Aalto University
School of Science
Master's Programme in Machine Learning and Data Mining

Jonathan P. F. Strahl

Patient appointment scheduling system:
with supervised learning prediction

Master's Thesis
Espoo, May 27, 2015

Supervisor: Professor Samuel Kaski, Aalto University
Advisors: Paul Wagner, M.Sc. (Tech.)
Marcus Dobrinkat, M.Sc

Aalto University
School of Science
Master's Programme in Machine Learning and Data Mining

ABSTRACT OF MASTER'S THESIS

Author:	Jonathan P. F. Strahl		
Title:	Patient appointment scheduling system: with supervised learning prediction		
Date:	May 27, 2015	**Pages:**	x + 126
Major:	Information and Computer Science	**Code:**	T-61
Supervisor:	Professor Samuel Kaski		
Advisors:	Paul Wagner, M.Sc. (Tech.) Marcus Dobrinkat, M.Sc		

Large waiting times at hospital outpatient clinics are a cause of dissatisfaction to patients, cause additional stress to hospital staff, increase the risk of contagion and add complications for patients with medical conditions. Reducing waiting times and surgeon idle time improves the quality of service and efficiency of a hospital: this is a recently growing focus in healthcare.

Oulu hospital[a] wants to identify and reduce large waiting times at their outpatient clinic. For the past few years the clinic has used a self-service system[b] whereby patients register on arrival and hospital staff use a patient call-in system. The past schedules are analysed using this data: information visualisations and performance measures are provided. The worst performing clinic sessions are the subject of the scheduling optimisation prototype system.

The scheduling optimisation focuses on predicting the duration of an appointment and the late arrival of the surgeon. These two factors have been identified as causes of long patient waiting times. The variance of the duration is identified to be high, therefore supervised-learning regression is used for both simple inference and prediction. The features that are good predictors and the results of the prediction accuracy are reported.

With the predicted appointment durations, and surgeon arrival times, a scheduling optimisation approach is used to improve the existing schedule; a simple greedy hill-climbing approach is evaluated.

It is found that using the historical data to simulate a real day, appointment rules and scheduling optimisation the patient waiting time is reduced with this method. Showing the system to be potentially promising.

[a] Oulun yliopistollisen sairaalan yhteystiedot
[b] X-akseli Oy self-service system

Keywords: Appointment scheduling, outpatient scheduling case study, stochastic scheduling, supervised learning, service time prediction, local search scheduling, static scheduling, offline scheduling, appointment scheduling with simulation

Language: English

Acknowledgements

Firstly let me say how grateful I am to be afforded this opportunity to apply an array of my interests and passions to potentially improve the quality of healthcare in Finland, if only a small bit, for now! For this I must thank Janne Argillander, Marcus Dobrinkat and the X-akseli team. I am fortunate that X-akseli Oy is a company that believes in developing through research: leading to the well-thought-out early design of their system to capture the data necessary for this thesis, and providing funding, support and guidance throughout this thesis work.

Additionally, without the collaboration of Oulu university hospital, and in particular without the many questions answered by Sanna Yrjänä, this thesis would not have been possible. Thank you for your time and patience.

With my first endeavour into scientific research Paul Wagner and Timo Honkela have supported and guided me. Timo's passion, vast knowledge of the field and ingenious visions in computational cognitive science opened my mind and confused me at times: what is research if you know what you are doing all the time? Paul Wagner has been a great mentor with his knowledge of reinforcement learning and experience of publishing scientific research. During our talks I often digressed onto all sorts of other topics and he often enlightened me. Thank you to you both and to the Department of Information and Computer science where I have learnt a lot. I extend my thanks to Samuel Kaski for his time to supervise this thesis.

Thank you Daniel Viera, for approaching me to join X-akseli, where I have taken on lots of exciting and rewarding projects: including this thesis.

Lastly, this work would not have happened without my wife Jessica. Thank you for introducing me to Finland, and thank you for understanding why I am a student again at my age.

Espoo, May 27, 2015

Jonathan P. F. Strahl

Symbols

$\neg$	logical negation
$\implies$	implies
α	scheduling machine setup
A	waiting time threshold
a_t	patient arrival time
β	scheduling restrictions and constraints
C	completion time
CS	cost
c_v	coefficient of variation
d	due date
e_t	appointment end time
γ	scheduling objective
L	lateness
O	overtime
p	process time (also service time or appointment duration)
r_t	patient reserved appointment time
s_t	appointment start time
S	machine (resource or surgeon) idle time
τ	transition state $\in \{true, false\}$
T	tardiness
U	weighted unit penalty
W	waiting time
w	cost function weight
ξ	placeholder for a statistical measure

Notation

a_j	arrival time of patient j
C_j	completion time of job (patient) j on last machine (of appointment)
d_j	due date of job j
$E()$	mean expected value
γ_{ws}	sum of weighted idle time
γ_{wt}	sum of weighted tardiness
L_j	lateness of job j
μ_{TE}	average transition-early time
μ_{TEW}	average transition-early waiting time
μ_{TLW}	average transition-late time
μ_{TLW}	average transition-late waiting time
μ_{ULW}	average unplanned-late waiting time
μ_{wt}	average weighted tardiness
$max\{\}$	sample maximum
p_j	process time of job (patient) j
S_j	machine (resource or patient) idle time for job (patient) j
T_j	tardiness for job (patient) j
U_j	weighted unit penalty for job (patient) j
U_{j_A}	weighted unit penalty for job (patient) j with threshold value A
W^{ul}	unplanned-late waiting time
W^{te}	transition-early waiting time
W^{tl}	transition-late waiting time
N_{T_A}	weighted sum of the number of jobs (patients) above the threshold time A

Abbreviations and Acronyms

ASP	appointment scheduling problem
BI	initial block: appointment rules
The clinic	Oulu university hospital neurological, orthopaedic and dermatological outpatient clinic
FR	first reservation: appointment rules
I	appointment interval: appointment rules
job	a patient to be seen
LQ	lower quartile
machine	a resource at a weekly session: a surgeon
nbr	number of jobs being scheduled
patient flow	the average time a patient is at the hospital, from arrival, to completing all reservations and leaving the hospital
prec	precedence constraint
rcrc	recirculation
RMSE	root mean square error
RSS	residual sum of squares
RSE	residual standard error
SE	standard error
TEW	transition-early waiting time
TLW	transition-late waiting time
surgeon	a neurological, orthopaedic or dermatological surgeon
UQ	upper quartile
ULW	unplanned-late waiting time
weekly session	the weekly usage of a clinic resource by a particular surgeon

Contents

Chapter 1

Introduction

1.1 Why reduce waiting times?

> I wasted time, and now doth time waste me.
>
> William Shakespeare, Richard II

With many healthcare-service options available, a recent need has grown to increase customer satisfaction at hospitals, according to a report in the United States [20]. One area where patients are often dissatisfied is in the time waiting to see a physician. The report states an average waiting time of 24 minutes, and shows how satisfaction declines as the waiting time increases, with 93.1% of satisfied patients when waiting time is under five minutes, dropping to 84.9% when waiting time is over ten minutes. Additional studies in the US have pointed out the need to improve efficiency in healthcare as there is a growing elderly population and hospital costs are rising [9]. Since 2000 the demand for outpatient services (ambulatory care services) has increased in Western countries [13, Sec. 3.1].

The national health service (NHS) in the UK wrote up a patient's charter to improve the quality of healthcare services [11]. It states that patients must be seen within thirty minutes of waiting, including waiting time between reservations (e.g. a diagnostic examination followed by a surgeon appointment): this is not by law but as a target for UK NHS hospitals. Additional to customer satisfaction, busy waiting rooms are a cause of stress for hospital staff.

In discussions with the hospital, a patient that had been followed through the clinic reported complications with medical conditions when waiting times were long; diabetic patients need to maintain insulin levels and patients with back problems found sitting for long periods very discomforting.

1.2 Previous studies on reducing waiting times

Appointment scheduling prior to the 1950s was focused on ensuring that the Doctor's wasted time was minimised with little regard for the amount of time the patients wait [1, 15]. A common practise was single-block scheduling, where all patients were booked in at the beginning of a session and served on a first-come, first-served basis. In his paper [1], Welch shows how patient waiting times can be significantly reduced without wasting the time of the Doctor. Using data acquired by a previous study and applying queuing theory analysis, he showed the relationship between patient waiting times and Doctor idle times for different scheduling approaches. Further studies on this relationship support his findings [13, pg. 136].

One relationship of interest is the effect of under-estimating and over-estimating the average duration length of the reservations (Fig. 1.1). Analysis of Oulu's outpatient clinic shows that the duration times vary a lot (see Section 3.4.2). This makes it very difficult to plan a good schedule, risking high waiting times or high idle times when under or over estimating the average appointment duration respectively [3]. The increased prediction accuracy of the appointment duration is in relationship with the improvement of the appointment schedule [15]. This study uses supervised-learning regression to make predictions more accurately than using the average duration time given the historical data available (see Section 4.5 for methods, Section 5.4 for implementation and Section 6.2 for results).

Another relationship of interest is how the waiting times change based on how many patients are booked in to the first appointment block (Fig. 1.2). From the analysis of Oulu's outpatient clinic it can be seen that the session will usually not begin until two or three patients have already arrived (see Section 3.4.5). By making a safe estimation of when the surgeon will arrive, the schedule can aim to start at the optimal time to reduce the patient waiting time while taking into consideration the risk of the surgeon being idle.

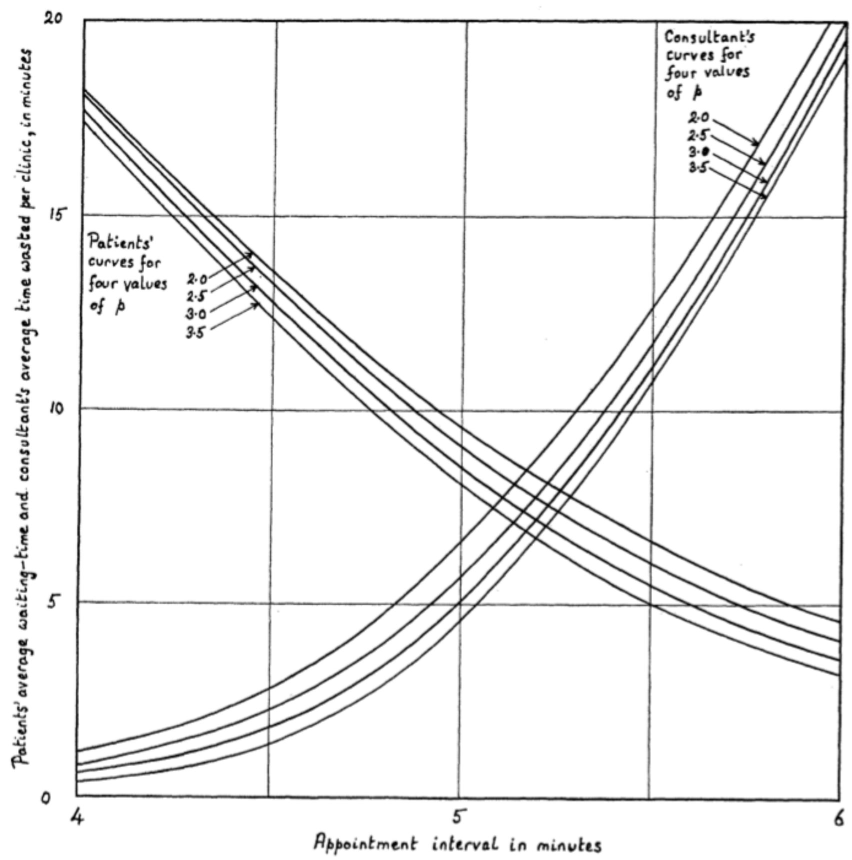


Figure 1.1: From Bailey [1]: Plot of the relationship between the scheduled interval time of an appointment, for a fixed-interval schedule, and the average waiting times for the patients and the resource consultant (surgeon). The study used Pearson Type III distributions to model the duration times of the consultations. The parameter p is a shape parameter that is changed to account for some overlap in the appointments where patients are seen simultaneously for a short period.

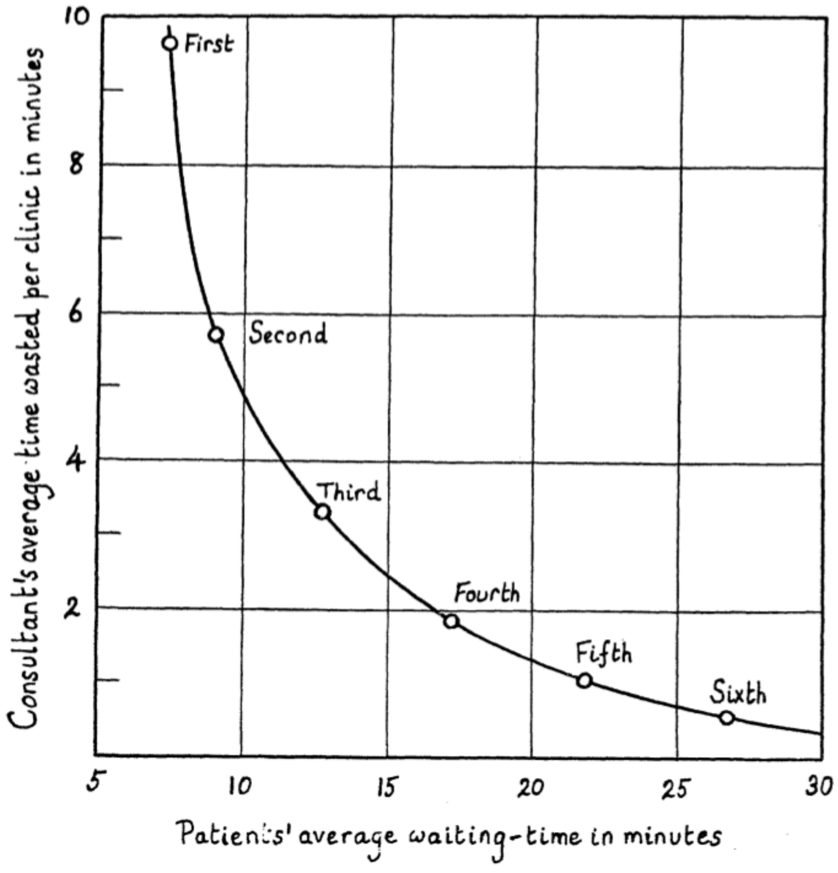


Figure 1.2: From Bailey [1]: Relation between average waiting times for the consultant (surgeon) versus the patient, when starting the session at the first, second, third, etc., patient.

1.3 Problem statement

The aim of this study is to reduce the patient waiting times at Oulu's outpatient clinic. The scheduling system receives an existing schedule for a future date. The job of the scheduler is to change the existing reservation times to reduce the amount of waiting time for the patients. This problem requires offline operational planning (with restrictions) [13]. It is worth noting that higher level, tactical and strategical planning can also have a large impact on the schedule.

The new schedule must not change: the start or end time if it causes the surgeon to arrive before the patients or if substantial overtime will be required, the number of patients to be seen, the availability of the surgeon. If a patient is arriving from a previous reservation the starting time of the appointment must aim to leave enough time for the patient to complete the previous reservation and travel to the outpatient clinic. Additional to this, if a patient's previous reservation was a diagnostic examination, enough time must aim to be given for the examination report to be completed. Once the session starts it is desirable that there is always a patient ready to be seen when the surgeon is available: surgeon idle time between appointments is undesirable.

1.4 The objectives of this thesis

This study first analyses the outpatient clinic to investigate the causes of waiting time and to rank the weekly sessions based on where the largest waiting times can be found. The surgeon unpunctuality and the high variance of the appointment duration are identified as large contributing factors and become the main focus of the optimisation. Trying to predict the duration of the appointment and when the surgeon will arrive is the first step in reducing the waiting times. After this the expected patient waiting time can be reduced while minimising the surgeon idle time during the active part of the session. The active part is from the start of the first until the end of the last patient appointment; the inactive part of the session is any time left at the end of the session after all patients have been seen.

existing schedule analysis

Analysis of Oulu's outpatient clinic is undertaken using the patient-flow data (Section 2.3.2.1): visualisations are produced, existing appointment schedule performance metrics are calculated and a report is generated for the hospital. The analysis highlights weekly sessions at the hospital where waiting times are worst by scoring them based on a set of performance measures. The weekly sessions with the poorest scores are subject to schedule optimisation.

waiting categories

A common mistake when measuring waiting times is that the time between a patient arriving and the patient being seen is measured as waiting time: this is inaccurate. Within the literature it is common to call waiting time any time past the reserved time [3]. How early a patient arrives is not within the direct control of the scheduler. It could be argued that a good schedule can indirectly improve patient punctuality through good reputation, by giving patients confidence in the schedule, but this is beyond the scope of this study. Even less obvious is that some patients that are seated in the waiting room may be there due to a previous reservation finishing much earlier than expected: this waiting is within the control of the schedule. Different waiting categories are defined and measured in Section 3.4.

appointment duration

One of the challenges when scheduling at the clinic is the large variance in appointment duration length. The duration of a patient's appointment is predicted as accurately as possible: using supervised learning regression. With less than a thousand samples, a dependent variable with high variance and with very few features available, the task is challenging. Discussions with the hospital contact, Sanna, to understand the clinic domain helped to extract features from the X-akseli data that could be influential in the duration of the surgery appointment: if a patient had a diagnostic scan in the morning, what scan they had, if they have had surgery in the recent past and if this is their first visit.

late start

late arrival

The analysis highlighted that surgeons often start late at the weekly sessions with the highest waiting times; **surgeon unpunctuality** is predicted with a simple linear regression model and the prediction interval is used to choose an estimated arrival with a high level of confidence. Occasionally a patient will not show for an appointment, or will arrive late; **patient punctuality** is modelled with a tall and narrow logistic distribution with high kurtosis and **no-shows** are so rare they are not modelled for this clinic.

