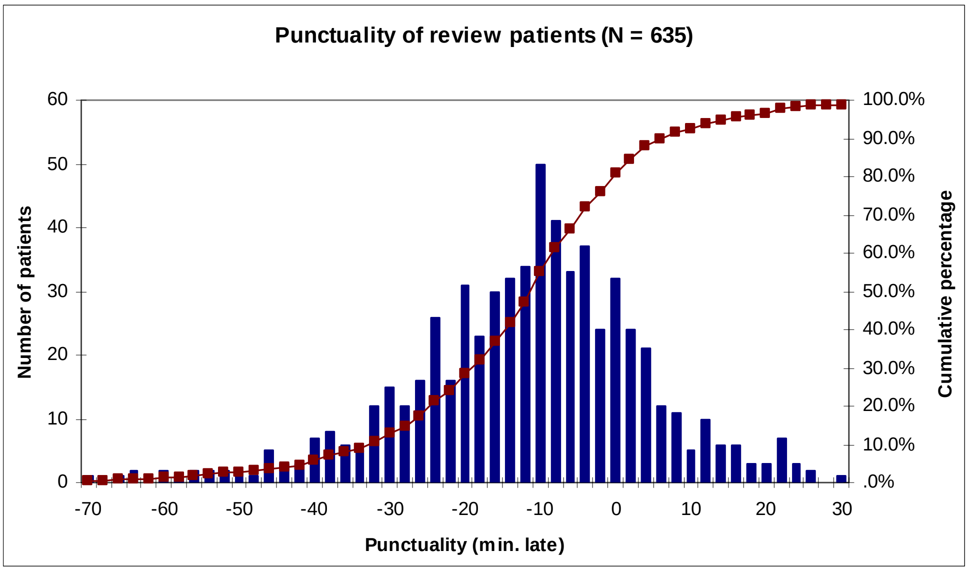


Figure 5.2 Histogram and cumulative frequency for the measured punctuality of review patients (N=635)

5.4 Doctor's processes

5.4.1 Punctuality of doctor for first consultation

The punctuality of doctors differs between the clinic sessions in the morning and those in the afternoon. Figure 5.3 shows the measured punctuality. Because of the relatively small sample size, we were unable to properly fit a probability distribution function. A normal distribution fits best, and fitted the patients' punctuality as well. Table 5.3 contains the descriptive statistics, and Appendix C contains the Q-Q plot (Figure C.2).

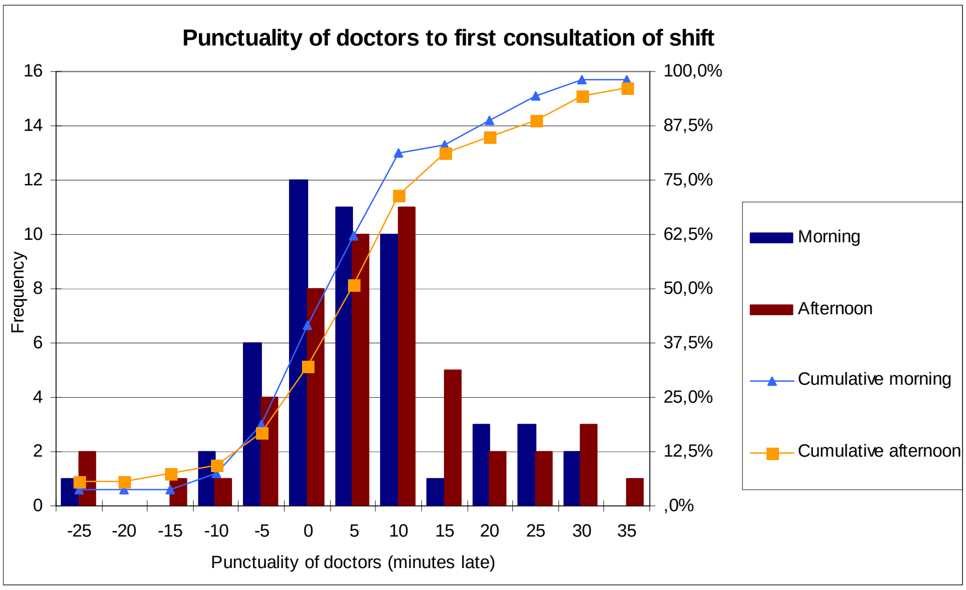


Figure 5.3 Histogram for measured punctuality of doctors (N = 109)

Table 5.3 Descriptive statistics of doctors' punctuality.

Punctuality of doctors	Morning	Afternoon	Total
N	53	56	109
Mean (minutes)	3.7	5.9	4.8
Median (minutes)	2	5	4
Standard deviation (minutes)	14.3	15.6	15.0
Coefficient of variation	3.9	2.7	3.7
Percentage of doctors on time	41.5%	32.1%	35.8%

We assume the doctors' punctuality has a normal distribution with a mean of 3.7 minutes late and a standard deviation of 14.3 minutes in the morning, and a mean of 5.9 minutes late and a standard deviation of 15.6 minutes in the afternoon.

5.4.2 Percentage of clinic sessions with a medical student

Based on interviews and observations, we estimate that medical students are scheduled to attend 50% of the resident's and non-oncology specialist's clinic sessions, where they perform their own consultations with every new patient prior to the doctor's consultation with this patient. However, medical students have an estimated no-show rate of 20%. Medical students do not attend oncology specialist's clinic sessions.

5.4.3 Consultation duration

Figures 5.4 and 5.5 show histograms for the net and gross consultation times respectively. We can fit them to a Gamma distribution, as the Q-Q plots in Appendix D show (Figures D.1 and D.2). As mentioned before, the consultation durations differ per consultation type and per doctor type. We assume that the net and gross consultation durations for all consultation types and doctors are Gamma distributed as well. Appendix E depicts the mean and variances of the net and gross consultation durations we measured, per consultation type and per doctor type.

To be able to estimate the gross consultation duration of a patient on basis of his consultation type and the treating doctor, we cluster the preparation times, net consultation durations and administration times. Each cluster contains the *consultation type - doctor type* combinations that have comparable characteristics. The mean duration for a cluster is the average duration over all measured consultations forming that cluster. The differences between any two clusters are statistically significant, while the differences between consultation types within one cluster are not. Appendix F contains the arrangement of *consultation type - doctor type* combinations into clusters, and gives the parameters (*alfa* and *beta*) for the Gamma distributions that we can fit to the respective cluster.

Upon scheduling a patient's, we use the mean preparation time, consultation duration, and administration duration of the respective clusters corresponding to this patient's consultation type. For example, we estimate the gross consultation time of a patient with consultation type 'follow-up consultation' who is to be scheduled for a resident doctor, as indicated in Table 5.4. The estimated gross consultation time (792.9 seconds, or 13.2 minutes for the example in Table 5.4) is rounded to the nearest multiple of 5 minutes to determine the patient's appointment interval (15 minutes or the example in Table 5.4).

Table 5.4 Example of the determination of the estimated gross consultation time (data from Tables F.1 to F.3).

Estimation for gross consultation duration	**Cluster**	**Mean (seconds)**	**Standard deviation (sec.)**
Estimation for preparation time	Cluster 3	153.5	137
Estimation for net consultation duration	Cluster 1	588.9	394.7
Estimation for doctor's administration time	Cluster 1	50.5	104.4
Estimation for gross consultation duration		792.9	430.6

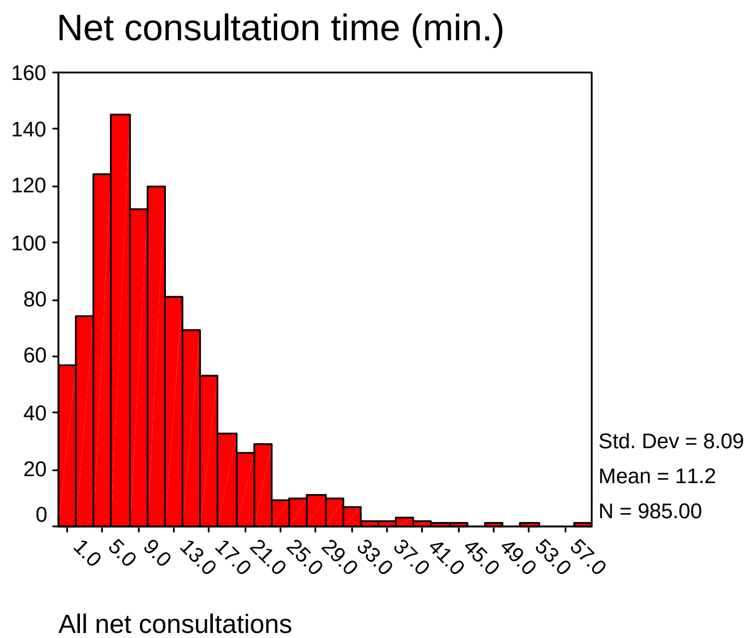


Figure 5.4 Histogram for the net consultation time for all doctors and all consultation types.

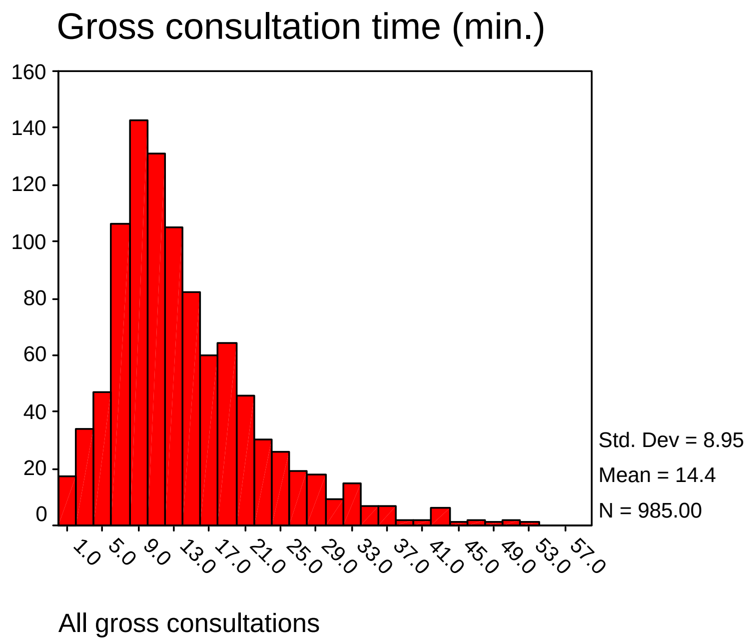


Figure 5.5 Histogram for the net consultation time for all doctors and all consultation types.

6. Simulation results

6.1 Evaluation approach

To evaluate the overall performance of the configurations, we formulate a scaled and weighed average of the performance indicators *doctor's utilization* and *doctor's overtime,* which we call the *doctor's performance* (*DP*) of a configuration. To compare clinic sessions with different durations, we determine the realized overtime rate (RO_{sj}) and realized utilization rate (RU_{sj}) for each doctor as a percentage of the scheduled duration of his clinic session in a particular shift.

$$RO_{sj} = \frac{\text{End of administration duration for the last patient} - \text{scheduled end of last consultation}}{\text{Scheduled duration of the clinic session}}$$

$$RU_{sj} = \frac{\text{Non-idle time during scheduled duration of the clinic session and overtime}}{\text{Scheduled duration of the clinic session}}$$

Since the doctors value the utilization of their time during the planned duration of a clinic session twice as important as the avoidance of overtime at the end of a clinic session (see Section 2.2), we set the weights as follows:

$$\alpha_{RO} = 1$$

$$\alpha_{RU} = 2$$

Now, we calculate the *doctor's performance* per configuration:

$$DP = \sum_{s}\sum_{j}\alpha_{RU}\left(1 - RU_{sj}\right) + \alpha_{RO}RO_{sj}$$

Note that the doctor's idle time is measured indirectly as well. A clinic session with a high utilization rate and a low overtime rate has a low idle time rate as well. The *doctor's performance* is a dimensionless indicator. A lower value means a better *doctor's performance*.

The *DP* does not include the access time and the internal waiting time. Instead, we introduce minimum patient service levels for these two factors, which were formulated in cooperation with a doctor and the outpatient clinic's manager:

- The average access time over all doctors is not allowed to exceed 10 working days.
- The average internal waiting time of patients is not allowed to exceed 20 minutes.

All configurations that violate at least one of these two minimum service levels, are not considered for further evaluation.

6.2 Simulation results for individual performance indicators

Irrespective of the *doctor's performance* (*DP*) value of the configurations, we rank the configurations on each performance indicator separately. However, configurations that do not meet the minimum patient service levels are not evaluated in this section. Appendix G contains a table with the performances of all configurations on all performance indicators.

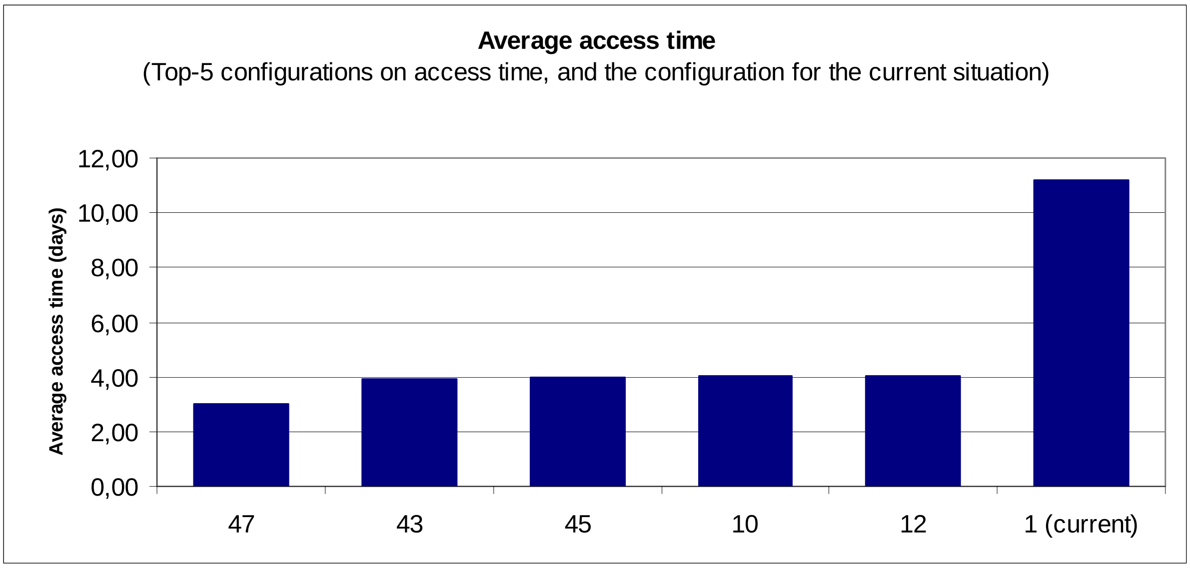
6.2.1 Access time

The majority of the configurations considerably shorten the access time. Figure 6.1 depicts the top-5 performing configurations on access time. All these five configurations include pile-up scheduling and the Bailey-Welch appointment rule. A short access time is an indicator for a flexible appointment system, because there are more possibilities to schedule a patient than in case of dedicated slot usage. When measuring the access time, we find these scheduling possibilities as well. These findings are in correspondence with our expectations and show the benefit of pile-up scheduling for this performance indicator.

