Final Project

Patient Flow Analysis in Pain Rehabilitation

A case study at rehabilitation centre Het Roessingh

Author:
N. Baër, BSc

Graduation Committee:	*On behalf of:*
Prof. Dr. R.J. Boucherie	University of Twente
Prof. Dr. J.L. Hurink	University of Twente
Drs. N. Kortbeek	University of Twente
Dr. N. Litvak	University of Twente
Dhr. O. Roukens	Het Roessingh

October 8, 2010

UNIVERSITY OF TWENTE.

Het Roessingh
centrum voor revalidatie

Contents

Summary

Het Roessingh is a rehabilitation centre located in Enschede, The Netherlands. Roughly 3000 patients are treated at Het Roessingh each year and the types of treatment that these patients require are very diverse. Each patient who is treated at Het Roessingh is placed in one of three groups: Adult Rehabilitation, Children's Rehabilitation or Pain Rehabilitation. In this thesis we will focus on the Pain Rehabilitation department.

The Pain Rehabilitation department treats patients who suffer from chronic pain or chronic fatigue syndrome for which no biomedical treatment is known. In this department the patients are taught to cope with their pain, thereby enabling them to fully participate in society. The treatment in Pain Rehabilitation is very effective and the department is considered to be among the best in The Netherlands. However, there are still minor problems in the department which have major consequences regarding patient flows. In the last five years Het Roessingh has undertaken several steps in improving their patient flows which resulted in the introduction of treatment plans. In these treatment plans the entire treatment is described for a specific group of patients. By using these treatment plans Het Roessingh can give patients more clarity about their treatment and the planning department should be able to create a timetable for each patient in which the entire treatment is scheduled. In total there are 18 different treatment plans and the patients that are treated according to a treatment plan are either outpatients or inpatients and are treated in groups or individually. We will use these 18 treatment plans to analyse the following four aspects in the patient flows at Het Roessingh.

1) Many patients in Pain Rehabilitation are individual outpatients. These patients receive an intensive treatment from an interdisciplinary team of practitioners. To make sure that each patient receives the correct treatment, meetings are scheduled to discuss a patient's progress. However, these meetings occur once every six weeks and the attendence at these meetings is very low because of the practitioners' busy schedules. Therefore, many meetings are cancelled and then the treatment of the patient in question is extended with an additional six weeks. By using a basic model and a defined strategy we show that the duration of a treatment increases significantly when the fraction of cancelled meetings grows. From these results we can conclude that the meetings must have top priority in these treatment plans to minimize the total lead time.

2) Almost all patients who are referred to Het Roessingh must undergo an intake in which several interviews take place to determine the best treatment plan for a patient. Every week, an average of 17.7 patients are referred to Het Roessingh and 16.6 of them apply for an intake. We use a simple queueing model to determine the number of intakes needed per week to meet specific waiting time requirements.

3) To understand the total patient flows, we create a queueing model which we analyse both by using a simulation model and an analytical model. These models are used to test the effects of several planning strategyies to improve the patient flows. We investigate stability of the system to figure whether the Pain Rehabilitation department has enough capacity (in resources and staff hours) to handle the offered load. Our main conclusion is that the current capacity will be insufficient if all patients are treated exactly as prescribed by the treatment plans and that the difference between the treatment plan's requirements and the existing capacity is in fact quite large. We will use the analytical model to find the bottelnecks in the system which cause the system to be unstable.

4) In addition to the simulation model we will give a staffing rule which results in a reasonably high workload for the practitioners and an efficient treatment for each patient. We show that by using this staffing rule the system can handle the same amount of patients as in the simulationmodel, 7.5 new patients per week, with 22% less FTE's and that the system can handle 30% more patients by rearranging the current number of FTE's. We will close this thesis by giving a staffing rule for the arrival rate of 17.7 patients per week, which is the actual arrival rate of patients in the Pain Rehabilitation department, and we will show that the system can handle 136% more patients with only 74% more FTE's.

Samenvatting

Het Roessingh is een revalidatiecentrum in Enschede. Jaarlijks worden er in dit centrum ongeveer 3000 patiënten behandeld welke allen een zeer variërende behandeling nodig hebben. Elke patiënt valt in een van de volgende drie categorieën: Volwassenenrevalidatie, Kinderrevalidatie of Pijnrevalidatie. In deze scriptie richten wij ons op de Pijnrevalidatie.

Mensen met chronische pijn en vermoeidheid waardoor geen biomedische behandeling meer mogelijk is kunnen terecht voor behandeling in de Pijnrevalidatie in Het Roessingh. Hier wordt de patiënten geleerd om te gaan met hun klachten om zo weer deel te nemen aan de maatschappij. De behandeling in de Pijnrevalidatie is efficiënt en behoort tot de beste van Nederland. Toch zijn er in het behandelproces problemen te vinden die grote gevolgen hebben voor de patiëntenstromen binnen de Pijnrevalidatie. Het Roessingh heeft daarom in de afgelopen vijf jaren meerdere stappen ondernomen om de logistiek binnen het centrum te verbeteren. Een van de resultaten was de introductie van zorgpaden, waarin de behandeling van een bepaalde soort patiënt beschreven wordt. Door deze zorgpaden te gebruiken zal Het Roessingh een beter beeld kunnen schetsen aan haar patiënten over de behandeling die zij krijgen. In totaal zijn er 18 verschillende zorgpaden binnen de Pijnrevalidatie en de patiënten die deze zorgpaden volgen worden in de polikliniek of in de kliniek behandeld en in groepen of individueel. In deze scriptie zullen we deze 18 zorgpaden gebruiken om de volgende vier punten in de patiëntenstromen binnen de Pijnrevalidatie te onderzoeken.

1) Een groot deel van de patiënten in de Pijnrevalidatie krijgt een individuele behandeling in de polikliniek. Deze patiënten krijgen een intensieve behandeling van een multidisciplinair team van behandelaars. Om te zorgen dat iedere patiënt de juiste behandeling krijgt worden er vergaderingen gehouden om de voortgang van de patiënt te bespreken. Deze vergaderingen worden echter maar eens in de zes weken gehouden en de opkomst tijdens deze vergaderingen is erg laag vanwege de drukke agenda's van de behandelaars. Hierdoor komen veel vergaderingen te vervallen met als gevolg dat de behandeling van een patiënt met zes weken wordt verlengd. Met behulp van een kansmodel en een voorgeschreven strategie die beschrijft welke uitgevallen vergadering voor verlenging zorgt, kunnen we aantonen dat de gemiddelde behandelduur significant zal groeien wanneer de kans op uitval van een vergadering bijna 1 is. Met behulp van deze resultaten kunnen we concluderen dat de vergaderingen erg belangrijk zijn in het behandelproces en dat deze de hoogste prioriteit moeten hebben bij de behandelaars.

2) Bijna elke patiënt die wordt doorverwezen naar de Pijnrevalidatie zal een intake moeten ondergaan waarin onderzocht wordt wat er met de patiënt aan

de hand is en wat de beste behandeling is om de patiënt te genezen. Elke week worden er gemiddeld 17.7 patiënten doorverwezen en 16.6 daarvan hebben een intake nodig. Met behulp van een simpel wachtrijmodel zullen we het aantal intakes bepalen welke nodig zijn om aan bepaalde eisen te voldoen voor de wachttijd.

3) Om de patiëntenstromen binnen de Pijnrevalidatie te begrijpen zullen we een wachtrijmodel bepalen welke we zullen analyseren aan de hand van een analytisch model en een simulatiemodel. Met behulp van deze twee modellen kunnen we veranderingen aanbrengen in het behandelproces om zo te achterhalen welke aanpassingen positieve effecten hebben op de patiëntenstromen. We zullen met deze modellen de stabiliteit van het wachtrijmodel testen en uitzoeken of de Pijnrevalidatie over voldoende middelen beschikt om de huidige patiëntenstromen aan te kunnen. De belangrijkste conclusie van deze scriptie is dat de huidige hoeveelheid beschikbare capaciteit binnen de Pijnrevalidatie te klein is om alle patiënten volgens de zorgpaden te behandelen en dat het verschil tussen de benodigde en beschikbare capaciteit schrikbarend groot is. Met behulp van het analytische model zullen we de bottlenecks in het system localiseren.

4) Als aanvulling op het simulatiemodel zullen wij een model geven wat een capaciteitsadvies geeft voor een systeem waar de werkdruk voor de behandelaars groot genoeg is en waar patiënten efficient worden behandeld. Met behulp van dit model zijn we in staat om dezelfde hoeveelheid patiënten te behandelen als in het simulatiemodel, namelijk 7.5 nieuwe patiënten per week, met 22% minder capaciteit. Daarnaast zijn we in staat om 30% meer patiënten te behandelen als we de huidige capaciteit opnieuw verdelen over de behandelaars. We zullen deze scriptie afsluiten met een capaciteitsadvies voor een aankomstintensiteit van 17.7 nieuwe patiënten per week, wat de daadwerkelijk aankomstintensiteit in de Pijnrevalidatie is, en we zullen zien dat we 136% meer patiënten kunnen behandelen met maar 74% meer capaciteit.

Preface

I am proud to present my Master Thesis in Applied Mathematics, at the chair of Stochastic Operations Research at the University of Twente, The Netherlands. The research for this thesis was performed at Rehabilitation Centre Het Roessingh in Enschede. The goal of this thesis is to analyse and improve patient flows at the Pain Rehabilitation department of Het Roessingh.

I want to thank Nelly Litvak and Nikky Kortbeek, my supervisors at the University of Twente for their mathematical guidance. All the weekly meetings resulted in a thesis which I am proud of.

I also want to thank Olaf Roukens, my supervisor at Het Roessingh for his guidance during my research. He was always there for me when I needed a helping hand in understanding the Pain Rehabilitation department. I also want to thank Patrick de Haan, Dirk van der Veen en Bert Jaspers for their contributions to my research.

Last but not least, I want to thank my family and friends for their love and support during my research and I want to thank Ellen in particular for her support during the period in which this thesis was written.

Niek Baër, BSc
October 8, 2010

Chapter 1

Introduction

Worldwide, millions of patients are admitted in a rehabilitation centre each year as a consequence of accidents, sickness or congenital disorders. In The Netherlands alone, more than 80.000 patients were treated in rehabilitation centres in 2009. In this year approximately 30.000 patients started their treatment in rehabilitation centres throughout The Netherlands and the expectations for 2010 are that this number will be increased by 13% to a total of 34.000 [18, 19].

In the current health care system in The Netherlands all patients are well insured and most of the time the costs for rehabilitation are compensated by insurance companies. The total costs of rehabilitation in The Netherlands are over 400 million euros per year and studies show [15] that each invested euro results in a revenue of five euros. However, in times of economical crisis it is important that this revenue is as high as possible.

An obvious solution is improving treatment, if we are able to cure a patient with less treatment then we can increase the revenue of each invested euro. Another possible solution is improving patient flows. By analysing the patient flows we can find ways to treat more patients in less time and therefore increase efficiency.

1.1 Rehabilitation care

Rehabilitation is the process in which a patient is assisted in improving or recovering lost functions after an event, illness or injury that causes functional limitations. Usually, this takes place in a rehabilitation centre in which a patient is treated by a multidisciplinary team of practitioners and a rehabilitation specialist for a period of time. During this rehabilitation, each member of such a team treats the patient in different areas of the rehabilitation process.

Rehabilitation centre Het Roessingh in Enschede, The Netherlands, is an example of a rehabilitation centre in which patients are treated by a multidisciplinary team. In the last five years Het Roessingh has developed treatment plans in order to improve their patient flows. In this thesis we will perform a patient flow analysis of the Pain Rehabilitation department in Het Roessingh based on their treatment plans.

1.2 Patient Flow Analysis

The treatment in the Pain Rehabilitation department in Het Roessingh is considered to be among the best in The Netherlands. However, there are still minor problems in the department which have major consequences in the patient flows. Meetings are cancelled, reports about patient progress aren't written on time and the time patients must wait for their treatment is long. All these problems are related to the high load of practitioners. Practitioners are always busy treating patients and don't have time available for other activities. All these problems suggest that we are dealing with an ad-hoc system where measures have to be taken to manage the patient flows.

In this thesis we present models with which we are able to improve the patient flows at the centre. We can show the effects of cancelled meeting on the average treatment time of one patient and we present a model which we use to analyse the intake procedure. Our main contribution is the design of a simulation model and an analytical model with which we can determine the critical load of the centre and the bottlenecks in the patient flows. Finally, we will present a staffing rule with which we are able to increase overall utilization of the practitioners in a bottleneck-free situation. These models are all based on the principle of treatment plans.

1.3 Overview

In this thesis we will present our analysis of the patient flows at the Pain Rehabilitation department in rehabilitation centre Het Roessingh. In Chapter 2 we will start by explaining how patients are treated at Het Roessingh and we formulate which problems arise during these treatments. At the end of Chapter 2 we will present a brief literature review concerning the problems under study. In Chapters 3 and 4 we will present two models which we use to analyse respectively the attendence of meetings and the intake structure at Het Roessingh. In Chapter 5 we will represent the Pain Rehabilitation department by a feedforward queueing system. We will analyse this queueing system with a simulation model and an analytical model for stability and we will determine the bottlenecks in the system. In Chapter 6 we present our fourth model in which we determine a staffing rule for the Pain Rehabilitation department, which we test with the simulation model presented in Chapter 5. In the last chapter of this thesis, Chapter 7, we will summarise the findings from each of the four preceding chapters and we will present our recommendations for further research.

Chapter 2

Het Roessingh

Het Roessingh is a rehabilitation centre located in Enschede and it was established in 1948, just after the Second World War. In the beginning the centre treated patients with severe physical injuries such as amputations. It was located in an old villa and there were only three employees. In the next years the centre grew bigger and a special department was founded to teach patients to walk with an artificial leg. It soon became clear that these patients also needed help on a psychological level, and a multidisciplinary rehabilitation centre was born.

Since the opening of the rehabilitation centre a lot has changed. What started out as a small establishment with 3 employees grew to a large centre with more than 750 employees. Nowadays, Het Roessingh is more than just a rehabilitation centre and with its Research & Development department it is constantly researching new ways to improve their methods of treatment.

2.1 Organisation

Roughly 3000 patients are treated in Het Roessingh every year. Each of these patients needs a specific treatment and is therefore categorized into one of the following groups:

- **Adult Rehabilitation -** Rehabilitating adults with a handicap are taught to take care of themselves and how to cope with everyday problems. Patients in the Adult Rehabilitation are treated as inpatients as well as outpatients. Rehabilitation specialists try their best to treat every patient as an outpatient but in many cases, for patients with a spinal cord lesion for example, this is not possible and the patient is admitted to the clinic.
- **Children's Rehabilitation -** Children and young adults (age 0 to 21 years) can also be admitted as inpatients or outpatients. In some cases extra care is needed and the choice is made to admit a child to the clinic. During their stay, the children can go to school at the Educational Centre of Het Roessingh.
- **Pain Rehabilitation -** In the Pain Rehabilitation department patients are treated who suffer from chronic pain or who are diagnosed with the

chronic fatigue syndrome for which no biomedical treatment is available. At Het Roessingh these patients are taught to cope with their pain and to minimize its negative effects on their life. Patients from this group can be admitted in three different ways: as a clinical patient, as a semi-clinical patient or as an outpatient. We will explain these terms in a later part of this section.

In these three groups there are multiple specialized teams which asure that each patient receives the correct treatment. In the adult rehabilitation for instance, we find teams specialised in spinal cord lesion and in the pain rehabilitation there are teams for chronic fatigue syndrome.

From an organizational point of view it is best to group certain teams together and form a so-called RVE, which stands for "Result Responsible Units" ("Resultaat Verantwoordelijke Eenheden" in Dutch). An RVE consists of practitioners, rehabilitation specialists and nursing staff. Each RVE is co-ordinated by a process manager and a medical co-ordinator. In this thesis we will restrict ourselves to RVE 6, the Pain Rehabilitation department.

Patients in the Pain Rehabilitation department are referred to the centre by their general practitioner or by a doctor in a hospital. Almost all patients start their treatment with an intake. During an intake session, which lasts for a few hours, the patients will be interviewed by different practitioners about their complaints. At the end of these interviews a rehabilitation specialist will decide which treatment is best for each patient. In this decision he will be advised by a multidisciplinary team of practitioners.