A patient with an appointment sometimes has another reservation prior to their appointment; the **preceding reservations** are analysed and included in the reports and the transition times are modelled for the scheduling system.

initial approach

Initially the approach was to include all the weekly sessions in the clinic into the appointment scheduling problem (ASP) to create a new schedule for the entire clinic. A lot of the initial work using job shop scheduling approaches, with shifting bottleneck heuristics, disjunctive graphs and branch and bound techniques had to be discarded. Taking into account the late-starts, unpredictability of the service duration, patients revisiting the same surgeon on the same session (recirculation) and the surgeon availability constraints for each of the fifty or so weekly sessions became very difficult. It was agreed with the hospital that a simpler approach of focusing on one weekly session per optimisation would be a good start. This work could lead to a larger encompassing approach in the future.

scheduling task

Therefore, the ASP will comprise a single surgeon's weekly session. An additional request from the hospital for the first new schedules is not to change the original schedule too much; the new schedule will not change the sequence of the reservations. A single machine problem (single surgeon) with a penalty for distancing itself from the original schedule strongly lends itself to a local search optimisation approach. Therefore a simple greedy hill-climbing approach is tried, this is similar to a Tabu search approach that has been reported in the literature to have good results for similar job shop scheduling problems [17].

1.5 Structure of the thesis

Here in **Chapter 1**, the negative effects of long waiting times are explained with reference to published reports. This leads to referring to previous studies to show typical causes of a bad appointment schedule as areas to focus on for improvement. An overview of how these areas are tackled follows. The background chapter, **Chapter 2**, introduces the real-world elements and the data; the hospital outpatient clinic where the schedule is being optimised and the X-akseli self-service system from where the data is gathered are both described. The raw data available from the system and preprocessing of the data is briefly explained. Lastly, the current schedule in use at the clinic is described in the context of appointment scheduling. **Chapter 3** is an analysis of the clinic. This identifies where waiting times are a problem and explains the different measures of performance. The different waiting categories are introduced. The analysis also shows the high variation in the appointment durations, the regular late starts at the weekly sessions, the punctuality of the patients, the rarity of a patient no-show and the effects of having a same-day reservation prior to the appointment. Covered in **Chapter 4** is the theoretical background. Supervised-learning regression is explained, with a focus on linear regression and k-nearest neighbour. Methods of selecting the best features and assessing the models are also explained. Appointment rules and local search algorithms for scheduling optimisation are explained. In **Chapter 5** the system implementation is described: an overview of all the components and how they work together. The empirical evaluation is in **Chapter 6** where results of the appointment duration (service time) predictions and testing optimised schedules are provided. In the discussion chapter, **Chapter 7**, the accuracy of the appointment duration prediction, observations during the work regarding the literature, future improvements and the system design are discussed. Finally in **Chapter 8** the paper sums up with the conclusions: the successes of the appointment duration prediction accuracy, addressing the late-starts of the surgeons in the simulation and using the appointment rules with the local search algorithm are pointed out.

Chapter 2

Outpatient appointment scheduling

2.1 Terminology

There are two main categories of patients: inpatients and outpatients [27, Ch. 1] [9, pg. 804]. An **outpatient** is a short term patient that typically stays for less than a day, especially not overnight. Throughout the study, patient refers to outpatient. All patients at the clinic are **elective patients** [27, Sec. 1.2.1] [9, p. 804] [13, pg. 143]; all patients have reservations and there are no urgent or drop-in patients that can show up without prior notice [3, pgs. 521-522]. Each appointment room is a **resource** at the outpatient clinic that is shared by the surgeons. A surgeon has access to the same resource once a week. This **weekly session** is how the clinic areas are divided: analysis is done for each resource on each day of the week. Within the general terminology of the X-akseli system a **reservation** is a time when a patient is booked to use a resource at the hospital. Within the clinic and within the appointment scheduling literature these reservations are more specifically **appointments**. In the context of this study the patients time from the first reservation of the day to the completion of the final reservation of the day is the **patient flow**. Patients can have other reservations on the same day as the clinic appointment. If a patient has a reservation before the appointment, the arrival to the clinic appointment is called a **transition** within this study. If a patient arrives early to the appointment from a previous reservation the waiting time from arrival is **early-transition waiting**, and if a patient arrives late due to a previous reservation, this is a **late-transition**.

outpatient

elective patient

resource

weekly session

reservation

appointment

patient flow

transition

Appointment scheduling can be built using **appointment rules** [3, Sec. 4.1]. A block-size is how many patients are booked into a single reservation and a begin-block (initial block) is the number of patients booked for the first reservation. A block size of one with an initial block of one is an **individual block** schedule. The appointment interval is the time between each appointment in the schedule. A **fixed-interval** is when the interval for each patient has an equal spacing. The combination of setting these variables describes an appointment rule.

individual block

fixed-interval

2.2 The hospital outpatient clinic

The Oulu university hospital neurological, orthopaedic and dermatological outpatient clinic (referred to as the clinic from here-on-in) is the focus of this thesis. The clinic is where the specialist surgeons (referred to as **surgeons** from here-on-in) meet with patients for discussions. The **appointments** are generally either pre-surgery discussions, or post-surgery control meetings to check-up on the patient's progress after surgery (see Fig. 2.1 for a flow diagram of the **patient access path** to the clinic). The type of meeting with the patient is not available in the existing data. Prior to an appointment a surgeon may want to see the results of one or several **diagnostic examinations**. These can be provided by the health centre from where the patient was sent to the clinic or they can be done at the hospital. If a patient has to travel far for the appointment the diagnostic examination is usually done on the same day as the appointment (see Section 3.4.4 for details of preceding reservations). The examiner completes a report that is required for the surgeon appointment.

surgeon

appointment

diagnostic exam

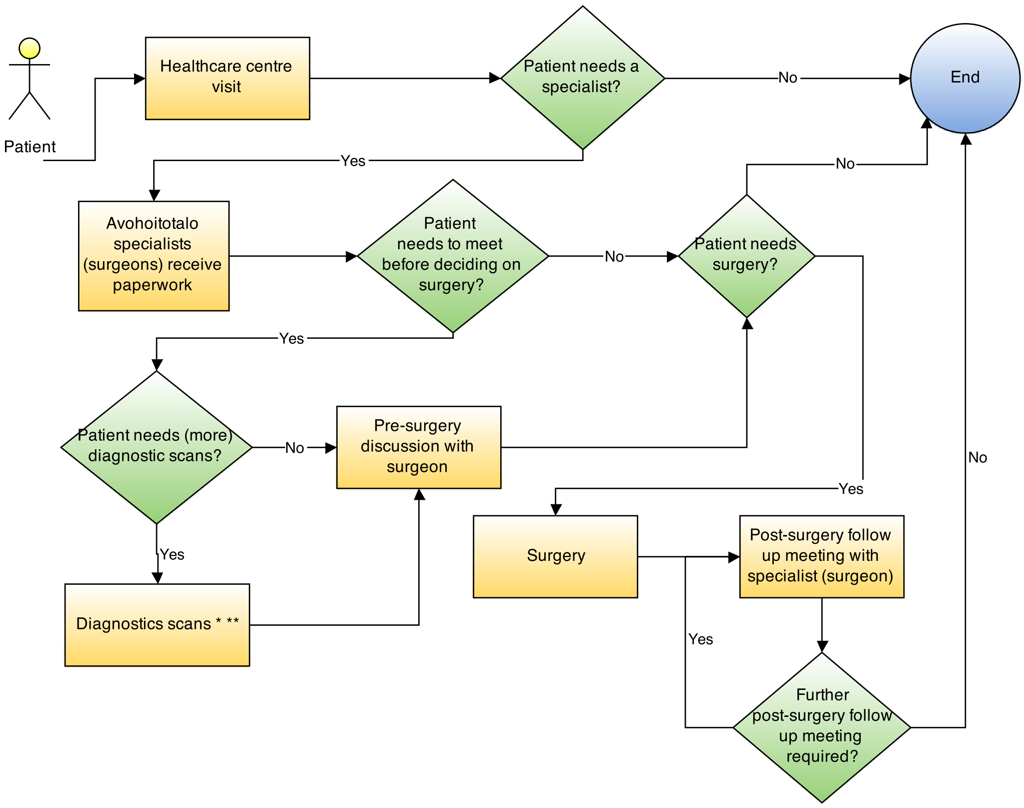
The clinic has used the X-akseli self-service system for several years. This provides a lot of data on patient movement through the clinic.

2.3 The X-akseli Oy system and data

2.3.1 The system

X-akseli Oy provides a patient self-service system for hospital departments, and other healthcare and welfare practises. A patient's reservation (appointment booking) includes a letter that contains a barcode: sent to the patient several weeks before the appointment date. On arrival to the hospital a patient presents the barcode to a self-service kiosk at the entrance; the kiosk registers the patient in the system, prints a ticket with a patient number and guides them to the correct waiting area. On arrival at the waiting area the patient presents the barcode to another kiosk. If the patient is at the cor-

self-service kiosk



* Either at the healthcare centre or at Avohoitotalo.
** For diagnostics in Avoihoitotalo, if patient travels a long distance the diagnostics are booked on the same day as the pre-surgery discussion.

Figure 2.1: Flow diagram: patient's access path to the clinic

rect location the patient is asked to sit and wait for their appointment: the patient is registered as waiting in the system. In the waiting area is a large screen, known as an information display, that calls patients by their patient number to the correct room. *infomation display*

A surgeon has a web page on their computer[1] that shows a list of patients and their status: from here the surgeon can see which patients are waiting to be seen. When the surgeon is ready to see a patient, they select an available patient from the list. On completion of the appointment, the surgeon presses 'completed' in the Doctor's view to signal the end of the appointment. *Doctor's view*

When a patient shows a barcode at a kiosk, when a surgeon calls in a patient and when a surgeon closes an appointment in the web application an event is stored in the system. These **events** contain date, time, user, reservation information and more. These three types of events are key to monitoring **patient flow** through the hospital. *events* *patient flow*

2.3.2 The data

Events in the X-akseli system (see Section 2.3.1), are recorded in a relational database management system. The events of interest are when a patient arrives in the clinic waiting area, when they are called in by the surgeon and when the appointment ends. Additionally, it is useful to know the same information for other reservations that patients at the clinic have on the same day as their clinic appointment. The arrival, call-in and completion events for all areas of the hospital are processed for use by the scheduling system. The **patient flow** through the clinic, the **waiting times**, **appointment durations** and **transition times** can all be measured with this data. The time when a **diagnostic examination** report is ready is not available. *event data* *patient flow*

Ethical considerations [23, pgs. 221-222] have been made for this data. The patient data is anonymous and resource names have been adjusted so that specific members of staff involved are anonymous. The hospital is a public hospital so a lot of the information used in this study is public information.

It is exciting to have so much patient-flow data available. The X-akseli system provides a-lot of data on all parts of the patient flow not typically seen: either the patient arrival and surgeon arrival are not available, only the duration is available [5, pg. 19]; or small samples of data collected through human monitoring of hospital clinics and surgeries are used [1] [11, Sec. 4]. With so much data this study looks at the feasibility of using machine learning and data mining for aiding in improving the scheduling at the clinic. It is

[1] X-akseli Doctor's view

worth noting that the self-service kiosk allows for automated data collection that often requires a lot of manual effort.

2.3.2.1 Preprocessing

patient flow table

A lot of work has gone into tidying the event data into an aggregate table that clearly shows when a patient arrives, starts and ends a reservation: this is the patient flow table. The X-akseli system event data was not readily suitable to create a clean aggregate table of the patient flow data. The data had a number of event repetitions, missing events and events that occured much sooner or much later than could have been possible, e.g., a patient that was in the waiting room for over two months. These inaccuracies and the missing data required some logical checks and filtering. The cause of the inaccuracies are largely unknown but were found to be partly due to human error, system error, or even system characteristics. A summary of the preprocessing steps taken is in Appendix A.

2.4 The existing clinic schedule

The existing schedule (as seen in the data and as communicated by hospital staff) uses individual-block fixed-interval scheduling. The session typically starts around 09:00 and patients are booked in at intervals, that are usually of twenty minutes, with a break for lunch and a shorter session starting after lunch (see Appendix B for example schedule visualisations). Without jumping into the analysis of the schedule, it is worth mentioning that the duration of each appointment varies largely. This makes it very challenging to produce a good schedule that keeps both patient waiting times and surgeon idle times low.

Another challenge in the scheduling is to predict when the same-day diagnostic examination report will be ready. To manage this uncertainty, patients that have an examination are only given the reservation time of the examination and are not given the time of the following surgeon appointment. A patient with an exam is instructed to go to the clinic waiting room after the examination and there they wait indefinitely until the surgeon has the report and an available time to see them. In the data it can be seen that the diagnostic examination is generally in the morning and the following appointment time reserved for patients tends to be several hours after the exam, but in reality the patient is ready much sooner, and may wait for a long time to be seen.

Chapter 3

The clinic scheduling environment

In this chapter the clinic schedule is analysed. First appointment scheduling performance measures are introduced and explained in the context of the clinic. Prior to analysis, the waiting time categories are described. The performance measures are then used to evaluate the weekly sessions at the clinic and rank them. The six worst performing weekly sessions are then analysed in more detail. For each selected weekly session the analysis looks at the different waiting times for each waiting-time category. The analysis extends into the preceding reservations, checking their relevance and what type of reservations precede the different weekly-session appointments, if any. The surgeon's punctuality of the weekly session is analysed; this was observed as a large cause of waiting times in the weekly session visualisations (see Appendix B). In the literature the patient punctuality and the no-shows are important for planning the schedule [3, Sec. 2], therefore these are also analysed.

3.1 Terminology

Within the broader scheduling terminology an appointment time can be seen as a **release time** r_j, where the job j relates to the patient. A single day at a resource at the clinic is a **session**, this can be defined as a period when a sequence of patients are seen at a specific resource with a planned start and end time. **Idle time** is any time during a session when a surgeon has no patients available to see. A planned session has a start and an end time. If the time to see all the patients requires more time than is given, the time spent seeing patients after the end time is **overtime**. The **duration** of an appointment is the **service time** in Appointment scheduling literature and the **process** time in scheduling theory, p_{ij} being the process time of an operation of job j on machine i.

3.2 Waiting time categories

When discussing waiting times it is important to give a clear exposition of the different categories of waiting at the clinic (see Fig. 3.1). The schedule has four events: arrival time of the patient a_t, reserved time of the appointment r_t, actual start time of the appointment s_t and end time of the appointment e_t. The intervals between these events fall into different waiting categories depending on the order of these events and whether a patient arrives from a preceding reservation or if this is the patient's first reservation: transition being true or false, $\tau \in \{true, false\}$. The waiting categories are colour coded in relation to their effect on the quality of service to the patient: green being a positive effect, blue neutral, yellow to be aware of, orange, pink and red being increasingly more negative.

If a_t precedes r_t the patient has arrived early. This is either **early arrival**, if this is the patient's first reservation of the day, or **early transition**, if the patient had a preceding reservation on the same day:

$$(a_t \leq r_t) \land \neg\tau \implies \text{early arrival} \quad , \tag{3.1}$$

$$(a_t \leq r_t) \land \tau \implies \text{early transition} \quad . \tag{3.2}$$

If a patient arrives early the appointment can start early, this is categorised as an **unplanned-early start**:

$$(a_t \leq r_t) \land (s_t \leq r_t) \implies \text{unplanned-early start} \quad . \tag{3.3}$$

If the patient arrives early, and the reservation starts after the reserved time, this is an **unplanned-late start**:

$$(a_t \leq r_t) \land (r_t < s_t) \implies \text{unplanned-late start} \quad . \tag{3.4}$$

Another scenario is if the patient is late for the appointment. The **late arrival** is unavoidable if it is the patient arriving late to the hospital; it could be avoidable if the patient is late due to a preceding reservation, **late transition**:

$$(r_t < a_t) \wedge \neg\tau \implies \text{late arrival} \quad , \tag{3.5}$$

$$(r_t < a_t) \wedge \tau \implies \text{late transition} \quad . \tag{3.6}$$

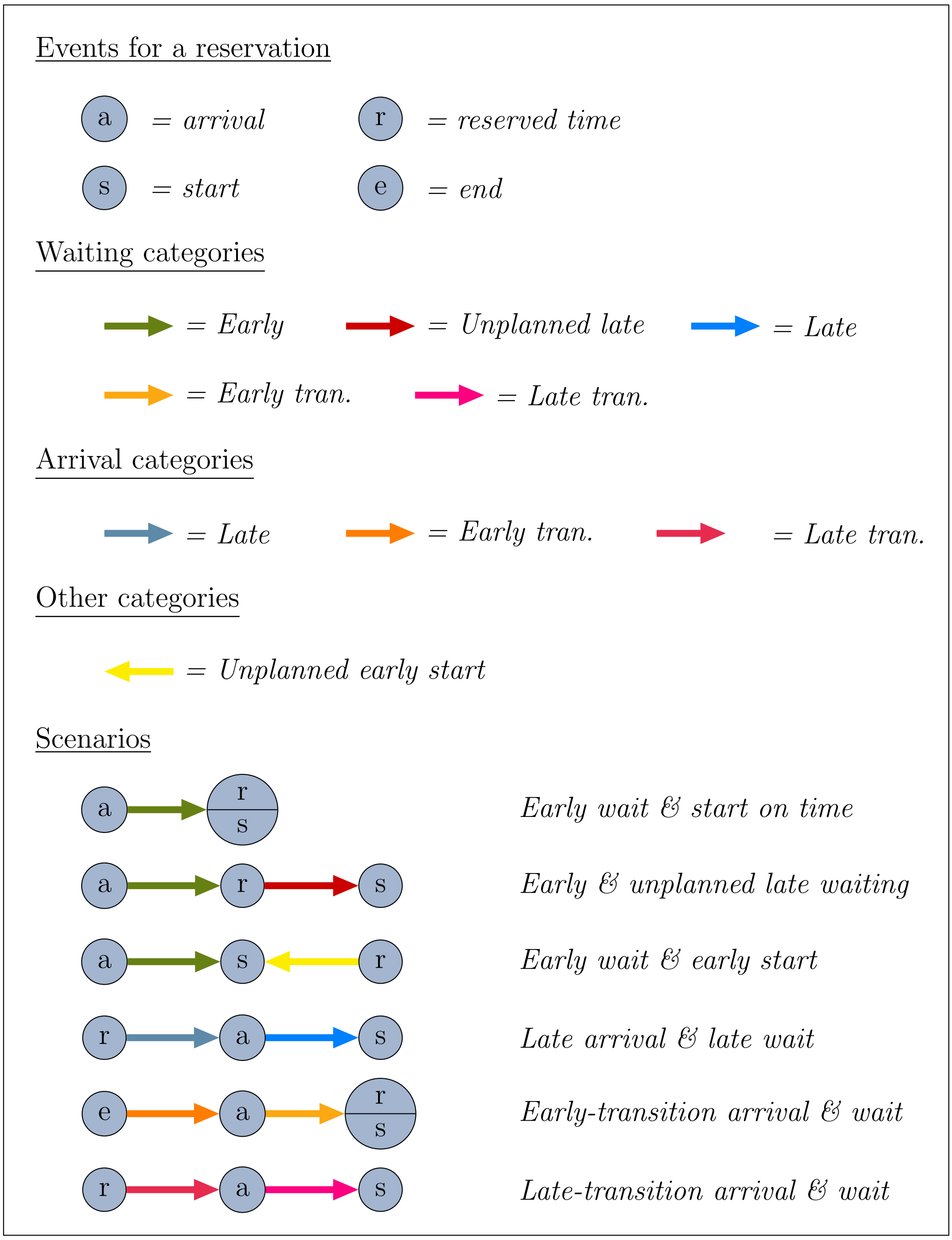


Figure 3.1: Waiting categories and the scenarios in which they occur.