There is a significant difference in the access time in the real outpatient clinic (recapitulate the average access times from Section 2.2.1) and the model outcome for the access time in the configuration representing the current situation. The difference is presumably the result of production-demand imbalances in the past in the real outpatient clinic. The model shows that an access time of approximately 11 working days is possible in the current situation's configuration. However, further decreases to less than a week are possible with pile-up scheduling and a Bailey-Welch appointment rule.

6.2.2 Internal waiting time

The performance measure of the internal waiting time shows the average internal waiting time over all patients of the 480 working days run length. The sequencing rule and the appointment rule have the largest impact on the internal waiting time. The shortest waiting times are achieved with an individual block appointment rule, because there is no waiting time 'competition' with another patient waiting for the doctor's first appointment slot. With regard to the sequencing rule, it appears that having low-variance consultations at the beginning of a clinic session reduces patients' internal waiting times. Disturbances in the execution of the appointments (e.g. longer consultations than expected) influence a large number of patients' waiting times negatively when they occur early in the clinic session. This supports the findings of Klassen and Rohleder (2004). Figure 6.2 depicts the top-5 results on internal waiting time.



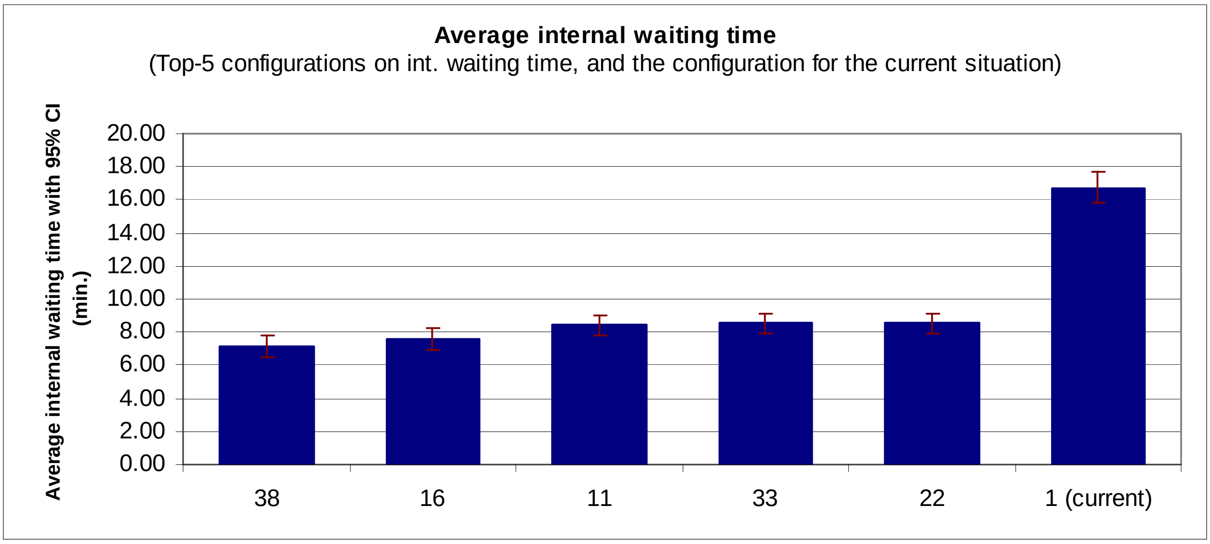
Configuration	47	43	45	10	12	1 (current)
Decisions	**static (A2)**	**static (A2)**	**static (A2)**	dynamic (A1)	dynamic (A1)	dynamic (A1)
SchHorizon	**10days (B3)**	30days (B2)	30days (B2)	70days (B1)	70days (B1)	70days (B1)
Slot usage	**pile-up (C2)**	**pile-up (C2)**	**pile-up (C2)**	**pile-up (C2)**	**pile-up (C2)**	dedicated (C1)
Seq. Rule	**FSCT (D3)**	**FSCT (D3)**	FLV (D6)	**FSCT (D3)**	FLV (D6)	FCFA (D1)
App. Rule	**BW rule (E2)**	**BW rule (E2)**	**BW rule (E2)**	**BW rule (E2)**	**BW rule (E2)**	Indiv.block (E1)
CMS	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	no (F2)
Access time (days)	3.0	3.9	4.0	4.0	4.0	11.2
DP	0.293	0.280	0.277	0.293	0.281	0.320

(Experimental factor values that are similar to the value of the best performing configuration are shown bold in the table above.)

Abbreviations

App. rule	Appointment rule
BW rule	Bailey-Welch appointment rule
CI	Confidence interval of the mean
CMS	Correction of new patient's appointment times for attendance of Medical Students
Config.	Configuration
Decisions	Method of decision-making
Dedicated	Time slots dedicated for consultation types
DP	Doctor's performance
FCFA	First-Call-First-Appointment
FLV	Schedule First patients with Low Variance of consultation time
FSCT	Schedule First patients with a Short expected Consultation Time
Indiv.block	Individual block appointment rule
SchHorizon	Length of the scheduling horizon
Seq. rule	Sequencing rule
Slot usage	Usage of time slots

Figure 6.1 Performance of top-5 configurations on access time, and the configuration for the current situation.



Configuration	38	16	11	33	22	1 (current)
Decisions	**static (A2)**	dynamic (A1)	dynamic (A1)	**static (A2)**	dynamic (A1)	dynamic (A1)
SchHorizon	**30days (B2)**	**30days (B2)**	70days (B1)	70days (B1)	**30days (B2)**	70days (B1)
Slot usage	**dedicated (C1)**	**dedicated (C1)**	pile-up (C2)	pile-up (C2)	pile-up (C2)	**dedicated (C1)**
Seq. Rule	**FLV (D6)**	**FLV (D6)**	**FLV (D6)**	**FLV (D6)**	**FLV (D6)**	FCFA (D1)
App. Rule	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**
CMS	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	no (F2)
Int. waiting time (min.)	7.11	7.56	8.42	8.53	8.54	16.73
95% CI int. waiting	(6.46, 7.77)	(6.89, 8.23)	(7.77, 9.06)	(7.91, 9.15)	(7.93, 9.15)	(15.79, 17.66)
DP	0.360	0.318	0.330	0.331	0.332	0.320

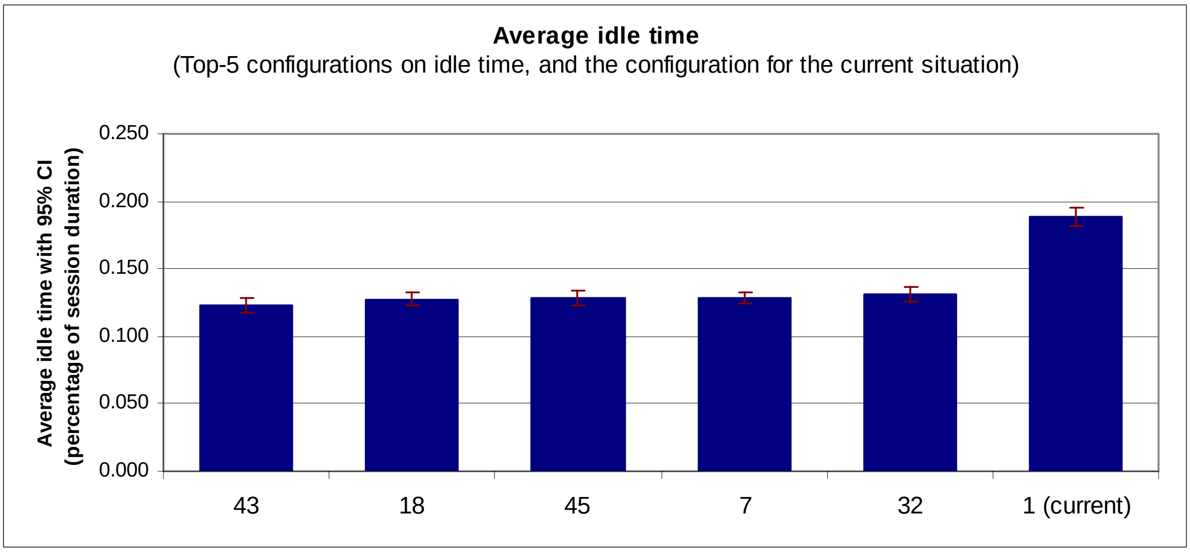
(Experimental factor values that are similar to the value of the best performing configuration are shown bold in the table above.)

Figure 6.2 Performance of top-5 configurations on internal waiting time, and the configuration for the current situation.

6.2.3 Doctors' idle time

Figure 6.3 shows the top-5 performance on the performance indicator of idle time. The usage of time slots appears to influence the idle time. The 'compactness' of an appointment schedule in case of pile-up scheduling, has a positive effect on the idle time. Patients are scheduled directly following each other in a pile-up schedule, resulting in a low probability that the doctor becomes idle. Empty time slots are spread out over the clinic session in case of dedicated time slot usage, which is *scheduled idle time.*

Furthermore, seven out of the ten best performing configurations on idle time contain the Bailey-Welch appointment rule. Scheduling two patients at the beginning of a clinic session contributes to the 'compactness' of the schedule and reduces idle time. These results are as we expected, and support the findings of Hutzschenreuter (2004). The three top-10 schedules on idle time with an individual block appointment rule all contain the FCFS sequencing rule in combination with pile-up scheduling, which ensures a concentration of patients at the beginning of a clinic session as well.



Configuration	43	18	45	7	32	1 (current)
Decisions	**static (A2)**	dynamic (A1)	**static (A2)**	dynamic (A1)	**static (A2)**	dynamic (A1)
SchHorizon	**30days (B2)**	**30days (B2)**	**30days (B2)**	70days (B1)	70days (B1)	70days (B1)
Slot usage	**pile-up (C2)**	**pile-up (C2)**	**pile-up (C2)**	**pile-up (C2)**	**pile-up (C2)**	dedicated (C1)
Seq. Rule	**FSCT (D3)**	FCFA (D1)	FLV (D6)	FCFA (D1)	**FSCT (D3)**	FCFA (D1)
App. Rule	**BW rule (E2)**	Indiv.block (E1)	**BW rule (E2)**	Indiv.block (E1)	**BW rule (E2)**	Indiv.block (E1)
CMS	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	no (F2)
Idle time (%)	12.3	12.8	12.8	12.8	13.1	18.9
95% CI idle time	(11.8, 12.9)	(12.3, 13.2)	(12.3, 13.3)	(12.4, 13.3)	(12.5, 13.6)	(18.1, 19.6)
DP	0.280	0.318	0.277	0.319	0.284	0.320

(Experimental factor values that are similar to the value of the best performing configuration are shown bold in the table above.)

Figure 6.3 Performance of top-5 configurations on the doctors' idle times, and the configuration for the current situation.

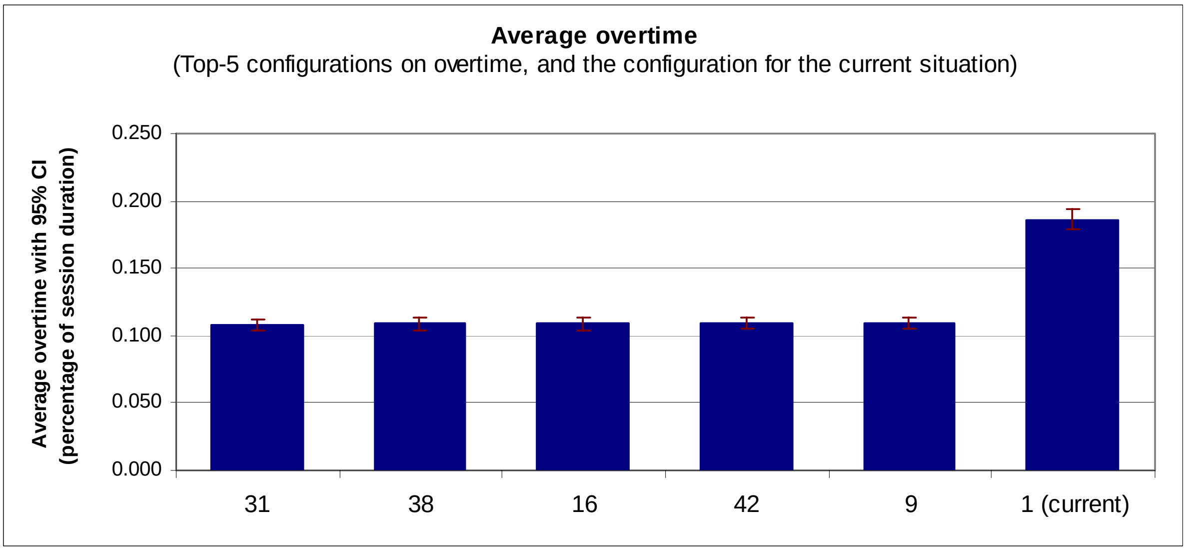
Remarkably, the correction of appointment times for the presence of medical students does not significantly influence the doctor's idle time. We expected a doctor to be idle when his medical student treats an intake patient at the scheduled appointment time. However, the randomness of consultation durations and arrival times corrects for this expected idle time.

6.2.4 Doctors' overtime

The six best performing configurations for the doctors' overtimes all contain dedicated time slots and an individual block appointment rule, but were excluded for extremely long access times. This refers to the trade-off between doctor's and patient's stakes once more, as Bailey (1952) already announced.

The remaining configurations contain individual block systems as well (Figure 6.4). For configurations with a Bailey-Welch rule, the model calculates overtime from the end of the scheduled consultation of the last patient of a clinic session, while the overtime is not corrected for the scheduled consultation duration of the extra patient at the beginning of the clinic session. This results in a conservative representation of overtime. However, all but two configurations still have less overtime than the configuration representing the current situation.

Pile-up scheduling and scheduling consultations with low variance early in the clinic session (FLV-rule), both result in low overtime values. For pile-up scheduling, there are less ‘gaps’ in the schedule, which has a comparable positive effect as for the idle time. Due to the FLV-rule, the largest deviations between the scheduled and the actual consultation durations occur towards the end of a clinic session. These deviations cause doctor’s idle time as well as patient’s internal waiting time. Towards the end of a clinic session, overtime is caused by these disruptions as well. The FLV-rule reduces this effect.



Configuration	31	38	16	42	9	1 (current)
Decisions	**static (A2)**	**static (A2)**	dynamic (A1)	**static (A2)**	dynamic (A1)	dynamic (A1)
SchHorizon	**70days (B1)**	30days (B2)	30days (B2)	30days (B2)	**70days (B1)**	**70days (B1)**
Slot usage	**pile-up (C2)**	dedicated (C1)	dedicated (C1)	**pile-up (C2)**	**pile-up (C2)**	dedicated (C1)
Seq. Rule	**FSCT (D3)**	FLV (D6)	FLV (D6)	**FSCT (D3)**	**FSCT (D3)**	FCFA (D1)
App. Rule	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**
CMS	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**	no (F2)
Overtime (%)	10.8	10.9	10.9	10.9	11.0	18.6
95% CI overtime	(10.4, 11.2)	(10.4, 11.3)	(10.4, 11.4)	(10.5, 11.3)	(10.5, 11.4)	(17.9, 19.3)
DP	0.339	0.360	0.318	0.341	0.339	0.320

(Experimental factor values that are similar to the value of the best performing configuration are shown bold in the table above.)