After the intake the patient will start his treatment. The patient will be treated in a group or individually and based on the treatment which he receives, he will be treated in one of the following ways:

- **Clinical Patient** - The patient spends 4 nights per week at the clinic and is treated throughout the week from Monday to Friday. After this week the patient returns home for a week of rest. The patient returns to the clinic in the third week for another week of treatment and is at home in the fourth. This process proceeds until the treatment is finished. A few weeks after the last week of treatment the patient returns to the centre for one week. In this final week checkups are scheduled to monitor the effects of the treatment. In this week the patient gets the same treatment as during a regular week of treatment.
- **Semi-Clinical Patient** - The patient spends 2 nights per week at the clinic and is in treatment for 2.5 days, from Monday morning until Wednesday noon or from Wednesday noon until Friday afternoon. Unlike the clinical patients, a semi-clinical patient is at the clinic every week. The patient returns to the centre a few weeks after the last treatment for one day of checkups.
- **Outpatient** - An outpatient does not stay at the clinic but has several appointments per week for which the patient must come to the rehabilitation centre.

During an intake, the decision can be made to admit a patient for observation. When a patient is admitted for observation he will receive a specific treatment with which the practitioners and the rehabilitation specialist can decide which treatment is best for the patient. The patient will start this treatment when his observation is completed.

The general patient flow described above is depicted by the bold line in Figure 2.1: patients arrive at the centre and they will undergo an intake. After this intake they will start a treatment in either a group or individually and when this treatment is finished they will leave the centre. A patient can also start an observation period after the intake. When this observation is completed he will be treated in a group or individually or he will leave the centre.

There are some deviations from this general patient flow which we depict by the dotted, dashed and dash-dotted line in Figure 2.1. The dashed line in this picture denotes patients which arrive at the centre and do not need an intake before they start their treatment.

By the dotted line we denote the patients who leave the centre after they had an intake. During this intake the practitioners and the rehabilitation specialist find out that the complaints of the patient cannot be treated at Het Roessingh and the patient is referred to another centre.

Although the intake is very effective, it is possible that patients are wrongly diagnosed and that they start the wrong treatment. This may become clear during the treatment and the decision is made to stop the current treatment and start a different one. This process is depicted by the dash-dotted line.

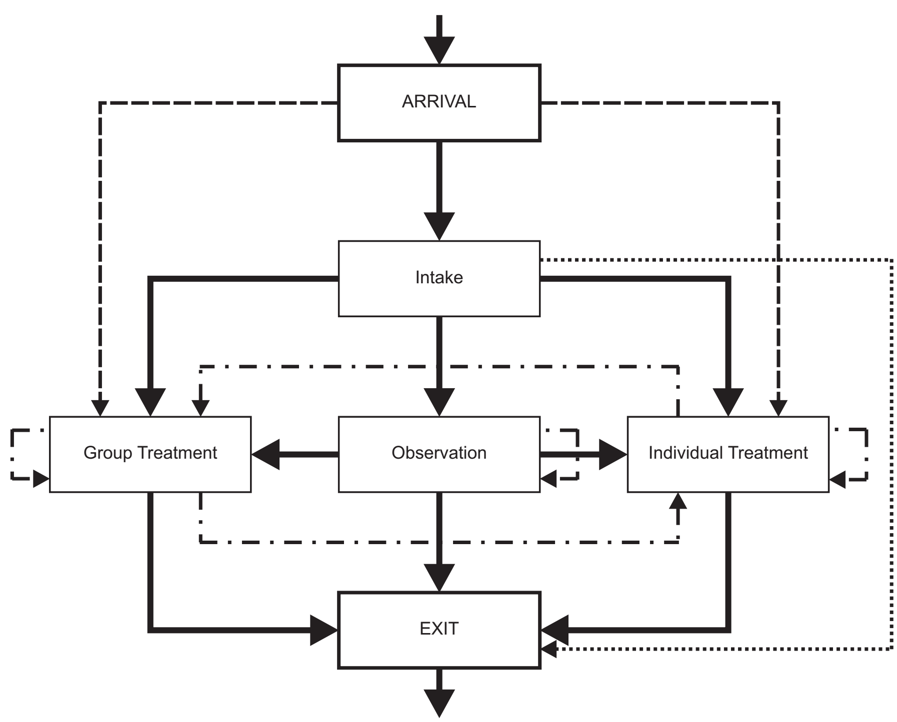

Figure 2.1: Overview

As we mentioned earlier, a patient can either be treated in a group or individually. In both cases we find that the treatment is effective and that patients are satisfied with the treatment. There is, however, a big difference between the two cases. In a group, multiple patients are treated at the same time and the treatment given is the same for each patient. Because there are multiple patients in a group, we need a clear agenda so the patients know which treatment is given at what time. This agenda is usually made for the entire treatment and this gives the patients a clear perspective about their treatment. This is different for the individually treated patients. Here we find that patients receive a treatment which is unique for every patient. Scheduling an appointment with these patients is easier because it only involves two persons: the patient and the practitioner. However, this also results in short-term planning in which the patient does not know which treatment is given over a longer period.

2.2 Treatment Plans

In the near future Het Roessingh will start to work with the EPD and DBC systems. These system are developed by the Dutch government in the last five years and must result in improving health care in The Netherlands. The EPD system is a large digital database containing medical records of all patients. With this database it will be easier for physicians, practitioners and other medical staff to access patient's records. Up to now this was only possible with the paper version of a patient's medical records. With the DBC system, which stands for "Diagnosis-Treatment Combination", there will be a lot more clarity about the treatment which a patient receives. The system was designed to offer insurance companies a better view of the costs of the treatment given to a patient. In this system each diagnosed patient is linked to a certain treatment with known costs, it must therefore be known which treatment is given to a patient with a certain diagnosis.

While the development of the EPD and the DBC systems was still in progress Het Roessingh has introduced two project groups called "Flow" and "Follow". The goal of these projects was to optimise the scheduling processes used in the planning department by applying methods used in optimising operations in industry. This resulted in the introduction of "Treatment Plans", a pre-defined plan of treatment for a group of patients with the same diagnosis. We must not confuse this with the internationally known clinical pathways or critical pathways [24]. An example of such a treatment plan is shown in Appendix C. Here we can see that a treatment plan contains information about the number of appointments each week, the duration of these appointments and which practitioner is needed. It does not, as opposed to clinical pathways, contain any information about the exact content of each appointment. Another difference with the clinical pathways is the total duration of a treatment plan. Some treatment plans last for a few weeks but others last for more than 6 months, while a clinical pathway usually last a lot shorter. There is also a difference between the type of appointments made in treatment plans and the type of appointments made in clinical pathways. In a treatment plan almost all appointments are recurrent for a few weeks, for instance fitness or swimming, while most appointments in clinical pathways only happen once. However, clinical pathways can be very useful in creating the treatment plans.

In 2000, several Belgian institutes started working on a process in which clinical pathways were created and evaluated. This 30 step program is based on the Plan-Do-Check-Act Cycle (also known as the Deming Cycle) in Figure 2.2. Using this cycle we can divide the 30 steps into the four phases of the Deming Cycle [28]:

- Plan: Define a patient population, create an interdisciplinary team en create a first version of the clinical pathway.
- Do: Collect data concerning the current procedure and Best Practice.
- Check: Interpret data collected in the Do-phase and adjust clinical pathway with results.
- Act: Implementation and continuous evaluation of the clinical pathway.

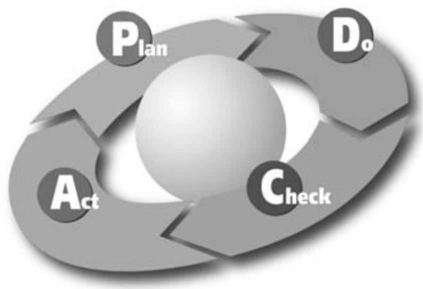

Figure 2.2: Deming Cycle

Based on this 30 step program and on the PDCA-cycle, a method was developed at Het Roessingh to evaluate treatment plans for certain patient groups [11]. In this method the evaluation is performed by comparing the predetermined treatment plan with the actual treatment. With the help of this model Het Roessingh can evaluate their treatment plans and alter them where necessary.

Another difficult task is the scheduling of the treatment plans. At first, this does not seem very obvious because a lot of patients are treated in a group and these groups are constantly present at Het Roessingh. Therefore we can reserve a part of the agenda of the partitioners every week at the same time. In this way the planning should not be that hard. However, it becomes harder when we want to schedule individual outpatient appointments. In this thesis we study how the scheduling of treatment plans affects the patient flows and what implication it has for staffing strategies.

2.3 Research Objectives and Methodology

There are three major problems at Het Roessingh that are common in Operations Research: the waiting list is too long, the treatment time per patient is often longer than necessary and the practitioners have a high working pressure. In this thesis we will analyse the patient flows at the Pain Rehabilitation department in Het Roessingh by using their treatment plans. In our patient flow analysis we will look at four problems which we will discuss in the next subsections.

2.3.1 Interdisciplinary meeting for RDB-patients

There are three treatment plans for individual outpatients, also known as RDB-patients. These treatment plans last for 15, 21 or 27 weeks and during these treatment plans meetings are scheduled to discuss a patient's progress. These meetings occur once every six weeks and the attendence at these meetings is very low because of practitioners' busy schedules. Therefore, many meetings are cancelled and the treatment of the patient in question is extended with an additional six weeks.

Our first objective is to analyse how patient flows are affected by cancellation of these meetings.

In order to analyse this problem we will present a model which determines the mean treatment time of an RDB-patient for a given probability that an arbitrary meeting will be cancelled.

2.3.2 Intakes at Pain Rehabilitation

Every week new patients are referred to the Pain Rehabilitation department. Most of these patients must undergo a general intake before they can start their treatment. Each week 15 intakes are held and these intakes are held in three sessions of five intakes each. Until one year ago, the Pain Rehabilitation held 18 intakes per week: three sessions of six intakes. On average 17.7 patients are referred to the Pain Rehabilitation department each week and 16.6 of them must undergo an intake.

Our second objective is to compare the two scenarios described above and determine which scenario performs better.

We will determine the mean waiting time for an intake at various arrival rates by using Lindley's Equation. We will use this model to show that the current structure is unstable and that a minimum of 17 intakes is needed each week.

2.3.3 Patient Flow Analysis

The three problems which we mentioned at the beginning of this section (long waiting times, long treatment times and high working pressure) are all characteristics of an overloaded system. Therefore our third objective is to analyse the patient flows in the Pain Rehabilitation department and to determine the bottlenecks and stability conditions.

In Chapter 5 we will represent the Pain Rehabilitation department by a feedfoward queueing system. We will analyse this queueing system with a simulation model and an analytical model for stability and we will determine the bottlenecks in the queueing system. We will eventually stabilise the queueing system by decreasing the arrival rate and use the simulation model to test the effects of four simple solutions to improve the patient flows.

2.3.4 Staffing Improvements

Our final objective is to present a staffing rule for the Pain Rehabilitation department. Using this staffing rule must result in a Pain Rehabilitation department where patients are treated efficiently and where the working load of practitioners is still high enough so that no capacity is wasted.

We will use the square root staffing rule to determine a staffing rule which satisfies these demands. The goal of this staffing is to stabilise the queueing system presented in Chapter 5 by increasing the amount of available capacity and maintain high utilization of the practitioners.

2.4 Literature

In this section we will present a brief literature review of the problems under study.

2.4.1 Scheduling treatment plans

A lot of literature is known about scheduling appointments in health care but scheduling entire treatment plans is yet a rather unfamiliar problem. In the literature we can find ways to schedule single appointments (both deterministic and stochastic) and maximize revenues [29]. In [5, 4] we learn how to schedule multiple appointments on the same day, to minimize waiting times and to maximize utilization at the same time. Unfortunately, this *offline* method needs all information about patients and equipment before it can create a planning. New studies have resulted in an *online* scheduling model in which a planning is provided at the moment a patient sends in an application [2].

2.4.2 Staffing Rules

Staffing problems are often considered in planning of call-centers where it is important to determine the amount of agents needed to answer the calls. In this process, the centre is looking for a staffing rule which satisfies its demands. An important result in staffing strategies is the square root staffing rule to determine the number of agents s to satisfy demands:

$$s = \frac{\lambda}{\mu} + \beta\sqrt{\frac{\lambda}{\mu}}$$

where λ/μ is the offered load (with λ the arrival rate and μ the service rate) and β respresents the service grade. This staffing rule is typical for the Halfin-Whitt regime or Quality and Efficiency Driven (QED) regime introduced by [9]. In this regime a balance is made between the Quality Driven regime, which focusses on the quality of service, and the Efficiency Driven regime which emphasizes on the utilization of servers. All three regimes are so called limitation regimes for heavy traffic queueing models in which the utilization is close to 1. An overview of these different regimes is found in [1] in which solutions for the regimes are shown and how these regimes are related to each other. In this paper the square root staffing rule is revisited and extended with costs for staffing and delay. In [14] we find another refinement of the square root staffing rule by expanding Erlang C, which is defined for an integer number of servers s, to $s \in \mathbb{R}$.

Impatient customers

In a call center model, customers will arrive with a certain task which must be completed by the server. If the waiting times are too long the customer will leave the system in order to return another time. We find solutions for the $M_t/G/s_t + G$-model with general customer abandonment ($+G$) in the form of a flexible simulation based *iterative-staffing algorithm* (ISA) in [7]. In this paper it is shown that for special Markovian $M_t/M/s_t + M$ cases this ISA converges. We also find extensive research on the $G/G/s + G$-model in [6].

Single skill and Multiple skill models

A customer arriving to the queue arrives with a certain task to be completed by a particular server. In the single skill model we find that each server is capable of completing one task whereas in a multiple skill model we see that a server is able to complete multiple tasks. In a single skill call center model, for example, the agents can speak only one language and in the multiple skill model the agents can speak several languages so they can serve more customers.

In [30] we find the approximation of a multi-server single skill $G/G/s + G$-model by analysis of the corresponding fluid model. The main objective in this paper is to determine an optimal number of agents such that profits are maximized. These profits depend on the revenue for throughput and costs concerning the number of servers in use, the abandonments and the waiting times.

Decoupling of a multiple skill model into a single skill model is discussed in [8] where we learn that staffing disregards the multiclass nature of the system and is analogous to the staffing of a single skill system with the same aggregate demand and a single global quality-of-service constraint.

We find in [10] the solution to the dynamic control problem in the multi-pool multi-class setting using traditional heavy-traffic limit theory. In this model the arrival rates and service rates grow linearly so that the system utilization approaches 1.

Canonical designs

We can define several canonical designs in the mutliple skill model which are named after their schematic appearance of the general model. The first design is a V-design in which 2 job classes arrive to a single queue. The server can serve jobs from both classes. This system is discussed in [21] in which the authors determine the required number of servers and an assignment policy so that desired expected waiting times can be satisfied.

The N-design describes a model with 2 job classes and 2 queues. In one queue we find a server which can handle both jobs, in the other queue we find a server which can handle only 1 job. This is discussed in [26] in which fixed priority policies are used. We find an approximation procedure to estimate the performance of this queueing system in [25].

The last design is an M-design in which there are 2 job classes and 3 queues. In two of these queues we find a specialized server which can handle one job type. Here these two servers can handle different jobs. At the third queue we find a generalist: a server which can handle both job classes. This model is analysed in [20] under the assumption that the specialists have a higher service rate than the generalist.

There are more designs possible like an I-design, which relates to the single skill model, and Λ-design, which we can separate into two I-designs, and a

W-design in which each server can handle 2 job classes. This last design can be separated into two V-designs. The combination of W-designs and M-designs is also known as a multi-skill multi-server system and it is the general setting of multi-skill call centers.

Discrete batch service

In [13] we find an alteration to the standard model for which the staffing rule is created. Instead of a continuous service rate we find that this paper emphazises on deterministic services, a $M/D/s$-model. By using Lindley's equation the authors find that this model relates to a Gaussian random walk. In this thesis we will use their paper to derive a model which we can use to determine a staffing rule for the network of treatment plans at the Pain Rehabilitation department.

Chapter 3

Cancelled meetings for RDB-patients

The individual outpatients, also called RDB-patients, where RDB stands for: "Roessingh Outpatients" ("Roessingh Dag Behandeling" in Dutch), can be treated in three different treatment plans. The main difference between these three treatment plans is the length of the treatment. The first is the shortest and lasts for 15 weeks, the second for 21 weeks en the last for 27 weeks. In these treatment plans a meeting is scheduled every six weeks starting in week three. During such a meeting information has to be provided by the practitioners to discuss a patient's progress and to make decisions for further treatment. However, internal records show that the attendence at such meetings is less than 40% and in many cases there is not enough information available to make a correct decision about a patient's treatment. When this occurs, the meeting is cancelled and the rehabilitation specialist must decide if the treatment will be extended with an additional six weeks or not. Such an extension will always take place at the end of the original treatment plan and will always be six weeks long. The rehabilitation specialists do not work according to a predefined policy, when a meeting is cancelled they will decide what is best for the patient. However, there is one situation in which all rehabilitation specialists make the same decision, when the last meeting of the treatment plan is cancelled, the treatment will always be extended.