3.3 Performance measures

Within the field of appointment scheduling there are typically three objectives a schedule attempts to minimise: patient waiting times, resource idle time and overtime [3, Sec. 3] [13, pgs. 136-137] [1, 4, 5, 15, 16, 25].

3.3.1 Patient waiting time

Five waiting categories have been identified in this case study: early-arrival, unplanned-late, late-arrival, early-transition and late-transition waiting. Of these one is out of the scope of the schedule: early-arrival. One is of low importance: late-arrival. Two are dependent on external factors: early-transition and late-transition waiting. One is directly effected by the scheduling and surgeon punctuality: unplanned-late waiting.

Using scheduling notation [19, Sec. 2.1] a waiting time for job (patient) j can be defined using the tardiness T_j of job j. To define tardiness as a waiting time, each job's due date is $d_j = r_j + p_j$ where release time (scheduled appointment time) is r_j and process time (duration of appointment) is p_j. Using the completion time of each job C_j the lateness of a job is $L_j = C_j - d_j$, from this the tardiness of the job is $T_j = max(L_j, 0)$. By minimising the tardiness in the static appointment schedule the waiting times are minimised. The scheduling objective to minimise and performance metrics are:

Waiting as job tardidness

$$\gamma_{wt} = \sum_{j=1}^{n} w_j T_j \quad , \tag{3.7}$$

$$\mu_{wt} = \frac{\gamma_{wt}}{n} = \frac{1}{n}\sum_{j=1}^{n} w_j T_j \quad , \tag{3.8}$$

for n jobs (patients), where w_j is the weight for job (patient) j. Note that the first waiting time is equal to the late start of the resource, taking the patients in sequential order. The late start is often not incorporated into appointment scheduling literature [4, 9]. Assuming each patient is seen immediately after the preceding patient, T_j is the expected waiting time for job (patient) j. By adding a weight w_j for each patient, patients with medical conditions that make waiting more difficult can have a higher weight and the optimised schedule will reduce their waiting more than the other patients with a smaller weighting.

This metric is not however fair, as it does not consider the amount of waiting per individual patient. A common target for hospitals is to reduce

Waiting threshold count

the number of patients that wait above a threshold waiting time, this reflects the quality of service [11, 20] [3, Sec. 3]. To introduce fairness a waiting time threshold A is introduce into the objective function that can have a cost associated depending on how large it is, e.g., a thirty minute threshold should have a much higher cost (weight) than a ten minute threshold. With a predefined threshold value A for the waiting time an objective similar to the number of tardy jobs [17, Sec. 3] [19, Sec. 2.1] can be constructed (Eq. 3.10):

$$U_{jA} = \begin{cases} 1, & \text{if } T_j > A \quad , \\ 0, & \text{otherwise} \quad , \end{cases} \tag{3.9}$$

$$N_{TA} = \sum_{j=1}^{n} w_j U_{j_A} \quad . \tag{3.10}$$

For a maximum threshold of thirty minutes, $A = 30$, N_{T30} is a weighted count of the number of jobs that exceed a waiting time of thirty minutes. By default the weight of the unit penalty is equal to the threshold value to give higher threshold values more importance, i.e., $w_j = A, j = 1, 2, ..., n$.

Another way to encourage fairness is to use the mean of the sum square of the waiting time [3, Sec. 3]. Penalising large individual waiting times more than the sum of an equal amount of smaller waiting times:

$$\gamma_{wt^2} = \sum_{j=1}^{n} w_j T_j^2 \quad , \tag{3.11}$$

$$\mu_{T^2} = \frac{\gamma_{wt^2}}{n} = \frac{1}{n} \sum_{j=1}^{n} w_j T_j^2 \quad . \tag{3.12}$$

3.3.2 Resource idle time and punctuality

In this study the number of patients and the session length are both fixed by the hospital; the scheduler cannot choose to construct the schedule by adding or removing patients nor can the session length be changed. For this reason the surgeons idle time over the session cannot be changed. However, if the idle time at the end of the session is deducted from the remaining idle time, this can give a measure of the idle time in between appointments. A large remaining idle time implies that either more patients could be seen during the session or the session length could be shortened; if there is no remaining idle time and the session over runs, this implies that too many patients (or too many patients with long appointment durations) are being seen or that the session should be made longer. The idle time S for a job j is:

$$S_j = a_j - a_{j-1} - p_{j-1} \quad , \tag{3.13}$$

where a_j is the actual start time of job j, and p_j is the duration of job j, $a_0 = p_0 = 0$. The first idle time in the session, S_1, is particularly interesting as it is the late start of the resource. Sessions starting late (resource unpunctuality) has been observed in the analysis of the existing schedules in the hospital (Section 3.4.5). An objective for the schedule is to minimise the weighted total idle time (subtracting the end-of-session idle time) of the resource (the surgeon):

resource unpunctuality

$$\gamma_{ws} = \sum_{j=1}^{n} w_j S_j \quad . \tag{3.14}$$

Interesting information on the performance of a schedule is to see the ratio of the surgeon's idle time per appointment $\mu_{ws} = \frac{\gamma_{ws}}{n}$ to the average patient waiting time μ_{wt}, $\frac{\mu_{ws}}{\mu_{wt}}$: this gives some idea of the value of the surgeon's time versus the patient's time. Historically the surgeon's time was many times more valued than the patient's [1, 15]; in some cases the cost of the patient waiting can be high, e.g. in a military hospital [25]. In public healthcare the importance of reducing the patient waiting time is increasing [11, 20]. This ratio can be defined as an objective for the scheduling to predefine the balance between the two or as an informative metric.

waiting time ratio

3.3.3 Overtime

In this study, a predefined number of patients are booked within a fixed session time, both are given and not changeable: only the appointment times can be changed. The scheduling system will give a prediction of the time required to see all the patients: this may be under or over the given fixed session length. If the session finishes early the surgeon will have free time at the end of the session; if the session finishes late the surgeon will be required to do overtime to see all the patients. This is seen within the new schedule predicted in this case study.

3.3.4 Further considerations

Another aspect of interest is the number of no-shows [3, Sec. 2]. When

no-shows

creating a real schedule this must be considered as it has a significant impact on the schedule, especially if the no show is during the beginning of the session. Showing how many people are in the waiting room throughout the schedule is an informative metric to get an idea of how many seats will be occupied, predicting this is made much more difficult by patient early arrivals.

number of people waiting

Early arrival patients (patient punctuality) helps to reduce the surgeons idle time at the cost of filling up the waiting room, but this is not within the control of the schedule. However, the hospital invites patients fifteen minutes prior to the reservation time to encourage early arrival. **Early transition** is managed by the scheduler: measuring the waiting time caused by early arrival shows how much waiting is caused by leaving too much time between the reservation prior to the appointment and the appointment.

Late transition is a bad situation where the patient cannot make it in time for the appointment due to the previous reservation completing too late or, in the case of a diagnostic examination, the examination report not being ready in time for the appointment. The number of these cases are analysed. **Late arrival** is due to the patient, and therefore the waiting time for this category of patient is of lower priority. This is analysed but is not considered of high importance.

3.4 Analysis

Prior to starting to optimise the schedule the current situation is analysed using the historical data available (see Section 2.3.2 for a description of the data). Schedule performance (Section 3.4.1), appointment duration lengths (Section 3.4.2), current waiting times (Section 3.4.3) and patient's preceding reservations (Section 3.4.4), surgeon unpunctuality (Section 3.4.5) and patient unpunctuality and no-shows (Section 3.4.6) are analysed. This analysis is informative to the hospital. It also identifies which weekly sessions have the largest waiting times, allowing the study to focus on those weekly sessions. Lastly, these results expose what are the larger causes of waiting time, so that these areas can be the main focus of the optimisation.

3.4.1 Schedule performance and ranking

The clinic has approximately seventy resources, the aim is to identify and optimise the schedule with respect to the categories of waiting times that can be improved, i.e., unplanned-late and transition waiting, for those sessions with the longest waiting times. The top-six ranked weekly sessions are reported on. The ranking is performed on the unplanned-late waiting time.

Ranking is done using the unplanned-late waiting time statistics: median, lower quartile and upper quartile percentiles; the mean and mode measures of central tendency; and the percentage of reservations that have waiting times above thirty minutes [6]. The first time the ranking was done it was found that some weekly sessions that were ranked highly had very few samples. I felt I should incorporate the number of samples into the ranking. Additionally, some of the weekly sessions had a reasonable amount of samples but the variance was very large: the results from these lacked confidence. I came up with an easy heuristic to penalise weekly sessions with low numbers of samples and high variance:

$$\xi_a = \xi(\max(\ln n - c_v, 0)) \quad , \tag{3.15}$$

where n is the number of samples, $c_v = \frac{\mu}{\sigma}$ is the coefficient of variation, ξ is one of the statistics (either a percentile or a measure of central tendency) and ξ_a is the adjusted statistic that is used for ranking.

For the percentage of late reservations the variance of the statistical value was not a factor: the percent is an absolute value without any variation. The equation only used the number of samples for ranking:

$$p_a = p \log_2 n \quad , \tag{3.16}$$

where p is the percentage (proportion) of reservations that start over thirty minutes late for all patients that arrive early or on time, n is the number of samples and p_a is the adjusted percentage used for ranking each weekly session.

For each statistic the weekly sessions were ranked, the top six are reported (see Tables 3.2 and 3.3). For comparison of their performance compared to the average values at the clinic the overall statistics for the clinic are presented (see Table 3.1). Notice how the maximum value is sometimes extremely high; even after much effort preprocessing, some bad data may have still remained.

An overall scoring measure is used to combine these statistics into an overall ranking. A simple method is used: applying points to the rankings in Tables 3.2 and 3.3 from six to one, highest to lowest, and summing the scores. Due to the significance of the percentage that are late, the score for the percentage is given a factor of three. The overall top six resources are the focus of the following analysis and optimisation.

3.4.2 Appointment duration length

A summary of the statistics for the durations lengths at the highest ranked weekly sessions (Table 3.4): comprising a five-number summary [6], the mean

Categories	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Total	0.00	15.00	35.00	43.67	62.00	537.00
Early arrival	0.00	6.00	14.00	17.81	25.00	205.00
Early transition	0.00	11.00	24.00	29.84	41.00	251.00
Unplanned late	0.00	12.00	26.00	34.26	47.00	525.00
Late arrival	0.00	6.00	19.00	29.17	39.00	403.00
Late transition	0.00	10.00	24.00	32.76	47.00	303.00
Unplanned early	1.00	6.00	14.00	24.28	28.00	288.00

Table 3.1: Clinic overall waiting times (in minutes) and unplanned early start time. See Section 3.4 for details of the categories. 'Min.' shows the minimum values, '1st Qu.' the first quartile, '3rd Qu.' the third quartile and 'Max.' the maximum value.

resource	day	mean	resource	day	mode	resource	day	median
N1	Tues.	64.49	N1	Wed.	58	N1	Wed.	54
N1	Wed.	65.12	O1	Fri.	67	N1	Tues.	48
N2	Mon.	54.16	O4	Wed.	46	P2	Thur.	47
P2	Mon.	47.12	N1	Thur.	46	P2	Mon.	41
P2	Thur.	48.98	O9	Thur.	51	O1	Fri.	51
N3	Wed.	62.72	N2	Thur.	47	O3	Mon.	33

Table 3.2: Highest ranked unplanned waiting times (in minutes) mean, mode and median, using ranking described in Eq. 3.15.

resource	day	LQ	resource	day	UQ	resource	day	percent
N1	Wed.	32	N1	Tues.	79	N1	Wed.	81.00
O1	Mon.	26	N1	Wed.	76.5	O1	Fri.	70.00
L1	Sat.	41	P2	Mon.	68	O1	Mon.	66.00
O1	Fri.	29	O3	Mon.	57.5	P2	Thur.	63.00
P2	Thur.	23	P2	Thur.	67.25	O5	Thur.	61.00
N1	Thur.	24	N2	Mon.	64	L1	Sat.	59.00

Table 3.3: Highest ranked unplanned waiting times (in minutes) lower quartile (LQ), upper quartile (UQ) and the percent of late reservations above the acceptable waiting time (30 minutes), using rankings described in Eqs. (3.15) and (3.16).

and the coefficient of variation; and the histogram plots with distribution estimates, show that the variance of the appointment durations are high. The coefficient of variation is a standardised measure of relative dispersion of the data, each unit is a factor of the mean as a standard deviation [6]. It is often used as a standard measure for comparing the variation of appointment durations [3, Sec. 2.5]. Empirical studies show a range that is approximately between 0.35 and 0.85. Looking at the results, five of the variances are above the high end of the range.

The twenty minute interval observed to be used is within six minutes of the average duration of the weekly sessions analysed (see Section 3.4.2); twenty minutes, however, is mostly above the average duration time, risking high surgeon idle time. This and the high variation in appointment duration could have caused surgeons that followed the original schedule to wait several minutes between each appointment: by arriving late and filling up the waiting room this idle time between appointments is removed at the cost of high patient waiting times, maybe this is why surgeons consistently arrive late?

Considering under or over estimation of an appointment duration will incur either large patient waiting times or waste a surgeons time, respectively [13, pg. 136], it is unlikely that using the mean, mode or median as an estimate fo the duration length would produce a good schedule [1, 3, 15] (see Section 1.2 for more details). A more accurate prediction will be made using supervised learning, Section 4.5.

rsrc_name	day	min	LQ	median	mean	UQ	max	CV
N1	Wed.	0	7	13	15.34	19.25	110	0.81
O1	Mon.	0	13	16	18.64	22	79	0.53
N1	Tues.	0	11	18	21.05	26	279	0.97
P2	Thur.	0	7	12	15.77	20	218	1.16
P2	Mon.	0	6	12	17.15	22	121	1.08
O5	Thur.	0	16	23	26.33	36	90	0.58

Table 3.4: Highest ranked weekly session duration lengths (in minutes) five-number summary, mean and coefficient of variation.

3.4.3 Waiting time

Waiting time can be categorised into parts that the schedule can and cannot influence (see Section 3.4). Waiting that the schedule can reduce are unplanned-late waiting, the waiting past the reservation time and transition waiting, the early or late arrival at a resource due to the time gap between the reservation before the appointment and the appointment reservation. Using statisical tools an analysis of the categories that can be influenced follows.

3.4.3.1 Unplanned-late waiting

Focusing on the top-six ranked weekly sessions with the highest waiting times, a five-number summary, the mean and variance is reported (Table 3.5) and histogram plots with density estimations are produced Fig. C.2 for the unplanned-late waiting times at the clinic.

resource	day	min.	LQ	median	mean	UQ	max.	mode	c_v	percent
N1	Tues.	0	21	48	64.49	79	506	30	1.17	0.51
N1	Wed.	1	32	54	65.12	76.5	525	58	1.04	0.70
O1	Mon.	1	26	37	41.44	52.5	220	31	0.67	0.61
O5	Thur.	0	24	43	48.03	64	185	31	0.67	0.63
P2	Mon.	0	20.5	41	47.12	68	231	24	0.74	0.54
P2	Thur.	0	23	47	48.98	67.25	310	55	0.70	0.59

Table 3.5: Highest ranked weekly session unplanned-late waiting (in minutes) five-number summary, mean, coefficent of variation and percentage of all on-time arrivals that waited over thirty minutes.

3.4.3.2 Transition-early waiting

To analyse the early-transition waiting a statistical summary of the early arrival times (see Table 3.6) and a histogram plot with density estimations Fig. C.3 are produced.

resource	day	min	LQ	median	mean	UQ	max	mode	cv	percent
P2	Mon.	0	24.5	41	54.39	69.5	269	41	0.90	0.08
P2	Thur.	0	28.75	55	63.91	76	213	76	0.75	0.07
N1	Wed.	1	22	51	58.87	80	325	8	0.84	0.41
N1	Tues.	2	36	69	79.89	96	302	2	0.79	0.36
O5	Thur.	0	15	26	30.45	43	110	23	0.69	0.68
O1	Mon.	0	16	28	36.17	48	343	16	0.96	0.69

Table 3.6: Highest ranked weekly session early-transition arrival prior to appointment time (in minutes) five-number summary, mean and coefficient of variation. Percent is the percentage of the early-transitions of all early arrivals at the resource.