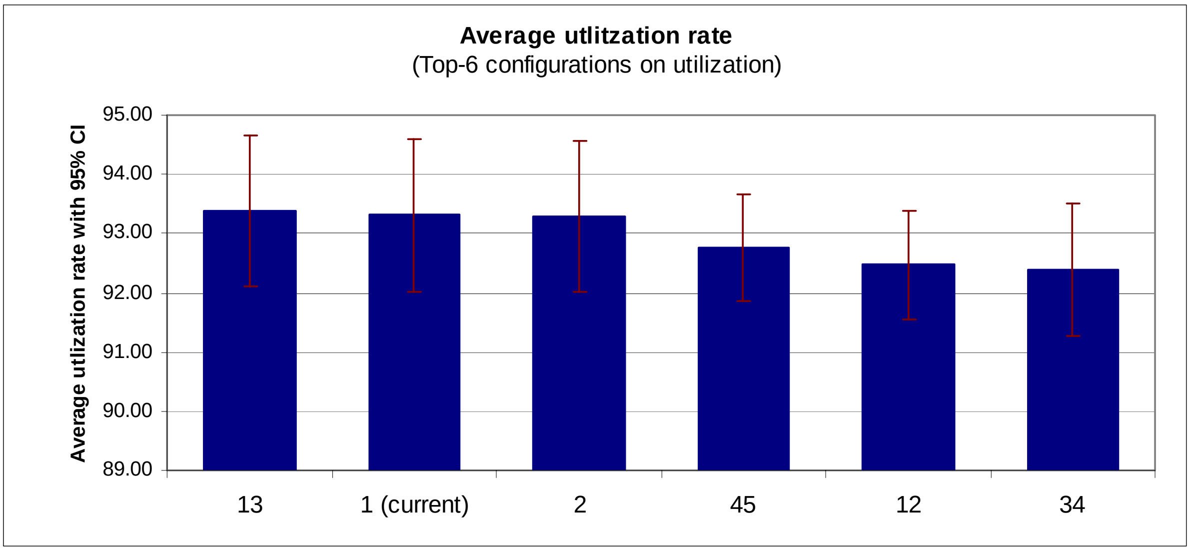
Figure 6.4 Performance of top-5 configurations on the doctors' overtime, and the configuration for the current situation.

6.2.5 Utilization rate

The results for the utilization rate show the same trends as those for the idle time, although the configuration for the current situation and its neighbors score very well on utilization (Figure 6.5). However, this is paid for by high overtime rates, long access times and long internal waiting times. Other well-performing configurations on the utilization rate were excluded for too long internal waiting times. Keeping the doctor busy goes at the costs of long waiting patients and full waiting rooms.

Apart from configurations 1, 2, and 13, all configurations with a utilization rate over 91% contain pile-up scheduling, a Bailey-Welch appointment rule and either a FSCT or a FLV sequencing rule. Again, to pile up patients and to ensure a full schedule at the beginning of the clinic session guarantees low idle time and therefore high utilization rates, not always at the costs of extraordinary long patient waiting times. Remarkably, the method of decision-making has very limited influence on the utilization rate. We expected the local-search heuristic (Section 4.7) included in the static scheduling appointment systems, to improve the utilization rate. Apparently, the eventually positive influence of this local search heuristic is dominated by the influence of the other experimental factors. We expect that applying the heuristic to the schedules of more doctors, and swapping between a larger number of patients, can improve the effect of this local search heuristic on the overall performances of appointment systems.

Configurations with a lower utilization rate are able to treat a lower number of patients. To obtain the yearly production targets, additional capacity is needed. We do not include configurations with extra capacity in the model, since we assumed the capacity to be fixed under all circumstances. However, the model results show that on average approximately one additional clinic session per week suffices for each 2 percent point decrease of utilization rate, compared to the configuration for the current situation. For example: the utilization rate of configuration number 34 is 92.4 %, whereas the utilization rate of the current situation is 93.3 % (Figure 6.5). To treat is many patients per year using the appointment system of configuration number 34 as in the current situation, the outpatient clinic needs to execute on average 0.45 extra clinic sessions per week (i.e. (93.3 – 92.4) / 2).



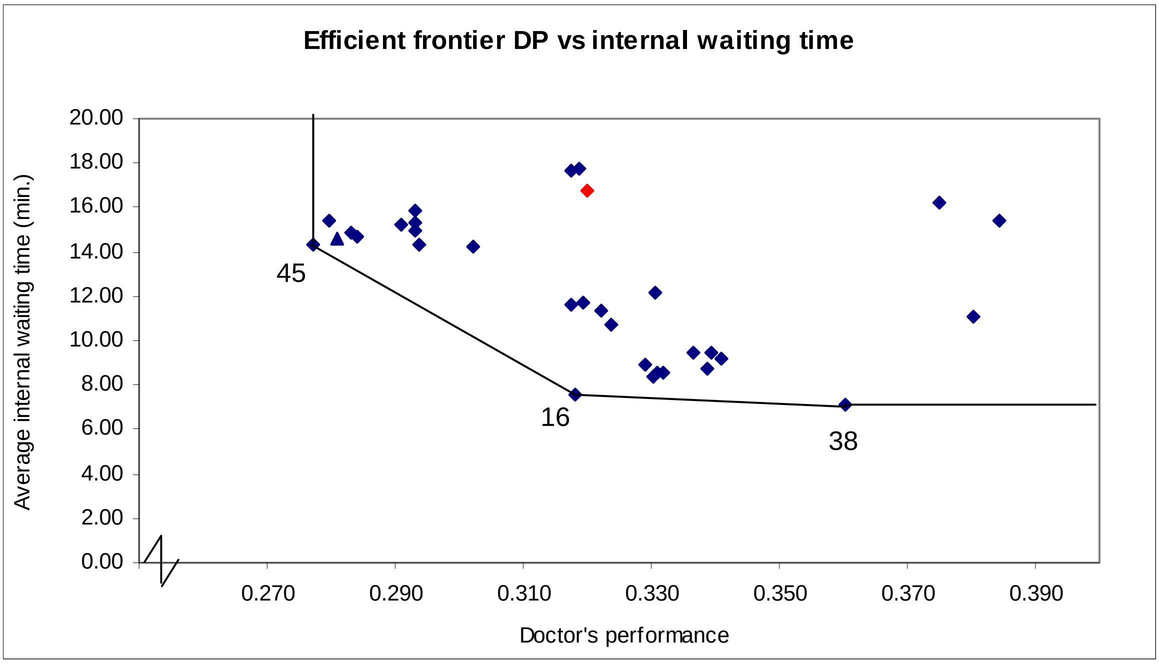
Configuration	13	1 (current)	2	45	12	34
Decisions	**dynamic (A1)**	**dynamic (A1)**	**dynamic (A1)**	static (A2)	**dynamic (A1)**	static (A2)
SchHorizon	**30days (B2)**	70days (B1)	70days (B1)	**30days (B2)**	70days (B1)	70days (B1)
Slot usage	**dedicated (C1)**	**dedicated (C1)**	**dedicated (C1)**	pile-up (C2)	pile-up (C2)	pile-up (C2)
Seq. Rule	**FCFA (D1)**	**FCFA (D1)**	**FCFA (D1)**	FLV (D6)	FLV (D6)	FLV (D6)
App. Rule	**Indiv.block (E1)**	**Indiv.block (E1)**	**Indiv.block (E1)**	BW rule (E2)	BW rule (E2)	BW rule (E2)
CMS	**yes (F1)**	no (F2)	**yes (F1)**	**yes (F1)**	**yes (F1)**	**yes (F1)**
Utilization (%)	93.4	93.3	93.3	92.8	92.5	92.4
95% CI utilization	(92.1, 94.7)	(92.0, 94.6)	(92.0, 94.6)	(91.9, 93.7)	(91.6, 93.4)	(91.3, 93.5)
DP	0.319	0.320	0.317	0.277	0.281	0.283

(Experimental factor values that are similar to the value of the best performing configuration are shown bold in the table above.)

Figure 6.5 Performance of top-6 configurations on the utilization rate.

6.3 Doctor – patient trade-off

The previous section showed different 'best performing' configurations for the patient-based performance indicators at the one hand and the outpatient clinic-based indicators at the other hand. The *doctor's performance* (*DP*) value indicates the weighed performance on the doctor-based indicators *overtime* and *utilization*. To compare the *DP* value with the patient-based *average internal waiting time*, we create an efficient frontier (Figure 6.7).



Configuration	45	16	38
Decisions	static (A2)	dynamic (A1)	static (A2)
PlanHorizon	30days (B2)	30days (B2)	30days (B2)
Slot usage	pile-up (C2)	dedicated (C1)	dedicated (C1)
Seq. Rule	FLV (D6)	FLV (D6)	FLV (D6)
App. Rule	BW rule (E2)	Indiv.block (E1)	Indiv.block (E1)
CMS	yes (F1)	yes (F1)	yes (F1)
Utilization	92.8%	89.5%	87.4%
Overtime	13.2%	10.9%	10.9%
Realized performance	0.277	0.318	0.360
Internal waiting time (min.)	14.3	7.6	7.1

Figure 6.7 Efficient frontier of doctor's performance vs. internal waiting time. The configuration for the current situation is shown with a red dot.

The table below Figure 6.7 shows the main characteristics of the three configurations on the efficient frontier. The best doctor performance is achieved with configuration number 45 (the utmost left point in the graph). Static pile-up scheduling ensures a low amount of unfilled 'gaps' in the appointment schedule, resulting in high doctor utilization (92.8 %). The Bailey-Welch appointment rule further concentrates the patients to the beginning of the clinic session, which also results in a low overtime rate (13.2 %, a reduction of 5.7 percent point compared to the current situation). With the majority of patients arriving at the beginning of the session, the doctor uses his idle time later on in the clinic session to compensate for the overtime. With the FLV sequencing rule, this configuration achieves a low number of disturbances to the first half of the schedule, and thereby results in a relatively low average internal waiting time (14.4 minutes, a reduction of 14%), compared to other high-utilization configurations. However, the utilization rate of the configuration for the current situation (93.3 %) will not be reached, resulting in the need to increase the capacity by 0.6 clinic sessions per week to be able to treat as many patients as in the current situation. Note that the experimental factor values of configuration number 45 are completely opposite to those of the configuration representing the current situation.

Configuration number 38 offers the lowest average internal waiting time of 7.1 minutes, a reduction of 58% compared to the current situation. As a result of its static dedicated scheduling routine, patients can be spread-out evenly over the available slots. Although this results in the fourth-lowest *DP*-value, the planned idle times between consultations offers short waiting times to patients and gives doctors the possibility to hold their appointment times and to reduce overtime to 10.9 %, which is a reduction of 8 percent point compared to the current situation.

However, with a small increase in average internal waiting time of 0.5 minutes (compared to configuration number 38) to 7.6 minutes, configuration number 16 offers a 2.1 percent point higher utilization rate (89.5%). This utilization rate is equivalent to the need of a capacity increase of 2 clinic sessions per week to treat as many patients per year as in the current situation. The only difference in experimental factors values with configuration number 38 is the method of decision-making. We explain the effect of dynamic scheduling on the utilization rate by the fact that the scheduling routine for dynamic scheduling always selects the first available appropriate slots of a day, thereby concentrating patients to the beginning of a clinic session. The local search heuristic in the static scheduling routine also optimizes for the expected internal waiting time.

Another advantage of the dynamic decision-making in configuration 16 is the counter personnel workload, since for static scheduling all patients have to be contacted again once their appointment has been scheduled. Balancing the clinic sessions over the week can improve the workload for the personnel further.

With respect to the ease of implementation, configuration number 16 outperforms configurations 38 and 45. Configuration number 16 has the same experimental factor values as the current situation, except for the length of the scheduling horizon, the sequencing rule, and the adjustment of new patients' appointment times in case medical students attend the clinic sessions. To implement the appointment system of configuration number 16, the outpatient clinic has to change the sequence of the dedicated time slots such that consultation types with a low variance of consultation duration are scheduled early in a clinic session. Combined with a reduction of the length of the scheduling horizon from 70 to 30 working days, the appointment system reduces the access time as well. Implementation of configuration number 45 requires the outpatient clinic to change all control parameters and mechanisms covered by the experimental factors.

Over all configurations, the length of the scheduling horizon and the correction of appointment times for the presence of medical students have little effect on the performance of the appointment system.

6.4 Data Envelopment Analysis

This section briefly describes the results of a Data Envelopment Analysis (DEA), a quantitative performance-evaluation and benchmarking method. We perform an explorative study to the contribution of a DEA to the analysis of the appointment system performances. We use a DEA model based on Coelli *et al.* (1998), as described by Drift (2006).

The basic idea of this sophisticated method for performance evaluation is its ability to determine the efficiency of a configuration for multiple outputs (performance indicators) without using arbitrary ratios to weigh the outputs (Drift, 2006). The DEA model finds for each configuration a virtual configuration by increasing (decreasing) the desirable (undesirable) outputs as much as possible, with the inputs remaining unchanged. The outputs are increased (decreased) within the feasible range of configurations only (Coelli *et al.*, 1998). Figure 6.8 visualizes this principle. In this figure, a virtual configuration B' is created for the real configuration B. X_1 and X_2 are undesirable outputs. Configurations A and C lie on the efficient frontier, meaning that there are no increases (decreases) in outputs possible within the feasible range for these configurations. In Figure 6.8, we call configurations A and C 100%-efficient. The efficiency rate of configuration B is determined by the ratio between the distance from the origin to B', and the distance from the origin to B. Configurations A and C are the *peers* of configuration B (Drift, 2006). Complementary to the efficiency rate, the *peer count* of a configuration is a measure for efficiency as well. The *peer count* is the number of times a 100%-efficient configuration is a *peer* for another configuration. For a detailed description of DEA, we refer to Drift (2006) and Coelli *et al.* (1998).

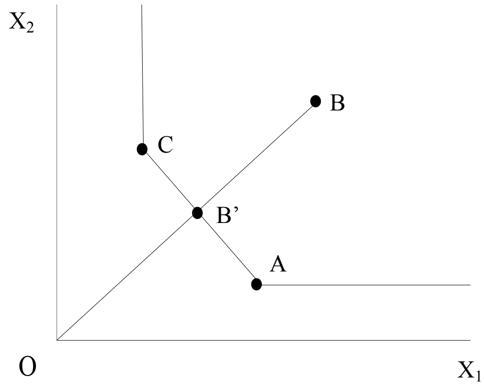


Figure 6.8 Visualization of the DEA model (Drift, 2006).

Since all configurations have the same inputs, we only evaluate the outputs. The performance indicators used as an output in the DEA are *average access time*, *average internal waiting time*, *average overtime rate*, and *average utilization rate*. The latter is a desirable output, while the others are undesirable. We do not include the average idle time, since this performance measure overlaps with the utilization rate.

We use the 'Undesirable Measure Model'-tool of DEA Excel Solver® by Zhu (2003), and Microsoft Excel® to compute the DEA results (Table 6.1).