Because there is no exact policy available we will present our results for the following three different policies:

- **P1 - Aggressive policy:** We will extend the treatment with six weeks for every cancelled meeting.
- **P2 - Passive policy:** We will only extend the treatment with six weeks if the last meeting of the original treatment plan is cancelled.
- **P3 - Hybrid policy:** We will extend the treatment with $n-1$ periods of six weeks when there are n consecutive cancelled meetings and we will extend the treatment with an extra period of six weeks if the last meeting of the original treatment plan is cancelled. Three consecutive cancelled meetings will result in extending the treatment which two periods of six weeks.

These policies are used during the original treatment plans and the number of weeks with which we will extend the original treatment plan will be determined at the end of the treatment plan. In week six of every extended period a meeting is scheduled. This meeting is considered to be very important and as a result we will always add another extended period of six weeks to the treatment whenever this meeting is cancelled.

The effects of the three policies during the original treatment plan are shown in Table 3.1. In this table we look at the treatment plan of 15 weeks where meetings are scheduled in week 3, week 9 and week 15. We denote these meetings with respectively M1, M2 and M3. The values in columns M1, M2 and M3 denote if the meeting is cancelled or not. A 1 indicates a successful meeting while a 0 stands for a cancelled meeting. The number of extended periods (each of six weeks) with which we extend the treatment according to each policy is shown in the columns P1, P2 and P3. In the last column we find the probability that this event occurs. In this column we use p to denote the probability that at an arbitrary meeting is cancelled.

M1	M2	M3	P1	P2	P3	Pr
1	1	1	0	0	0	$(1-p)^3$
1	1	0	1	1	1	$p(1-p)^2$
1	0	1	1	0	0	$p(1-p)^2$
0	1	1	1	0	0	$p(1-p)^2$
1	0	0	2	1	2	$p^2(1-p)$
0	1	0	2	1	1	$p^2(1-p)$
0	0	1	2	0	1	$p^2(1-p)$
0	0	0	3	1	3	p^3

Table 3.1: Extension treatment plans RDB-patients.

3.1 Model formulation

In this section we will create a model to show the effect of the attendence level at meetings on the lead time, or treatment time, of a patient. We will do this by determining the mean lead time for an RDB-patient with a treatment plan of 15 weeks and for policy P3. The other policies and treatment plans can be evaluated in the same manner.

We denote the total lead time of a patient in weeks by X. This total lead time is the sum of the original treatment time in weeks and the total number of weeks with which we extend the original treatment plan. We have seen in Table 3.1 that in some cases the original treatment plan is extended with three periods of six weeks. However, during these three periods it is also possible that the treatment is extended due to low attendence at a meeting in an extended period. We therefore define Y_i to be the total number of weeks with which we extend the original treatment plan, given the fact that we initially extended the original treatment plan with i periods. The policy during an extended period prescribes that we continue extending the treatment until we find a successful meeting. Therefore, Y_i follows a Negative Binomial distribution, we will extend the treatment until we find i successful meetings. This results in:

$$P(Y_i = 6j) = \binom{j-1}{i-1}(1-p)^i p^{j-i}, \quad i > 0, j \geq i$$

with $P(Y_0 = 0) = 1$. In this formula we use j to denote the total amount of **periods** with which we extend the original treatment plan. Because a period is six weeks long we find that the total amount of **weeks** with which the original treatment plan is extended is equal to $6j$.

Finally, we define q_i to be the probability of extending the original treatment plan with i periods. These values are determined by the chosen policy and can be obtained from Table 3.1. The values of q_i belonging to policy P3 are stated in Table 3.2.

i	q_i
0	$(1-p)^2(1+p)$
1	$p(1-p)(1+p)$
2	$p^2(1-p)$
3	p^3

Table 3.2: Probability q_i.

With the definition of Y_i and q_i we can now state that

$$P(X = 15 + 6j) = \sum_{i=0}^{3} q_i P(Y_i = 6j), \quad j \geq 0,$$

and we can determine $E(X)$ by

$$E(X) = 15 + \sum_{i=0}^{3} q_i E(Y_i).$$

Since Y_i is Negative Binomially distributed we find:

$$\begin{aligned} E(Y_i) &= iE(\text{\# periods until successful meeting}) \\ &= i\frac{6}{1-p}. \end{aligned}$$

For $i = 0$ we find $E(Y_0) = 0$. With $E(Y_i)$ known we can now determine $E(X)$ by:

$$\begin{aligned} E(X) &= 15 + \sum_{i=0}^{3} q_i E(Y_i) \\ &= 15 + \frac{6p(1+2p)}{1-p}. \end{aligned}$$

3.2 Results

By using the calculations from Section 3.1 we can determine the mean lead time for all three treatment plans and all three policies. The results are stated in

	P1	P2	P3
15 weeks	$15 + \frac{18p}{1-p}$	$15 + \frac{6p}{1-p}$	$15 + \frac{6p(1+2p)}{1-p}$
21 weeks	$21 + \frac{24p}{1-p}$	$21 + \frac{6p}{1-p}$	$21 + \frac{6p(1+3p)}{1-p}$
27 weeks	$27 + \frac{30p}{1-p}$	$27 + \frac{6p}{1-p}$	$27 + \frac{6p(1+4p)}{1-p}$

Table 3.3: Mean lead times for different policies.

Table 3.3 where the rows represent the original length of the treatment plan in weeks and the columns denote the different policies.

The results in the last column (P3) and the first row (15 weeks) are plotted in respectively Figure 3.1 and Figure 3.2. It is clearly seen that the mean lead time grows very rapidly with the probability p, no matter which treatment plan or which policy is used. This can be explained by the Negative Binomial distribution of Y_i where we will continue extending the treatment until we find i successful meetings. As a result we find that the mean lead time $E(X) \rightarrow \infty$ when $p \rightarrow 1$.

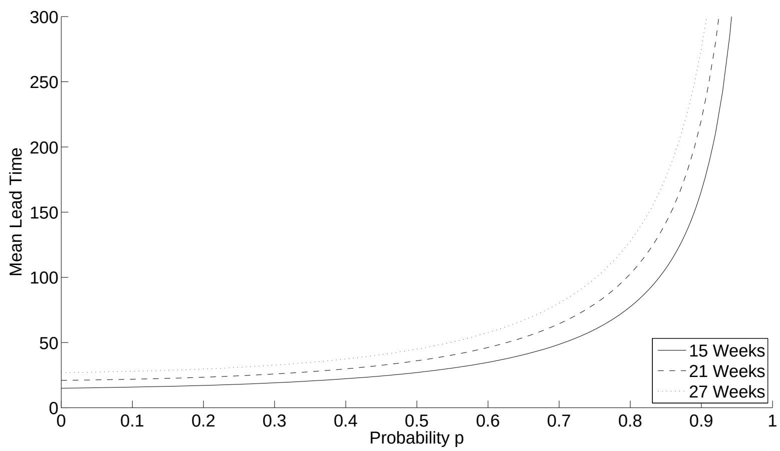

Figure 3.1: Mean lead time per treatment plan for policy P3.

This result also has a large impact on the number of patients we are able to treat in the same week. For instance, when we take the treatment plan of 15 weeks and we use policy P3, we can see that by increasing p from 0.4 to 0.8 the mean lead time increases by a factor 3.5. Now by using Little's Law [17], $EL = \lambda EW$, we can also see that the mean number of patients that we are able to treat in one week time also increases by a factor 3.5. So whenever we have enough capacity to treat all RDB-patients at $p = 0.8$, we could treat 3.5 times as many patients when we would decrease p to a value of 0.4.

3.3 Conclusions

By looking at the plots in Figure 3.1 and Figure 3.2 we can state that it is essential that the attendence at the meetings is as high as possible. We showed

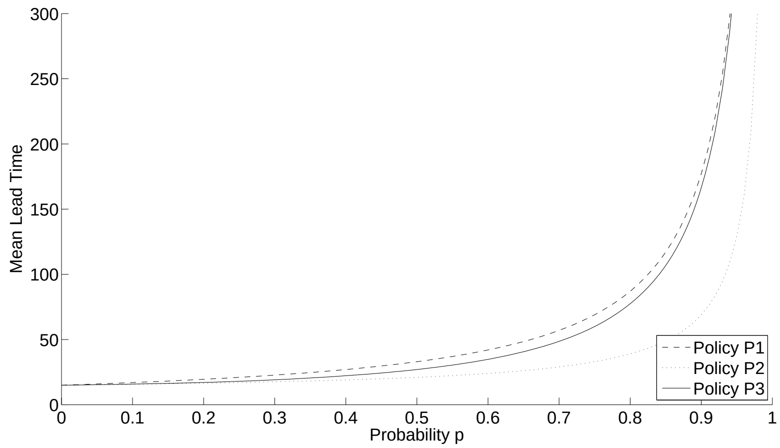

Figure 3.2: Mean lead time per policy for treatment plan of 15 weeks.

that by an increment of factor two in the probability p, we were able to treat more than three times as many patients.

The results from Section 3.2 can be extended to an arbitrary policy as long as this new policy complies with the two basic rules. These rules describe that the treatment plan must be extended with one period of six weeks when the last meeting of the original treatment plan is cancelled and that the treatment must be extended with six weeks whenever a meeting is cancelled during an extended period. The passive policy works solely with these rules which means that an arbitrary policy will perform worse or equal to the passive policy. In the aggressive policy, each cancelled meeting ends up in an extension of six weeks. An arbitrary policy will therefore always perform better or equal to the aggressive policy.

Chapter 4

Intakes at the Pain Rehabilitation department

Almost all patients who are referred to Het Roessingh must undergo an intake before they can be treated. These intakes are organised in three intake sessions per week in which a group of five patients is served per session. We will only allow an intake session to be scheduled where there are enough patients available to fill the group. In this way we will keep the efficiency of an intake session as high as possible. Het Roessingh keeps records of every patient referred to them from January 2005 up until now. From these records we can obtain the number of referrals to Het Roessingh each month and determine the mean number of arrivals each week. These records are shown in Table 4.1. In this table we can also see that there is no clear seasonality present in the data and we therefore assume that the arrival rate is time invariant.

Referrals	2005	2006	2007	2008	2009
January	13	53	76	59	79
February	76	72	67	73	77
March	73	64	73	67	98
April	61	67	84	80	75
May	65	70	79	78	72
June	83	111	70	62	81
July	72	60	55	55	66
August	50	51	48	65	60
September	43	80	64	67	82
October	66	79	80	69	72
November	65	87	78	80	84
December	51	68	50	62	77
Total Per Year	718	862	824	817	923
Average Per Week	13.8	16.6	15.8	15.6	17.8

Table 4.1: Number of referrals in the past.

However, not all patients have to undergo the same general intake. There are two treatment plans with their own intake procedure and one treatment plan for which no intake is needed. 93.9% of the patients must undergo an intake, which results in an average of 16.6 patients requiring an intake each week.

4.1 Model formulation

We will create a queueing model which can be used to determine the mean queue length and the mean waiting time for an intake. This model is designed to cope with the general setting where there are a intake sessions and where exactly b patients undergo their intake per session. Here we will only schedule an intake session when there are at least b patients available. If we do not have enough patients available to fill the entire session, we will not schedule the session at all.

In the queueing model we will denote the number of patients in the queue at the beginning of week t by L_t, the number of arrivals in week t by A_t and the number of intakes held in week t by s_t. We now use Lindley's equation [16]:

$$L_{t+1} = \max\{L_t + A_t - s_t, 0\},$$

to analyse the queueing model. In this model we assume that the arrival rate is time invariant and Poisson distributed with mean λ:

$$P(A_t = j) = \frac{e^{-\lambda}\lambda^j}{j!}, \quad \forall t.$$

Each intake session is a batch process of b patients and we therefore know that s_t depends on L_t:

$$s_t = \begin{cases} jb, & jb \leq L_t < (j+1)b, \quad 0 \leq j < a \\ ab, & ab \leq L_t. \end{cases}$$

By the definition of s_t we note that $L_{t+1} \geq 0$. With A_t and s_t known we can define a Markov Chain in which the states represent the number of patients in the queue. We will define the transition matrix Q as:

$$Q_{(i,j)} = \begin{cases} P(A_t = j - (i - kb)), & kb \leq i < (k+1)b, \quad j \geq i - kb, \quad 0 \leq k < a \\ P(A_t = j - (i - ab)), & ab \leq i, \quad j \geq i - ab \\ 0, & \text{otherwise.} \end{cases}$$

In order to obtain the stationary distribution of the number of patients in the queue we must solve

$$\pi = \pi Q,$$

together with $\sum_i \pi_i = 1$, to a unique solution where π_i is the probability that i patients are waiting in the queue. With this stationary distribution we can determine the mean queue length by

$$E(L) = \sum_i i\pi_i$$

and by using Little's Law we find

$$E(W) = \frac{E(L)}{\lambda}.$$

In theory, the queue can hold infinitely many patients but because of computational limitations we must truncate the queue so that it can hold a finite number of patients. By truncating the queue we must adjust the transition matrix Q by:

$$Q_{(i,N)} = 1 - \sum_{k=1}^{N-1} Q_{i,k}$$

where N specifies the total number of states in the truncated queueing model. We will use this truncated queue to find an approximation of the stationary distribution of the original queue. An appropriate value for N can be found in an iterative way. We start with $N = 50$ and we increase N by 25 in each iteration step i until:

$$E(L_i) - E(L_{i-1}) \leq \epsilon.$$

4.2 Results

In this section we will present the results in terms of mean queue length and mean waiting time for varying λ. We will present these results for two different cases:

- Three intake sessions per week for five patients each, $a = 3$ and $b = 5$.
- Three intake sessions per week for six patients each, $a = 3$ and $b = 6$.

Het Roessingh worked according to the second case until roughly one year ago. At that time the intake procedure changed and the current situation of 15 intakes per week was introduced. The mean queue length and the mean waiting time for both cases are shown in respectively Figure 4.1 and Figure 4.2.

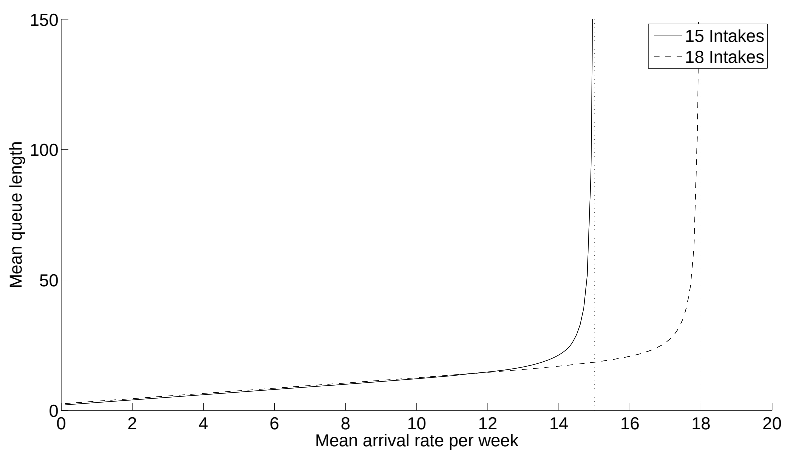

Figure 4.1: Mean queue length for given λ.

We can see that when λ is close to the maximum number of intakes per week, the mean queue length becomes very large. This effect is also visible in

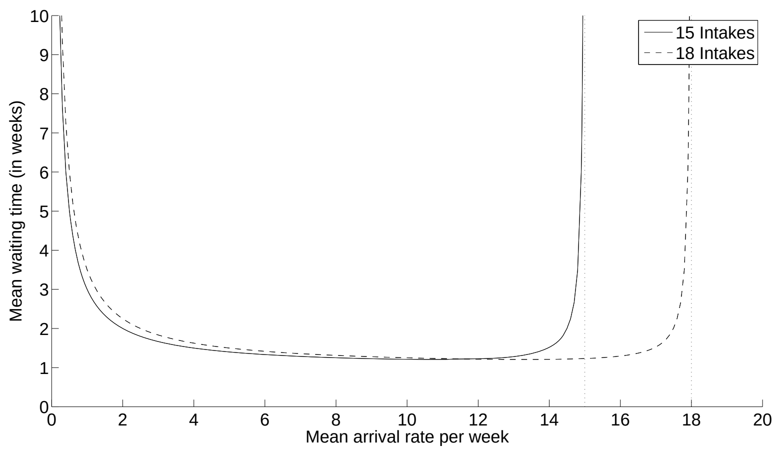

Figure 4.2: Mean waiting time for given λ.

the mean waiting time. There are on average 17.7 referrals to Het Roessingh each week and 93.9% of these patients must undergo an intake before they can start their treatment. This results in an average of 16.6 patients per week which must undergo an intake. The current situation is not capable of handling this amount of patients. The old situation on the other hand, should be able to cope with this amount and result in a mean queue length of 23.1 patients and a mean waiting time of 1.4 weeks. In this situation of 18 intakes per week we find a utilization of 92%.