Sometimes the patient arriving early from a previous reservation (transition-early) will see the surgeon much earlier than the scheduled reservation time. The hospital have explained that the transition-arrival patients from the diagnostic examinations are not given an appointment time with the surgeon, they only receive a time for the diagnostic examination, with instructions to then go to the waiting room and wait to be seen by the surgeon. A statistical summary of the actual waiting times of the early-transition patients (see Table 3.7) and a histogram plot with a density estimation for the early-transition waiting times (see Fig. C.4) are produced.

resource	day	min	LQ	median	mean	UQ	max	mode	cv	percent
P2	Mon.	0	21.5	41	48.61	58.5	251	41	0.92	0.65
P2	Thur.	0	24	40	47.69	65.75	172	24	0.79	0.72
N1	Wed.	1	21	46	51.21	76	159	83	0.71	0.66
N1	Tues.	0	28	46	55.57	78	185	2	0.72	0.83
O5	Thur.	0	14	26	29.66	42	110	23	0.70	0.44
O1	Mon.	0	14	26	32.78	45	148	16	0.79	0.46

Table 3.7: Highest ranked weekly session early-transition waiting time (in minutes) five-number summary, mean and coefficient of variation. Percent is the percentage of the early-transitions that were above thirty minutes.

3.4.3.3 Transition-late waiting

A patient arriving late from a preceding reservation is an occurrence that could be avoided or reduced with improvements in the schedule. Currently the scheduling is done so that this is unlikely to occur at the cost of large waiting times for the patient; leaving overly large time gaps between the preceding reservation and the appointment, it is rare that a patient will not have completed the preceding reservation prior to the appointment time. A statistical summary (see Table 3.8) and visualisation (see Fig. C.5) of how late patients arrive shows that it is rare for late transitions to occur for P resources, not common for N resources but more common for O resources.

resource	day	min	LQ	median	mean	UQ	max	mode	cv	percent
P2	Mon.	0	24.5	41	54.39	69.5	269	41	0.90	0.08
P2	Thur.	0	28.75	55	63.91	76	213	76	0.75	0.07
N1	Wed.	1	22	51	58.87	80	325	8	0.84	0.41
N1	Tues.	2	36	69	79.89	96	302	2	0.79	0.36
O5	Thur.	0	15	26	30.45	43	110	23	0.69	0.68
O1	Mon.	0	16	28	36.17	48	343	16	0.96	0.69

Table 3.8: Highest ranked weekly session late-transition arrival prior to appointment time (in minutes) five-number summary, mean and coefficient of variation. Percent is the percentage of all transition arrivals that are late.

resource	day	min	LQ	median	mean	UQ	max	mode	cv	percent
P2	Mon.	0	21.5	41	48.61	58.5	251	41	0.92	0.65
P2	Thur.	0	24	40	47.69	65.75	172	24	0.79	0.72
N1	Wed.	1	21	46	51.21	76	159	83	0.71	0.66
N1	Tues.	0	28	46	55.57	78	185	2	0.72	0.83
O5	Thur.	0	14	26	29.66	42	110	23	0.70	0.44
O1	Mon.	0	14	26	32.78	45	148	16	0.79	0.46

Table 3.9: Highest ranked weekly session late-transition waiting time (in minutes) five-number summary, mean and coefficient of variation. Percent is the percentage of all late arrivals that are due to transitions.

3.4.4 Preceding reservations

There are a number of reasons why the preceding reservation is of interest when scheduling appointments: the preceding reservation imposes a constraint on the scheduling; patients are typically assigned too much time between the preceding reservation and the clinic appointment, as shown above in early-transition analysis (Section 3.4.3.2). By accurately predicting the end time of the previous reservation this gap can be reduced. The X-akseli system is in many parts of the hospital, so often there is historical data available for the preceding reservation resources. Supervised learning could be used to make a prediction. A third reason for interest for this information is as a feature for predicting the duration of the clinic appointment or for categorising the patients based on which diagnostics exam they have had prior to the appointment.

An overview of the previous reservations for the weekly sessions of interest shows what reservations typically precede a surgeon appointment (see Fig. 3.2). For the six weekly sessions that we are focusing on in this study the statistics for the preceding reservations show what reservations precede these specific weekly sessions (see Table 3.10 and Figs. 3.2 and 3.3). A large portion of the orthopedic surgeon appointments (O resources) have a preceding diagnostic examination, the plastic-surgeon appointments (P resources) have very few preceding reservations and the neuro-surgeon appointments (N resources) have a mixture of diagnostic and other preceding reservations. This information is logical, as orthopedic surgery largely requires x-rays and other diagnostic examinations, as oppose to plastic-surgery that rarely requires such exams. Neuro-surgery is complicated and requires different types of resources for patients.

The diagnostic examinations produce a report that is required by the surgeon for the appointment, this data is not available so no statistics can be generated. Discussing this with a student that monitored patients through the clinic the examination was ready twice within fifteen minutes but maybe in some cases it takes much longer.

resource	day	Any(%)	Dg.(%)	Cln.(%)
N1	Tues.	39	17	2
N1	Wed.	47	14	2
O1	Mon.	76	61	1
O5	Thur.	78	64	8
P2	Mon.	16	10	3
P2	Thur.	13	10	2

Table 3.10: Same-day preceding reservations percentages. 'Any' refers to any reservation, 'Dg.' are the percentage of diagnostic-examination reservations on floor R reservations, and 'Cln.' are out-patient clinic reservations on floor one.

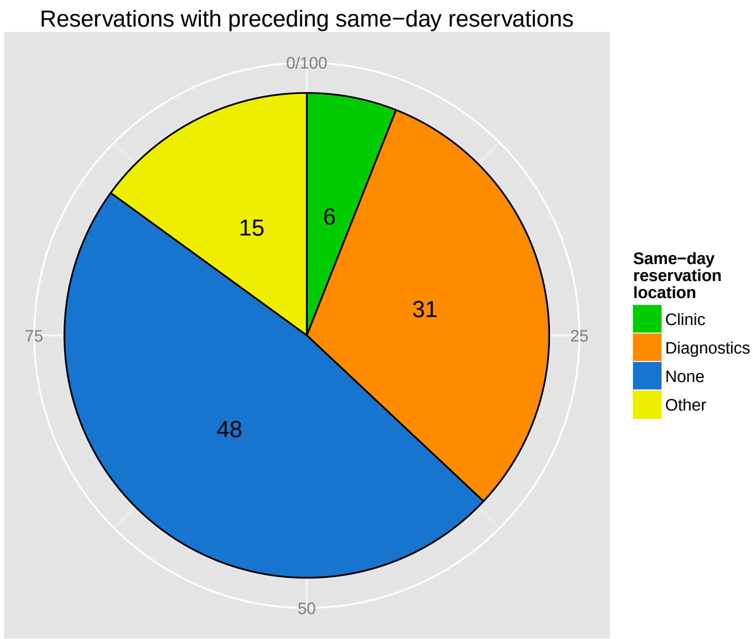


Figure 3.2: Same-day preceding reservations percentages. Clinic refers to the outpatient clinic and diagnostics refers to the diagnostics resources on Floor R.

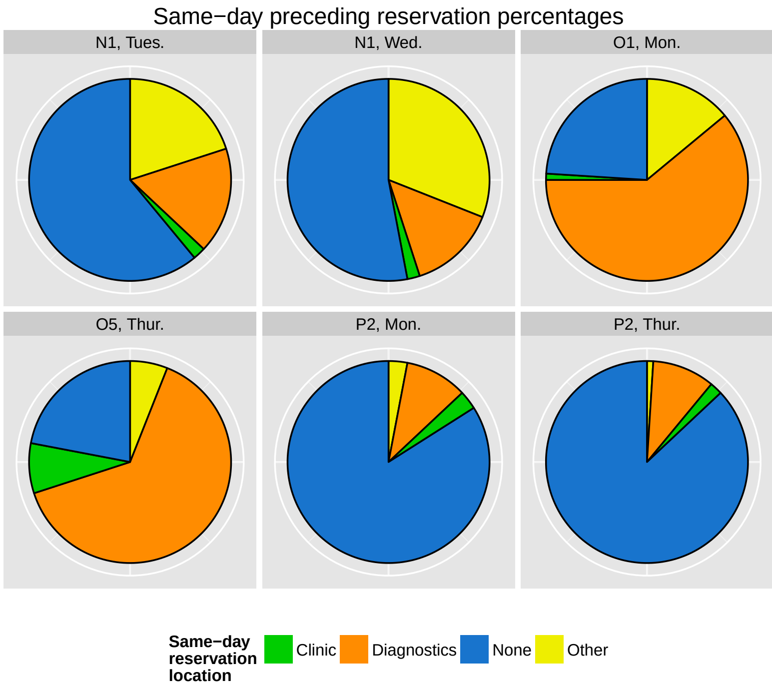


Figure 3.3: Same-day preceding reservation percentages per weekly session. Clinic refers to the outpatient clinic and diagnostics refers to the diagnostic examinations.

3.4.5 Resource unpunctuality

Resource (surgeon) unpunctuality is important, a good schedule can be vitiated if the resource starts significantly late [15]. Analysis has shown that some resources often begin late into the session. This will cause an initial waiting time for the patient that propagates throughout the session [1, 3]. A statistical summary (see Table 3.11) and visualisations (see Appendix C.4) show when surgeons arrive for the weekly sessions. During scheduling this late start is to be considered and used in the new schedule to reduce the patient waiting times.

Resource	Day	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	mode
P2	Mon.	-186	19	49	45.36	72	149	59
P2	Thur.	-50	11	27	32.54	51.5	174	12
N1	Wed.	16	34	44	67.63	100	177	41
N1	Tues.	-67	6	40.5	55.02	117	169	12
O5	Thur.	-15	1.75	7.5	11.47	15.75	118	2
O1	Mon.	-9	9	26	23.34	34	67	31

Table 3.11: Late start five-number statistical summary, the mean and the mode.

3.4.6 Patient unpunctuality and no-shows

Patients will occasionally arrive late or not at all. This can quickly turn a good schedule into a very bad schedule where the surgeon is made to wait a long time for patients to arrive. In order to avoid large surgeon idle time patient unpunctuality and no-shows must be considered in the schedule. If patient no-show is high it will be very difficult to produce a good schedule due to the large amount of uncertainty [3]. Patient unpunctuality is also detrimental, but less so.

An analysis of the patient no-shows was not easy given the data. If a patient did not arrive at the hospital at all, then the data could indicate this, so the percentage of non-arrivals is given in the table, Table 3.12. Less clear were patients that seemed to arrive, looking at the data, but then did not have enough information to determine the full appointment. This could be due to a data issue, a software user issue, a change of plan at the hospital (maybe the patient was sent back or told to go somewhere else) or that they really didn't make it to the appointment even though they arrived at the hospital. An analysis of the patient arrivals (omitting transition arrivals) shows that few patients arrive late and so the impact of late arrivals is small, see Table 3.13 and Fig. C.13.

Resource	Day	Non-Arrival(%)	Non-Processed(%)
N1	Wed.	3	11
O1	Mon.	1	7
N1	Tues.	3	15
P2	Thur.	3	7
P2	Mon.	5	13
O5	Thur.	1	15

Table 3.12: Highest ranked weekly session patient no-show percentages. 'Non-Arrival' is the percentage of patients that did not arrive at the hospital, 'Non-Processed' is the percentage of patients that were not completely processed.

resource	day	min	LQ	median	mean	UQ	max	cv	P()
P2	Mon.	-223.9	-25.03	-13.73	-14.67	-3.8	159.1	-1.80	0.83
P2	Thur.	-289.3	-25	-13.86	-19.35	-4.887	113.6	-1.84	0.87
N1	Wed.	-204.8	-25.05	-14.98	-15.74	-7.112	180.1	-1.71	0.90
N1	Tues.	-212.7	-25.4	-14.02	-9.796	-5.096	263	-4.53	0.84
O5	Thur.	-137.6	-21.99	-11.82	-12.8	-3.888	292.9	-2.80	0.86
O1	Mon.	-144.6	-27.75	-14.27	-18.54	-4.65	67.98	-1.53	0.85

Table 3.13: Highest ranked weekly session patient punctuality (in minutes) five-number summary, mean and coefficient of variation. Percent is the percentage of patients that arrive on time.

3.4.7 Unplanned-early starts

This can have positive and negative effects and the decision to start early needs consideration of the current status of the schedule. Taking patients in early that changes the sequence of the schedule can often introduce unwanted waiting time that could have been avoided [3]. It has been observed that patients who arrive earlier than a patient before them in the schedule can be called in before the preceding patient, even if that patient is in the waiting room, see Appendix B for example observations.

Chapter 4

Outpatient appointment scheduling

4.1 Literature

Appointment scheduling is a sub-field of scheduling that grew since early studies in the fifties, of which Bailey [1] made a large impact showing how patient waiting could be reduced with minimal cost of overtime or idle-time to the surgeon, simply by applying good scheduling rules. Bailey used queueing theory and simulation with distributions to show relationships of patient waiting time and surgeon idle time. Since then a large amount of work has been done in this area [1, 3–5, 9, 11, 15–17, 25]. A lot of the work will make assumptions on patients being punctual, always showing for appointments and having no constraints for the resource availability, so they are often not directly applicable to this problem: they are however very relevant. Methods applied include appointment rules, global search, local search, evolutionary algorithms, mathematical programming and heuristics.

In the initial study by Bailey [1] and in other studies [3, 13, 15] the importance of the appointment duration (appointment interval) being accurate is discussed. The study shows patient waiting time and the surgeon idle time is inversely related; patient waiting grows in relation to underestimating the appointment interval, and surgeon idle time grows in relation to overestimating the appointment interval Fig. 1.1. The less variance in the scheduling, the easier it is to use historical data to estimate an appointment duration, unfortunately the variance of the durations at this clinic are very high leading to difficulties in constructing a good schedule. A relation between good appointment scheduling and the appointment duration (service time) variance has been shown [3]. Supervised learning regression [6, 12, 14, 24], using

appointment duration

the X-akseli historical data, is used to try to find a relationship between features in the X-akseli data and future appointment durations. Simple machine learning and statistical methods are applied with the two-fold aim of understanding what features are useful in prediction and in trying to find a prediction that is accurate.

The same study [1] showed the relationship of the patient waiting time and the surgeon idle time with different numbers of patients arriving at the beginning of a session Fig. 1.2, known as the initial block size [3]. An analysis of the arrival times of the surgeons at the clinic, Section 3.4.5, showed that many resources consistently start late. Often several patients have already arrived and some have already been waiting for over thirty minutes before the surgeon arrives. By inviting multiple patients at a later start time patient waiting time can be reduced while not increasing the risk of the surgeon being idle. Linear regression is used to predict the surgeon's arrival. Relying solely on prediction accuracy was too risky in this situation, considering how starting the session too late would cause the surgeon to arrive to an empty clinic. The linear regression prediction interval [12, 14] is used to find a start time with a high level of confidence that the surgeon will not have already arrived.

surgeon punctuality

4.2 Dealing with uncertainty

A number of aspects of the schedule are stochastic: appointment duration lengths, arrival time of a patient, arrival time of a surgeon to the session, time of transition from the previous resource, and when the report is ready from a previous resource. The impact on the schedule and the variance is considered when deciding how to deal with uncertainty.

The appointment duration lengths have high variance, see Section 3.4.2, and inaccuracies have a large detrimental impact on the schedule [1, 3]. Therefore, a lot of effort is made to predict this value accurately: using domain-specific knowledge, relevant features are extracted from the X-akseli data and supervised learning regression is used for prediction.

Patients rarely arrive late and even more rarely do not show for a reservation, see Section 3.4.6. Therefore simple approaches to dealing with these uncertainties are used: sampling from a suitable distribution [12, 14] is used to estimate patient unpunctuality and no-shows are not considered.

The time of transition for a patient is important to reduce the transition-early waiting, the previous reservation completion time is estimated with a simple linear regression model. For preceding diagnostic examinations a report must be ready before the appointment can start. The X-akseli data has no historical data available for this report, a fifteen minute period can be used, this is a value that was observed by a fellow student who followed a patient through the hospital. The report readiness data is planned to be collected by X-akseli and can be used for future scheduling.

4.3 Statistical modelling

From Bailey [1] and other recent papers a gamma distribution is used to model the service time (appointment duration) [3, 4], sometimes truncated gamma [25]. Using a histogram to visualise the frequency distributions, Appendix C.1, it can be observed the historical data seems to be gamma distributed: having a postive skew with a long tail and all values are zero or greater.

Maximum likelihood estimation (MLE) is a method for parameter estimation, but the gamma function makes differentiation difficult: the method of moments estimation is quite straight forward for the gamma distribution [22]. MLE is often preferred as a less biased estimate but for this simple case the simpler method of estimation is chosen. Method of moments fits distribution moments with data sample moment estimates. For the gamma distribution the scale parameter α and shape parameter λ can be found from the sample data:

$$Gamma(\alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}, \quad \alpha, \lambda > 0 \tag{4.1}$$

The method of moments matches the mean and shape parameters of the gamma distribution with the first and second moments of the sample data:

$$\mu_1 = E[X] = \frac{\alpha}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \quad , \tag{4.2}$$

$$\alpha = \bar{x}\lambda \quad . \tag{4.3}$$

$$\mu_2 = E[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \mu_1^2 + \frac{\mu_1}{\lambda} \quad , \tag{4.4}$$

$$\hat{\lambda} = \frac{\mu_1}{(\mu_2 - \mu_1^2)} = \frac{\bar{x}}{\hat{\sigma}^2} \quad . \tag{4.5}$$

$$\hat{\alpha} = \frac{\mu_1^2}{(\mu_2 - \mu_1^2)} = \frac{\bar{x}^2}{\hat{\sigma}^2} \quad . \tag{4.6}$$

To test the distribution fit a normal distribution is also fitted to the sample data using MLE and compared with the gamma distribution fit. The normal distribution method of moments estimation is:

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2a^2}\right) \quad , \tag{4.7}$$

$$\mu_1 = E[X] = \mu \quad , \tag{4.8}$$

$$\mu_2 = E[X^2] = \mu^2 + \sigma^2 \quad , \tag{4.9}$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \quad , \tag{4.10}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad . \tag{4.11}$$

Distributions are fitted to the existing duration data.