Table 6.1 shows that 15 configurations are 100% efficient on the selected outputs. The relatively low number of configurations compared to the number of output measures, results in a low level of discriminating power (Drift, 2006). Most efficiency scores lie close to 100%. However, the peer count differs considerably among the efficient configurations. When compared to the patient-doctor trade-off in Section 6.3, the same configurations perform well: numbers 16 and 45. Number 38 (being the third configuration on the efficient frontier in Section 6.3) is efficient as well, but this configuration is not a peer to other configurations. We conclude that configurations 16 and 45 score clearly better than the other configurations, irrespective of the weighing factors for performance indicators.

Table 6.2 shows a summary of Table 6.1, depicting partial configurations that lead to an efficient configuration. The blanks the rows can be filled by any evaluated value to create an efficient configuration. The last three columns show, respectively, the number of efficient configurations that can be created with each partial configuration, the total peer count of these configurations, and the average peer count per configuration to be formed. The first and the fourth row have the largest total and average peer counts, meaning these partial configurations perform best.

Table 6.1 DEA results. Efficient configurations are highlighted green. For the abbreviations, refer to Table 6.1.

Configuration	Efficiency score	Peer(s)	Peer count	Configuration's experimental factor values					
				Decision	Sch horizon	Slot usage	Seq.rule	App. rule	CMS
1	1.000	1.000 x '1'	1	dynamic (A1)	70days (B1)	dedicated (C1)	FCFA (D1)	Indiv.block (E1)	no (F2)
2	1.000	1.000 x '2'	2	dynamic (A1)	70days (B1)	dedicated (C1)	FCFA (D1)	Indiv.block (E1)	yes (F1)
7	0.995	0.218 x '11'; 0.204 x '16'; 0.578 x '45';	0	dynamic (A1)	70days (B1)	pile-up (C2)	FCFA (D1)	Indiv.block (E1)	yes (F1)
9	1.000	1.000 x '9'	0	dynamic (A1)	70days (B1)	pile-up (C2)	FSCT (D3)	Indiv.block (E1)	yes (F1)
10	0.994	0.037 x '31'; 0.925 x '43'; 0.037 x '47';	0	dynamic (A1)	70days (B1)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)
11	1.000	1.000 x '11'	6	dynamic (A1)	70days (B1)	pile-up (C2)	FLV (D6)	Indiv.block (E1)	yes (F1)
12	0.999	0.004 x '16'; 0.005 x '22'; 0.237 x '43'; 0.754 x '45'	0	dynamic (A1)	70days (B1)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)
13	1.000	1.000 x '13'	1	dynamic (A1)	30days (B2)	dedicated (C1)	FCFA (D1)	Indiv.block (E1)	yes (F1)
16	1.000	1.000 x '16'	14	dynamic (A1)	30days (B2)	dedicated (C1)	FLV (D6)	Indiv.block (E1)	yes (F1)
17	0.994	0.281 x '2'; 0.708 x '16'; 0.011 x '24';	0	dynamic (A1)	30days (B2)	dedicated (C1)	FLV (D6)	BW rule (E2)	yes (F1)
18	0.996	0.222 x '11'; 0.218 x '16'; 0.561 x '45';	0	dynamic (A1)	30days (B2)	pile-up (C2)	FCFA (D1)	Indiv.block (E1)	yes (F1)
20	1.000	1.000 x '20'	0	dynamic (A1)	30days (B2)	pile-up (C2)	FSCT (D3)	Indiv.block (E1)	yes (F1)
21	0.995	0.012 x '16'; 0.037 x '31'; 0.951 x '43';	0	dynamic (A1)	30days (B2)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)
22	1.000	1.000 x '22'	6	dynamic (A1)	30days (B2)	pile-up (C2)	FLV (D6)	Indiv.block (E1)	yes (F1)
23	0.990	0.081 x '22'; 0.212 x '43'; 0.612 x '45'; 0.095 x '47'	0	dynamic (A1)	30days (B2)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)
24	1.000	1.000 x '24'	1	static (A2)	70days (B1)	dedicated (C1)	FCFA (D1)	Indiv.block (E1)	yes (F1)
29	0.994	0.460 x '11'; 0.145 x '16'; 0.395 x '45';	0	static (A2)	70days (B1)	pile-up (C2)	FCFA (D1)	Indiv.block (E1)	yes (F1)
31	1.000	1.000 x '31'	2	static (A2)	70days (B1)	pile-up (C2)	FSCT (D3)	Indiv.block (E1)	yes (F1)
32	0.999	0.037 x '16'; 0.037 x '22'; 0.702 x '43'; 0.224 x '45'	0	static (A2)	70days (B1)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)
33	1.000	0.538 x '11'; 0.010 x '16'; 0.441 x '22'; 0.012 x '45'	0	static (A2)	70days (B1)	pile-up (C2)	FLV (D6)	Indiv.block (E1)	yes (F1)
34	0.998	0.056 x '16'; 0.038 x '43'; 0.906 x '45';	0	static (A2)	70days (B1)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)
35	0.967	0.621 x '1'; 0.085 x '16'; 0.294 x '45';	0	static (A2)	30days (B2)	dedicated (C1)	FCFA (D1)	Indiv.block (E1)	yes (F1)
38	1.000	1.000 x '38'	0	static (A2)	30days (B2)	dedicated (C1)	FLV (D6)	Indiv.block (E1)	yes (F1)
39	0.966	0.163 x '2'; 0.810 x '16'; 0.027 x '45';	0	static (A2)	30days (B2)	dedicated (C1)	FLV (D6)	BW rule (E2)	yes (F1)
40	0.992	0.324 x '11'; 0.168 x '16'; 0.508 x '45';	0	static (A2)	30days (B2)	pile-up (C2)	FCFA (D1)	Indiv.block (E1)	yes (F1)
42	1.000	1.000 x '42'	0	static (A2)	30days (B2)	pile-up (C2)	FSCT (D3)	Indiv.block (E1)	yes (F1)
43	1.000	1.000 x '43'	7	static (A2)	30days (B2)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)
44	0.998	0.865 x '11'; 0.016 x '16'; 0.038 x '22'; 0.081 x '45'	0	static (A2)	30days (B2)	pile-up (C2)	FLV (D6)	Indiv.block (E1)	yes (F1)
45	1.000	1.000 x '45'	14	static (A2)	30days (B2)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)
47	1.000	1.000 x '47'	2	static (A2)	10days (B3)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)
49	0.993	0.036 x '16'; 0.019 x '22'; 0.297 x '43'; 0.647 x '45'	0	static (A2)	30days (B2)	pile-up (C2)	FSCT (D3)	BW rule (E2)	no (F2)
50	0.995	0.090 x '13'; 0.910 x '45'	0	dynamic (A1)	70days (B1)	pile-up (C2)	FLV (D6)	BW rule (E2)	no (F2)

Table 6.2 DEA results: partial configurations that lead to an efficient configuration.

Decision	Sch horizon	Slot usage	Seq.rule	App. rule	Number of configurations	Total peer count	Avgerage peer count
static	<=30 days	pile-up	non-FCFA	BW-rule	3	23	7.7
dynamic		dedicated	FCFA	Indiv.block	3	4	1.3
	70 days	dedicated	FCFA	Indiv.block	3	4	1.3
dynamic	30 days		FLV	Indiv.block	2	20	10.0
dynamic		pile-up	FLV	Indiv.block	2	12	6.0
		pile-up	FSCT	Indiv.block	4	2	0.5

6.5 Sensitivity analyses

We perform sensitivity analyses to evaluate the impact of changes to the input data on the performance indicators. If the changes to the input data have little impact on the performance of a configuration, the system is not sensitive to environmental changes and performs well under many circumstances. On the other hand, a large impact on the performance indicates a system that is vulnerable to small changes in the environment. We evaluate the impact of those input parameters we expect have the largest impact on the performances. Only first order effects are considered, which means that we change one input parameter at a time, leaving the other input parameters unchanged. We perform the sensitivity analyses on the configuration representing the current situation, and on the two efficient configurations with the highest peer count, resulting from the DEA: configurations 16 and 45. The subsections below describe the input parameters on which we perform a sensitivity analysis.

6.5.1 Arrival rate of appointment requests

We base the number of appointment requests per year on the realized number of scheduled appointments in the year 2006. Trends suggest an increase in the usage of hospital health care services (StatLine, 2007). We evaluate the sensitivity of the appointment system configurations to an increase in the number of appointment requests by 10%. We expect the utilization rate and the access time to increase under these circumstances.

6.5.2 No-show rate

In the model, we assume the no-show rate to be constant. However, the management of the outpatient clinic considers introducing a financial penalty for patients who do not show up. This policy is likely to reduce the no-show rate. Therefore, we evaluate the impact of a reduction of the no-show rate by 25%. This influences the arrival rate of appointment requests as well, since 59% of the no-show patients requests a new appointment. Therefore, the arrival rate of appointment requests reduces by 0.8% for a 25% lower no-show rate. We expect this reduction in no-show rate influences the utilization rate positively, because a larger percentage of the scheduled patients actually receives treatment. Meanwhile, we expect the average internal waiting time to increase with a reduced no-show rate. A patient who does not show up creates a 'time buffer' for the next patient, resulting in a lower internal waiting time for the latter. With a reduced no-show rate, the occurrence of these 'time buffers' reduces.

6.5.3 *Patients' punctuality*

The punctuality of patients has been measured many times before (e.g. by Blanco White and Pike, 1964; Chen and Robinson, 2005; Harper and Gamlin, 2003; Hutzschenreuter, 2004), resulting in mean punctualities that vary from 0 to 15 minutes early. Our measurements show an average punctuality of 12.9 minutes early (recall Section 5.3.1). In the sensitivity analysis, we evaluate the effects of changing the mean patient punctuality by -2.5 minutes and by +2.5 minutes. We expect this mainly affects the internal waiting times, since this performance measure is a direct result of the patients' punctuality.

6.5.4 *Doctors' punctuality*

As analyzed in Subsection 5.4.1, doctors arrive on average late for their clinic sessions. Doctors' punctuality is only partly controllable by management interventions and discipline, since doctors' activities outside the outpatient clinic may have overtime and cause doctors' late arrivals to the outpatient clinic. We perform a sensitivity analysis on the effects of changes to the doctors' punctuality. We evaluate the effects of changing the mean doctor's punctuality by -2.5 minutes and by +2.5 minutes, on the utilization rate and the internal waiting times. We expect the average internal waiting time to change proportionally to the doctors' punctuality, since the complete clinic session is moved a minute backwards, for each minute the doctor arrives later.

Table 6.3 summarizes the sensitivity analyses. The subsections below describe the performances per indicator of the three evaluated configurations under the different sensitivity analyses.

Table 6.3 Sensitivity analyses.

Input parameter	Change to input parameter	Number
Arrival rate of appointment requests	* 110 %	A
No-show rate	* 75 %	B
Patients' punctuality	- 2.5 minutes	C
Patients' punctuality	+ 2.5 minutes	D
Doctors' punctuality	- 2.5 minutes	E
Doctors' punctuality	+ 2.5 minutes	F

6.5.5 Access time

Figure 6.9 depicts the average access times of the three evaluated configurations for the sensitivity analysis. The ‘Base’ performance shows the performance of the configurations without changes to the input parameters. The letters A and B refer to the first and second sensitivity analysis in Table 6.1 respectively. Sensitivity analyses C to F show no changes at all to the access time, and are therefore not included in Figure 6.9.

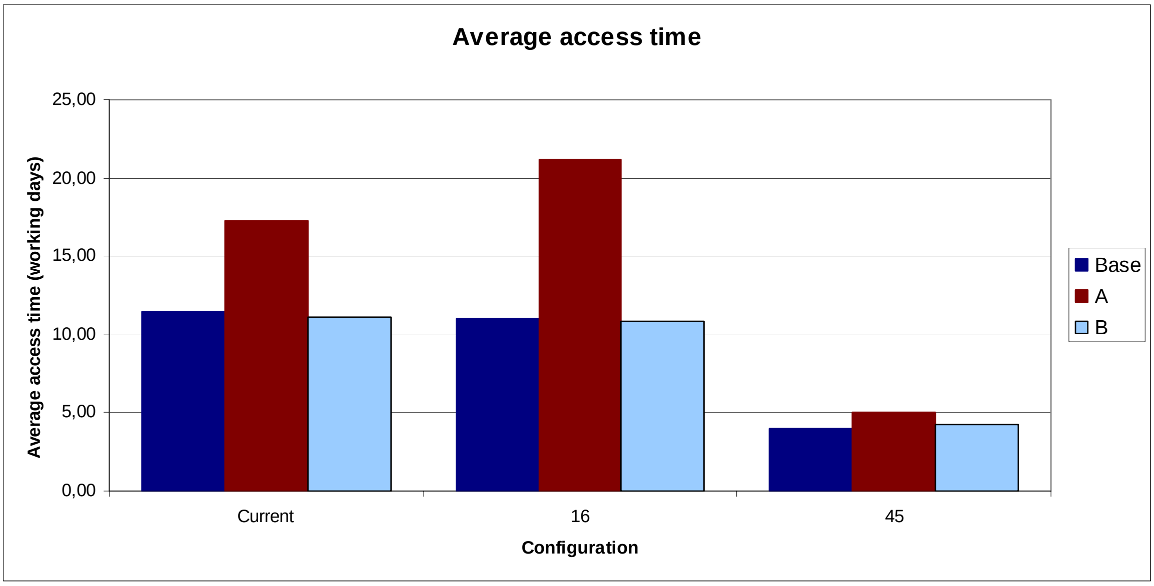
The access time performance of the configurations is very sensitive to an increase in the number of appointment requests. Since the capacity of the outpatient clinic does not change accordingly, the utilization increases, as well as the number of failed appointment requests and the length of the waiting list. The outpatient clinic is unable to treat this large number of patients within reasonable time. The access time increase is especially large for the current configuration and for number 16, which both include dedicated time slot usage and dynamic decision-making. Those factors result in a low level of flexibility to schedule appointments, creating ‘gaps’ in the schedule and long access times at the same time. An appointment system with pile-up scheduling (such as number 45) is less vulnerable to the increase in the number of appointment requests. However, in both cases the outpatient clinic is advised to adjust its capacity according to the arrival rate of appointment requests, in order to balance capacity and demand, and ensure lower access times.

The changes to the access time are smaller for a reduction of the no-show rate (sensitivity analysis B). Contrary to the expectations, the access time increases for configuration number 45. We expected the access time to decrease, because the number of appointment requests decreases for a lower no-show rate. We have no explanation for the 7.3% increase for configuration number 45.