4.3 Conclusions

In the previous section we have shown the effect on the mean queue length and the mean waiting time for λ ranging from 0 to 18 in two different cases. By looking at the plots in Figure 4.1 and Figure 4.2 we can conclude that the old situation of 18 intakes per week performs better than the current situation of 15 intakes per week. This current situation results in an instable queueing system when there are on average 16.6 patients who apply for an intake each week. To stabilise the queueing system we must have capacity for at least 17 intakes per week. However, we must bear in mind that by increasing the number of intakes per week we will decrease the utilization of the intakes. Increasing the number of intakes per week will also result in a higher probability that no intake was needed in a particular week.

Chapter 5

Patient Flow Analysis

In this chapter we will analyse the patient flows in the entire network of treatment plans in the Pain Rehabilitation department in Het Roessingh. For this purpose we will model the network of treatment plans as a feedforward queueing system which we will study analytically and by simulation. The problems arising at Het Roessingh (long waiting lists, high working pressures and unfinished jobs) are all characteristics of an overloaded system and with these two approaches we want to answer the most important question of this thesis, namely whether or not the queueing system is stable and if the Pain Rehabilitation department is indeed overloaded.

We will use both models to answer this question and the results can be found in Section 5.4.

With the simulation model we will also test the effect of four aspects on the patient flows. We have seen one of these aspects before in Chapter 3, namely the cancellation of meetings. In Chapter 3 we have seen that the mean lead time of an RDB-patient grows rapidly when the probability of cancelled meeting increases. With this simulation model we can also see the effects of these cancelled meetings on the entire queueing system.

The second aspect is the introduction of an artificial buffer. Patients in the Pain Rehabilitation department follow a unique treatment plan while they are diagnosed with an average treatment plan. This difference is the result of the uniqueness of each patient. In the queueing system we will denote this difference by stochastic fluctuations and we will introduce an artificial buffer which can absorb these fluctuations.

A treatment plan contains all the information of a patient's treatment and it specifies the amount of capacity needed for each week of treatment of a patient. Usually we will only schedule a patient whenever there is enough capacity available for each week of the treatment plan, which we will call **complete planning**. However, we speak of **partial planning** when we allow patients to be scheduled when there is not enough capacity available for at least one week of the treatment plan. At this moment Het Roessingh allows patients to be partially planned but by the introduction of treatment plans this should not be allowed anymore. With the simulation model we can test this third aspect by comparing the difference between complete planning and partial planning in the queueing system.

Patients who are referred to Het Roessingh will spend a long time in the queue until they can start their treatment. In the worst case scenario a patient will visit three different queues, one before intake, one before observation and one before the actual treatment. We want to minimize the waiting time between the observation and the actual treatment and we will try to achieve this by giving priority to patients who completed their observation or by planning these patients partially. This last aspect can give some insight in how the waiting times for these patients can be improved by changing scheduling procedures.

The results found by the simulation model on each of these aspects can be found in Section 5.5.

5.1 Model definition

The system consists of a network of 18 treatment plans and it can be seen as a network of discrete time queues. Each treatment plan is represented by a queue in this network. When a patient arrives at a queue he must wait until his treatment will begin. The service needed at each queue is described in the treatment plan belonging to this queue. An example of a treatment plan can be found in Appendix C. In this treatment plan we can find how often each treatment is given each week, how long each treatment lasts and if the treatment is given in a group or individually. In this treatment plan we can also find if a treatment is given to all patients who follow this treatment plan, Prob$=1$, or only to a fraction, Prob<1. Each patient will be served in a batch of size n. When $n=1$ we say that the patient is treated individually. The service in each treatment plan is divided in k weeks and in each of these weeks service is needed from several practitioners, for example physiotherapists and psychologists. Each practioner has a certain amount of capacity available each week to treat patients and the planned treatment time of each practitioner cannot exceed this capacity. In this model we will group all practitioners of the same discipline so the total capacity available is given per discipline. This available capacity is measured in Full Time Equivalents (FTE's) per week and is depicted in Table 5.1, the abbreviations in this table are explained in Appendix A. In this table we also included the number of beds available at the Pain Rehabilitation department. This capacity is used for all 18 treatment plans. In the simulation model we will not distinguish between individual practitioners in a discipline. In this way we only have to look at the total amount of available capacity per discipline.

Discipline	Available capacity
BA	2.58
ET	3.55
FT	7.38
MW	4.69
PS	4.53
RA	4.00
Beds	28 Beds

Table 5.1: Available capacity per discipline.

When a patient has completed treatment plan i he will transfer to treatment plan j with probability $Q_{(i,j)}$. This transition only depends on i and j. By adding a source node s and a sink node r we can also define arrivals to the network and departures from the network. We say that a patient arrives to the network to start his treatment in treatment plan j with probability $Q_{(s,j)}$ and that a patient leaves the network after completing treatment plan i with probability $Q_{(i,r)}$. This process is a Markov Chain with transition matrix Q and absorbing state r. The transition matrix Q is stated in Appendix B.

In Figure 5.1 we give a example of the network described above with 4 treatment plans. In this figure we can see that treatment plan 2 and 3 consist of three blocks. These blocks represent the weeks in the treatment plans and in each of these weeks capacity is needed from different disciplines.

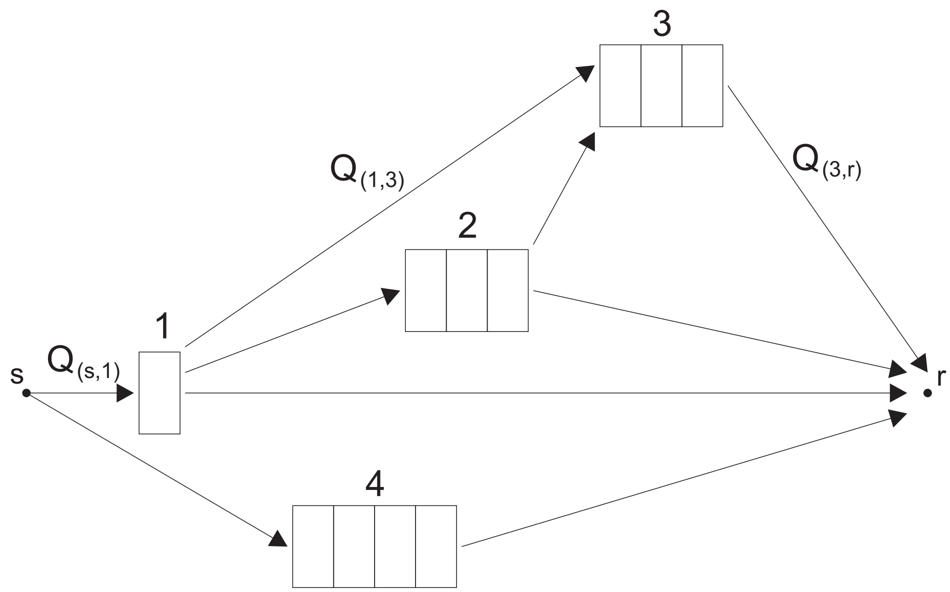

Figure 5.1: Example of a network of four treatment plans.

5.2 Simulation Model

In order to present a clear overview of the simulation model we will use an approach stated in Robinson [22] and we divide this section in five subsections:

- Model Objectives
- Input
- Output
- Content
- Assumptions and Simplifications

In each of these five subsections we will explain one of the building blocks of the simulation model.

5.2.1 Model Objectives

The most important aspect of our simulation is that is must represent reality. This means that it must be able to:

- Handle a random number of arrivals each week.
- Determine a treatment plan for a patient and be able schedule it.
- Put patients in the queue belonging to their treatment plan according to the First-Come-First-Serve policy.
- Put patients in a queue with priority over other patients.

To summarize, the simulation model must be able to deal with patients and patient scheduling as it is dealt with in Het Roessingh. Furthermore, it must be able to operate under various circumstances which are defined as interventions. The purpose of this simulation model is to analyse the system in several situations by means of the performance indicators, namely:

- Waiting time for the intake, observation and treatment.
- Utilization for all the disciplines and for the beds available at Het Roessingh.
- Total service time for each patient.
- Probability of extending the treatment plan.
- Mean number of arrivals and discharged patients each week.

With the last performance indicator we can answer an important question concerning stability of the system. In a stable system the mean number of arrivals must be equal to the mean number of departures each week. Due to the importance of this question we will discuss this subject seperately in Section 5.4.

5.2.2 Input

In the simulation model we use two different types of input. The first type is input which is stated in a database. An example of this type of input is the data belonging to a treatment plan. The second type is input are interventions which we change in each simulation run to analyse its effects. We will first discuss input of the first type.

Treatment Plans. The treatment plans are the fundamentals of our simulation model. An example of a treatment plan of five weeks is shown in Appendix C. In a treatment plan we can find all the data needed in order to treat a patient, we can find which treatments are given and how long each treatment lasts, how many patients are treated at the same time and we can find in which week the treatment is given. It is also known with which probability a treatment is needed. Some patients need extra care compared to other patients in the same treatment plan. Therefore we find that some treatments have a probability less than 1, for example “Sport/Spel” given by the BA practitioners in Appendix C. We see that the probability that an arbitrary patient needs this treatment is equal to 0.5.

Discipline	Direct time	Indirect time
BA	80%	20%
ET	68%	32%
FT	79%	21%
MW	68%	32%
PS	50%	50%
RA	68%	32%

Table 5.2: Direct and Indirect time per discipline.

In the last row of this example we can find that the patient following this treatment plan needs a bed for half a week. This is typical for a semi-clinical patient who spends two nights per week in the clinic.

Transition Matrix. In Section 5.1 we defined a transition matrix Q. This matrix is shown in Appendix B and gives the probability $Q_{(i,j)}$ that a patient who completed treatment i will start treatment plan j. It also provides us with the probabilities of arriving to the system and start with treatment j, denoted by $Q_{(0,j)}$, or leave the system after completing treatment plan i, denoted by $Q_{(i,19)}$.

Arrival Process. New patients arrive to the system every week and we assume that the number of patients entering the system are Poisson distributed with mean 17.7, which is the mean number of patients which arrived to the system each week in 2009, see Table 4.1.

Indirect Time. The time each practitioner works is divided in direct and indirect time. Here, direct time indicates treatment time with a patient and indirect time refers to the time it takes to write reports or other patient-related activities without the patient being present. The amount of indirect time is different for each discipline but is stated in Table 5.2. The total treatment of a patient requires both direct and indirect time but only direct time is mentioned in the treatment plans. We must therefore correct each treatment plan by adding the correct amount of indirect time for each treatment. The corrected treatment plans can then be used in the simulation model.

Priorities. For each treatment plan a queue is created with patients waiting for starting that particular treatment plan. The patients in this queue are served according to the First-Come-First-Serve policy. The order in which we schedule the treatment plans is on priority base. The priorities used in the simulation model are shown in Appendix D. In this priority table we can see that Intake has highest priority and that ZP-20A has second highest priority. We will therefore first schedule the patients from the queue belonging to Intake according to an FSFC-policy. When this is finished we will schedule the patients from the queue belonging to ZP-20A according to an FCFS-policy.

Groups. Some patients are treated in a group and some are treated individually, depending on their treatment plan. Knowing the size of these groups is important in order to schedule patients. Patients who are treated in a group are treated in a batch of several patients at once. The treatment cannot start when the batch is not full. It is also important to know how many groups can be treated in the same week. These group constraints are mentioned in Appendix E.

Capacity. For each discipline it is known how much Full Time Equivalents, FTE's, are available per week to treat patients. It is also known how many beds are available and how many groups can be treated at the same time. The amount of capacity available determines whether a patient will be scheduled or not. In the basic model we demand that a patient will only be scheduled if there is enough capacity available for each week of the treatment plan. The available capacity used in the simulation model is shown in Table 5.1.

With this simulation model we want to test four interventions:

Buffer. Every time we schedule a patient's treatment plan, we will schedule the average treatment plan. However, each patient is unique and has a unique treatment plan which may differ from the average treatment plan. In reality we know in week i the actual amount of treatment needed in week $i+1$. We will therefore update the scheduled treatment to the unique treatment plan one week in advance. Because of the stochastic nature of the unique treatment plans this results in fluctuations in the capacity needed each week. With the simulation model we want to test the effect of a so called artifical buffer on this phenomenom. By introducing an artificial buffer we will put aside a part of the available capacity for special purposes. This artificial buffer cannot be used to schedule new patients and is only used for the stochastic fluctuations.

Partial Planning. The treatment plans are introduced to improve the patient flows at Het Roessingh and to create more clarity for the patients. An important aspect of these treatment plans is that it is known in advance how much capacity is needed on average for a paticular treatment plan. This can be used by the planning department by only scheduling a treatment plan whenever there is enough capacity available for each week of the average treatment plan. It is also possible that we allow patients to start their treatment whenever there is at least one week in which there is not enough capacity available. We call this *Partial Planning*. In the current system, where we do not work with treatment plans yet, a lot of the patients are partially planned. This should yield a higher utilization but it also results in a very complex treatment plan in which the patient does not have a clear view of his treatment.

In the simulation we will only allow individual outpatients to be partially planned. Every other patient will either be part of a group or his treatment requires a bed. In both cases it is not possible to allow partial planning.

Decency Rule. During his treatment a patient can encounter a maximum of three different queues in which he must wait before starting his treatment. One for the intake, one for observation and one for his actual treatment. In this case, a patient cannot be diagnosed during the intake and he is admitted for observation. After this observation the patient will start his actual treatment. In the basic simulation, these patient will enter the queue belonging to the treatment plan according to an FCFS policy before they will start their treatment. With the simulation model we want to test the effect of using a different decency rule, namely partial planning of priority queueing. In the first rule we will partially plan these patients so they can start their treatment immediately when the observation is finished. With the second rule the patient will enter the queue belonging to the treatment plan with top priority.

Cancelled meetings for RDB-patients. In Chapter 3 we discussed the effect of the level of attendence at the interdisciplinary meetings for RDB-patients. In this simulation model we want to test the effect of this level of attendence on the entire system. We will therefore consider two possibilities for the probability of cancelled meetings, namely a constant probability or a probability depending on the number of patients in the system. The last possibility is based on observations in Het Roessingh, when there are more patients in the system, more meetings will be cancelled.

The use of Partial Planning and the Decency Rule is depicted in the flow chart in Figure 5.2. This chart starts in the triangle with First Treatment Plan (1st TP) and ends in one of the rectangles denoting the action which will be executed by the simulation model. In this flow chart we denote the two possibilities for Partial Planning by a 1 (allowing) and a 0 (not allowing), the three possibilities for the Decency Rule are 1 (normal queueing), 2 (priority queueing) and 3 (partial planning).

5.2.3 Output

With this simulation model we want to test the effects of the interventions mentioned in Section 5.2.2, namely: the use of an artificial buffer, partial planning, the use of a decency rule or the cancellation of meetings on the patient flows. We will do this by means of the following five output variables.

Waiting Time. We will look at three different average waiting times: the waiting time before an intake, the waiting time before treatment and the waiting time between observation and treatment. The first waiting time is also discussed in Section 4. The second waiting time can be categorized into two cases, either a patient starts his treatment or observation immediately without an intake, or he undergoes an intake first. We will use the Decency Rules to decrease the third waiting time.

Service Time. The mean service time is defined as the sum of the service times of the treatment plans and the waiting times in the queues. This is the total time between referral to the centre and discharge from the centre.

Utilization. We determine the utilization for each discipline per simulation. This utilization is defined as the total time a discipline works with patients divided by the total available time per discipline.