4.4 Predicting the appointment duration

In many simulation approaches the duration of an appointment (service or process time) uses a fitted distribution [1, 3, 4, 9, 15, 16, 25], typically

Gamma. However, the typical coefficient of variance is 0.35 to 0.85 [3], much lower than what has been observed in the analysis of the clinic duration times in Section 3.4.2, Table 3.4. Therefore, supervised learning regression is applied to predict the duration time, in the hope this has better results than using the mean or a fitted distribution.

In the scheduling literature a stochastic scheduling problem has stochastic job processing times (random appointment durations), scheduling problems in this category have fewer algorithms than for deterministic scheduling problems [19]. Predicting the process times changes the problem to a deterministic scheduling problem, for which there are a wider choice of algorithms for optimisation.

4.5 Prediction with supervised learning

This section describes what supervised learning is and how a good model is chosen. Considerations when extracting features and methods of dummy coding are described.

4.5.1 Terminology

A set of N input vectors with P dimensions are represented as an N x P matrix X. Each instance is indexed $i = 1, 2, ..., N$ and each column (feature, predictor) is indexed $j = 1, 2, ..., P$. The numerical output vector is y of length N.

4.5.2 The supervised learning problem

Within the domain of machine learning, supervised learning can be used for prediction, when the variable to be predicted is numeric (quantitative) this is know as regression [12, 14, 24]. Supervised-learning regression can be seen as a function approximation from a set of predictors (features) to a response variable (dependent variable output). Input variables, X, are predictors, independent variables or features, the output variable, $y \in \mathcal{R}$, is the response or dependent variable. The general regression equation is then:

$$y = f(X) + \epsilon \quad , \tag{4.12}$$

where $f(X)$ is a regression function and ϵ is the error term. The error term is assumed to have zero mean and a Gaussian distribution, $\epsilon \sim \mathcal{N}(0, \sigma)$. Prediction approximates the function $\hat{f}(X)$ and predicts the response $\hat{y} =$

$\hat{f}(X) + \epsilon$. The goal is to learn the function mapping but not the error term, this can be shown in terms of the residual:

$$E(y - \hat{y}) = E[f(X) + \epsilon - \hat{f}(X)]^2 \tag{4.13}$$

$$= \underbrace{[f(X) - \hat{f}(X)]^2}_{\text{Reducable}} + \underbrace{\text{Var}(\epsilon)}_{\text{Irreducable}} \quad . \tag{4.14}$$

4.5.3 Choosing a model

When choosing a regression model there is a choice between better model interpretability or better model accuracy [14]. This study is focused on both model inference, to report to the hospital and X-akseli, and on prediction accuracy to provide the scheduler with the most accurate estimate of the appointment duration. Following the principle of Occam's razor a simple model with few assumptions is used as a first approach. Linear regression is a simple and often powerful parametric regression model that is easy to interpret with few parameters to learn. A limitation is the assumption that the relationship between the predictors and the response is linear [12, 14].

Another consideration is choosing between parametric and non-parametric models. Regression approximates a regression function $f(X)$. Parametric models make some assumptions on the form of the function, e.g., linear regression assumes the function f is linear. The fewer parameters, the larger the assumption of the form of the function, and the less observations required for learning those parameters. If the model assumption is wrong however, this model bias will cause underfitting, resulting in poor prediction accuracy. On the other hand, having too many parameters will risk too high variance in the model causing overfitting, this is the bias-variance trade off. A non-parametric model, similar to a model with many parameters, is much more flexible and able to learn the form of the regression function more closely, but with higher complexity in learning and at the risk of overfitting. A non-parametric model in general requires more observations for training than parametric models [6, 14, 24].

Choosing the best model is difficult, "there is no free lunch". This is to say that no model is best, they all work differently on different types of problems. Therefore, trying different models and having a comparative measure of quality is essential. Once a choice of what type of models will be used, cross validation and the mean square error is used for model selection.

4.5.4 Feature extraction

Good predictors can help to accurately predict the response variable. Feature extraction is the collection of these good features, in some cases existing features can be transformed. Continuous valued inputs can be converted to interval, ordinal or categorical variables. For example the patients previous surgery can be represented as a continuous value by counting the number of days (with decimal places for fractions of a day). This variable can be converted to an interval by binning the values into a finite set of ranges, e.g., from zero to 7 days, 8 to 14 days, 15 to 21 days, etc. An ordinal variable could be used to abstract the period further, into either 'very recent','recent','not recent','not relevant', where the different categories have a clear ordering from highest to lowest. To abstract the variable further a categorical variable could be 'had surgery' with the domain true and false, in statistics this variable is a **factor** and the domain true and false are the **levels**. A two value categorical variable is a dichotomous variable, also a binary variable or bit. Other transformations can also be done to the data if it is believed the new feature is good for learning.

continuous

interval

ordinal

categorical

dichotomous

4.5.5 Dummy coding

With supervised learning, models often require that categorical variables (factors) are dummy coded [12]. The supervised learning models used require a binary string representation of the categories (Linear regression and K-nearest neighbour, with the distance metrics used here, both require this). With k levels of a factor, $k-1$ dummy variables are created, the value zero representing the first level. Each dummy variable is a dichotomous (binary) variable and this type of variable is represented well in linear regression (adjusting the weight of each coefficient) and in K-nearest neighbour regression (Using symmetric or asymmetric binary dissimilarity to measure binary vector distances). Take the example of a factor for gender with three levels, male, female and unknown:

factor level	$dummy_1$	$dummy_2$
male	0	0
female	1	0
unknown	0	1

Table 4.1: Dummy coding categorical variables (factors).

4.6 Linear regression

Linear regression, as the chosen method of supervised learning regression, is described. Calculating the p-values for the model predictors is explained as this is used later for feature selection. The different methods of correlation and ANOVA are described as these are used later to analyse the relationship between the predictors and the response variable. Finally, the prediction interval is explained as this is used to find a prediction that gives a higher level of confidence than the mean response.

Linear regression using the method of least squares is a traditional method founded in the early 19th century by Legendre and Gauss [14]. It is a simple supervised learning approach, the assumption is that the underlying function is approximately linear. The Linear regression parametric model has the benefit of being interpretable, allowing for easy understanding of what feature effects the prediction; in an additive model the coefficient relates directly to the feature, explaining the effect of one unit of the predictor to the response variable, additionally a parametric model has a small number of parameters to be learnt to model the data.

A linear model is parametric, where $P+1$ coefficients must be learnt to fit the model to the data. Given an input vector $x^T = (x_1, x_2, ..., x_P)$ the output response y is predicted:

$$y \approx \beta_0 + \sum_{j=1}^{P} \beta_j x_j \quad . \tag{4.15}$$

A typical method of fitting the parameters is by minimizing the least squares criterion. Using simple linear regression as an example, where $P = 1$, for each prediction $\hat{y}_i$, $i = 1, ..., N$, given a training observation, x_i :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad . \tag{4.16}$$

Least squares approach fits the line to minimise the residual sum of squares (RSS). A residual is the difference between the prediction and the actual value $e_i = y_i - \hat{y}_i$ and $RSS = \sum_i^n e_i^2$. To minimise the RSS, we take the partial derivate for each coefficient and set it to zero. For a simple linear regression problem, with two parameters, an intercept and a slope:

residual sum of squares

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2 \quad . \tag{4.17}$$

The least squares coefficient estimates are:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad , \tag{4.18}$$

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\mathrm{Cov}(x_i, y_i)}{\mathrm{Var}(x_i)} \quad , \tag{4.19}$$

where $\mathrm{Cov}(x, y)$ is the covariance between the predictor and the response variables and $\mathrm{Var}(x)$ is the variance of the predictor.

prediction interval

It is of value to know the uncertainty of the prediction. For each prediction in linear regression a prediction interval, similar to a confidence interval, gives a range for a given level of confidence. The prediction interval combines uncertainty of the irreducible error ϵ and the reducible error, the uncertainty in the function approximation of $f(x)$.

4.6.1 Feature selection

Each feature in linear regression has a coefficient that estimates the linear regression line. Using the standard error of a coefficient, a t-statistic is found, from which a p-value for each coefficient can be derived. A null-hypothesis test for each coefficcient uses the p-value. The null-hypothesis $H_0 : \beta = 0$ states that the predictor X has no relationship with the output response y. If the p-value is low enough the null-hypothesis can be rejected and we can say that the coefficient is statistically significant. The level of significance is the degree of risk that a null-hypothesis is rejected when it is actually true, this is a type-one error, [23]. A minimum acceptable value is dependent on the data.

A good feature for prediction can have a strong correlation between a predictor and the response variable. For numerical predictors a test for linear correlation shows the strength of this relationship. For categorical and dichotomous variables analysis of variance (ANOVA) shows the strength of a categorical variable at predicting the response.

4.6.1.1 Coefficient p-values

Using the estimates the least squares line is found [12, 14]. The standard error test can give an approximation of how close to the true value the coefficient estimates are given the data, below are the standard errors for the intercept and slope coefficient of the simple linear regression example:

$$\widehat{SE}(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right] \quad , \tag{4.20}$$

$$\widehat{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad , \tag{4.21}$$

where σ^2 is Var(ϵ) and as it is unknown it is estimated from the data. This is estimated by the residual standard error (RSE):

$$RSE = \sqrt{\frac{1}{n-2} RSS} \quad . \tag{4.22}$$

The standard error can be used to get a confidence interval for the parameters. A 95 % confidence interval is approximately:

$$\hat{\beta} \pm 2\widehat{SE}(\beta) \, , \tag{4.23}$$

where β is any of the coefficients. Calculations for the simple linear regression β_0 and β_1 coefficients' standard errors are in Eq. 4.21.

The standard error is used in computing the null hypothesis. A t-statistic determines how many standard deviations the coefficient value is from β_0:

$$t = \frac{\beta_1 - \beta_0}{SE(\beta_1)}. \tag{4.24}$$

From this a p-value is obtained, this is the probability that β_i is 0: the probability that the null-hypothesis is true. It says how much of a difference β_i makes in the relationship between the predictors and the output.

4.6.1.2 Correlation

Another interesting indicator of the relationships of pairs of the variables, are the correlation coefficients. A correlation coefficient is a normalised measure $[-1, 1] \in \mathcal{R}$ of a bivariate relationship, where -1 is a perfect negative relationship, 1 is a perfect positive relationship and 0 is no relationship, it can be seen as a measure of the dependence between pairs of variables [23]. A high correlation does not prove causality but it does show that one variable can be predicted from a change in another variable. A high correlation between a predictor variable and the response indicates a good feature for prediction. A typical method for continuous valued variables is the Pearson product-moment correlation, it is a linear correlation measure of Eucidean distance between the variables:

$$r = \frac{\mathrm{Cov}(x, y)}{\mathrm{s}_x \mathrm{s}_y} . \tag{4.25}$$

If changes in value differ alot in the correlation, a better measure of correlation is the rank correlation measures. This describes the relationship of the same increase or decrease of two variables, ignoring the quantity of change in the variable. Two popular measures are described.

Kendall's Tau measures the *symmetric difference distance* between pairs of variables $d_\Delta(\mathcal{P}_1, \mathcal{P}_2)$: *symmetric difference distance*

$$\tau = \frac{\frac{1}{2}N(N-1) - d_\Delta(\mathcal{P}_1, \mathcal{P}_2)}{\frac{1}{2}N(N-1)} \tag{4.26}$$

$$= 1 - \frac{2[d_\Delta(\mathcal{P}_1, \mathcal{P}_2)]}{N(N-1)} , \tag{4.27}$$

where $\mathcal{P}$ is the list of ordered pairs for and $\frac{1}{2}N(N-1)$) is the total number of pairs. The measure is counting the number of concordant and discordant pairs from the ordered lists, giving a count for the number of pairs not in the union of both lists [23].

Spearman's Rho (ρ) is a non-parametric rank correlation coefficient that replaces the values of the variables with their rank, so (x_i, y_i) becomes (Rank(x_i), Rank(y_i)), where Rank is the position of the variable in an independently ordered list of the values for each variable. The squared distance between each pair d_i^2 is used in the equation:

$$r_s = 1 - \frac{6\sum d_i^2}{(n^3 - n)} . \tag{4.28}$$

In the case of tied values within the ranking of either variable, the values are given the midrank score, in equation Eq. 4.28 this prevents a perfect relationship from yielding a 1 or -1, the formula needs to accomodate for this:

$$r_s = \frac{n^3 - n - 6\sum d_i^2 - 6(t' + u')}{sqrt(n^3 - n - 12t')sqrt(n^3 - n - 12u')}, \tag{4.29}$$

where $t' = (\sum t^3 - sumt)/12$, t being the number of tied scores in the x ranks, and $u' = (\sum u^3 - sumu)/12$, u being the number of tied scores in the y ranks. Spearman's Rho does not assume any distribution of the variables, unlike Pearson's product-moment coefficients.

4.6.1.3 ANOVA

Many of the variables used in predicting the appointment durations for the schedule use categorical variables. To find the association of the categorical variables with the output reponse a one-way analysis of variance (ANOVA) can be carried out. Each categorical variable is a factor, and the categories for each factor are levels. Applying one-way ANOVA can give a p-value for the mean of each level with respect to the response. ANOVA tests if the difference in means are significant. ANOVA works by comparing the within group variance and the between group variance.

4.6.2 Prediction interval

For any linear regression mean response there is an element of uncertainty. A prediction interval uses the irreducible and reducible errors (described in Section 4.5.2) to give an interval, larger than the confidence interval, for a desired level of confidence [18, pg. 57]:

$$PI = \hat{y} \pm t_{1-\alpha/2,n-2} SE_{\hat{y}} \ , \tag{4.30}$$

where t is the t-statistic, and $SE_{\hat{y}}$ is the estimated standard error. From the analysis it can be seen that the response is approximately normal, so this interval will be approximately correct.

4.7 K-nearest neighbours regression

The relationship between the response and the predictors is found to likely be non-linear. K-nearest neighbour (k-NN) is explained a simple non-linear alternative to the linear regression model.

A non-parametric approach can often yield better prediction results as it assumes no parametric form for the regression function f. A popular and simple non-parametric supervised learning method is k nearest neighbours regression [14, 24]. Given the predictor x_0 the k nearest observations are used to calculate a weighted average for the response:

$$\hat{f}(x_0) = \frac{1}{K} \sum_{x_i \in \mathcal{N}_0} w_i y_i \quad . \tag{4.31}$$

The addition of a weight w_i can be used to add decay based on the increasing distance of an observation in the neighbourhood. A typical method is to use the distance d in the weight decay $w_i = 1/d_{i,j}$, where i is the index of the observation in the neighbourhood of predictors and j is the new observation.

Unlike eager learners, e.g., linear regression, k-NN is a lazy learner also known as an instance-based learner, during training the observations are stored and when given a new set of predictors a generalisation model is constructed, this is in contrast to eager learners that construct a generalisation model during training and do little work during prediction [10]. This comes at a cost of storing the observations as part of the model and computation requirements during each prediction. *lazy learner*

An important consideration with k-NN is the distance matrix, this determines the neighbourhood of observations and their weighting. Continuous predictor variables will commonly use a Euclidean distance metric. For categorical values the Euclidean distance can be misleading, so an alternative metric is required. Given two binary strings the asymmetric Eq. 4.33 and symmetric Eq. 4.32 binary dissimilarity is a distance measure between them Table 4.2. The asymmetric binary dissimilarity gives more importance to positive values [10]: *distance matrix* *symmetric binary dissimilarity*

	$x_i = 1$	$x_i = 0$	$\sum$
$x_j = 1$	q	r	q+r
$x_j = 0$	s	t	s+t
$\sum$	q+s	r+t	p

Table 4.2: Contingency table for binary (dichotomous) variables, where p is the total number of variables.

$$d_{symm} = \frac{r+s}{q+r+s+t} \tag{4.32}$$

$$d_{asyym} = \frac{r+s}{q+r+s} \quad . \tag{4.33}$$

For a distance matrix with a combination of quantitative and qualitative types Gower's generalized coefficient of dissimilarity is used [8]:

$$d(i,j) = \sum_{k=1}^{P} \left(\frac{d_{ijk}\delta_{ijk}}{\sum_{k=1}^{P} \delta_{ijk}} \right) \quad , \tag{4.34}$$

where P is the number of features, i and j are the two observation being compared. $\delta_{ijk} \in 0,1$ is one if the variable is relevant and zero otherwise, in the case of asymmetric binary comparison two variables that are zero for that feature would have a delta of 0. All features are standardised into the range $[0,1]$ and d_{ijk} is the standardised distance for that feature, measured in with the relevant metric for that type.

4.8 Model assessment and selection

A score to use for comparing different model's performance is needed. A test of the predicting power on unseen data is also a requirement. Here a measure to make this comparison is explained.

In supervised learning regression, where a model is built to capture the relationship between predictors X and a response y, it is important to consider the generalisation of the model. Generlisation is the ability of a model to make predictions on unseen data. During training there is a bias / variance trade-off. As the model complexity increases it is less bias and the variance in the model increases. A model that is overly bias will underfit and fail to predict accurately. As a model is trained with the training data the variance increases, the more complex a model the more variance it can learn, if a model is tuned too well to the training data it will too closely resemble the training data, this is known as overfitting. A model that overfits the data loses it's ability to generalise, failing to predict unseen samples accurately. In order to avoid overfitting model training can use cross validation. In cross validation the training data is divided into training and validation sets, the model is trained using the training data and prediction accuracy is tested on the unseen validation set.

A measure of accuracy to compare different regression models is the mean squared error (MSE) that uses the residual [24]:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 \quad . \tag{4.35}$$

The MSE can be viewed as a bias-variance decomposition, a sum of the variance and bias of the model and the irreducible error:

$$E(y_0 - \hat{f}(x_o)) = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}([\hat{f}](x_0))]^2 + \mathrm{Var}(\epsilon) \quad . \tag{4.36}$$

4.9 Appointment scheduling

4.9.1 Terminology

The goal of reducing a set of appointment scheduling objectives is known as an appointment scheduling problem (**ASP**). **Elective patient** is a patient that is scheduled prior to the appointment date, unlike urgent patients that often need service as soon as possible and show up on the day [3, 27].