6.5.6 Internal waiting time

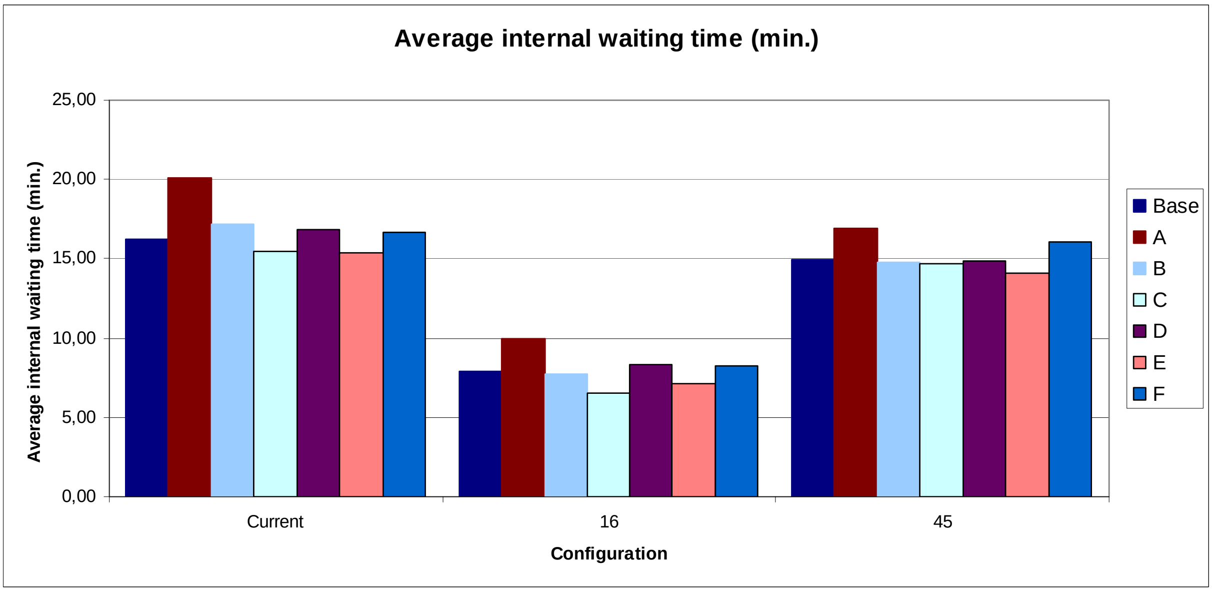
The change to the number of appointment requests results in an increase in internal waiting time (sensitivity analysis A), as Figure 6.10 shows. With more patients requesting an appointment, the clinic sessions are better utilized. A high utilization rate often results in longer internal waiting times, as Section 6.3 shows. A lower no-show rate (B) results in an increase of the internal waiting time for the current situation, as expected. However, the average internal waiting time shows a small decrease for configurations 16 and 45 with a lower no-show rate (B). This is contrary to our expectations. With respect to the punctualities, the average internal waiting time decreases as patients and doctors arrive earlier, and vice versa. The change to the average internal waiting time is restricted to less than 1.5 minute, for a 2.5-minute change to the punctuality.

We conclude that the performances on internal waiting time are vulnerable to changes to the arrival rate of appointment requests, but quite robust to changes to the no-show rate and punctualities.



Configuration	Current	16	45
A	51,0%	92,4%	25,9%
B	-3,0%	-1,2%	7,3%

Figure 6.9 Sensitivity analysis results on the average access time, and a table containing the values and relative changes to the base scenario.

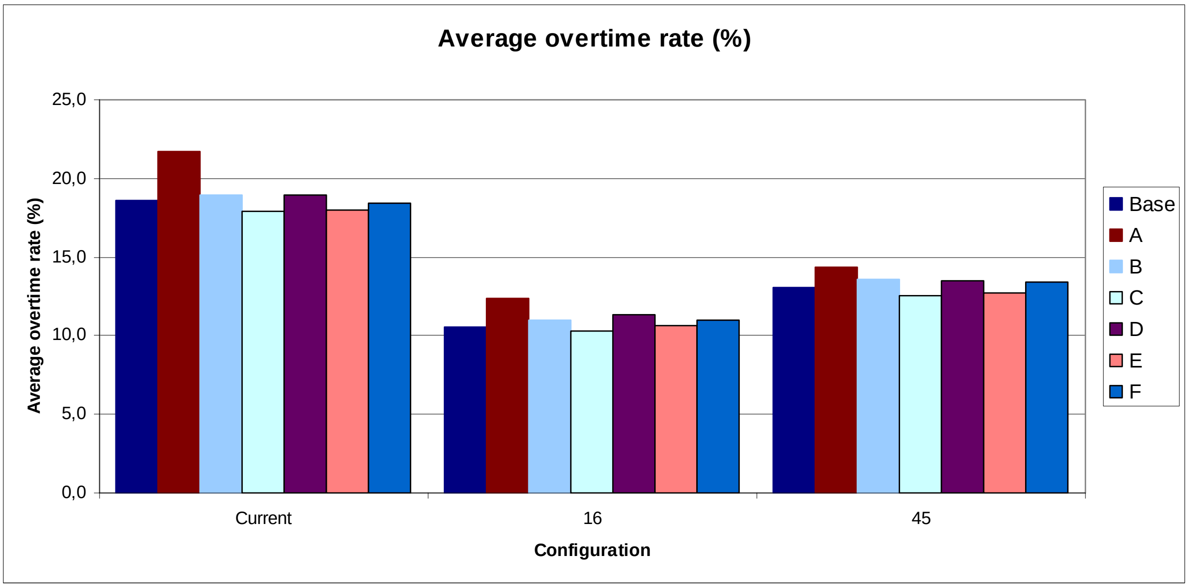


Configuration	Current	16	45
A	23,7%	25,9%	13,2%
B	6,0%	-2,1%	-0,9%
C	-4,5%	-17,3%	-1,7%
D	3,6%	6,2%	-0,4%
E	-5,5%	-10,1%	-5,9%
F	2,7%	4,1%	7,6%

Figure 6.10 Sensitivity analysis results on the average internal waiting time, and a table containing the relative changes to the base scenario.

6.5.7 Overtime rate

The sensitivity analyses results on overtime are comparable to those on internal waiting time, as Figure 6.11 shows. The increase of the arrival rate of appointment requests (A) results in a considerable increase in overtime. The configurations are far less vulnerable to changes to the other input parameters included in the sensitivity analyses.

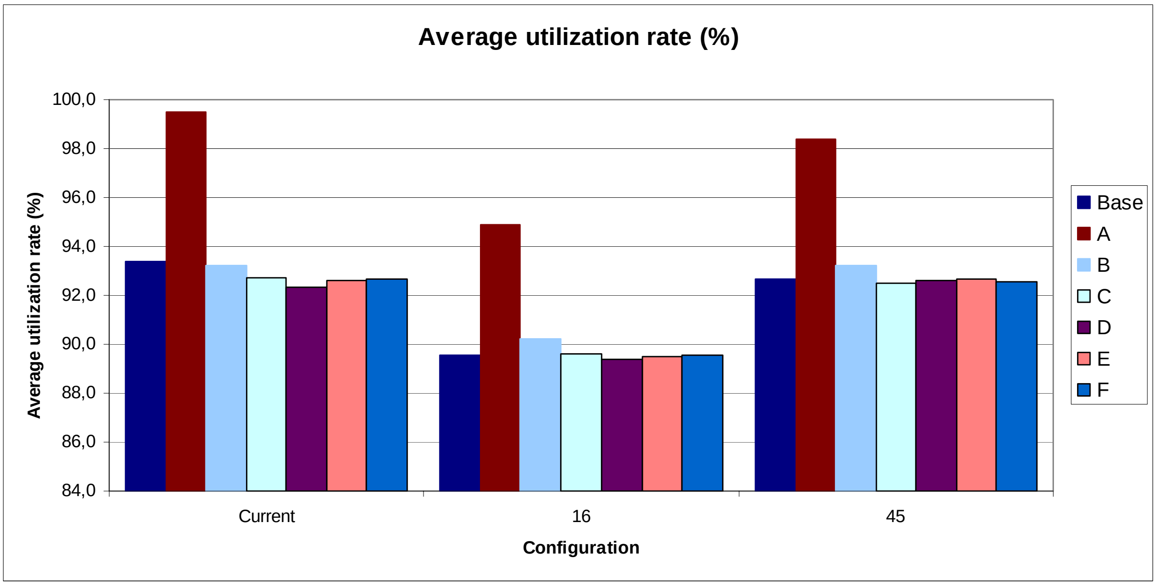


Configuration	Current	16	45
A	16,8%	16,4%	10,2%
B	1,8%	3,5%	3,9%
C	-3,7%	-2,4%	-3,5%
D	1,9%	6,8%	3,5%
E	-3,4%	0,5%	-2,8%
F	-0,7%	4,1%	3,0%

Figure 6.11 Sensitivity analysis results on the average overtime rate, and a table containing the relative changes to the base scenario.

6.5.8 Utilization rate

As we expected, the utilization rate increases for an increasing number of appointment requests (Figure 6.12). The performances are very sensitive to this increase. However, the utilization rate is almost stable for changes to the punctualities. The number of patients per clinic session and the amount of 'gaps' in the schedule do not change for changing punctualities, resulting evidently in stable utilization rates. A decreasing no-show rate results in a light increase of the utilization rate, as expected. This increase is at most 0.7 percent point (for configuration 16). We conclude that the utilization performances are not sensitive to changing input parameters, except for an increasing number of appointment requests.



Configuration	Current	16	45
A	6,5%	5,9%	6,1%
B	-0,1%	0,7%	0,6%
C	-0,7%	0,0%	-0,2%
D	-1,2%	-0,2%	-0,1%
E	-0,8%	-0,1%	0,0%
F	-0,8%	0,0%	-0,1%

Figure 6.12 Sensitivity analysis results on the average utilization rate, and a table containing the relative changes to the base scenario.

6.5.9 Efficient frontier

An important aspect is the sensitivity of the efficient frontier, as presented in Section 6.3. What changes to the input parameters cause other configurations to lie on the efficient frontier? Figure 6.13 shows the efficient frontier of Section 6.3, to which the results of the sensitivity analyses are added. The yellow triangles represent the sensitivity-analysis results of the current situation (red dot), the pink squares represent the sensitivity-analysis results of configuration number 16, and the light-blue crosses represent the sensitivity-analysis results of configuration number 45. The light-blue crosses interfere with the performances of other configurations (dark-blue dots). This means configuration number 45 outperforms other configurations for a limited set of input parameters. However, the pink squares (for configuration 16) remain on the efficient frontier.

The purple stars represent sensitivity analysis results for configurations 22, 33 and 44, lying closest to number 16. One purple star exactly overlaps the dark blue dot of configuration 16, meaning that configuration number 33 with patients arriving on average 2.5 minutes earlier (sensitivity analysis C) performs equally on overtime, utilization and internal waiting time as configuration 16 with ‘standard’ patient punctuality.

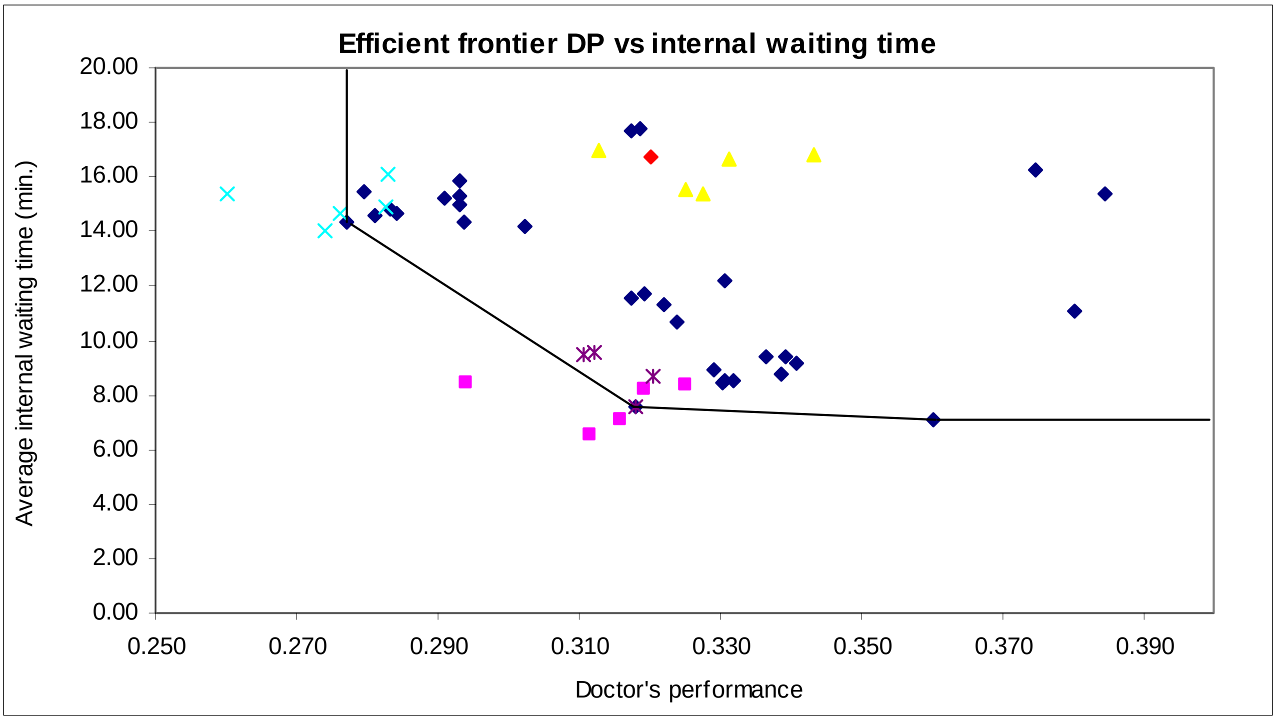


Figure 6.13 Sensitivity analysis results and their effects on the efficient frontier.

Concluding the sensitivity analysis, we state that the performance of configuration number 16 is least vulnerable to changes to the input parameters, except for its performance on access time. Changes to punctualities will not affect the performances on a large scale. The outpatient clinic's management must be aware of changes to the no-show rate, since they affect the overtime rate and internal waiting time. Capacity adjustments are necessary to cope with changes to the amount of appointment requests.

7. Conclusions and recommendations

7.1 Conclusions

We refer back to the aim of this study, as stated in Chapter 1:

The aim of this study is to improve the appointment system for the Erasmus MC ENT outpatient clinic such that patients' internal waiting times and access times are shortened, the personnel's workload is balanced, and the outpatient clinic's utilization rate is increased at the same time.

The simulation model of the Erasmus MC ENT outpatient clinic shows that a number of changes to the appointment system account for significant improvements in average access times, average internal waiting times, overtime, and utilization.

An appointment system with dedicated time slot usage and an individual block appointment rule, combined with a sequencing rule that assigns patients with low variance of consultation duration to the beginning of a clinic session (configuration number 16), is able to achieve a 55 % reduction of average internal waiting times, and an average overtime that falls from 19 % to 11 %. This appointment system involves a relatively small number of changes to the current appointment system. However, the utilization rate is 89.5 %, whereas the configuration for the current situation has a utilization rate of 93.3 %. Therefore, on average 2 additional clinic sessions per week are needed to treat as many patients as in the current situation. Sensitivity analyses show vulnerability of the performance of this appointment system to an increasing number of appointment requests.

On the other hand, an appointment system where patients are scheduled statically to a pile-up appointment schedule with a Bailey-Welch appointment rule, and a sequencing rule that assigns patients with low variance of consultation duration to the beginning of a clinic session (configuration number 45), reaches a high utilization of the doctor's time in the outpatient clinic (92.8 %), while the average internal waiting time drops by 14 % to 14 minutes, and thereby remains within the acceptable range. The performance of this appointment system is vulnerable to an increasing number of appointment requests as well. Implementation of this appointment system requires the outpatient clinic to change all the values of all evaluated control parameters and mechanisms. To treat as many patients as in the current situation, the outpatient clinic has to execute on average 0.6 extra clinic sessions per week.