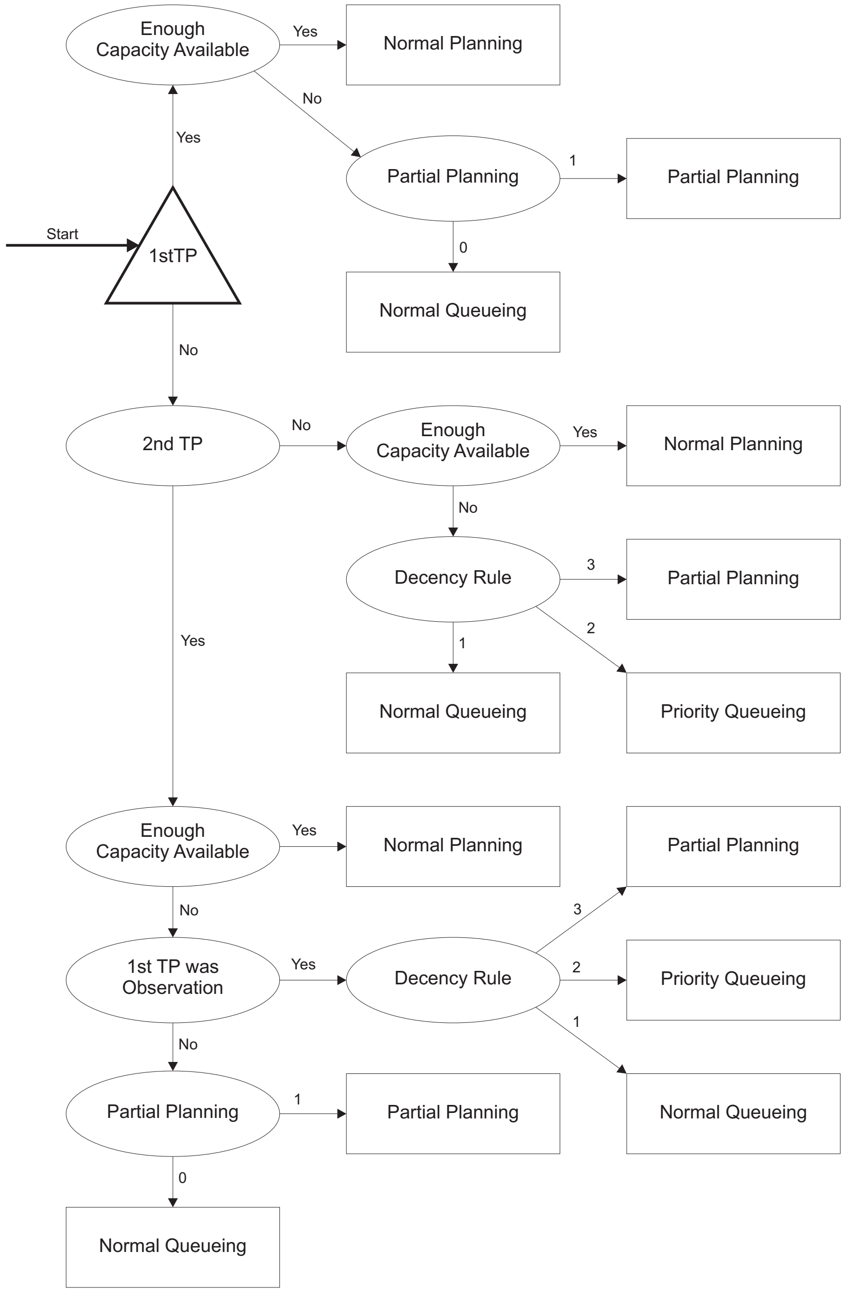

Figure 5.2: Scheduling Protocol.

Number of Arrivals and Departures. Each week we keep track of the number of arrivals and the number of discharged patients in the system. These numbers are used to analyse the stability of the system.

Probability of overtime. In updating the average treatment plan to the unique treatment plan it is possible that there is not enough capacity available to complete the treatment plan before the due date. If this happens, we will extend the treatment plan to make up for missed treatments. We will consider this as overtime. An extension due to cancelled meetings for RDB-patients is also considered as overtime.

5.2.4 Content

The content of the simulation model is explained with the help of the patient flow depicted in Figure 5.3. Each station is shortly explained in the list below:

- **Arrival -** Patients arrive at the system following a Poisson distribution.
- **Determine Treatment Plan -** Given the patient's history, a treatment plan is determined.
- **Scheduling Decisions -** Given the protocol in Figure 5.2 a scheduling decision is made and the patient will either enter a queue or he will start his treatment.
- **Treatment -** In week k we will decide whether or not a patient will start his treatment in week $k + 1$. We will decide to schedule a patient when there is enough capacity available for his entire treatment plan. In other words, when we schedule a patient's treatment, he will start in one week from now. For each patient the avarage treatment plan is scheduled and during the treatment we update the scheduled treatment plan with the unique treatment plan. We do this one week in advance. During the treatment, we also check whether or not the patient is an RDB-patient and if a meeting is cancelled or not. We will extend these treatment plans with an additional six weeks according to the hybrid policy described in Chapter 3.
- **Exit -** When a patient is cured he will leave the system. At this time all output variables for the patient are collected and stored.

5.2.5 Assumptions and Simplifications

The network of treatment plans at Het Roessingh is a complex system and in order to create a reasonably fast simulation, assumptions and simplifications must be made. These are listed below.

- The simulation model works with disciplines instead of practitioners. The amount of the available capacity per discipline is the sum of the available capacity per practitioner in this discipline.
- We do not worry about the planning department, whenever there is enough capacity available for the treatment plan of a patient, we assume that this treatment plan can be scheduled.
- We assume the arrival rate to be time invariant although we see some deviations in Table 4.1. A detailed research should point out whether or not the assumption of time invariance is correct.

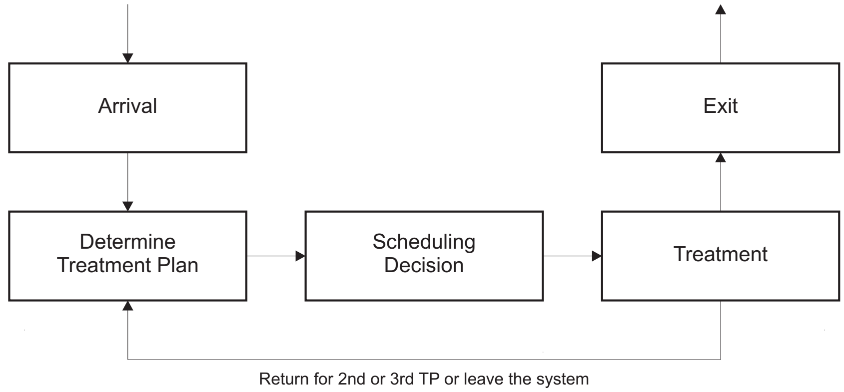

Figure 5.3: Patient Flow Diagram

- Patients are always perfectly diagnosed. The dash-dotted line in Figure 2.1 depicts patients with an incorrect diagnosis. These patients have to switch treatment plans to receive the correct treatment. This happens to a very small fraction of patients and is therefore ignored in the simulation model.
- The treatment plans for RDB-patients are very hard to predict and we find a lot of deviation when we compare the actual treatment plan with the average treatment plan. At Het Roessingh it is not clear how large these deviations actually are. We will therefore assume that these deviations are uniformly bounded with the average treatment plan as its mean. The lower bound is stated at 50% of the mean and the upperbound at 150%. Say, for instance, that a treatment lasts for 60 minutes in the average treatment plan, then the treatment can last somewhere between 30 and 90 minutes in the unique treatment plan.
- Patients do not miss a treatment. In reality, patients get sick, go on holiday or will sometimes not show up for their appointments without reason.
- We make a similar assumption for the available practitioners at Het Roessingh. We will not take into account sudden changes in the available capacity due to sickness, etc.

5.2.6 System specifications

With this simulation model we want to test different aspects of the system and this means that we must declare a run-time and warm-up period for each simulation and the number of repetitions. We use Welch's method to find the warm-up period and to this end we apply a moving average of size 199 to the mean number of departures each week. The result is shown in Figure 5.4 and from this figure we find that the warm-up period must be 200 weeks. At this point the mean number of departures will be close to its equilibrium point. However, we can also see that the moving average will not result in a flat line.

This is a result of the size of the moving average and of the nature of the mean number of departures. It is possible that the difference in the number of arrivals between two weeks is larger than 30, causing the moving average to oscillate.

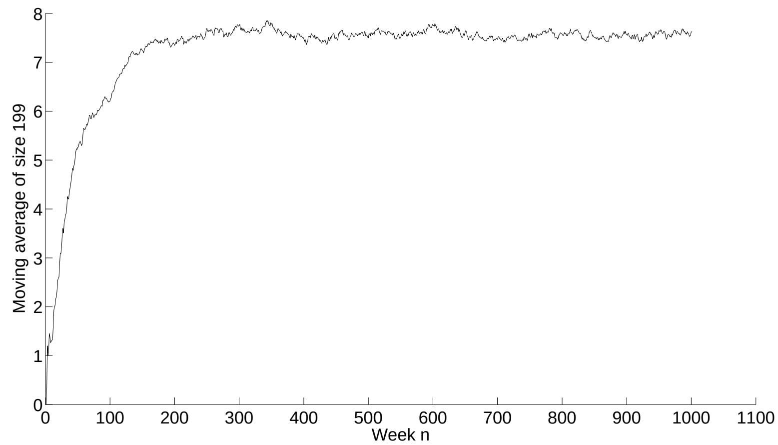

Figure 5.4: Moving average of size 199 on mean number of departures for $\lambda = 7.5$.

We will choose the total run-time including warm-up period to be 1100 weeks. The created simulation is very complex and if we choose the total run time to be larger than 1100 weeks we will run into computational limitations. The 95% confidence interval of the mean number of departures now results in 7.6 ± 0.3. If we want to reduce this interval we must choose a larger run-time. However, this is not possible with the currently used software. Due to time constraints we will choose to run only one repetition.

5.3 Workload analysis

In this section we will present an analytical model with which we can determine the mean amount of capacity needed in the system each week. This can then be used to determine the utilizations of the disciplines in the network. In this section we will also determine the variance of the amount of capacity needed in the system each week. We will not use this until Chapter 6, but the derivation is similar to the derivation of the mean.

To determine the mean and the variance we will first simplify the network defined in Section 5.1. Instead of looking at a network of treatment plans we will introduce the concept of a route which is defined as the combination of one or more consecutive treatment plans describing the entire treatment of a patient from start until finish.

By using this definition we can also look at the network as if it were a tree of A routes instead of a graph of treatment plans. This process is depicted in Figure 5.5. In this figure we can also see that each treatment plan is divided in blocks which denote the weeks of a treatment plan and because a route is a combination of treatment plans we can also divide a route into weeks. A route consisting of three treatment plans of five weeks each will results in a route of 15 weeks. These routes are shown in the bottom half of Figure 5.5. In this figure we colored the third week of each route black.

From the treatment plans we know that each week requires a certain amount of capacity from each discipline, so by the definition of a route we can say that each week in a route denotes a week of treatment and in this week a certain amount of capacity is needed. To avoid nasty formulas we will add dummy weeks to the end of each route so that each route has length T and create $N = AT$ states in which n reprensents the pair (a, t). This dummy week represents a week in which 0 FTE's of work are required from each discipline. In Figure 5.5 we colored one of the states dark grey. In this figure we can see that each route consists of seven weeks, including dummy weeks, and that the entire system consists of 35 states.

We are now interested in the total amount of capacity needed in the system each week for each discipline. The formulas in this section are derived for one single discipline and can be used for each discipline seperately. We will denote the amount of capacity needed by C and we want to find $E(C)$ and $Var(C)$. Before we derive the model we will first introduce the variables which we use besides C. For each state n we will define three variables. The total capacity needed is denoted by X_n, the number of patients is denoted by L_n and the capacity needed for one patient is denoted by U_n. The capacity needed in state n is the same for all patients in this state, so U_n is independent of the patient. The last variable we need represents the probabilty that a patient will be in a certain route. For this purpose we will define a set $\mathcal{R}_y$ which contains all states n in route y. This route y is a combination of three treatment plans i, j and k and we define the variable q_n for $\forall n \in \mathcal{R}_y$ to be:

$$q_n = q_{i \to j \to k} = Q_{s,i} \cdot Q_{i,j} \cdot Q_{j,k},$$

where s is the source node. In this notation we allow i, j and k to be the sink node r. This variable q_n is equal for $\forall n \in \mathcal{R}_y$ and we state that q_n is the probability that an arbitrary patient will flow into route y when he arrives to the system. This variable q_n will be needed when we want to determine the mean and variance of the number of patients in a certain state L_n.

We can now formulate the model and start by stating that

$$C = \sum_{n=1}^{N} X_n.$$

The number of arrivals to the system is Poisson distributed with mean λ. However, because of the batching process during the intake we will find that the number of patients on routes starting with the intake follow a multinomial distribution. Therefore

$$\begin{aligned} E(C) &= E\left(\sum_{n=1}^{N} X_n\right) \\ &= \sum_{n=1}^{N} E(X_n). \\ Var(C) &= Var\left(\sum_{n=1}^{N} X_n\right) \end{aligned}$$

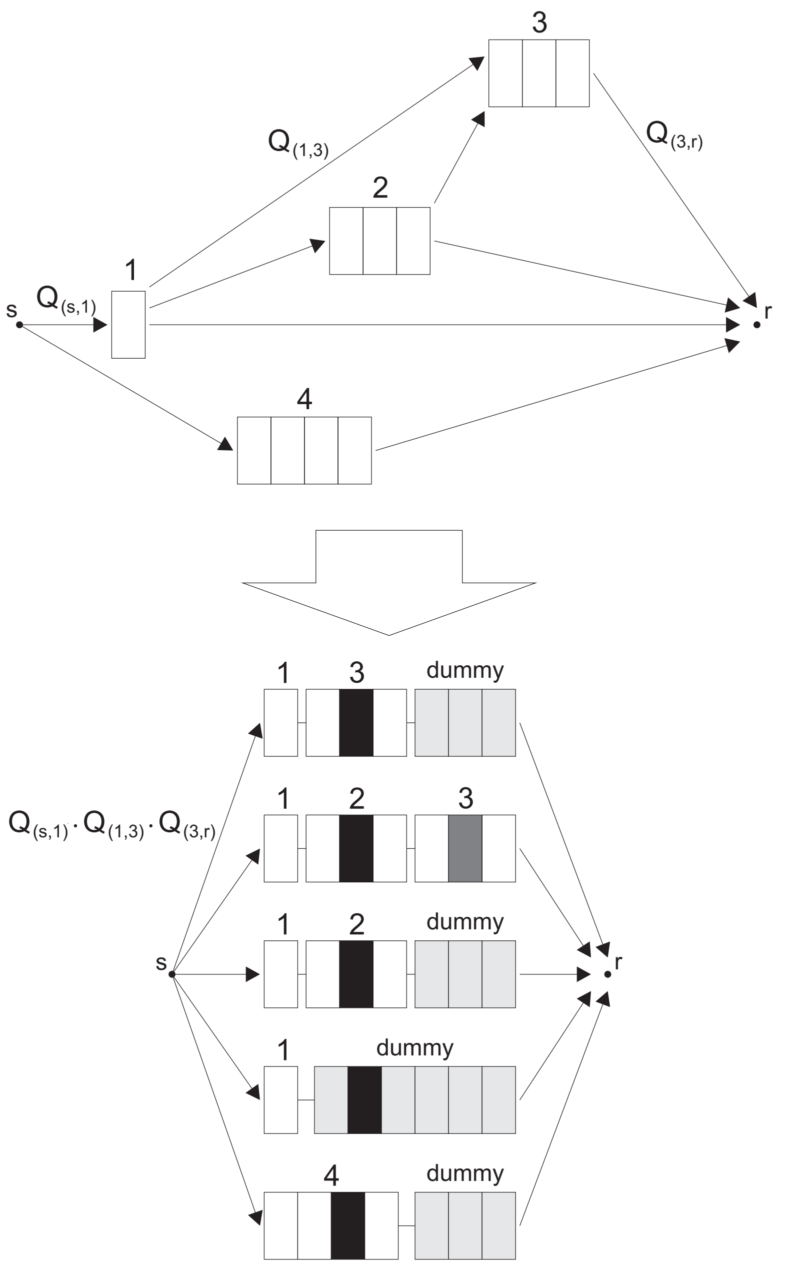

Figure 5.5: Creating a Tree of Routes out of a Network of Treatment Plans.

$$= \sum_{n=1}^{N} Var(X_n) + 2 \sum_{n=1}^{N} \sum_{m=n+1}^{N} Cov(X_n, X_m).$$

Because X_n is the random sum of L_n i.i.d. random variables U_n we can find

$E(X_n)$ by using Wald's equation [23]:

$$\begin{aligned} E(X_n) &= E\left(\sum_{i=1}^{L_n} U_n\right) \\ &= E(L_n)E(U_n). \end{aligned}$$

By conditional expectations we can also find according to Ross [23]:

$$\begin{aligned} Var(X_n) &= Var\left(\sum_{i=1}^{L_n} U_n\right) \\ &= E(L_n)Var(U_n) + E(U_n)^2 Var(L_n). \end{aligned}$$

For the covariance we can find

$$Cov(X_n, X_m) = E(X_n X_m) - E(X_n)E(X_m),$$

where

$$\begin{aligned} E(X_n X_m) &= E\left(\sum_{i=1}^{L_n} U_n \sum_{j=1}^{L_m} U_m\right) \\ &= \sum_{l_n=0}^{\infty} \sum_{l_m=0}^{\infty} l_n l_m E(U_n U_m) P(L_n = l_n,\ L_m = l_m) \\ &= E(U_n)E(U_m) \sum_{l_n=0}^{\infty} \sum_{l_m=0}^{\infty} l_n l_m P(L_m = l_m,\ L_n = l_n) \\ &= E(U_n)E(U_m)E(L_n L_m), \end{aligned}$$

and therefore

$$\begin{aligned} Cov(X_n, X_m) &= E(X_n X_m) - E(X_n)E(X_m) \\ &= E(U_n)E(U_m)E(L_n L_m) - E(L_n)E(U_n)E(L_m)E(U_m) \\ &= E(U_n)E(U_m)Cov(L_n, L_m). \end{aligned}$$

All that remains is determining the mean and variance of L_n and U_n. The mean and variance of U_n can be obtained from the treatment plans. These treatment plans say very specifically which treatment is definitely needed, $p = 1$, and which treatment is not always needed, $p < 1$. We can use this information to determine the mean and the variance of the capacity needed per patient in state n.