4.9.2 Literature

The broad aim of appointment scheduling is to find a sequence of appointment times that minimises a set of objectives that are commonly resource idle time, patient waiting time and session overtime. There are two broad categories in appointment scheduling, static and dynamic [3]. A **static ASP** is an offline problem where the schedule is prepared for the session a priori. Dynamic appointment scheduling adjusts the schedule throughout the day making decisions online as patients arrive. This problem is a static ASP. The initial schedule handles a single surgeon's schedule, so it is a **single-server ASP**. The presence of **no-shows**, **patient unpunctuality** and **surgeons' unpunctuality** are considered. The clinic does not cater for walk-ins, only **elective patients**. Service times (appointment durations) are stochastic, and analysis shows (see Section 3.4.2) that the coeeficient of variation is generally very high compared to previous empirical studies [3]. Service times are independent though in reality this may not be the case, it has been observed that service times tend to be shorter toward the start and end of the day, making a dome shape [3].

4.9.3 Appointment rules

A common approach to managing a schedule is using a set of rules. The rules comprises three variable the block-size n_i, the begin-block n_1 and the appointment interval a_i. Adjusting these variables a number of different rules can be constructed. The simplest, commonly used prior to the initial scheduling studies of the 1950s, is a single block schedule where all N patients are invited at the start of the session, $n_1 = N$. The existing schedule at the hospital is single-block, fixed interval (individual block [16]), $n_i = 1, \forall i =$

$1, 2, ..., N$ and $a_i = c$, where c is a constant that is often twenty minutes. The individual block divides the session into N equally spaced segments. The Bailey-Welch rule is a small modification of this moving the last appointment to the first, having two patients arrive at the first scheduled time. The Bailey-Welch rule often performs well [16].

4.9.4 Local search

Local search is a general search procedure that starts with a given solution and makes incremental changes to optimise the schedule. The given solution is the starting state, the incremental change that can be made is the search process and all states that can be reached with an incremental change are the neighbouring states. In general a local search algorithm requires a definition for a problem representation, the neighbourhood, the search process and the acceptance-rejection criteria.

As the hospital will provide an existing schedule, this lends naturally to this type of optimisation. Additionally, the hospital have imposed a restriction on not making too many changes to the original schedule, this approach supports that the optimal schedule will be close to the original schedule. A local search method used for optimisation is described.

4.9.5 Hill climbing

Hill climbing is a greedy method of local search optimisation. Starting at an existing state (node) u; the cost of the state is recorded. The algorithm tries all neighbouring states (nodes) $v \in N_u$, selecting the best until no improvement can be made.

The problem with this approach is the algorithm will not move past the local optima. With a small search space and with the limitation of not being able to change the sequence of the patients this may find the global optimal for this problem.

Algorithm 1 Hill climbing algorithm

```
procedure HILL CLIMB
    cost_best = score(u)
    v_best = u
    repeat
        better = FALSE
        u = v_best
        for each node v ∈ N_u do
            if v is rejected then
                continue
            end if
            cost_v = score(v)
            if cost_v < cost_best then
                cost_best = cost_v
                v_best = v
                better = TRUE
            end if
        end for
    until better==FALSE
end procedure
```

Chapter 5

Scheduling System Implementation

5.1 Problem formulation

The existing situation has a fixed number of N patients that are booked several weeks ahead of the date of the appointment by employees of the hospital. The scheduling approach uses single block, fixed intervals. This proposed system will take the existing schedule and prior to sending the invitations out to the patients the schedule will be optimised with the objective of reducing patient waiting time, maintaining the number of patients and the working hours of the surgeon.

The problem is: a **static** appointment scheduling problem (ASP); a **stochastic** ASP, as the process times of the jobs are not known; a single-server ASP (or **single-machine** scheduling problem) as one surgeon is scheduled independently; the ASP has recirculation (**rcrc**), where a patient may revisit the same machine. There are also availability constraints (**other**): for the surgeon currently reduced to a lunch break, the arrival time of the surgeon in case of lateness and the end of session for the day and for the patient where there are preceding and succeeding reservations that are not within the control of the scheduling system. The multi-criteria objective is to minimise the average waiting time W_j, the count of patients that wait above certain threshold waiting time U_j, and minimise surgeon idle time between appointments, with the constraints on surgeon availability. In the notation for deterministic scheduling the problem can be broadly defined as:

$$\alpha \quad | \quad \beta \quad | \quad \gamma \tag{5.1}$$

$$\alpha = 1,\ \beta = rcrc, other,\ \gamma = \sum w_j W_j, \sum w_j Uj \sum w_j S_j \tag{5.2}$$

$$1|rcrc, other|\sum w_j W_j, \sum w_j Uj \sum w_j S_j \quad , \tag{5.3}$$

where α is the machine setup, β are the constraints and γ is the objective. The scheduling is made more challenging with surgeon and patient upunctuality and patient no-shows [3, 13]. One aspect that makes the scheduling less challenging is the absence of urgent walk-in patients [27].

5.2 Overview

The schedule optimisation uses the historical data to simulate a real day at the hospital and a local search algorithm to find the appointment rules that perform the best. The simulation uses a combination of sampling from distributions fitted to the historical data and supervised learning prediction models. The local search algorithm searches through an appointment-rule parameter space, where the first reservation (FR), appointment interval (I) and initial block (BI, number of patients invited to the first reservation) are the parameters. The scoring of each new appointment rule uses one hundred iterations, each with new samples, to find an average score.

Data is loaded into the free and open source statistical programming language environment R [21]. All statistical analysis, machine learning and scheduling optimisation is performed with tools in R. Plots were generated with ggplot [26].

5.2.1 Component overview

A flow chart of the system shows the components and their interactions (see Fig. 5.1). Preprocessing of the data is described in an earlier chapter (see Section 2.3.2).

Distribution fitting and prediction models are explained in prior sections (see Sections 4.3 and 4.5), linear regression with cross validation are used to find good prediction models. During simulation of the appointment schedule it is not known which of the waiting patients the surgeon will call in next, therefore dispatching rules are used to simulate the call-in process

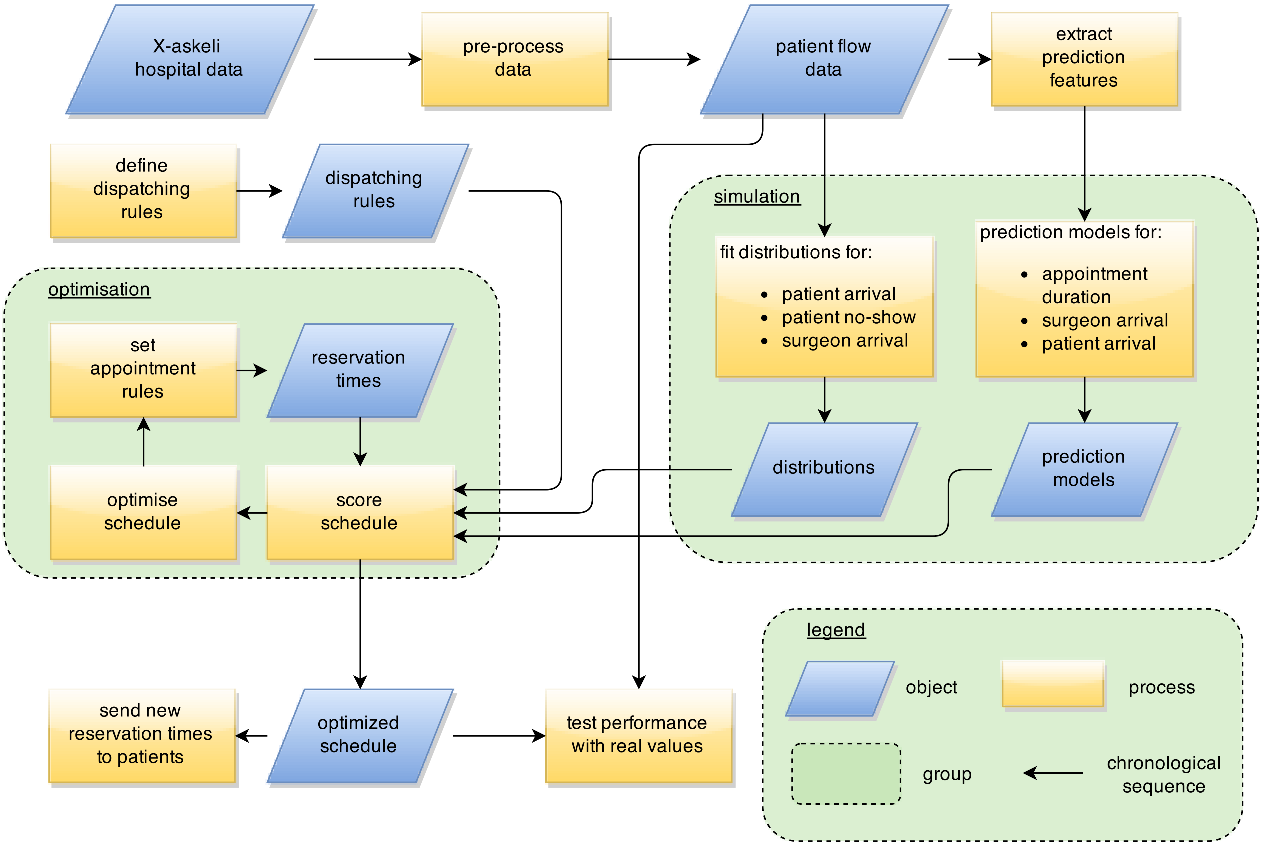


Figure 5.1: Overview of the scheduling system components

(Section 5.5). Optimisation uses the local search methods described in (Section 5.6.1), by changing the appointment rules within the defined neighbourhood.

5.3 Feature extraction

The X-akseli system contains data on the movement of patients over the past few years in parts of the hospital. Understanding the problem and extracting the features is the beginning to training a supervised learning model for prediction.

5.3.1 Appointment duration

Some information can be generally useful for any appointment prediction: demographic information like gender and age and time when the appointment was reserved. For more specific features it is best to discuss with the hospital to understand the purpose of the appointments and find the features in the limited data available that are relevant.

The appointments are explained in the environment chapter (see Section 2.2). Appointments relate to the previous diagnostic examination, previous appointments and previous surgeries: and how recent they were. This data was extracted from the available data (see Appendix E for a full list with descriptions). The purpose of the visit was not available, nor the type of examination, nor the results of the examination nor any information on examinations taken outside of the area of the hospital where the x-akseli systems are in place: I believe this information could have improved the predictions. The time of day of the reservation is relevant due to the dome shape phenomena [9, pg. 811], with shorter durations during the start and end and longer during the middle.

5.3.2 Surgeon arrival

The analysis of surgeon arrival (see Section 3.4.5) shows the arrival time likely relates to the first appointment time. The first appointment time is used to build a simple linear regression model to predict the arrival of the surgeon. A prediction interval (see Section 4.6.2) of seventy percent is used and the lower bound of the interval is given as the predicted time of arrival, this means that for 85% of predictions the Doctor will arrive after this time. Using this lower bound decreases the risk that a surgeon arrives when no patients have yet been scheduled, this would be a very bad situation for the system and so it is better if patients arrive some minutes early than late at all. This time is used as an initial starting point for the local search optimisation.

5.3.3 Patient transition arrival

Predicting the arrival of a patient from a previous reservation depends on the completion time of the previous reservation. To get an approximate estimate for this a simple linear prediction model uses the previous reservations resource as a predictor, and the time difference between the previous reservation's scheduled time and the actual end time. As it is a higher cost if a patient is late on arrival, the prediction interval is used to find a predicted time that would occur approximately 85% of the time. In case a patient is coming from a new resource, the mean of all the resources, plus two standard deviations is used instead. This prediction is used in combination with sampling from a distribution over all resources in a linear combination with 0.5 weighing for each. Finally a ten minute travel time is added to all the predictions.

5.4 Simulation

5.4.1 Distribution fitting

Any distribution that fits well to the data can be used. Typically a Gamma distribution is used for arrival times and duration times; the method of moments is a convenient way to fit the Gamma distribution (see Section 4.3 for an explanation).

The QQ-plot can show how well a distribution fits to the data points. To demonstrate a good and bad fit, QQ-plots for a Gaussian distribution and a Gamma distribution fitted to the same appointment duration data using the methods of moments are produced (see Figs. 5.2 and 5.3). For the appointment duration the QQ-plot shows that the Gamma distribution fits the data well, while the Gaussian distribution is a poor fit. The Gamma distribution is fitted to the duration times of the selected weekly sessions (see Appendix D). As the coefficient of variation is high for the appointment duration, distributions were not used for scheduling simulation, instead supervised learning is used to predict the duration for each appointment more accurately.

Simulation of the patient arrival and surgeon arrival use generalized logistic distributions that were found to fit well and R software had good packages to support working with them.

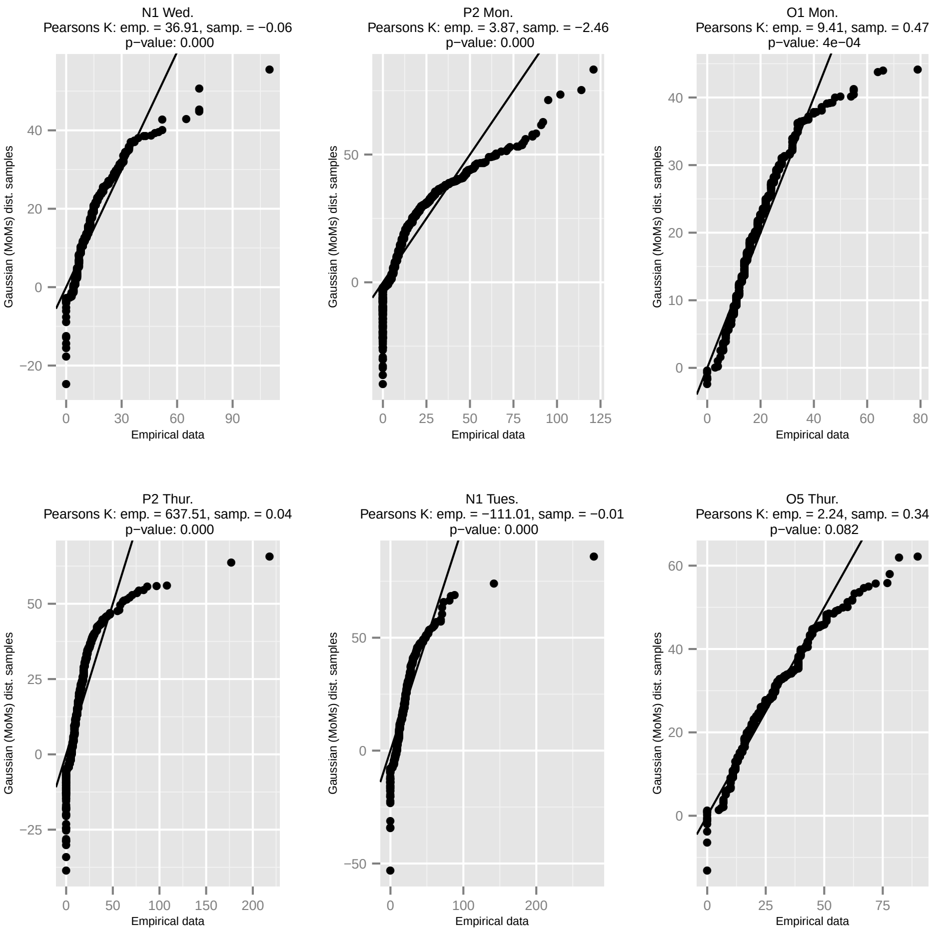


Figure 5.2: QQ-plot for the Gaussian distribution fitted to the appointment duration.

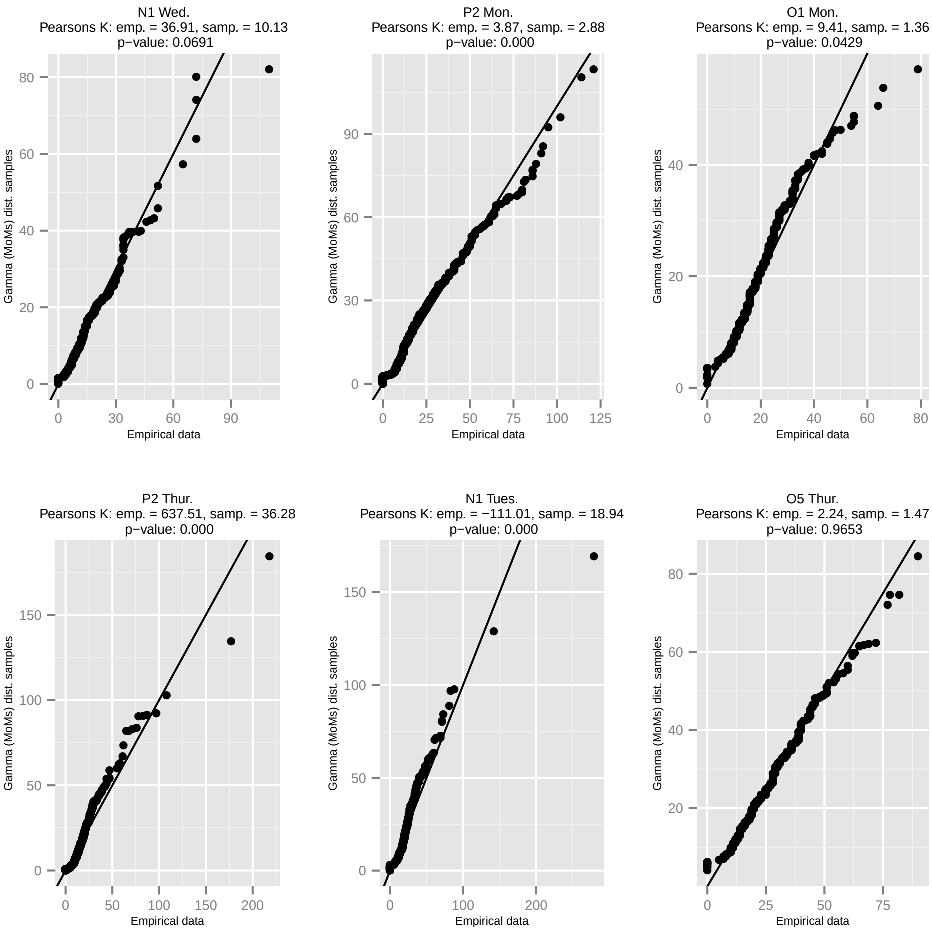


Figure 5.3: QQ-plot for the Gamma distribution fitted to the appointment duration.