The length of the scheduling horizon, and the correction of appointment times for the presence of medical students, have little effect on the performance of the appointment systems. The effects of the other experimental factors depend heavily on the combination of experimental factor values.

Under the formulated patient service level restrictions, the utilization rate of the current situation (93.3%) will only be reached by one other configuration, which differs slightly from the current appointment system. Restrictions to ensure a minimum patient service level urge the outpatient clinic to accept a lower utilization rate, although idle time and overtime rates can decline. These lower utilization rates have to be compensated for, by increasing the capacity to be able to treat the same number of patients as in the current situation. The capacity can be increased by adjusting the number of clinic sessions per week, or the length of clinic sessions. Results show that every additional clinic session per week compensates for a utilization rate decrease of approximately 2 percent point.

The direct contribution of the new classification of consultation types to variance reduction of consultations is unclear. However, the classification helps the schedulers to identify which patients have short or long expected durations or variances of consultation times, and is therefore very useful for sequencing purposes.

An appointment system with dynamic decision-making will not further increase the pressure on the workload for the counter and scheduling personnel, whereas static scheduling involves an extra call-back to each patient for the counter personnel, thereby raising the workload. A better balancing of clinic sessions over the week can improve the workload.

7.2 Recommendations

We recommend the management and doctors of the Erasmus MC ENT outpatient clinic the following:

- We recommend the implementation of the appointment system of configuration number 16, which consists of dynamic decision-making, a scheduling horizon of 30 days, dedicated time slot usage, a sequencing rule that schedules low-variance consultations at the beginning of a clinic session, individual appointment blocks, and a correction of the appointment times of new patients incase a medical student attends the clinic session. This appointment system yields a reduction of the average internal waiting time of 55% and a reduction of the overtime rate of 19% to 11%. The utilization rate drops from 93.3% to 89.5%. Implementation of this appointment system requires changes to the length of the scheduling horizon, the sequencing rule, and the correction of new patients' appointment times in case of the attendance of medical students.

- If the doctors, management, and personnel of the outpatient clinic prefer a higher utilization rate, we recommend implementing the appointment system of configuration number 45. This appointment system yields a utilization rate of 92.8%, which is almost as high as in the current situation. The internal waiting time reduces by 14%, while the overtime rate falls to 13%. However, the implementation of this appointment system requires the outpatient clinic to change all control parameters and mechanisms covered by the experimental factors of this study.

- We recommend to increase and decrease the outpatient clinic's capacity flexibly, based on forecasts for the arriving number of appointment requests. Sensitivity analyses show that the performances of appointment systems are vulnerable to an increasing number of appointment requests. If capacity can be adjusted accordingly, these negative effects may be reduced.

- We recommend to develop a questionnaire for the schedulers to determine a patient's consultation type when he requests an appointment. A good estimation of a patient's consultation duration and variance reduces waiting times and overtime. Doctors can give an indication for the consultation type and expected duration when he asks a patient to request a review appointment at the counter.

- We recommend standardizing the counter processes, to shorten the process time at the counter to avoid unnecessary doctor idle time, in case a doctor is available and the patient still queues for the counter. Improvements have been made with an electronic calling system to direct patients to the appropriate consultation room. Further standardization regarding the creation of medical status records and incoming telephone calls can increase the counter's availability.

- A reduction of the return rate reduces the number of appointment requests of review patients. Capacity can be reduced accordingly, to avoid a decreasing utilization rate. A lower number of clinic sessions per week results in improved availability of doctors for other activities. The reduction of the return rate is part of medical policy, as well as a decision of a doctor for every individual patient. Reduction of the return rate is only possible if medically justified.

- A collaborative approach of the doctors to balance their clinic sessions over the week can further improve the workload of the counter personnel. Since the work schedules of doctors are subject to several political and personal considerations, the doctors must play a proactive role to balance the production of the outpatient clinic over the week.

- We recommend to repeat time measurements and management information analyses on a regular basis to monitor the appointment-system performance. The averages and variances of the consultation types can be adjusted according to recent time measurements, resulting in better estimates of consultation durations. A changing percentage of new patients, and employing a new doctor are other reasons to perform time measurements again.

We recommend the following subjects for further research:

- This study is highly customized to the ENT outpatient clinic in the Erasmus MC. A further study on the generalization of the factors found in this study can contribute to a better understanding of the performance of appointment systems under realistic environmental conditions.

- A further study to the influence of the length of the scheduling horizon and the correction of appointment times to the presence of medical students on the performance of an appointment system contributes to the field of outpatient scheduling. We were unable to identify the contribution of these factors in the Erasmus MC ENT outpatient clinic with our model.

- We assume a stable capacity in this study. A fluctuating number of clinic sessions per week may influence the access time and utilization rate. Flexible capacity in periods of high or low patient demand may have positive effects on these performance indicators. We recommend additional research on the effects of unstable capacity.

- An outpatient clinic is usually part of a larger clinical pathway of patients in a hospital. The capacity, access time and return rate of an outpatient clinic influence the patient demand and waiting lists of inpatient departments and operating room departments. We recommend further research on this relationship, as well as on strategic planning of a clinical pathway.

References

Bailey, N. (1952). A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting times. *Journal of the Royal Statistical Society*, A14, 185-199.

Bailey, N. (1954). Queuing for medical care. *Applied Statistics* 3, 137-145.

Blanco White, M. and M. Pike. (1964). Appointment systems in Out-patients' Clinics and the Effect of Patients' Unpunctuality. *Medical Care* 2, 133-145.

Brahimi, M and D. Worthington. (1991). Queueing Models for Out-patient Appointment Systems: A Case Study. *Journal of the Operational Research Society* 42, 9, 733-746.

Cayirli, T and E. Veral (2003). Outpatient scheduling in health care: a review of literature. *Production and Operations Management Society* 12, 4, 519-549.

Cayirli, T, E. Veral, and H. Rosen. (2004). Assessment of patient classification in appointment systems. *1st Conference of the POMS College of Service Operations*, New York, NY, USA.

Cayirli, T, E. Veral, and H. Rosen. (2006). Designing appointment scheduling systems for ambulatory care services. *Health Care Management Science* 9, 47–58.

CBO (2004). Meetplan Werken zonder Wachtlijst. CBO Kwaliteitsinstituut voor de Gezondheidszorg, Utrecht. *In Dutch.*

CBO (2006). Werkmap Doorbraakproject Werken Zonder Wachtlijst SB p3 2006-2007. CBO Kwaliteitsinstituut voor de Gezondheidszorg, Utrecht. *In Dutch.*

Chen, R. and L. Robinson. (2005). Scheduling doctor's appointments with unpunctual patient arrivals. *Working paper, Davis Graduate School of Management, University of California, USA.*

Clague, J., P. Reed, J. Barlow, R. Rada, M. Clarke, and R. Edwards. (1997). Improving outpatient clinic efficiency using computer simulation. *International Journal of Health Care Quality Assurance* 10, 5, 197-201.

Coelli, T., D. Prasada Rao, and G. Battese. (1998). *An introduction to efficiency and productivity analysis.* Norwell, MA, USA: Kluwer Academic Publishers Group.

Denton, B. and D. Gupta (2003). A Sequential Bounding Approach for Optimal Appointment Scheduling. *IIE Transactions* 35, 11, 1003-1016.

Drift, M. van der. (2006). Performance Evaluation of Operating Rooms: A benchmark study in eight University Hospitals. *Master's thesis, University of Twente, Enschede, The Netherlands.*

Fetter, R. and J. Thompson. (1966). Patients' waiting time and doctors' idle time in the outpatient setting. *Health Services Research* 1, 1, 66–90.

Fries, B. and V. Marathe. (1981). Determination of optimal variable-sized multiple-block appointment systems. *Operations Research* 29, 2, 324-345.

Harper, P. and H. Gamlin. (2003). Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. *OR Spectrum* 5, 2, 207-222.

Ho, C, ,H. Lau, and J. Li (1995). Introducing variable-interval appointment scheduling rules in service systems. *International Journal of Production & Operations Management* 15, 6, 59-68.

Houdenhoven, M. van, G. Wullink, E. Hans, and G. Kazemier. (2006). A framework for Hospital Planning and Control. *Working paper, Erasmus Medical Center, Rotterdam, The Netherlands.*

Hutzschenreuter, A. (2004). Waiting Patiently: An analysis of the performance aspects of outpatient scheduling in health care institutes. *BMI Paper, Vrije Universiteit, Amsterdam, The Netherlands*.

Huang, X. (1994). Patient attitude towards waiting in an outpatient clinic and its applications. *Health Services Management Research,* 7, 2-8

Kaandorp, G, and G. Koole (2007). Optimal outpatient appointment scheduling. *Working paper, Department of Mathematics, Vrije Universiteit Amsterdam, The Netherlands.*

Klassen, K, and T. Rohleder (1996). Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management* 14, 2, 83-101.

Klassen, K, and T. Rohleder (2004). Outpatient appointment scheduling with urgent clients in a dynamic, multi-period environment. *International Journal of Service Industry Managemen*t 15, 2, 167-186.

Lau H. and A. Lau (2000). A fast procedure for computing the total system costs of an appointment schedule for medical and kindred facilities. *IIE Transactions* 32, 9, 833-839.

Lehaney, B., S. Clarke and R. Paul. (1999). A case of intervention in an outpatient department. *Journal of the Operational Research Society* 50, 9, 877-891.

Liu, L. and X. Liu. (1998). Block appointment systems for outpatient clinics with multiple doctors. *Journal of the Operational Research Society* 49, 1254-1259.

OECD (2005). Health at a Glance – OECD Indicators 2005. Available at *http://www.oecd.org*.

Patel, H., C. Luxman, T. Bailey, J. Brunning, D. Zemmel, L. Morrell, M. Nathan, and R. Miller. (2002). Outpatient clinic: where is the delay? *Journal of the Royal Society of Medicine* 95, 604-605.

Robinson, L. and R. Chen (2003). Scheduling doctors' appointments: optimal and empirically-based heuristic policies. *IIE Transactions* 35, 298-307.

Robinson, S. (1994). *Successful Simulation: A Practical Approach to Simulation Projects*. Maidenhead, UK: McGraw-Hill.

Rohleder, T. and K. Klassen. (2000). Using client-variance information to improve dynamic appointment scheduling performance. *Omega* 28, 3, 293-302.

Rohleder, T. and K. Klassen. (2002). Rolling Horizon Appointment Scheduling: A Simulation Study. *Health Care Management Science* 5, 201–209.

StatLine (2007). *Gebruik Medische Voorzieningen: Specialist.* 1981-2006. Voorburg, The Netherlands: Centraal Bureau voor de Statistiek. *In Dutch*. Available at *www.cbs.nl*.

TPG (2004). Het kan écht: betere zorg voor minder geld. Sneller Beter – Logistiek in de zorg. Eindrapportate TPG. Hoofddorp: TPG. *In Dutch.* Available at *http://www.snellerbeter.nl*

Vanden Bosch, P. and D. Dietz. (2000). Minimizing expected waiting times in a medical appointment system. *IIE Transactions* 32, 841-848.

Vanden Bosch, P., D. Dietz and J. Simeoni. (1999). Scheduling Customer Arrivals to a Stochastic Service System. *Naval Research Logistics* 46, 549-559.

Vissers, J, and J. Wijngaard. (1979). The outpatient appointment system: design of a simulation study. *European Journal of Operations Research* 3, 6, 459-463.

Wang, P. (1999). Sequencing and scheduling N customers for a stochastic server. *European Journal of Operational Research*, 119, 729-738.

Welch, J. and N. Bailey. (1952). Appointment systems in Hospital Outpatient Departments. *The Lancet*, 259, 1105-1108.

Zhu, J. (2003). *Quantitative models for performance evaluation and benchmarking.* Dordrecht, The Netherlands: Kluwer Academic Publishers Group.

Appendix A Definitions

Access time	The period between a patient's request for an appointment and his arrival at the outpatient clinic.
Administration time	Doctor's time for administration and cleaning when the patient has left the consultation room. Part of gross consultation time.
Appointment interval	The scheduled interval between two successive appointment times (Cayirli and Veral, 2003).
Appointment system	Combination of control parameters and mechanisms that determines the way of scheduling patients for outpatient doctor's appointments.
Appointment time	The scheduled starting time of a patients' net consultation time.
Block-size	The number of patients scheduled for the same appointment time for the same doctor.
Consultation type	Typology, indicating the phase in the clinical pathway of a patient's treatment. Two consultations of one doctor with the same consultation type, are assumed to have the same expected gross consultation time.
Control parameter or mechanism	Controllable factor that forms a part of an appointment system.
Desirable inter-consultation time (DICT)	Necessary time between two consultations, which cannot be related to any one patient directly.
Doctors' idle time	The total time during the clinic session when the doctor is not consulting, and there are no patients waiting to be seen. (Cayirli and Veral, 2003).
Doctors' punctuality	The time difference between the appointment time of the first patient in a doctor's clinic session, and the doctor's actual arrival time at the outpatient clinic.

Dynamic scheduling	Scheduling patients throughout the day without knowledge of the type and number of clients that will call for an appointment later (Klassen and Rohleder, 2004). Dynamic scheduling is a particular method of decision-making.
Gross consultation time	All the time during which a patient is claiming a doctor's attention, or at least preventing him from seeing the next patient (Bailey, 1952).
Initial block	The number of patients that are scheduled for the first appointment time of a doctor's clinic session.
Input parameters	Factors that serve as input for the modeled outpatient clinic and its appointment system, and that are not controllable by the outpatient clinic's management of medical staff.
Internal waiting time	The period between the scheduled starting time and the actual starting time of a patient's consultation.
Medical student	Student who consults new patients separately from a resident doctor or specialist doctor, but who is not allowed to diagnose or to perform medical actions himself. In Dutch: *co-assistent.*
Net consultation time	All time during which the patient is in the consultation room. Part of gross consultation time.
New patient	A patient who either (1) visits the outpatient clinic for the first time, or (2) visits the outpatient clinic for a new medical complaint.
No-show rate	The percentage of patients that do not show up at the outpatient clinic for their appointments, without canceling at least one day in advance.
Overtime	The positive time difference between the scheduled completion time of the clinic session and the actual end of the doctor's administration time for the last patient.
Patients' punctuality	The time difference between a patient's appointment time and his actual arrival time at the outpatient clinic (Cayirli and Veral, 2003).