To determine the mean, variance and covariance of L_n, we will create two sets $\mathcal{N}_A$ and $\mathcal{N}_B$. All states from the routes starting with the general intake are part of $\mathcal{N}_A$ and all other states belong to $\mathcal{N}_B$. Because of the batching process at intake we find that L_n is Multinomially distributed if $n \in \mathcal{N}_A$ and it is Poisson distributed if $n \in \mathcal{N}_B$.

To compute the means and the variances of L_n we use q_n and we define

$$E(L_n) = q_n \lambda,$$

and

$$Var(L_n) = \begin{cases} q_n(1 - q_n)\lambda, & \forall n \in \mathcal{N}_A \\ q_n \lambda, & \forall n \in \mathcal{N}_B. \end{cases} \tag{5.1}$$

Finally we introduce the set $\mathcal{T}_x$ which contains all states n from different routes which have the same week x in common. For instance, $\mathcal{T}_1$ contains all states which correspond to the first week of every route and $\mathcal{T}_{10}$ contains all the states which correspond to the 10^{th} week of every route. We can also relate this notation to the patients. When patients arrive to the network on $t = 0$ they will be scattered throughout the network 10 weeks later. The set $\mathcal{T}_{10}$ contains all states where these patients can be on $t = 10$.

Since we translated the network of queues into a tree of routes we will find that every state n will be contained in one particular set $\mathcal{T}_x$. Now due to independence between states from different sets $\mathcal{T}_{t1}$ and $\mathcal{T}_{t2}$, and because of independence in the Poisson distribution we have

$$Cov(L_n, L_m) = \begin{cases} -\lambda q_n q_m, & n, m \in \mathcal{T}_t, \quad n, m \in \mathcal{N}_A, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Which states that we will only find a covariance between patients who arrive at the centre at the same time and undergo an intake.

Finally we find for $E(C)$ and $Var(C)$:

$$\begin{aligned} E(C) &= \lambda \left(\sum_{n=1}^{N} q_n E(U_n) \right), \\ Var(C) &= \sum_{n=1}^{N} \lambda q_n Var(U_n) + Var(L_n) E(U_n)^2 + \\ & \quad 2E(U_n) \sum_{m=n+1}^{N} E(U_m) Cov(L_n, L_m), \end{aligned}$$

where $Var(L_n)$ and $Cov(L_n, L_m)$ are stated in (5.1) and (5.2). By looking at the formula for $E(C)$ we can see that this is a linear function in λ. When we know the mean amount of capacity needed for $\lambda = 1$, which is easily calculated by the model, and when we know the amount of available capacity we can calculate the maximal arrival rate to the system, λ_{max}, per discipline by dividing the available capacity by the capacity needed for $\lambda = 1$. We will use this in the next section where we determine the bottlenecks of the network.

5.4 Bottlenecks and Stability

In this section we will present the results concerning stability and bottlenecks by using the models presented in the previous two sections. With these results we will show that the queueing system is unstable and we will show where this instability is coming from. We answer the stability question by using the simulation model. With this model we determined the mean number of arrivals and departures for different λ. The results from these simulations are shown in Figure 5.6.

In this figure we can see that for an arrival rate of $\lambda = 17.7$ the mean number of arrivals is larger than the mean number of departures resulting in an unstable queueing system. This means that the number of patients in the system grows to infinity and that the queueing system "blows up".

The system looks stable for $\lambda = 7.5$ which is less than half of the current arrival rate. Questions arise how this is possible and how this should be solved.

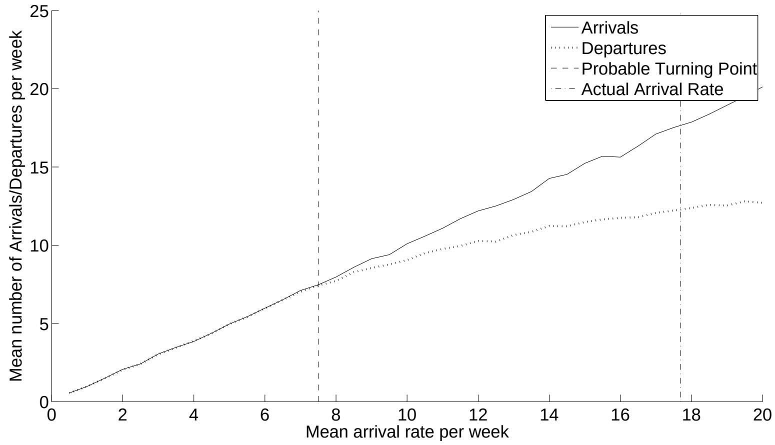

Figure 5.6: Mean number of arrivals and departures for various λ.

A first step in solving this problem is finding the bottlenecks, which we will do with the help of the analytical model of Section 5.3. With this model we will determine the mean amount of work needed each week for different arrival rates and we will compare this with the amount of capacity which is currently available. Before we use this model we want to demonstrate that the two models yield same results for average capacity demand. We will do this for two different λ's for which the queueing system is stable. The results are shown in Table 5.3. In this table we denote the results found by using the simulation model by “Sim” and the results found by the analytical model by “Ana”. The units in which the results are given are denoted in the last column. The results for the BA practitioners, for example, are given in FTE's

We can see that there is a minor difference between the simulation model and the analytical model. This difference can be explained by the fact that the simulation model is a realisation of a stochastic process whereas the analytical model works with the exact average. We will therefore see a little variation between the two results.

From the results in Table 5.3 we can conclude that the simulation model and the analytical model give similar results for a stable queueing system. In Table 5.4 we show the capacity needed for $\lambda = 1$, the available capacity in the Pain Rehabilitation department and the mean arrival rate which the system can handle at the most. This λ_{max} is obtained by dividing the available capacity by the capacity needed for $\lambda = 1$. When we look at the ET practitioners we can see that $\lambda_{max} = 3.55/0.44 = 8.0$, which means that this discipline has enough capacity available to treat 8.0 new patients each week. The group constraint for ZP-20B results in $\lambda_{max} = 20.2$, which means that if we only look at this particular constraint we are able to treat 20.2 new patients each week.

In Table 5.4 we can see that when $\lambda = 17.7$ only four out of nine group constraints are met and only one out of seven capacity constraints are met. We can see that the system can only take on a maximal arrival rate of 7.1 because of the group constraints of treatment plan ZP-20A. When we only look at the capacity constraints we find a maximum arrival rate of 8.0 because of the

	$\lambda = 3$		$\lambda = 5.5$		
	Sim	Ana	Sim	Ana	Unit
BA	0.47	0.45	0.80	0.83	FTE's
ET	1.35	1.33	2.39	2.44	FTE's
FT	2.57	2.58	4.61	4.73	FTE's
MW	1.13	1.11	1.98	2.03	FTE's
PS	1.24	1.19	2.15	2.19	FTE's
RA	0.44	0.43	0.76	0.78	FTE's
Intake	2.78	2.82	5.04	5.16	# Patients
Obs-Ach	0.47	0.46	0.83	0.85	# Patients
Obs-Voo	0.58	0.54	0.94	0.99	# Patients
ZP-20A	0.46	0.42	0.74	0.77	# Patients
ZP-20B	1.01	1.04	1.93	1.90	# Patients
ZP 20C	1.19	1.11	1.93	2.03	# Patients
ZP-20D	2.88	2.81	4.96	5.15	# Patients
ZP-20K	0.28	0.30	0.53	0.55	# Patients
ZP-20L	0.03	0.02	0.04	0.03	# Patients
Bed	4.85	5.02	8.93	9.21	# Beds

Table 5.3: Comparison between Simulation model and Analytical model.

critical load of occupational therapists (ET). Although we can see that these arrival rates are very low compared to the actual arrival rate, we must make a remark that these values of λ_{max} will still not result in a stable system. For each of the 16 constraints we can find that the utilization will become 100% when we use λ_{max}. If we want the system to be stable we must demand that the arrival rate will be even lower than λ_{max}.

5.5 Numerical Results

In this section we will present the numerical results found with the simulation model. These results are based on the four aspects described in Section 5.2.2. In order for these results to be usefull for this research we made two modifications to the simulation model. In this thesis we investigate how we can improve the patient flows by using the available capacity more efficiently. We will therefore drop the group constraints which specify how many groups can be treated in the same week. These group constraints are stated in Appendix E. Without these group constraints only rely on the capacity available at each of the disciplines. In the previous section we showed that $\lambda_{max} = 8.0$ when the system only relies on the capacity constraints. However, we also mentioned that if we choose $\lambda = 8.0$ that the utilization of ET will be 100% and that the system will still be unstable. We will therefore choose $\lambda = 7.5$ so that the utilization is less than 100% for all disciplines.

With this slightly modified simulation model we will perform a total of 10 simulations. With these simulations we will test the effects of partial planning, the decency rule, the artificial buffer and the probability of cancelled meetings. This simulation model gives us information on all points discussed in Section 5.2.3 but we will only present those results that provide us with useful informa-

	Available Cap.	Cap. Needed for $\lambda = 1$	λ_{max}
BA	2.58	0.15	17.1
ET	3.55	0.44	8.0
FT	7.38	0.86	8.6
MW	4.69	0.37	12.7
PS	4.53	0.40	11.4
RA	4.00	0.14	28.2
Intake	15	0.94	16.0
Obs-Ach	2	0.15	13.0
Obs-Voo	3	0.18	16.6
ZP-20A	1	0.14	7.1
ZP-20B	7	0.35	20.2
ZP-20C	6	0.37	16.2
ZP-20D	18	0.94	19.2
ZP-20K	6	0.10	59.6
ZP-20L	6	0.01	1065.0
Bed	28	1.67	16.7

Table 5.4: Determining λ_{max} for each constraint.

tion on each of the four aspects.

5.5.1 Cancelled meetings for RDB-patients

In the first simulation run we will test the effect of cancelled meetings for RDB-patients. We will perform four simulations in which we use different probabilities in each simulation namely: $p = 0$, $p = 0.3$, $p = 0.7$ and p is a monotone increasing function depending on the number of patients z in the system. The results of these four simulation are shown in Table 5.5.

	$p = 0$	$p = 0.3$	$p = 0.7$	$p = f(z)$
Mean # arrivals	7.5	7.4	7.3	7.6
Mean # departures	7.5	7.2	6.0	6.2

Table 5.5: Results on cancelled meetings

In this table we present the mean number of arrivals and the mean number of departures per week. These results show that the system is unstable for $p = 0.3$. However, when p increases we know that the mean lead time increases, so if the system is unstable for $p = 0.3$ it will also be unstable for $p > 0.3$. All other results will have no meaning since the system is unstable. This again confirms that the presence at the meeting is crucial.

5.5.2 Partial Planning

In the second simulation run we will test the effect of partial planning presented in Table 5.6. With this simulation we want to test if partial planning results in improving the patient flows. In Table 5.6 we denote the results from the simulation run where we allow partial planning by $Partial = 1$. When $Partial = 0$ we do not allow partial planning.

	$Partial = 0$	$Partial = 1$
Waiting time for intake	1.4 weeks	1.5 weeks
Waiting time for first treatment	4.9 weeks	6.7 weeks
Waiting time after observation	4.6 weeks	5.8 weeks
Total service time	27.0 weeks	28.5 weeks
Utilization BA	0.44	0.44
Utilization ET	0.94	0.96
Utilization FT	0.87	0.90
Utilization MW	0.59	0.61
Utilization PS	0.66	0.67
Utilization RA	0.27	0.27
Utilization Bed	0.45	0.45
Overall system Utilization	0.67	0.68
Probability of overtime	0.04	0.07
Mean # Arrivals	7.5	7.6
Mean # Departures	7.5	7.6

Table 5.6: Results for partial planning.

When we allow partial planning we will also schedule a patient's treatment plan when there is not enough capacity available for at least one week of the treatment plan. That means that in this particular week we do not have capacity left to absorb any stochastic fluctuations of the unique treatment plan. This often results in an extension of the treatment plan.

In Table 5.6 we can see that the mean number of arrivals is larger for $Partial = 1$. This is a result of the coinfidence interval and with a longer simulation run and better software this can be solved. However, in the results in Table 5.6 we must take this difference into account since it affects all values in the table. Although all values are a bit higher we see a significant increase in the probability of overtime. When we allow partial planning we will plan patients whenever they need to be planned, even when there is not enough capacity available. This results in a complex treatment plan in which several treatments are moved to a weeks further in the treatment plan. Together with the stochastic fluctuations in every treatment plan this often results in an extension of the treatment plan to make up for missed treatments.

With these results we can conclude that allowing partial planning does not lead to an improvement of the patient flows.

5.5.3 Decency Rule

The third aspect is the use of a decency rule. This decency rule determines how patients are handled between observation and treatment. With this simulation run we will test which decency rule results in the smallest waiting time between observation and treatment. The results are shown in Table 5.7. In this table we denote the three different decency rules, normal queueing, priority queueing and partial planning by respectively $Decency = 1$, 2 or 3.

In this table we can see that we can decrease the waiting time between observation and treatment by more than one week without seriously affecting the other waiting times and service time. We can see that partial planning

	Decency = 1	*Decency* = 2	*Decency* = 3
Waiting time for intake	1.4 weeks	1.4 weeks	1.4 weeks
Waiting time for first treatment	4.9 weeks	5.0 weeks	4.7 weeks
Waiting time after observation	4.6 weeks	3.5 weeks	3.4 weeks
Total service time	27.0 weeks	26.8 weeks	26.4 weeks

Table 5.7: Results for three decency rules.

as decency rule performs better than priority queueing by a small amount. However, before we can choose the best decency rule we must also consider the difficulties in the scheduling process of partial planning. We must therefore choose priority queueing as a decency rule. This rule decreases the waiting time between observation and treatment and does not increase the complexity of the scheduling process.

5.5.4 Artificial Buffer

In the last simulation run we will test the effect of an artificial buffer. The current arrival rate is 7.5, while calculations show we have enough capacity to handle an average of 8.0 new patients each week. The difference between these two values is used by the system as a natural buffer. The queueing system uses this buffer to schedule new patients and to absorb stochastic fluctuations of the treatment plans.

In Section 5.2.2 we explained that we will set aside a part of the natural buffer solely for the absorption of the stochastic fluctuations of the treatment plans. In this simulation run we will test three different sizes for this artificial buffer, namely 180 minutes, 360 minutes and 540 minutes. The size of this artificial buffer is the same for each discipline. The results of this simulation run is shown Table 5.8.

	0 min.	180 min.	360 min.	540 min.
Utilization BA	0.44	0.44	0.43	0.43
Utilization ET	0.94	0.94	0.93	0.92
Utilization FT	0.87	0.87	0.87	0.86
Utilization MW	0.59	0.59	0.59	0.58
Utilization PS	0.66	0.66	0.65	0.64
Utilization RA	0.27	0.27	0.26	0.26
Utilization Bed	0.45	0.45	0.44	0.43
Overall system Utilization	0.67	0.67	0.66	0.65
Probability of overtime	0.0439	0.0038	0.0006	0.0002
Mean # arrivals	7.5	7.5	7.5	7.6
Mean # departures	7.5	7.5	7.5	7.6

Table 5.8: Results for using an artificial buffer.

By using an artificial buffer we see that we manage to reduce the probability of overtime to almost 0 but by increasing the size of the buffer we also decrease the utilization per discipline. It is therefore important to determine the correct size of the artificial buffer. An artificial buffer can be used to decrease the probability of overtime, but because we cannot use the artificial buffer to schedule new patients, this will effect the utilization.

5.6 Conclusions

In this chapter we have analysed the patient flows using a feedforward queueing system. By using both the simulation model and the analytical model we can conclude that the network of treatment plans in the Pain Rehabilitation department is overloaded with the current arrival rate of 17.7 new patients per week. We can also conclude that the system can handle 7.1 new patients per week according to the group constraints and 8.0 according to the capacity constraints.

The conclusions drawn from the numerical results are listed below. These conclusions are based on the system with an arrival rate of 7.5 new patients per week.

- The probability of cancelled meeting should be as low as possible. In Chapter 3 we saw that the mean lead time per patient increase considerably, but with the simulation model we found that a cancelled meeting will effect the entire system. We showed that a queueing system which was stable for $p = 0$, turned out to be unstable for $p = 0.3$.

- The Pain Rehabilitation department should not use partial planning to schedule their patients. In Table 5.6 we can see that the system where we do not allow partial planning performs better than the system where we do allow partial planning.