5.4.2 Prediction models

The prediction model is presented with the extracted features (for a list and details see Appendix E). To improve the efficiency, performance and comprehensibility of the supervised learning model, and the effectiveness of future data collection, feature selection is used to find a good subset of the predictors [24, Feature Selection]. Each feature selection technique used here requires a parameter. Using cross validation the optimal feature-selection algorithm and it's parameters are found.

5.4.2.1 Feature selection

Three methods of feature selection [14, Ch. 3] for linear regression are compared and the method that returns the highest accuracy is then used. Each method is presented with a set of all predictors $p_a, |p_a| = m$

Forward selection takes one user-defined parameter, the feature subset size $k, k < m$ and starts with an empty set of selected predictors $p_s = \emptyset$. For each iteration of the algorithm every remaining predictor $p_r = p_a \setminus p_s$ is temporarily added to p_s and the RSS is compared for all; the predictor that returns the lowest RSS is then added to p_s. This repeats until the set is of size k. **Backward selection** starts with a set of all the features $p_s = p_a, |p_s| = m$, and a user-defined subset size parameter $k, k < m$. For each iteration the feature with the highest p-value is removed from p_s until $|p_s| = k$. As new features are added, the p-values of the existing features can change. **Mixed selection** takes a p-value threshold as a parameter, the method combines forward and backward selection. Starting with an empty set of selected predictors $p_s = \emptyset$, like in forward selection, the predictor not in the selected set $p_a \setminus p_s$ with the lowest RSS is added to p_s. On adding each new predictor all the p-values are checked and any predictors with a p-value above the user-defined threshold are removed. This repeats until adding any new variable will have a p-value that is above the threshold.

5.4.2.2 Cross validation

Cross validation is used to find the accuracy for different choices of the feature selection parameter for each feature selection method. There are a number of approaches to cross-validation, the general principle being to set aside a portion of the data available for training the model (the training set) and to use the other portion for testing the predictive ability of the model (the validation set). Methods include leave-one-out, k-fold and others. In this scenario these methods would use future unseen data to train the model, and validate it on past data, this is not logical. As this data is chronologically ordered, four most recent sessions were taken as the validation set, and one year past data from those session dates were used as training data. This is a good simulation of the real application of this system. For each session all the appointment durations were predicted, and the average root mean square error (RMSE) was calculated for each model on all four sessions (each session having between approximately 6 and 12 appointments). The model with the lowest average RMSE was selected. To see how the overall performance was against the twenty minute appointment duration used at the clinic for the existing schedule, the RMSE for a prediction of twenty minutes for each duration was used. In the literature the average mean value of the duration is sometimes used, the RMSE for the average duration value is also compared (Fig. 6.2).

5.5 Dispatching rules for simulation

A dispatching rule [19] has been designed for the outpatient clinic (see Fig. 5.4). The rule will see patients in order of reservation, and if the next reservation has a large enough gap another patient will be seen using a first-come-first-served policy (FCFS) within each waiting-category group: prioritising the groups. The group priorities are: transition-late, transition-early, early arrival, late-arrival.

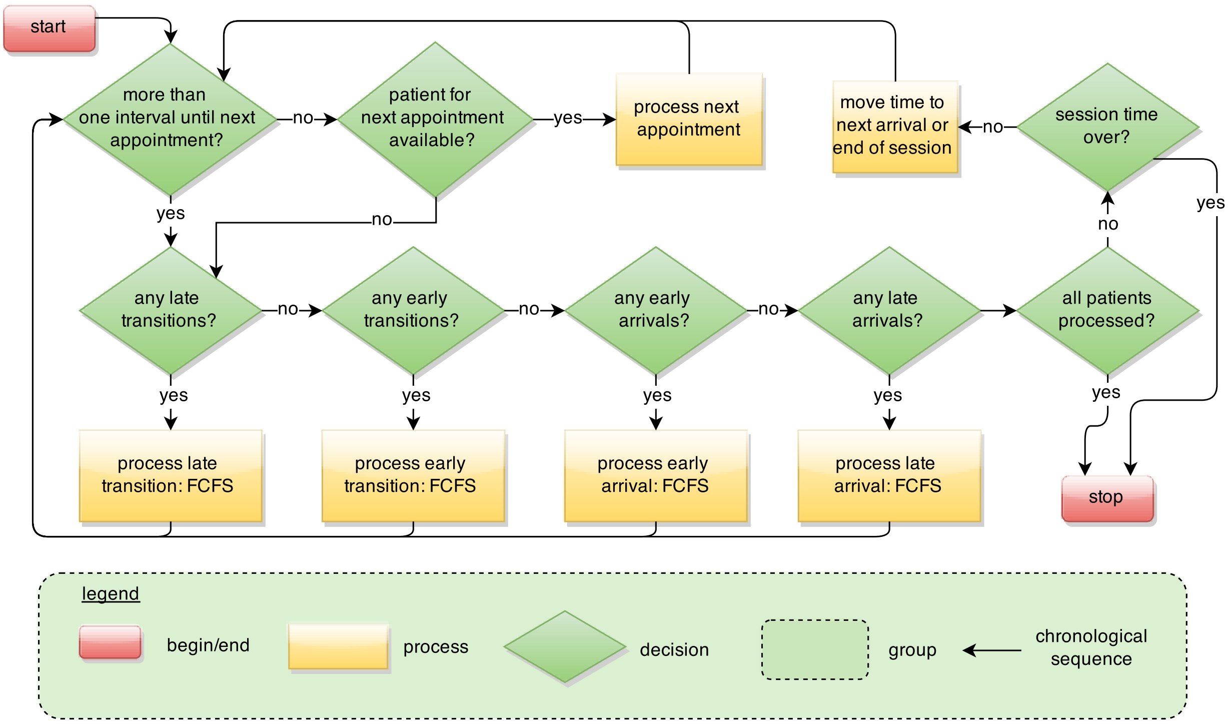


Figure 5.4: Dispatching rules used for the simulation of a real schedule. FCFS is first-come first-served.

5.6 Optimisation

5.6.1 Local search neighbourhood

For this problem, the state is represented by a tuple {FR, I, BI}: consisting of the first reservation time (FR), the appointment interval (I) for all appointments and the initial block (IB) that is the number of patients booked into the first reservation time. The search process can increase or decrease any part of the tuple by one defined unit: units are fifteen minutes for FR, five minutes for I and one patient for BI. Changing any of these single variables to create a new state defines the neighbourhood. To be accepted all variables must be greater than zero for any state: FR > 0, I > 0, BI > 0 (see Fig. 5.5 for an example state with the neighbourhood).

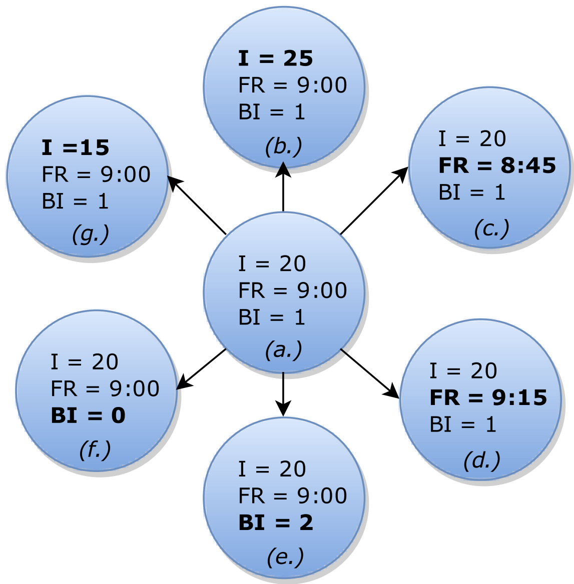


Figure 5.5: Local search neighbourhood example for schedule optimisation. Notice how setting the BI to zero (node f) is invalid, this node will be rejected during optimisation.

5.6.2 Local search cost function

The cost function requires manual tuning with respect to the important factors for the clinic (see [3, Sec. 3] for example cost functions). The cost function is a weighted sum of single objectives: the mean unplanned-late

$E(W^{ul})$, transition late $E(W^{tl})$ and transition early $E(W^{te})$ waiting times; number of patients above threshold A waiting times N_{TA}; the surgeon's mean idle time $E(S)$; the maximum unplanned-late waiting time $max(W_j^{ul})$ and overtime O:

$$CS = CS(W) + CS(N_T) + w_7E(S) + w_8max(W_j^{ul}) + w_9O \tag{5.4}$$

$$CS(W) = w_1E(W^{ul}) + w_2E(W^{tl}) + w_3E(W^{te}) \tag{5.5}$$

$$CS(N_T) = w_4N_{T10} + w_5N_{T20} + w_6N_{T40} \quad . \tag{5.6}$$

It is required to manually tune the weight vector w to find the values that give the correct balance of importance to the different scheduling objectives. For example, setting w_1 to one and all others to zero will change the schedule in any way possible in order to minimise only unplanned-late weighting; this might have unwanted results: large amounts of overtime and long periods of surgeon idle time. Therefore a balance between all the objectives needs to be found.

It is not clear what **overtime** is when there is not a definite end time for each session; overtime is defined as any time beyond the last original reservation time plus two intervals. For an interval of twenty minutes, and a last reservation of 14:00, overtime is any time after 14:40. **Idle time** is any time when the surgeon is waiting between appointments. If the session requires a lunch break, there should be approximately sixty minutes of idle time available.

Chapter 6

Empirical evaluation

6.1 Overview

The correlation of the numerical predictors linear relationship with the output response are analysed with correlation tests, for categorical variables ANOVA is used. Here there is some hope that some of the categorical predictors are significant. The accuracy of the supervised-learning prediction of the appointment duration is evaluated by comparing the error of the predictions with that of the interval used at the hospital and the mean duration time as recommended in the literature. A test for non-linearity shows that the data is likely to be non-linear. Finally, new schedules are constructed using the exact duration predictions to assess the results of how well the direct prediction works to predict the session's durations. A hill-climbing approach with a single objective demonstrates the ability of the local search optimisation for adjusting the appointment rules to achieve a very specific goal. Lastly, the method to be used for the real optimisation, using a multi-objective cost function is evaluated.

6.2 Supervised learning prediction

6.2.1 Feature selection

To get an idea of how good the features will be for prediction, correlation analysis, for numerical features, and analysis of variance (ANOVA), for categorical features, are applied to a randomly selected weekly session.

Each of the numerical predictor variables were tested for **correlation** with the response variable (appointment duration) (see Table 6.1). The correlation between the continuous (quantitative) variables gives some idea of how good they will be as predictors for the duration, but correlation does not imply causation. The results were not promising as there were no strong correlations. The correlation is weak for all the variables, it is unlikely they will make good linear predictors. The reservation time has the best result, but the patient age is not a good predictor.

patient age	appointment time	method
-0.07	0.23	Pearson correlation coefficients
-0.02	0.13	Kendall Tau correlation coefficients
-0.01	0.19	Spearman Rho correlation coefficients

Table 6.1: Correlation of numerical predictors with response variable.

ANOVA shows a relationship between each quantitative predictor variable (categorical feature) and the response (appointment duration). Box plots of the features with the highest ANOVA p-values (see Fig. 6.1) shows the effect of conditioning the response on these categorical variables [6]. An interesting result is how surgery in the past three to six months has a longer appointment duration time in general; if a patient is returning after this long period to speak to the surgeon it could be that there are complications which require long to discuss. These stronger predictors have very low p-values that is promising for the supervised learning accuracy.

6.2.2 Linear regression

Linear regression was used for prediction. The forward, backward and mixed feature selection methods were used to find the best features. Cross validation was used to get a root mean square error (RMSE) average over four future sessions that is used to compare the performance of each of the models. To measure the success of the prediction the results were compared with the existing method (using twenty minute intervals) and a method said to be commonly used in the literature by simpy using the mean of the duration times. The RMSE between each appointment duration and the twenty minute prediction.

The results show that the linear regression model performs much better overall than using a twenty minute fixed interval or the average duration fixed interval. The best results for feature selection was using mixed feature selection with a p-value threshold of 0.36. Note that the number of features shown are much higher than those listed because dummy coding the categorical features produced a large number of new dummy variables.

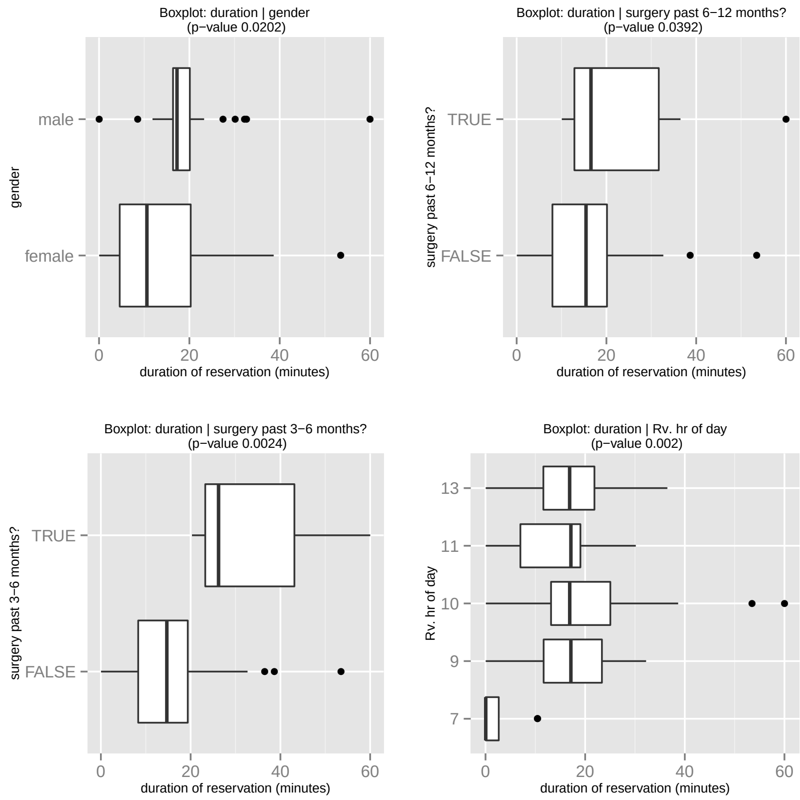


Figure 6.1: Boxplots of the most significant factor predictors for predicting the duration of a reservation. Each image shows the duration conditional on one of the factors, the p-value is attained from ANOVA f-statistic.

6.2.2.1 Test for non-linearity

A parametric model can introduce model bias by making false assumptions about the underlying structure of the regression function $f(x)$. A plot of the residual values [14] against the response can show if $f(x)$ is non-linear or not. If the residuals form a roughly straight line, with some evenly distributed noise, it would seem that the relationship is linear; if there is another pattern in the plot, this implies the relationship is non-linear.

The residual plot of the linear regression results shows that the relationship between x and y is likely to be non-linear (see Fig. 6.3). The pattern in the residual plot shows that the prediction starts below the true value and as the duration increases the errors increase. By taking the logarithm of the prediction $log(\hat{y} = f(X))$ the increase in the error can be reduced . This im-

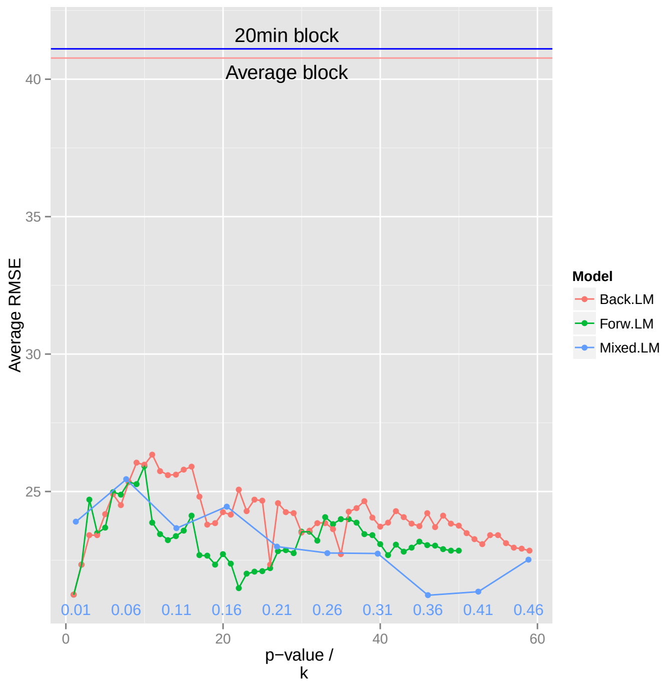


Figure 6.2: Root mean square error (RMSE) results for the different models and parameters using cross validation. A comparison of the RMSE of predicting each appointment as twenty minutes long '20min block'(as the current hospital schedule does for this resource) and for taking the average duration time from the past year appointments 'Average block', as an approach to a simple prediction. The mixed selection uses a p-value threshold, shown above the x-axis, forward and backward selection use a limit of k features.

plies a non-linear supervised learning model could perform better than linear regression. It seems that an exponential regression model may fit the data better.

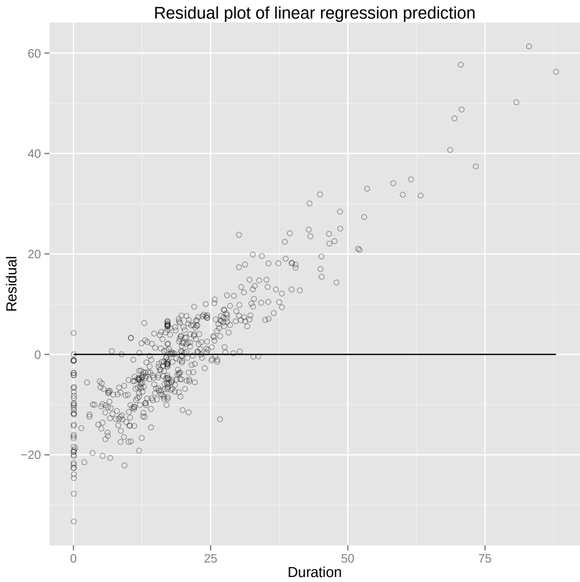


Figure 6.3: Residual plot for the linear regression model for predicting the duration of a reservation. The pattern in the data implies that the relationship between x and y is non-linear.