Preparation time	Time in which the doctor reads the patient's status record and test results and/or prepares material requirements for treatment. Part of gross consultation time.
Queue discipline	The order of calling patients from the waiting room by the doctor. First-come-first-serve or first-appointment-first-serve.
Resident doctor	Doctor with a medical degree (M.D.), who is not a specialist doctor (yet). In Dutch: *arts-assistent*.
Return rate	The number of review patients compared to new patients.
Review patient	A patient who has visited the outpatient clinic before with the same medical complaint.
Scheduling horizon	The number of shifts for which appointment can be scheduled in advance.
Scope	The number of successive shifts that are considered in an appointment-scheduling model.
Sequencing rule	The order in which patients, who call with an appointment request, are assigned to blocks based on a particular patient classification scheme (Cayirli, Veral, and Rosen, 2006).
Shift	Morning or afternoon of a working day, in which one or more parallel clinic sessions take place.
Slot	Appointment interval with a pre-defined length that is designated to patients with a certain consultation type.
Specialist doctor	Doctor, who is specialized in one sub-specialization of ENT.
Stages	The succeeding processes that are considered in the appointment-scheduling model (counter process, consultation with medical student, consultation with specialist, etc).

Static scheduling	Scheduling a finite number of patients simultaneously (Wang, 1999) at the beginning of each shift, using all appointments requests that arrived after the last moment of scheduling. Static scheduling is a particular method of decision-making.
Walk-in patient	A patient who arrives at the outpatient clinic without having an appointment.

Appendix B Extended formal problem description

Sets

$s \in S$	Doctor, an element of the set of all doctors *S*.
$p \in P$	Patient, an element of the set of all patients *P*.
$j \in J$	Scheduling shift index. $j = \{1,\ldots,J\}$. Each shift represents half a day. In every shift, one or more parallel clinic sessions (of different doctors) are performed.
$t \in T$	Interval index within a shift.

Parameters

T	Total number of intervals per shift.
q	Length of one interval in minutes
$\rho_i^{patient}$	Probability that a patient cancels his appointment *i* shifts in advance ($\rho_0^{patient}$ = no show rate)
ρ_i^{doctor}	Probability that a doctor cancels his clinic session *i* shifts in advance ($\rho_i^{doctor} = 0 \quad \text{for} \quad 0 \le i \le 60$, following a management decision).
π	Rolling scheduling horizon (number of shifts)

Decision variables

$x_{psjt} \in \{0,1\}$	Binary variable indicating whether patient *p* has an appointment with doctor *s* in shift *j*, starting at the beginning of interval *t*.
$f(x_{psjt})$	Appointment time of patient *p*.
$n_{jts} \in IN$	Number of patients whose appointments are scheduled to start at the beginning of the same interval *t* for the same doctor *s* in shift *j*.
$y_{ps} \in IN$	Number of successive intervals for which patient *p* is scheduled for his appointment with doctor *s*.

Stochastic variables

r_p	Arrival time of patient's *p* appointment request. Stochastic variable with an exponential distribution with mean inter-arrival time of λ minutes.

b_p	Arrival time of patient p at the outpatient clinic in the shift of his appointment. Stochastic variable with a normal distribution with mean $\left(f(x_{psjt})+\mu^{patients}\right)$ and standard deviation $\sigma^{patients}$.
b_{sj}	Arrival time of doctor s at the outpatient clinic in shift j. Stochastic variable with a normal distribution with mean $\left(f(x_{psj1})+\mu^{doctors}\right)$ and standard deviation $\sigma^{doctors}$.
L_p	Stochastic variable for the counter process duration for patient p, which starts at time l_p. L_p follows a gamma probability distribution with mean $\alpha_p^L\beta_p^L$ and variance $\alpha_p^L\left(\beta_p^L\right)^2$. The parameters of this gamma distribution depend on the patient's consultation type c, diagnosis type d and the treating doctor s.
V_p	Stochastic variable for the preparation duration for consultation with patient p, which starts at time v_p. V_p follows a gamma probability distribution with mean $\alpha_p^V\beta_p^V$ and variance $\alpha_p^V\left(\beta_p^V\right)^2$, equivalent to L_p.
H_p	Stochastic variable for the duration of net consultation of patient p, which starts at time $v_p + V_p$. H_p follows a gamma probability distribution with mean $\alpha_p^H\beta_p^H$ and variance $\alpha_p^H\left(\beta_p^H\right)^2$, equivalent to L_p.
A_p	Stochastic variable for the administration duration of the doctor for the consultation with patient p, which starts at time $v_p + V_p + H_p$. A_p follows a gamma probability distribution with mean $\alpha_p^A\beta_p^A$ and variance $\alpha_p^A\left(\beta_p^A\right)^2$, equivalent to L_p.
G_p	Gross consultation duration of patient p, which is determined by $V_p + H_p + A_p$.

Because of the stochastic nature of L_p, V_p, H_p and A_p, the variables l_p, v_p, and G_p are stochastic variables as well.

$DICT_{pp^*s}$	Stochastic variable for the duration of inter-consultation time the doctor consumes between his consultation with patients p and p^*. $DICT_{ps}$ follows a gamma distribution with mean $\alpha_{pp^*}^D\beta_{pp^*}^D$ and variance $\alpha_{pp^*}^D\left(\beta_{pp^*}^D\right)^2$, which depend on the consultation type of patient p and the consultation type of the next patient p^* on the appointment schedule, as well as on the treating doctor s.

Relationship

$y_{ps} = [E(G_p)/q]$, rounded to the nearest integer.

Planned performance measures

PI_s Planned idle time of doctor s

PO_s Planned overtime of doctor s

PIW_p Planned internal waiting time of patient p

PPL Planned peak load for counter

$$PIW_p = \begin{cases} 0 \quad for\ n_{t's} \leq 1 \\ \dfrac{1}{n_{t's}} \sum_{m \in P} E(G_m) \quad for\ n_{t's} > 1 \end{cases}$$ with $t' = \dfrac{f(x_{psjt'})}{q}$ and $m \in P$, all patients for which $x_{msjt} = x_{psjt}$.

$$PO_s = \max\left\{0, \sum_{p' \in P} \left(f(x_{p'sjt}) + PIW_{p'} + E(G_p) + E(DICT_{p'0s}) - Tq \right) \right\}$$ with $p' \in P'$, all patients with the latest appointment time of their clinic session.

$$PI_s = \sum_j \left(f(x_{p'sjt}) + PIW_{p'} + E(G_p) + E(DICT_{p'0s}) - (j-1)Tq \right) - \sum_p \left(y_{ps} + \alpha_p^D \beta_p^D \right) - PO_i$$ with $p' \in P'$, all patients with the latest appointment time of their clinic session.

$$PPL = \max n_{ts} (\forall t, s)$$

Scheduling objective

$$SO = \min\left\{ \gamma_{PO} \sum_{s=1}^{S} PO_s + \gamma_{PI} \sum_{s=1}^{S} PI_s + \gamma_{PIW} \sum_{p=1}^{P} PIW_p + \gamma_{PPL} PPL \right\}$$

Appendix C Punctuality Q-Q Plots

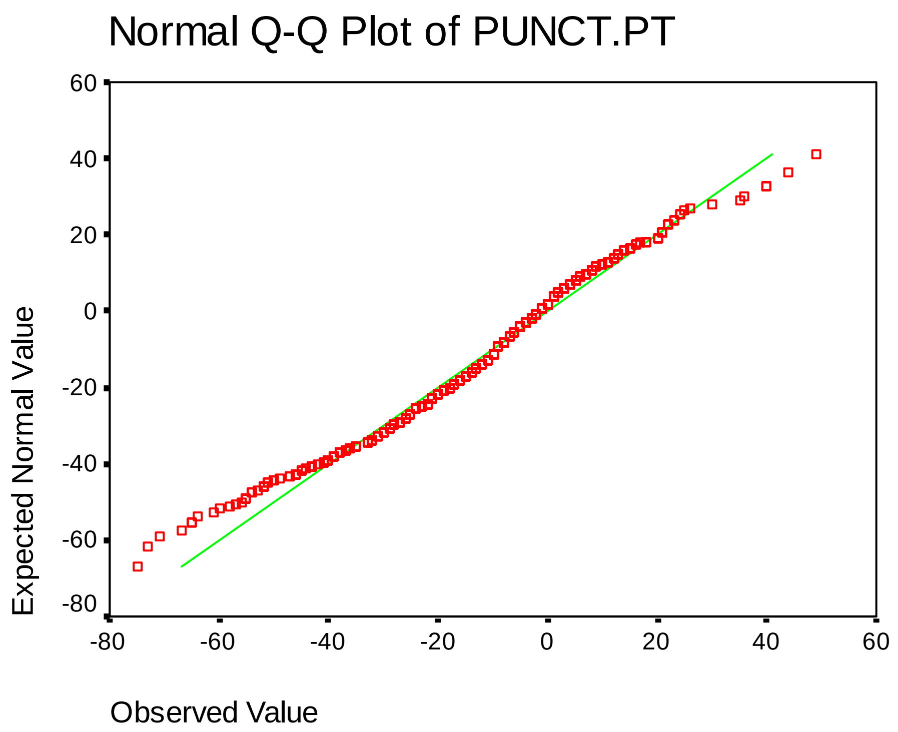


Figure C.1 Normal Q-Q Plot of patients' punctuality (N=748).

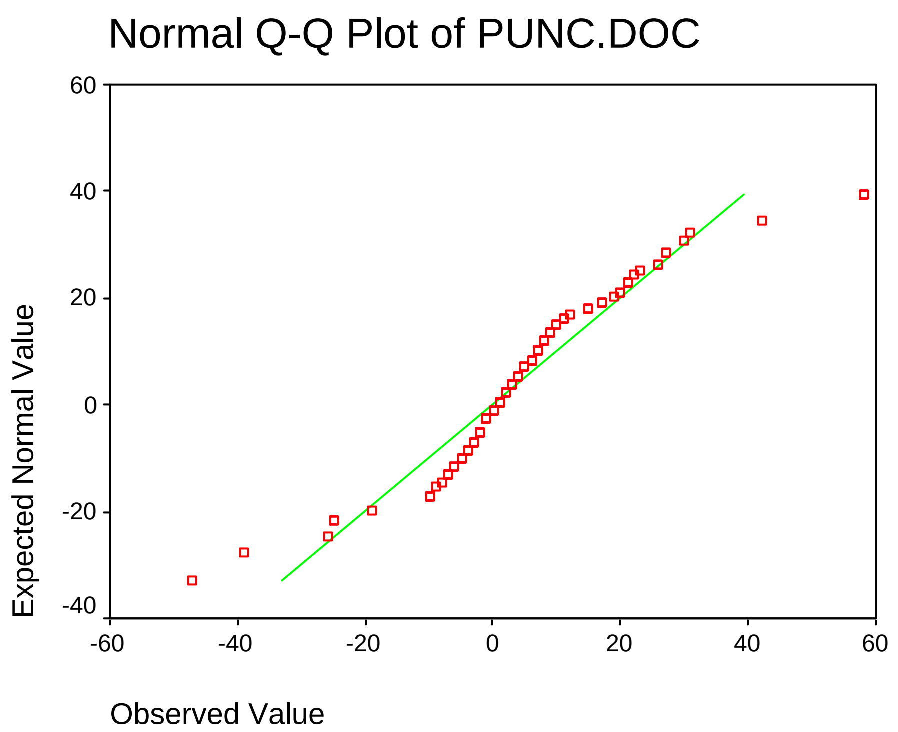


Figure C.2 Normal Q-Q Plot of doctors' punctuality (N=109).

Appendix D Consultation duration Q-Q plots

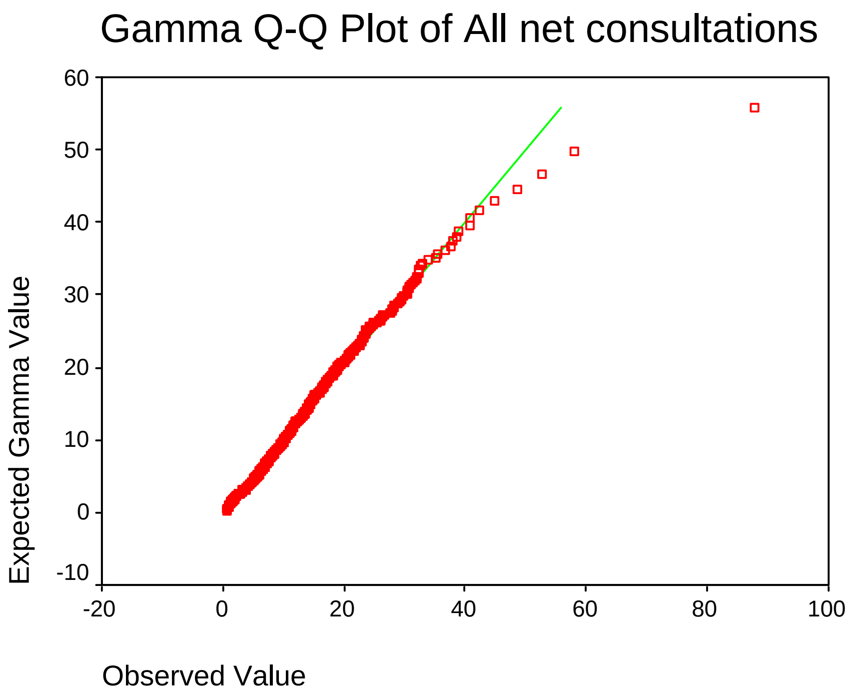


Figure D.1 Gamma Q-Q Plot of all net consultation durations (N = 985).

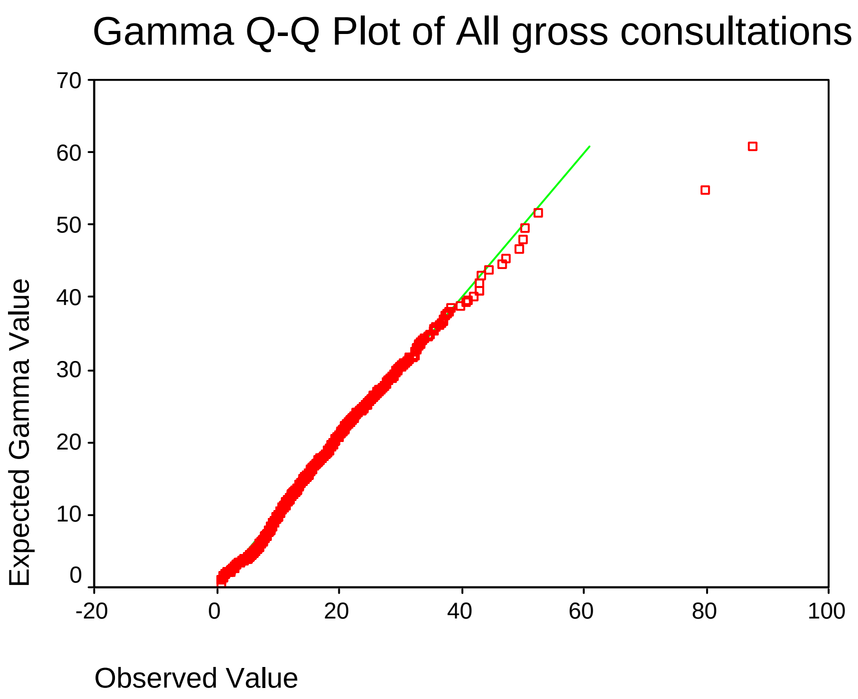


Figure D.2 Gamma Q-Q Plot of all gross consultation durations (N = 985).