- Priority planning should be used as decency rule. There is a minor difference between the partial planning decency rule and priority queueing, but because partial planning results in a difficult scheduling process it is best to choose priority queueing.

- An artificial buffer has a positive effect on the probability of overtime but it is important to use a correct sized artificial buffer. When this buffer is too large it will effect the system performance and when this buffer is too small the effect on the probability of overtime is unnoticeable.

Chapter 6

Staffing Improvements

In Section 5.4 we have seen that the network of treatment plans at the Pain Rehabilitation department is overloaded. In this chapter we will present a staffing rule which makes sure that waiting times and service time are small and utilization of the practitioners is high. We will use the analytical model from Section 5.3 to determine this staffing rule. With this model we can determine the mean amount of capacity needed each week for a given arrival rate λ. However, choosing capacity equal to demand results in a utilization of 100% for all disciplines and the queueing system will still be unstable. We must therefore create a buffer such that the queueing system is stable and utilization is high. When we are able to determine this buffer, we are also able to give a staffing rule which satisfies small waiting and service times and high utilization.

6.1 Model formulation

In this subsection we will present our model in determining the correct buffer. We will do this by means of the square root staffing rule as introduced in [9]. This staffing rule is normally used in an $M/M/s$ heavy-traffic model to determine the number of servers needed to result in a certain probability of delay:

$$s = \frac{\lambda}{\mu} + \beta\sqrt{\frac{\lambda}{\mu}},$$

in which s is the number of servers needed, λ/μ is the offered load and where β determines the size of the buffer. This formula is designed such that the utilization approaches 100% when $\lambda \to \infty$. We will use this formula to determine a staffing rule for the network of treatment plans at The Pain Rehabilitation department. In order to use this square root staffing rule we must first define how we model the network of treatment plans.

We use the model formulation stated in Section 5.1 where we model the network of treatment plans as a feedforward queueing system. We use the analytical model of Section 5.3 to determine the mean and the variance of the amount of capacity needed in the queueing system each week. In this model we summarise the entire network of treatment plans into one single queue. We assume that the amount of work that arrives to this queue every week follows a Normal distribution. This amount of work is served by a number of deterministic servers, the practitioners. With this staffing rule we want to determine the amount of practitioners (or FTE's) needed for a stable queueing system.

The assumption of a Normal ditribution can be justified by the fact that the amount of capacity needed is the sum of n stochastic variables which are either Multinomially distributed or Poisson distributed. When n is large enough this sum is Normally distributed.

For each discipline of practitioners let $Y_k(t)$ denote the amount of work in the queue in week t for a given arrival rate of k patients per week and let $C_k(t)$ denote the amount of work arriving to the queue in week t for an arrival rate of k patients per week (in FTE's). Finally let s denote the amount of available capacity in FTE's for the discipline in question. By Lindley's equation we can now define:

$$Y_k(t+1) = \{Y_k(t) + C_k(t) - s\}^+, \quad s \geq 0.$$

When we now let $t \to \infty$ we get

$$Y_k \stackrel{d}{=} \{Y_k + C_k - s\}^+.$$

And by applying the square root staffing rule for a normally distributed system

$$s = E(C_k) + \beta\sqrt{Var(C_k)},$$

we find

$$\begin{aligned} Y_k &\stackrel{d}{=} \left\{Y_k + C_k - (E(C_k) + \beta\sqrt{Var(C_k)})\right\}^+ \\ \frac{1}{Var(C_k)}Y_k &\stackrel{d}{=} \left\{\frac{1}{Var(C_k)}Y_k + \frac{C_k - E(C_k)}{Var(C_k)} - \beta\right\}^+. \end{aligned}$$

Now if we define

$$Y^* = \lim_{k\to\infty} \frac{Y_k}{\sqrt{Var(C_k)}},$$

we find

$$\begin{aligned} Y^* &\stackrel{d}{=} \{Y^* + N(-\beta, 1)\}^+ \\ &\stackrel{d}{=} \max\{0, X_1, X_1 + X_2, \ldots\} \\ &= M_\beta, \end{aligned}$$

where X_1, X_2, ... are independent and $N(-\beta, 1)$-distributed random variables. By [13], we have

$$P(Y_k = 0) = P(\text{no waiting time}) \approx P(M_\beta = 0),$$

and with the help of [12, 13, 3] we find

$$P(M_\beta = 0) = \sqrt{2}\beta e^Z,$$

with

$$Z = \frac{\beta}{\sqrt{2\pi}} \sum_{r=0}^{\infty} \frac{\zeta(\frac{1}{2} - r)}{r!(2r+1)} \left(\frac{-\beta^2}{2}\right)^r,$$

where $\zeta(x)$ is the Riemann zeta function.

We use $P(M_\beta = 0)$ to determine a buffer which is fair for each discipline because in this way, each discipline experiences the same probability of delay. This choice of buffer results in a β which is equal for each discipline but also results in a utilization which is unique for each discipline. In the next section we will give two staffing rules, one with equal β's and one with equal utilizations and we will compare these by using the simulation model stated in Chapter 5. We will also determine the arrival rate which the system can handle by rearranging the current available capacity in the system. Finally, we will show the staffing rule for the actual arrival rate to the system.

6.2 Results

In this section we will present the results found with the staffing model. With this model we performed four different experiments. In the next subsection we will show an interesting result about the dependence between practitioners from different disciplines. In Section 6.2.2 we will compare two different staffing rules: the staffing rule as explained in Section 6.1 where we focus on equal β and a staffing rule in which we focus on equal utilization of the disciplines. We will show that we can handle more patients per week by rearranging the current available capacity in Section 6.2.3. We will close this section by giving a staffing rule for the current arrival rate of 17.7 patients per week in Section 6.2.4.

6.2.1 Dependence between disciplines

In this subsection we will show the effect of dependency between disciplines on the staffing rule. The staffing rule which we defined in the previous section is originally designed for $M/M/s$-queues in which an arriving patient requires one task which can be completed by one server. In the network of treatment plans we find that each patient requires an amount of work which is spread out over several weeks and this work must be completed by multiple disciplines each week. Because a patient is treated by multiple disciplines we will see that the disciplines are correlated.

In the following results we show two distinct cases, each with arrival rate $\lambda = 7.5$. In the first case we will set the utilization of one discipline equal to 95% whereas all other disciplines will have infinite capacity. By doing so we will look at each discipline in isolation. This results in six staffing rules which we will test by using the simulation model.

In the second case we will set the utilization of each discipline equal to 95% and we look at the entire system. This scenario is also tested with the simulation model from Chapter 5. In addition, we will also perform a simulation run where we set the utilization of each discipline equal to 90%.

The staffing rules according to the 95% utilization and the 90% utilization are shown in Table 6.1. In this table we denote the overall system utilization by ρ. We will use the staffing rule for $\rho = 0.95$ also for the six staffing rules where we will observer the disciplines in isolation. For example, when we observe the FT practitioner in isolation we will set the staffing for FT equal to 6.78 and we set the staffing for all other disciplines equal to a large number M.

The results found by the simulation model for the first case (disciplines in isolation) are shown in Table 6.2 and the results for the second case are shown

	E(C)	Staffing Rule for $\rho = 0.95$	Staffing Rule for $\rho = 0.90$
BA	1.13	1.19	1.25
ET	3.33	3.50	3.70
FT	6.45	6.78	7.16
MW	2.77	2.91	3.07
PS	2.99	3.14	3.32
RA	1.06	1.12	1.18

Table 6.1: Staffing rules in FTE's for $\lambda = 7.5$.

in Table 6.3.

	Mean # Arrivals	Mean # Departures
BA	7.5	7.5
ET	7.4	7.4
FT	7.4	7.4
MW	7.3	7.3
PS	7.5	7.5
RA	7.6	7.6

Table 6.2: Results for disciplines in isolation with 95% utilization.

	Mean # Arrivals	Mean # Departures
95% utilization	7.6	6.8
90% utilization	7.6	7.4

Table 6.3: Results for entire system.

In Table 6.2 we can see that a utilization of 95% for each discipline in isolation results in a stable queueing system whereas in Table 6.3 we can see that setting the utilization of each discipline equal to 95% will result in instability of the queueing system. Even when we set the utilization equal to 90% we can see that the system is unstable.

If the discplines were independent of each other we would find a stable system in both cases. Since this is not the case there must be some dependence between the disciplines. Because of this dependence we cannot achieve a utilization close to 100% for $\lambda = 7.4$.

6.2.2 Comparing equal β's with equal utilizations

In this subsection we will test if we can find a better staffing rule than the staffing rule proposed in this chapter by choosing a different strategy. In this chapter we chose to set all β's equal per discipline and we want to compare this with the strategy in which we choose equal utilizations of all disciplines. This utilization is chosen such that the analytically determined utilization of the overall system is equal in both cases. These utilizations are shown in Table 6.4 where we give the overall system utilization for a given β. In this table we denote the utilization by ρ. We can see that ρ decreases as β increases. This is due to the staffing rule in which the size of the buffer is determined by β. If we increase β we will increase the size of the buffer. This will then result in a decrease in ρ since the mean amount of capacity needed will remain the same.

In Table 6.4 we find for $\beta = 1.75$ that $\rho = 0.85$ which results in two different staffing rules which we want to compare. In the first staffing rule we will set β for each discipline equal to 1.75 and in the second staffing rule we will set the utilization of each discipline equal to 85%.

β	ρ
1.000	0.91
1.416	0.875
1.750	0.85
2.500	0.80

Table 6.4: Analytical system utilization ρ determined for given β and $\lambda = 7.5$.

In this subsection we test four different values of β and four values of ρ. The first results are shown in Figure 6.5 and Table 6.6 in which we test the eight simulations for stability. The staffing rule for each simulation is stated in Appendix F to preserve clarity in this section.

	$\beta = 1$	$\beta = 1.416$	$\beta = 1.75$	$\beta = 2.5$
Mean # Arrivals	7.6	7.5	7.4	7.6
Mean # Departures	7.1	7.5	7.4	7.6

Table 6.5: Stability for different β's.

	$\rho = 0.91$	$\rho = 0.875$	$\rho = 0.85$	$\rho = 0.80$
Mean # Arrivals	7.5	7.5	7.4	7.5
Mean # Departures	7.4	7.4	7.4	7.5

Table 6.6: Stability for different ρ's.

In these tables we can see that the system is stable for $\beta = 1.416$ where all β's are equal whereas the system is unstable if we set all utilizations equal to 87.5%. In both cases we find an overal system utilization of 87.5%. That means that in a system which is close to being unstable the staffing rule with all β's equal performs better than the staffing rule where all $\rho's$ are equal. In Table 6.7 we state the results of the four tests where both staffing rules resulted in a stable system. In this table we abbreviated Waiting Time to WT for graphical reasons.

We can see that the system behaves similar for the two staffing rules. We can see that the waiting times and service time are a little longer for $\rho = 0.85$ but the overall utilization is higher than for $\beta = 1.75$. This can also be said of $\beta = 2.50$, here we see again that the waiting times and service time are a little longer but the utilization is also increased compared to $\rho = 0.80$. By comparing these results to the results of the basic model in Table 5.6 we can see that by using a staffing rule we can increase the system utilization from 67% to 84%. We can also see that the waiting times and the service time are similar to those in the basic model in Table 5.6.

By looking at the staffing rules for $\beta = 1.75$ and $\beta = 2.50$ in Appendix F

	$\beta = 1.75$	$\rho = 0.85$	$\beta = 2.50$	$\rho = 0.80$
WT for intake	1.4 weeks	1.4 weeks	1.4 weeks	1.4 weeks
WT for first treatment	5.4 weeks	5.7 weeks	4.8 weeks	4.9 weeks
WT after observation	5.1 weeks	5.3 weeks	4.5 weeks	4.4 weeks
Total service time	27.3 weeks	27.6 weeks	27.2 weeks	26.9 weeks
Utilization BA	0.80	0.84	0.76	0.81
Utilization ET	0.85	0.86	0.83	0.79
Utilization FT	0.83	0.87	0.82	0.79
Utilization MW	0.84	0.86	0.83	0.80
Utilization PS	0.83	0.84	0.80	0.81
Utilization RA	0.83	0.84	0.81	0.80
System Utilization	0.84	0.86	0.81	0.80
Prob. of overtime	0.02	0.03	0.02	0.02

Table 6.7: Results for different β and ρ.

we can see that we can treat the same amount of patients as in the queueing system from Chapter 5, where $\lambda = 7.5$, but with 22% less FTE's.

In the next two subsections we will use the staffing rule as introduced in this chapter by using equal β for all disciplines.

6.2.3 Staffing by rearranging current capacity

By using the staffing rule we can rearrange the current capacity available at the Pain Rehabilitation such that the system can cope with a higher arrival rate. We will find this λ by applying the staffing rule and the analytical model from Chapter 5 until the staffing sums up to the current available capacity of 26.73 FTE's. The staffing rules and results from the simulation model are shown in Table 6.8 and Table 6.9 for $\beta = 1.75$ and $\beta = 2.50$. In these tables we have included the results of the queueing system from Chapter 5 where $\lambda = 7.5$. The results of this system are shown in the last column and is denote by "Chaper 5 system"

	$\beta = 1.75$	$\beta = 2.50$	Chaper 5 system
λ	9.800	9.217	7.5
BA	1.78	1.82	2.58
ET	4.95	4.92	3.55
FT	9.69	9.68	7.38
MW	4.14	4.12	4.69
PS	4.56	4.58	4.53
RA	1.61	1.61	4.00
Total	26.73	26.73	26.73

Table 6.8: Staffing rule for rearranged capacity.

In Table 6.9 we can see that by using the staffing rules from Table 6.8 the queueing system is able to cope with a higher arrival rate than without the staffing rule. We can also see that by using the staffing rule we can increase the system utilization to 85% and at the same time decrease the waiting times

	$\beta = 1.75$	$\beta = 2.5$	Chaper 5 system
λ	9.800	9.217	7.5
WT for Intake	1.3 weeks	1.3 weeks	1.4 weeks
WT for first treatment	3.9 weeks	3.7 weeks	4.9 weeks
WT after observation	3.5 weeks	3.2 weeks	4.6 weeks
Total service time	25.7 weeks	24.1 weeks	27.0 weeks
Utilization BA	0.82	0.79	0.44
Utilization ET	0.87	0.86	0.94
Utilization FT	0.86	0.85	0.87
Utilization MW	0.86	0.84	0.59
Utilization PS	0.84	0.81	0.66
Utilization RA	0.85	0.83	0.27
System Utilization	0.85	0.83	0.67
Mean # arrivals	9.6	9.3	7.5
Mean # departures	9.6	9.3	7.5

Table 6.9: Results for rearranged capacity.

and service time. This decrease in the waiting times and service time is due to the fact that more patients enter the system and groups are filled quicker which reduces the waiting times.

If the Pain Rehabilitation department would rearrange their available capacity we find that they could handle 30% more patients without negative effects on waiting times, service time or utilization.

6.2.4 Staffing rule for current arrival rate

We close this section by giving the staffing rule for the actual arrival rate of 17.7. We show the staffing rule for $\beta = 1.75$ and $\beta = 2.50$ in Table 6.10 and the results from the simulation model in Table 6.11.

	$\beta = 1.75$	$\beta = 2.50$
BA	3.08	3.26
ET	8.66	9.00
FT	16.91	17.64
MW	7.23	7.53
PS	7.93	8.31
RA	2.81	2.94
Total	46.62	48.68

Table 6.10: Staffing Rules for $\lambda = 17.7$.

With these results we can see that we can create a stable system with the actual arrival rate by increasing the available capacity. When $\beta = 2.5$ we see that waiting times and service time are very low and that the overall utilization is 85%. However, when we choose $\beta = 1.75$ we see that waiting times and service time are still very short and that the overall utilization is equal to 90%. But most important, we see that the staffing rule for $\beta = 1.75$ needs 2 FTE's less than the staffing rule for $\beta = 2.5$.

	$\beta = 1.75$	$\beta = 2.5$
WT for Intake	1.2 weeks	1.1 weeks
WT for first treatment	2.4 weeks	2.0 weeks
WT after observation	2.7 weeks	1.6 weeks
Total service time	24.4 weeks	24.0 weeks
Utilization BA	0.87	0.80
Utilization ET	0.90	0.87
Utilization FT	0.90	0.86
Utilization MW	0.90	0.86
Utilization PS	0.89	0.83
Utilization RA	0.90	0.84
System Utilization	0.90	0.85
Mean # arrivals	17.8	17.4
Mean # departures	17.8	17.4

Table 6.11: Results for $\lambda = 17.7$.

In Table 6.11 we can see that we can achieve an overall utilization of 90% for $\lambda = 17.7$. However, in Section 6.2.1 we have seen that for $\lambda = 7.5$ the queueing system was unstable at an overall utilization of 90%. This is due to the square root staffing rule. This rule is designed so that the utilization approaches 100% when $\lambda \to \infty$. So for $\lambda = 17.7$ we are able to achieve a higher overall system utilization than for $\lambda = 7.5$.