Appendix E Measured consultation durations

Table E.1 Gross consultation times

Residents	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)	955	846	765	702	805	252	962
Variance (seconds^2)	271155	280466	300894	155000	251481	22585	1348214
Alfa	3.361	2.6	1.9	3.2	2.577	2.8	0.7
Beta	284.0	331.4	393.3	220.9	312.4	89.8	1401.4
Oncology specialists	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)		1446	1202	825	753	341	635
Variance (seconds^2)		836117	466348	180827	195802	38946	388728
Alfa		2.5	3.1	3.8	2.894	3.0	1.0
Beta		578.4	388.0	219.1	260.1	114.1	612.6
Non-oncology specialists	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)	696	1132	976	691	709	356	620
Variance (seconds^2)	284667	331364	216852	111418	130676	90511	120903
Alfa	1.7	3.9	4.4	4.3	3.8	1.4	3.2
Beta	409.0	292.7	222.2	161.2	184.4	254.4	195.0
Special clinic sessions	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)		1739	1706				
Variance (seconds^2)		561147	1025925				
Alfa		5.4	2.8				
Beta		322.6	601.5				

Table E.2 Net consultation times

Residents	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)	831	810	583	542	686	177	956
Variance (seconds^2)	195917	219914	228931	172625	235854	19346	1164236
Alfa	3.5	3.0	1.5	1.7	2.0	1.6	0.8
Beta	235.8	271.6	392.7	318.3	344.0	108.3	1218.1
Oncology specialists	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)		1242	829	664	561	238	544
Variance (seconds^2)		586090	288649	184594	155729	53642	344135
Alfa		2.6	2.4	2.4	2.0	1.1	0.9
Beta		471.8	348.4	277.9	277.5	225.6	633.1
Non-oncology specialists	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)	646	996	777	558	559	241	533
Variance (seconds^2)	179065	268934	163197	107105	103186	35469	102167
Alfa	2.3	3.7	3.7	2.9	3.0	1.6	2.8
Beta	277.3	269.9	210.1	191.9	184.7	147.3	191.7
Special clinic sessions	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Mean (seconds)		1564	1426				
Variance (seconds^2)		85931	692549				
Alfa		28.5	2.9				
Beta		54.9	485.8				

Intake2 = Intake: second echelon

Intake3 = Intake: third or fourth echelon

Appendix F Clustered consultation durations

Clustering of preparation times	Consultation type						
Doctor	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Residents	Cl1	Cl1	Cl2	Cl3	Cl1	Cl3	Cl1
Oncology specialsts		Cl2	Cl2	Cl1	Cl1	Cl3	Cl1
Non-oncology specialsts	Cl1	Cl1	Cl1	Cl3	Cl3	Cl2	Cl1
Special clinic sessions		Cl2	Cl1		Cl3		

(Cl = cluster)

Gamma distribution parameters of preparation times per cluster (seconds)

Cluster	Alfa	Beta	Mean	St.dev
Cl1	0.832	275.3	229.0	251.1
Cl2	0.919	364.8	335.2	349.7
Cl3	1.256	122.2	153.5	137.0

Table F.1 Clustered preparation durations.

Clustering of net consultation durations	Consultation type						
Doctor	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Residents	Cl2	Cl2	Cl1	Cl1	Cl1	Cl3	Cl2
Oncology specialsts		Cl4	Cl2	Cl1	Cl1	Cl3	Cl1
Non-oncology specialsts	Cl1	Cl2	Cl2	Cl1	Cl1	Cl3	Cl1
Special clinic sessions		Cl4	Cl4		Cl4		

(Cl = cluster)

Gamma distribution parameters of net consultation durations per cluster (seconds)

Cluster	Alfa	Beta	Mean	St.dev
Cl1	2.226	264.6	588.9	394.7
Cl2	2.320	370.5	859.8	564.4
Cl3	1.375	165.9	228.0	194.5
Cl4	3.744	362.4	1357.0	701.3

Table F.2 Clustered net consultation durations.

Clustering of doctor's administration times	Consultation type						
Doctor	Intake2	Intake3	Diagnostics	Follow-up	Periodical	Telephonic	Other
Residents	Cl1	Cl3	Cl2	Cl1	Cl1	Cl2	Cl1
Oncology specialsts		Cl3	Cl3	Cl2	Cl1	Cl1	Cl2
Non-oncology specialsts	Cl2	Cl1	Cl1	Cl1	Cl1	Cl1	Cl2
Special clinic sessions		Cl3	Cl3		Cl2		

(Cl = cluster)

Gamma distribution parameters of doctor's administration times per cluster (seconds)

Cluster	Alfa	Beta	Mean	St.dev
Cl1	0.234	215.8	50.5	104.4
Cl2	0.264	91.1	24.1	46.8
Cl3	0.331	424.1	140.2	243.9

Table F.3 Clustered administration durations.

Appendix G Simulation results

Table G.1 depicts the aggregated results of all tested configurations. This table shows configurations that violate the minimum patient service levels in *italic* font. The top-5 values for the doctor's performance (DP) and for the internal waiting time are shown in **bold** font.

The following abbreviations are used in the table:

App. rule	Appointment rule
BW rule	Bailey-Welch appointment rule
CMS	Correction of new patient's appointment times for attendance of Medical Students
Config.	Configuration
Decisions	Method of decision-making
Dedicated	Time slots dedicated for consultation types
DP	Doctor's performance
FCFA	First-Call-First-Appointment
FLV	Schedule First patients with Low Variance of consultation time
FSCT	Schedule First patients with a Short expected Consultation Time
Indiv.block	Individual block appointment rule
SchHorizon	Length of the scheduling horizon
Seq. rule	Sequencing rule
Slot usage	Usage of time slots

Table G.1 Simulation results

Config. Nr.	Decision	Sch horizon	Slot usage	Seq. ule	App. rule	CMS	DP	Avg access time (days)	Avg. Internal waiting time (min.)	Avg. Overtime	Avg. Idle time	Avg. Utilization	Return rate	Yearly throughput intake	Yearly throughput review	Length of waiting list
1 (current)	dynamic (A1)	70days (B1)	dedicated (C1)	FCFA (D1)	Indiv.bock (E1)	no (F2)	0.320	11.17	16.73	18.6%	18.9%	93.3%	5.61	3993	18704	916.24
2	dynamic (A1)	70days (B1)	dedicated (C1)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.317	11.17	17.69	18.3%	18.4%	93.3%	5.61	3993	18704	916.24
3	dynamic (A1)	70days (B1)	dedicated (C1)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.351	34.96	7.21	10.5%	20.9%	87.7%	5.48	3896	17543	2960.29
4	dynamic (A1)	70days (B1)	dedicated (C1)	FSCT (D3)	BW rule (E2)	yes (F1)	0.359	34.55	11.46	12.3%	15.8%	88.2%	5.48	3901	17547	2918.35
5	dynamic (A1)	70days (B1)	dedicated (C1)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.313	21.74	7.62	10.6%	18.9%	89.6%	5.40	3976	17936	1920.17
6	dynamic (A1)	70days (B1)	dedicated (C1)	FLV (D6)	BW rule (E2)	yes (F1)	0.331	22.13	11.93	13.1%	14.4%	90.0%	5.41	3965	17937	1962.26
7	dynamic (A1)	70days (B1)	pile-up (C2)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.319	5.67	11.71	13.6%	12.8%	90.9%	5.47	3979	18037	1558.70
8	dynamic (A1)	70days (B1)	pile-up (C2)	FCFA (D1)	BW rule (E2)	yes (F1)	0.282	5.05	20.42	15.9%	9.0%	93.9%	5.50	4042	18500	1474.39
9	dynamic (A1)	70days (B1)	pile-up (C2)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.339	4.55	9.43	11.0%	14.9%	88.5%	5.90	3642	18292	1924.55
10	dynamic (A1)	70days (B1)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)	0.293	4.01	15.29	12.5%	13.3%	91.6%	5.88	3726	18729	1221.96
11	dynamic (A1)	70days (B1)	pile-up (C2)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.330	4.70	**8.42**	11.7%	15.3%	89.3%	5.57	3870	18080	1659.41
12	dynamic (A1)	70days (B1)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)	**0.281**	4.04	14.55	13.0%	13.5%	92.5%	5.70	3902	18701	900.51
13	dynamic (A1)	30days (B2)	dedicated (C1)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.319	7.15	17.76	18.6%	18.7%	93.4%	5.60	4026	18671	638.11
14	dynamic (A1)	30days (B2)	dedicated (C1)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.345	20.35	7.07	10.7%	20.5%	88.1%	5.44	3912	17651	1702.49
15	dynamic (A1)	30days (B2)	dedicated (C1)	FSCT (D3)	BW rule (E2)	yes (F1)	0.356	20.91	11.38	12.5%	15.7%	88.5%	5.45	3949	17636	1753.52
16	dynamic (A1)	30days (B2)	dedicated (C1)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.318	11.21	**7.56**	10.9%	19.0%	89.5%	5.40	3958	17875	963.25
17	dynamic (A1)	30days (B2)	dedicated (C1)	FLV (D6)	BW rule (E2)	yes (F1)	0.331	11.25	12.17	13.1%	14.2%	90.0%	5.44	3973	17901	964.87
18	dynamic (A1)	30days (B2)	pile-up (C2)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.318	5.76	11.58	13.6%	12.8%	90.9%	5.43	3979	18031	825.34
19	dynamic (A1)	30days (B2)	pile-up (C2)	FCFA (D1)	BW rule (E2)	yes (F1)	0.284	4.95	20.49	16.1%	9.4%	93.9%	5.45	4044	18483	763.76
20	dynamic (A1)	30days (B2)	pile-up (C2)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.336	4.73	9.42	11.0%	14.6%	88.7%	5.88	3709	18176	1323.11
21	dynamic (A1)	30days (B2)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)	0.291	4.12	15.23	12.4%	13.2%	91.7%	5.87	3792	18724	957.56
22	dynamic (A1)	30days (B2)	pile-up (C2)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.332	4.65	**8.54**	11.5%	15.0%	89.1%	5.57	3863	18069	1111.99
23	dynamic (A1)	30days (B2)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)	0.302	4.05	14.21	12.9%	14.1%	91.3%	5.59	3915	18238	636.27
24	static (A2)	70days (B1)	dedicated (C1)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.375	13.86	16.22	18.6%	20.3%	90.6%	5.62	4040	18221	1087.01
25	static (A2)	70days (B1)	dedicated (C1)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.416	45.12	6.64	10.1%	23.1%	84.3%	5.43	3844	16723	3809.45
26	static (A2)	70days (B1)	dedicated (C1)	FSCT (D3)	BW rule (E2)	yes (F1)	0.426	46.05	10.96	12.1%	18.2%	84.7%	5.48	3831	16784	3894.40
27	static (A2)	70days (B1)	dedicated (C1)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.365	22.58	6.89	10.4%	20.6%	86.9%	5.30	3948	17317	1812.73
28	static (A2)	70days (B1)	dedicated (C1)	FLV (D6)	BW rule (E2)	yes (F1)	0.379	22.24	11.16	12.4%	15.7%	87.2%	5.33	3974	17326	1815.86
29	static (A2)	70days (B1)	pile-up (C2)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.324	5.42	10.68	12.8%	13.6%	90.2%	5.43	4032	17897	1529.04
30	static (A2)	70days (B1)	pile-up (C2)	FCFA (D1)	BW rule (E2)	yes (F1)	0.239	3.35	30.49	19.0%	9.4%	97.6%	5.56	4113	19071	176.61
31	static (A2)	70days (B1)	pile-up (C2)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.339	5.24	8.75	10.8%	14.9%	88.4%	5.79	3748	18071	1787.11
32	static (A2)	70days (B1)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)	**0.284**	4.26	14.64	12.6%	13.1%	92.1%	5.84	3821	18793	960.51
33	static (A2)	70days (B1)	pile-up (C2)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.331	4.73	**8.53**	11.6%	15.4%	89.2%	5.57	3880	18074	1739.58
34	static (A2)	70days (B1)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)	**0.283**	4.43	14.82	13.1%	13.2%	92.4%	5.63	3944	18553	923.05
35	static (A2)	30days (B2)	dedicated (C1)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.384	9.26	15.36	18.1%	20.7%	89.8%	5.57	3971	18153	784.37
36	static (A2)	30days (B2)	dedicated (C1)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.368	22.61	7.29	10.4%	21.0%	86.8%	5.42	3979	17320	1912.81
37	static (A2)	30days (B2)	dedicated (C1)	FSCT (D3)	BW rule (E2)	yes (F1)	0.376	22.72	11.93	12.5%	16.0%	87.5%	5.42	4001	17314	1909.45
38	static (A2)	30days (B2)	dedicated (C1)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.360	11.58	**7.11**	10.9%	20.5%	87.4%	5.38	3931	17244	929.65
39	static (A2)	30days (B2)	dedicated (C1)	FLV (D6)	BW rule (E2)	yes (F1)	0.380	11.14	11.04	12.4%	15.7%	87.2%	5.44	3798	17385	916.75
40	static (A2)	30days (B2)	pile-up (C2)	FCFA (D1)	Indiv.bock (E1)	yes (F1)	0.322	5.51	11.35	12.9%	13.3%	90.4%	5.53	3894	17970	810.31
41	static (A2)	30days (B2)	pile-up (C2)	FCFA (D1)	BW rule (E2)	yes (F1)	0.235	3.37	29.97	18.7%	9.7%	97.6%	5.48	4226	18920	180.23
42	static (A2)	30days (B2)	pile-up (C2)	FSCT (D3)	Indiv.bock (E1)	yes (F1)	0.341	4.76	9.19	10.9%	15.0%	88.4%	5.76	3793	18054	1371.39
43	static (A2)	30days (B2)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)	**0.280**	3.93	15.43	12.5%	12.3%	92.3%	5.63	3934	18554	639.11
44	static (A2)	30days (B2)	pile-up (C2)	FLV (D6)	Indiv.bock (E1)	yes (F1)	0.329	4.76	8.90	11.8%	14.9%	89.4%	5.69	3749	18052	1068.51
45	static (A2)	30days (B2)	pile-up (C2)	FLV (D6)	BW rule (E2)	yes (F1)	**0.277**	4.00	14.35	13.2%	12.8%	92.8%	5.75	3801	18647	629.88
46	static (A2)	10days (B3)	pile-up (C2)	FCFA (D1)	BW rule (E2)	yes (F1)	0.237	3.13	29.21	18.4%	9.6%	97.4%	5.56	4190	18932	174.26
47	static (A2)	10days (B3)	pile-up (C2)	FSCT (D3)	BW rule (E2)	yes (F1)	0.293	3.02	15.00	12.5%	13.3%	91.6%	5.51	4074	18150	414.17
48	static (A2)	30days (B2)	pile-up (C2)	FCFA (D1)	BW rule (E2)	no (F2)	0.232	3.37	28.71	18.9%	9.9%	97.8%	5.48	4226	18920	180.23
49	static (A2)	30days (B2)	pile-up (C2)	FSCT (D3)	BW rule (E2)	no (F2)	0.294	4.33	14.34	12.9%	14.1%	91.8%	5.72	3898	18572	761.93
50	dynamic (A1)	70days (B1)	pile-up (C2)	FLV (D6)	BW rule (E2)	no (F2)	0.293	4.34	15.87	14.1%	14.3%	92.4%	5.62	3958	18506	931.22