In the stable queueing system used in Chapter 5 we found that with a total of 26.73 FTE's we are able to handle 7.5 new patients each week. By using the staffing rule from this subsection with $\beta = 1.75$ we can see that we can handle 17.7 new patients each week with 46.62 FTE's. That means that we can handle 136% more new patients each week with only 74% more FTE's.

6.3 Conclusions

In this chapter we introduced a staffing rule for the Pain Rehabilitation department at Het Roessingh. With this staffing rule we are able to increase utilization and stabilize the network. By looking at the results from the previous section we can conclude:

- There is a dependency between the disciplines. For $\lambda = 7.5$ we are able to increase the utilization for each discipline in isolation to a value of 95%, whereas the overall utilization of 90% resulted in instability.

- In a stable system we see that the staffing rule where all β's are equal and the staffing rule where all ρ's are equal perform in a similar way. We see a difference between the two rules when we are dealing with a system which is almost unstable. In this case the staffing rule where all β's are equal performs best.

- By rearranging the amount of capacity which is currently available we are able to give a staffing rule in which 30% more patients can be treated.

- A high performance can be achieved for the current arrival rate of 17.7 patients per week with 74% more FTE's. With this increase we are able to handle 136% more patients each week compared with the stable queueing system from Chapter 5 where the arrival rate was 7.5 patients per week.

Chapter 7

Conclusions and Recommendations

Throughout this thesis we made several conclusion which we will repeat and summarise in this chapter. We will start by restating the conclusion from each chapter of this thesis and will close this chapter by presenting our recommendations for further research.

7.1 Conclusions

In this section we will repeat the conclusions from each chapter.

Cancelled meetings for RDB-patients In Chapter 3 we presented a model which we used to analyse the effects of cancelled meetings on the patient flows. We have seen that the mean lead time of a patient grows rapidly when the probability of a cancelled meeting increases. It is therefore crucial that the attendence at these meetings is as high as possible.

Intakes at the Pain Rehabilitation department In Chapter 4 we have seen that the current situation of 15 intakes per week in not enough for the number of patients which apply for an intake each week. The number of intakes each week must be increased to a minimum of 17 to avoid a critical load of the intake procedure.

Patient Flow Analysis In Chapter 5 we found that the Pain Rehabilitation department is indeed overloaded. The feedforward queueing system which we used to model the department was able to handle an arrival rate of 7.5 patients per week while the actual arrival rate is 17.7 patients per week. In this chapter we found that the ET practitioners appear to be a bottleneck for the entire queueing system. With the simulation model we tested four aspects in the queueing system and we can conclude that:

- The cancelled meetings are a big problem for the queueing system. With the simulation model we found that these cancelled meetings van turn a stable system with an arrival rate of 7.5 patients per week into an unstable system.

- Het Roessingh should stop with partial planning. By using the simulation model we can see that partial planning does not improve the system performance.
- By using a decency rule we can decrease the waiting time between observation and treatment. Both priority queueing and partial planning resulted in a decrease in the waiting time.
- The use of an artificial buffer can decrease the probability that an treatment plan will take longer than predicted. This buffer must be chosen in a correct way so that it will not affect the regular patient flows.

Staffing Improvements In Chapter 6 we introduced a staffing rule which results in a reasonably high working load for the practitioners and an efficient treatment for all patients. By using this staffing rule on the queueing system from Chapter 5, where the arrival rate was 7.5 per week, we are able to:

- Treat the same amount of patients with 22% less FTE's.
- Treat 30% more patients by rearranging the current amount of capacity.
- Treat 136% more patients, resulting in an arrival rate of 17.7 patients per week, with only 74% more FTE's.

7.2 Recommendations for further research

In this section we will present our recommendations for further research.

- The fundamentals of this thesis are the treatment plans and the transition matrix. It is therefore important that this data is kept up to date so that the models and conclusions presented in this thesis can still be used in the future.
- In Appendix D we can see that the Intake has highest priority. It is interesting to see what will happen if we give the Intake a lower priority.
- Some treatment plans are designed for a clinical patients. These patients will be treated for 10 weeks at the clinic (one week at the clinic, one week at home et cetera) followed by a period of 10 weeks in which the patient is at home. After this period the patient return to the clinic for 1 final week of treatment. In Figure 7.1 we show the treatment of four patients in parallel as a block of 32 weeks. Each week will be colored to denote which patient is treated in this week. We will leave the week blanc if no patient is treated. We can see in this figure that we will end up in 8 weeks of treatment which we cannot assign to a new patient since we are working with Complete Planning. It is useful to analyse the effects of this process on the patient flows in order to improve efficiency.

- The coinfidence interval in the simulation model is large due to computational limitations. It is therefore important that better software is used to improve the simulation model.

- In Chapter 5 we grouped all practitioners and started analysing the system per discipline. In this process the uniqueness of each practitioner is lost and patients preferences to be treated by the same practitioner each week cannot be satisfied. By looking at each practitioner instead of each discipline we are able to present more accurate results.

- In Chapter 5 we looked at the average waiting time and the average service time of a patient. It is interesting to see how these waiting times and service time are distributed.

Figure 7.1: Parallel treatment of four clinical patients

Appendix A

Abbreviations

In this chapter we will explain all abbreviations used in this thesis.

Abbreviation	Dutch meaning	English meaning
RDB-patient	Roessingh DagBehandeling	An individual outpatient
RVE	Resultaat Verantwoordelijke Eenheid	Part of the centre is in charge of a particular part of treatments
BA	Bewegingsagoog	Kinesiologist
ET	Ergotherapeut	Occupational therapist
FT	Fysiotherapeut	Physiotherapist
MW	Maatschappelijk Werker	Social Worker
PS	Psycholoog	Psychologist
RA	Revalidatiearts	Rehabilitation Specialist
EPD	Electronisch Patienten Dossier	Digital database of medical records
DBC	DiagnoseBehandelCombinatie	System which gives more clarity in the costs of treatment

Appendix B

Transition Matrix

In this chapter we will present the transition matrix used in our calculations. In this matrix we can find that the transition rate from the source node to the intake is 0.939 and combining this with an arrival rate of 17.7 we find that on average 16.6 patients apply for an intake each week.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0 - Source	0	0.939	0.011	0	0	0	0	0	0	0	0	0	0	0	0.030	0.020	0	0	0	1
1 - Intake	0	0	0	0.082	0.064	0	0.041	0.062	0.103	0.085	0.153	0.102	0.041	0.028	0	0	0.046	0.008	0.001	0.184
2 - NBT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3 - Obs Ach	0	0	0	0	0	0.280	0.080	0.020	0.100	0.025	0.045	0.030	0	0.010	0	0	0	0.050	0	0.360
4 - Obs Voo	0	0	0	0	0	0.030	0.080	0.030	0.210	0.055	0.099	0.066	0.010	0	0	0	0	0.090	0	0.330
5 - ZP 20A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
6 - ZP 20B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
7 - ZP 20C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8 - ZP 20D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9 - ZP 20E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10 - ZP 20E+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11 - ZP 20E++	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12 - ZP 20F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13 - ZP 20G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14 - ZP 20H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15 - ZP 20I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16 - ZP 20J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
17 - ZP 20K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18 - ZP 20L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
19 - Sink	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure B.1: Transition Matrix for treatment plans

Appendix C

Treatment Plan

In this chapter we will show an example of a treatment plan. In these treatment plans we can find all the information we need to analyse the patient flows. It contains information about the number of appointments each week per treatment, the duration of these appointments and which practitioner is needed. It also shows which treatment are definitely needed for a patient, Prob=1, and which treatment are only needed for a fraction of the patients, Prob¡1. In these treatment plans we can also find if a treatment is given in a group or individually. We can also find the number of beds needed for this treatment plan. In this example we see that we need a bed for half a week, which corresponds to a semi-clinical treatment.

					Frequency per Week				
Practitioner	Treatment (in Dutch)	Duration (in min)	Groupsize	Prob.	1	2	3	4	5
BA	Fitness	30	3	1	1	2	1	0	0
	Zwemmen	30	3	1	1	1	1	0	0
	Sport/Spel	30	1	0.5	1	1	1	0	0
	Overleg**	15	1	1	1	1	1	1	0
ET	ET	60	1	1	1	1	1	0	0
	Overleg**	15	1	1	1	1	1	1	0
FT	FT	60	1	1	1	1	1	0	0
	Overleg**	15	1	1	1	1	1	1	0
GL	Mentor	60	1	1	1	1	1	0	0
	Overleg**	15	1	1	1	1	1	1	0
MW	MW	60	1	1	1	1	1	0	0
	Partnergesprek	60	1	1	1	0	0	0	0
	Overleg**	15	1	1	1	1	1	1	0
PS	PS	60	1	0.75	1	3	2	0	0
	PDW	60	1	1	1	1	0	0	0
	Overleg**	15	1	1	1	1	1	1	0
	Uitslaggesprek*	30	1	1	0	0	0	0	1
RA	RA	20	1	1	1	1	1	0	0
	Overleg**	15	1	1	1	1	1	1	0
	Uitslaggesprek*	30	1	1	0	0	0	0	1
Bed	Bed	-----	1	1	0.5	0.5	0.5	0	0

Figure C.1: Example of a Treatment Plan

Appendix D

Priority Table

In this chapter we present the priority table used by the simulation model. In this table we can find that Intake has highest priority and that treatment plan ZP-20E++ has lowest priority.

Treatment Plan	Priority
1 Intake	1
2 NBT	15
3 Obs Ach	3
4 Obs Voo	4
5 ZP-20A	2
6 ZP-20B	5
7 ZP-20C	6
8 ZP-20D	7
9 ZP-20E	16
10 ZP-20E+	17
11 ZP-20E++	18
12 ZP-20F	8
13 ZP-20G	9
14 ZP-20H	10
15 ZP-20I	11
16 ZP-20J	12
17 ZP-20K	13
18 ZP-20L	14

Table D.1: Priority Table

Appendix E

Group Constraints

In this chapter we will present the group constraints used in the simulation model. In this table we can find the groupsize and the amount of groups which can be treated at the same time.

Treatment Plan	Groupsize	# Groups allowed
1 Intake	5	3
2 NBT	1	∞
3 Obs Ach	2	1
4 Obs Voo	3	1
5 ZP-20A	1	1
6 ZP-20B	7	1
7 ZP-20C	6	1
8 ZP-20D	6	3
9 ZP-20E	1	∞
10 ZP-20E+	1	∞
11 ZP-20E++	1	∞
12 ZP-20F	6	∞
13 ZP-20G	6	∞
14 ZP-20H	9	∞
15 ZP-20I	6	∞
16 ZP-20J	7	∞
17 ZP-20K	6	1
18 ZP-20L	6	1

Table E.1: Group Constraints

Appendix F

Staffing rules used in Section 6.2.2 in FTE's

In this chapter we will present the staffing rules used in Section 6.2.2. In total we used eight different staffing rules where four are based on the principle of equal β's for all disciplines and the other four are based on the principle of equal utilization or ρ's for each disciline.

	$\beta = 1$	$\beta = 1.416$	$\beta = 1.75$	$\beta = 2.5$
BA	1.28	1.35	1.40	1.51
ET	3.63	3.75	3.85	4.08
FT	7.08	7.34	7.55	8.03
MW	3.03	3.14	3.22	3.42
PS	3.31	3.45	3.56	3.81
RA	1.17	1.22	1.26	1.34
Total	19.50	20.25	20.84	22.19

Table F.1: Staffing Rule for different β's.

	$\rho = 0.91$	$\rho = 0.875$	$\rho = 0.85$	$\rho = 0.80$
BA	1.24	1.29	1.33	1.41
ET	3.66	3.80	3.92	4.17
FT	7.09	7.37	7.57	8.07
MW	3.05	3.16	3.26	3.47
PS	3.29	3.41	3.51	3.74
RA	1.17	1.22	1.25	1.33
Total	19.50	20.25	20.84	22.19

Table F.2: Staffing Rule for different ρ's.

Bibliography

[1] S. Borst, A. Mandelbaum, and M.I. Reiman. Dimensioning large call centers. *Operations Research*, 52(1):17–34, 2004.

[2] A. Braaksma. Integrale planning van multidisciplinaire behandeltrajecten. Master's thesis, Universiteit Twente, Enschede, 2010.

[3] J.T. Chang and Y. Peres. Ladder heights, Gaussian random walks and the Riemann zeta function. *The Annals of Probability*, 25(2):787–802, 1997.

[4] C.F. Chien, Y.C. Huang, and C.H. Hu. A hybrid approach of data mining and genetic algorithms for rehabilitation scheduling. *International Journal of Manufacturing Technology and Management*, 16(1-2):76–100, 2009.

[5] C.F. Chien, F.P. Tseng, and C.H. Chen. An evolutionary approach to rehabilitation patient scheduling: A case study. *European Journal of Operational Research*, 189(3):1234–1253, 2008.

[6] J.G. Dai and S. He. Customer abandonment in many-server queues. *Mathematics of Operations Research*, 35(2):347–362, 2010.

[7] Z. Feldman and A. Mandelbaum. Staffing of time-varying queues to achieve time-stable performance. *Management Science*, 54(2):324–338, 2008.

[8] I. Gurvich, M. Armony, and A. Mandelbaum. Service-level differentiation in call centers with fully flexible servers. *Management Science*, 54(2):279–294, 2008.

[9] S. Halfin and W. Whitt. Heavy-traffic limits for queues with many exponential servers. *Operations Research*, 29(3):567–588, 1981.

[10] J.M. Harrison and M.J. López. Heavy traffic resource pooling in parallel-server systems. *Queueing Systems*, 33:339–368, 1999.

[11] W. Hartholt. Patienten en zorgpaden - De overeenkomsten en de verschillen. Internship Report, Universiteit Twente, Enschede, 2009.

[12] A.J.E.M. Janssen and J.S.H. van Leeuwaarden. On Lerch's transcendent and the Gaussian random walk. *The Annals of Applied Probability*, 17(2):421–439, 2007.

[13] A.J.E.M. Janssen and J.S.H. van Leeuwaarden. Back to the roots of the M/D/s queue and the works of Erlang, Crommelin and Pollaczek. *Statistica Neerlandica*, 62(3):299–313, 2008.

[14] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, and B. Zwart. Refining square root safety staffing by expanding Erland C. *Operations Research*, 2010. In press.

[15] L. Kok, A. Houkens, and N. Niessen. Kosten en baten van revalidatie. http://www.revalidatie.nl/resultaat/kosten-en-baten.

[16] D.V. Lindley. The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(2):277–289, 1952.

[17] J. D. C. Little. A proof for the queuing formula: $L = \lambda W$. *Operations Research*, 9(3):383–387, 1961.

[18] Revalidatie Nederland. Jaarbeeld 2008. http://www.revalidatienederland.nl/actueel/publicaties.

[19] Revalidatie Nederland. Jaarbeeld 2009. http://www.revalidatienederland.nl/actueel/publicaties.

[20] E.L. Ormeci. Dynamic admission control in a call center with one shared and two dedicated service facilities. Unpublished.

[21] M. Perry and A. Nilsson. Performance modeling of automatic call distributors: Assignable grade of service staffing. *14. International Switching Symposium*, 2:294–298, 1992.

[22] S. Robinson. *Simulation: The Practice of Model Development and Use*. John Wiley and Sons, Ltd, 2004.

[23] S. M. Ross. *Stochastic Processes*. John Wiley and Sons, Inc, 1996.

[24] W. Sermeus and K. Vanhaecht. Wat zijn klinische paden. *Acta Hospitalia*, 3, 2002.

[25] R.A. Shumsky. Approximation and analysis of a call center with flexible and specialized servers. *OR Spectrum*, 26:307–330, 2004.

[26] D. A. Stanford and W. K. Grassmann. Bilingual server call centres. *Analysis of Communication Networks: Call Centres, Traffic and Performance*, 28, 2000.

[27] H. C. Thijms. *A first course in stochastic models*. John Wiley and Sons Ltd, 2003.

[28] K. Vanhaecht and W. Sermeus. Draaiboek voor de ontwikkeling, implementatie en evaluatie van een klinisch pad. 30 stappenplan van het Netwerk Klinische Paden. *Acta Hospitalia*, 3, 2002.

[29] R. Yedehalli Venkataramu. A revenue management framework for appointment allocation in rehabilitation outpatient clinics. Master's thesis, Wichita State University, Wichita, 2005.

[30] W. Whitt. Staffing a call center with uncertain arrival rate and absenteeism. *Mathematics of Operations Research*, 15(1):88–102, 2006.

[31] W. L. Winston. *Operations Research: Applications and Algorithms*. Brooks/Cole, Belmont, USA, 2004.