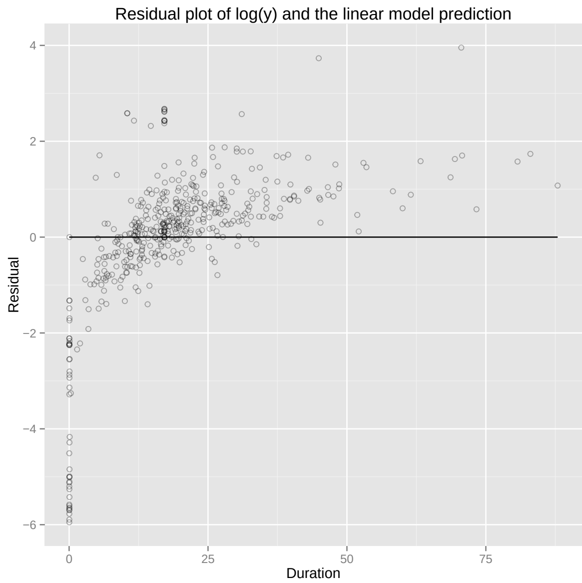


Figure 6.4: Residual plot for the exponential transformation of y in the linear regression model for predicting the duration of a reservation. The residuals now form a better fit than without the logarithmic transformation of the response variable.

6.3 Optimization

For each method developed here the final reservation times are scored by taking the actual times for that particular reservation and applying them with the new reservation times.

6.3.1 Using the predictions

A crude approach to test the quality of the predictions directly on the schedules is to start consecutive appointments at intervals given by the predicted appointment duration. This approach is very specific to a single schedule. This and the fact that reservation times are at very precise times, overcomplicating the schedule, make it a bad solution and it is only used for assessing the results (see Table 6.2 for the results of this, prediction chaining, approach).

weekly session	μ_{ULW}	N_{T10}	N_{T15}	N_{T20}	N_{T30}	N_{T40}	max.	S	patients
N2,Thur.	58	7	7	7	7	6	107	18	7
N2,Thur.*	109	7	7	7	7	7	199	51	7
N2,Mon.	17	4	1	1	1	0	40	4	5
N2,Mon.*	11	3	1	1	0	0	29	184	5
N1,Wed.	45	7	6	6	6	5	80	4	9
N1,Wed.*	3	2	1	0	0	0	16	195	9
P2,Mon.	27	3	3	3	2	2	98	59	8
P2,Mon.*	16	2	2	2	2	2	69	129	8
N1,Tues.	64	5	5	5	5	4	189	3	9
N1,Tues.*	9	4	2	0	0	0	19	0	9
N3,Wed.	17	6	4	4	1	0	34	68	8
N3,Wed.*	23	5	4	4	3	2	51	64	8

Table 6.2: Original unplanned waiting time score vs optimised (*) schedule waiting time score using prediction chaining. All time units are in minutes, measurements are the average unplanned-late waiting μ_{ULW}, the number of patients waiting above the threshold values N_{TA}, the maximum unplanned-late waiting time, max., the surgeon's idle time S, and the number of patients. Note that session N3, Wed, seems to process two patient simultaneously making it difficult to compare the optimised results with the original schedule (see Appendix G.1 for the session visualisations).

6.3.2 Hill climbing local search: single objective

A greedy hill-climbing approach starts with the existing schedule provided by the hospital and searches the local neighbourhood of appointment rules for better schedules until no better neighbouring schedule can be found (see Section 5.6.1 for a description of the neighbourhood and the search process). Each schedule's goodness is evaluated using a manually tuned cost function (see Section 5.6.2 for an exposition of the individual objectives).

To evaluate the performance of the optimisation, first a single objective of reducing the unplanned-late waiting was tried and then the more realistic objective of minimising a weighted combination of multiple objectives was evaluated. During local search optimisation each new appointment rule configuration is evaluated by running one hundred simulations of the day and taking the average score from the cost function.

Notice how the single objective optimisation only failed to minimise the waiting time for P2, Mon,. This session had a very long appointment duration that was most likely not predicted accurately leading. Also notice how the surgeon idle time has gone up by a large amount, beyond what is reasonable. Looking at the Gantt charts for the original and optimised schedules (see Appendix G.2 for the session visualisations) it can be seen that the overtime has also gone up very high in order to minimise the waiting time.

6.3.3 Hill climbing local search: multi objective

For the multi-objective cost function, each single objective weight had to be carefully chosen (see Table 6.4). Starting at the highest weighted objective a quick explanation for the choice of weight follows. The average idle time is undesirable for the surgeons, the surgeons do not want to wait between appointments not knowing when they will be interupted by the next patient, therefore each average minute of idle time is weighed heavily in the cost function. Reducing the waiting time should not be done at the cost of one or a few patients who carry the majority of the waiting, therefore the maximum unplanned-late waiting is given the second highest weighting. In order of severity, the number of patients waiting above the threshold are given weights equal to the threshold minutes. This gives priority to reduce the number of patients waiting above the larger thresholds than it does the lower thresholds. Transition-late waiting should be avoided, so this is given a higher weighting than the average late waiting. Transition-early waiting is not accurately simulated, nor is it as severe (per minute) as unplanned-late waiting, so it is giving a low weighting. A small amount of overtime is acceptable if large improvements can be made in the schedule, so this is included with an equal

weekly session	N2,Thur.	N2,Thur.*	N2,Mon.	N2,Mon.*	N1,Wed.	N1,Wed.*	P2,Mon.	P2,Mon.*	N1,Tues.	N1,Tues.*	N3,Wed.	N3,Wed.*
μ_{ULW}	58	9	22	21	47	1	27	8	64	0	58	0
N_{T10}	7	2	4	3	7	0	3	2	5	0	8	0
N_{T20}	7	2	2	3	7	0	3	1	5	0	8	0
N_{T40}	6	0	0	0	5	0	2	1	4	0	5	0
max.	107	27	40	40	80	6	98	47	189	0	442	4
S	18	55	4	184	4	202	59	138	3	214	101	129
patients	7	7	5	5	9	9	8	8	9	9	15	15
O	66	104	-27	-63	75	61	49	163	69	90	-3	23
μ_{TL}	0	0	0	91	8	0	0	0	0	0	1	0
μ_{TLW}	0	0	0	5	2	0	0	0	0	0	3	0
μ_{TE}	0	0	32	14	30	68	0	0	67	108	41	110
μ_{TEW}	0	0	17	13	15	13	0	0	34	26	34	44
lunch break	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FR	09:00	09:10	08:30	08:40	09:00	09:30	10:00	10:50	09:00	08:50	07:15	09:15
I	NA	40	NA	10	NA	40	NA	30	NA	45	NA	20
BI	NA	1	NA	2	NA	1	NA	1	NA	1	NA	1

Table 6.3: Performance metrics: original and hill climbing optimised (single-objective) schedules. Measurements include: unplanned-late waiting (ULW); transition-late (TL) that is the time between the reservation time and the end of the previous reservation; transition-late waiting (TLW); transition-early (TE) that is the time between the end of the previous reservation and the start of the appointment and transition-early waiting (TEW)

weighting as the average unplanned-late weighting (see Appendix G.3 for the session visualisations).

metric description	objective	weight
average late waiting	μ_{ULW}	10
overtime	O	10
idle time	S	100
# waiting over 10 mins.	N_{T10}	10
# waiting over 20 mins.	N_{T20}	20
# waiting over 40 mins.	N_{T40}	40
average transition-late waiting	μ_{TLW}	20
average transition-early waiting	μ_{TEW}	5
maximum late waiting	max.	50

Table 6.4: Cost function parameter settings for hill climbing optimisation

performance measure	N2,Thur.	N2,Thur.*	N2,Mon.	N2,Mon.*	N1,Wed.	N1,Wed.*	P2,Mon.	P2,Mon.*	N1,Tues.	N1,Tues.*	N3,Wed.	N3,Wed.*
μ_{ULW}	58	89	22	20	47	0	27	18	64	1	58	0
N_{T10}	7	7	4	3	7	0	3	2	5	1	8	0
N_{T20}	7	7	2	3	7	0	3	2	5	0	8	0
N_{T40}	6	5	0	0	5	0	2	2	4	0	5	0
max.	107	172	40	35	80	1	98	71	189	12	442	7
S	18	51	4	184	4	195	59	118	3	83	101	55
patients	7	7	5	5	9	9	8	8	9	9	15	15
overtime	66	99	-27	-63	75	56	49	138	69	-40	-3	-51
μ_{TL}	0	0	0	89	8	3	0	0	0	0	1	2
μ_{TLW}	0	0	0	5	2	0	0	0	0	0	3	3
μ_{TE}	0	0	32	13	30	40	0	0	67	68	41	70
μ_{TEW}	0	0	17	12	15	12	0	0	34	36	34	44
lunch break	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
FR	09:00	09:00	08:30	08:35	09:00	09:35	10:00	10:45	09:00	08:55	07:15	09:10
I	NA	25	NA	10	NA	30	NA	20	NA	25	NA	15
BI	NA	1	NA	1	NA	1	NA	1	NA	1	NA	1

Table 6.5: Performance measures and appointment rule configuration: original and hill climbing (multi-objective) optimised (*) schedules. Measurements include: average unplanned-late waiting (μ_{ULW}); average transition-late (μ_{TL}), that is the time between the reservation time and the end of the previous reservation; average transition-late waiting (μ_{TLW}); transition-early (μ_{TE}) that is the time between the end of the previous reservation and the start of the appointment and transition-early waiting (μ_{TEW}). The appointment rule parameters used are reported: the first reservation time (FR), the fixed appointment interval (I) and the initial block count (BI) (see Appendix G.3 for the session visualisations).

Chapter 7

Discussion

appointment duration variance

In the literature it is mentioned how large variance in the appointment durations makes it difficult to make a good schedule [3]. An example of this is in the optimisation of N2, Thursday, weekly session (see the Gantt chart in Fig. G.13). It can be seen in the original schedule that a long appointment duration on the fourth appointment, just before a lunch break, caused a large delay for the following reservations (even longer than the preceding delays). During optimisation a similar issue occurs, where the prediction for the duration during simulation seems to have been underestimated and so the reservation intervals are underestimated. Improving this prediction would improve this scheduling optimisation.

Bailey-Welch rule

The literature also claimed that the Bailey-Welch rule was often near to optimal [17], but in experiments with this data it performed poorly: this was likely due to almost no late patients and the fact that all patients showed for the appointments. It can also be seen that the optimised schedule rarely started the initial block with two patients and the appointment interval was often not the average.

Local search optimisation

When optimising a schedule by changing the sequence of the appointments, something asked not to do by the clinic for this problem, the literature reported that Tabu search [17] was the best performing optimisation method. Here, a similar, but more greedy, local search method is able to find a good appointment rule configuration.

waiting-idle relation

The relationship between the patient waiting time and surgeon idle time [1, 3] for different interval lengths can be seen when optimising the schedule for unplanned-late waiting time as a single objective. To reduce the patient waiting time the optimisation algorithm largely increased the idle time of the surgeon.

Future implementation

The hospital have started taking action to put together a report from the analysis of the waiting times produced in this study. This highlights the situations in which patients are being made to wait. A meeting with the surgeons and other staff at the hospital is organised where this study, the analysis and the optimisation, will be presented and discussed. The nurse in charge of the administration in the clinic, Sanna, is in discussions about how to begin implementing parts of this system. The parts of the infrastructure for overriding the appointment times to be sent to customers is all in place: X-akseli currently read the reservation times prior to the session date, and X-akseli have an SMS messaging system (to be used to message the new reservation times) setup in the hospital used for other parts of the system.

Some aspects need to be agreed with the hospital: The cost function will need to be tuned to the desires of the hospital. It was seen that unplanned-late waiting can be largely reduced, but at the cost of surgeon idle time, overtime and transition waiting patients: is this what they want? The current system is myopic: it focuses mainly on the single surgeon's appointments not able to change reservation times prior to the appointment, nor considering reservations succeeding the appointment. The system could be extended to manage these. During the meeting these things will be discussed.

System future development

Each component of the system (see Section 5.2.1) is able to be focused on for improvement in isolation: finding better features to extract; improving methods of feature selection, and with a lot of features, dimensionality reduction; finding supervised learning models that are more accurate; improved optimisation algorithms to find better appointment rules and with more flexibility (the ability to change the order of the appointments for example); improvements to the dynamic scheduling during simulation to replace the dispatching rules, e.g., using reinforcement learning (RL) to adapt the dispatching rule during the session. There is a lot of potential for future improvements.

patient categorisation

To improve the duration predictions, patients could be categorised into similar duration time groups. With the existing data, efforts could be made to categorise the data for the duration based on the previous resource, any prior surgeries. Clustering techniques may work here to find groups. These categories could be used at the hospital when booking in patients.

Extract more features

A future improvement for feature extraction would be to acquire details of the appointment, more patient demographic information, patient history and any other patient details that can effect the duration of the appointment. This data is accessible to X-akseli but more understanding is required to have access to these details for reasons of confidentiality.

Non-linear supervised learning

The tests for non-linearity showed that the data is likely to be non-linear. Using other methods of supervised learning regression for prediction may improve the prediction accuracy, for example, exponential regression, k-nearest neighbour (kNN, see Section 4.7), support vector regression (SVR), artificial neural networks (ANN) or Gaussian processes (GP). However, with the limitations on the features available it may be more effective to try to access more patient appointment details to improve the prediction of the duration.

no-shows

The literature explains how patient's late arrivals and no-shows should be managed in the schedule by overbooking as these vitiate the schedule, but as these were found to be very low at the clinic the no-show eventuality has not been addressed in this system. In other areas of the hospital these could require addressing, and so the system could be extended to simulate the no-shows.

transition ready prediction

The prediction of the transition arrival is very general. Further improvements could be made by more accurately predicting these. The issue of the diagnostic examination report is also not addressed well as the data for this is lacking. This data is planned to be aquired: with this data a more accurate transition ready state could be predicted.

dispatching rules

The existing dispatching rule takes into consideration the different waiting categories that each patient falls into and prioritises the patients for call-in on this basis. This is recommended to be used by the surgeons when calling in patients at the clinic.

local search

More state-of-the-art local search methods could be evaluated for the scheduling optimisation. For example, a simple but effective recent local search algorithm is step-counting hill-climbing [2].

simulation

The simulation could also be improved by predicting the durations again each time the reservation times change, as the time of day can effect the duration of the schedule. Reservations at the start or end of the session tend to be shorter than those in the middle, this is dome shape phenomena [9, pg. 811].

publicly available R package

This scheduling system is implemented in R. In the future a package could be made that requires data to be available in a certain format (the patient flow table for example), and it could work for any data given. I am in discussion with the company about this.

Chapter 8

Conclusions

Using appointment rules, optimised using simulation and local search methods, new schedules for the clinic were produced with a focus on fairly reducing patient waiting times. A system to continually optimise future clinic schedules is demonstrated and tested. Optimising future schedules using only historical data and testing the schedule with the real observed times for the optimised session, the schedules are shown to perform well with some room for improvement.

Late-starts, high variance in appointment durations and the transition arrivals are the main factors causing long waiting times. Using supervised learning and scheduling optimisation algorithms the schedule is improved (based on the performance measures in Table F.1).

Late-starts were analysed and found to be reasonably consistent so a simple linear regression model was used to predict the surgeon's arrival. Starting the session after the surgeon arrives is worst than starting early and making the patient wait, therefore the mean response is not used directly, a prediction interval is used and the earlier time from the interval is used. During optimisation a distribution fitted to the surgeon arrival times is sampled from to find the start time that gives the lowest cost for the schedule. Looking at the results sessions are not starting late and patients are waiting less for the session start.

The high variance of the appointment duration and lack of strong features makes prediction very difficult, but predictions were better than the hospital's twenty minute block and better than taking the average duration time.

The final evaluation of the schedules optimised with multi-objective optimisation largely reduced the overtime in comparison to the single-objective optimisation, the idle time, however, remained high, the cost of idle time could be increased more to reduce this further for future changes to the cost function.

Bibliography

[1] Bailey, N. A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society A14* (1952), 185–189.

[2] Bykov, Y., and Petrovic, S. A step counting hill climbing algorithm. *Nottingham University Business School Research Paper*, 2013-10 (2013).

[3] Cayirli, T., and Veral, E. Outpatient scheduling in health care: a review of literature. *Production & Operations Management 12*, 4 (2003), 519–548.

[4] Denton, B., and Gupta, D. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions 35*, 11 (2003), 1003–1016.

[5] Denton, B., Viapiano, J., and Vogl, A. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science 10*, 1 (2007), 13–24.

[6] Dodge, Y. *The Concise Encyclopedia of Statistics (Springer Reference)*, 1 ed. Springer, Apr. 2008.

[7] Glover, F. Tabu search: A tutorial. *Interfaces* (1990).

[8] Gower, J. C., and Gower, J. C. A general coefficient of similarity and some of its properties. *Biometrics* (1971).

[9] Gupta, D., and Denton, B. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions 40*, 9 (2008), 800–819.

[10] Han, J., Kamber, M., and Pei, J. *Data Mining: Concepts and Techniques, Second Edition (The Morgan Kaufmann Series in Data Management Systems)*, 2 ed. Morgan Kaufmann, Jan. 2006.

[11] Harper, P., and Gamlin, H. Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. *OR Spectrum 25*, 2 (2003), 207–222.

[12] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer, 2009.

[13] Hulshof, P., Kortbeek, N., Boucherie, R., Hans, E., and Bakker, P. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems 1*, 2 (2012), 129 – 175.

[14] James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*, 1st ed. Springer Publishing Company, Incorporated, 2013.

[15] J.D.Welch. Appointment systems in hospital outpatient departments. *Operational Research Quarterly 15*, 3 (1964), 224–232.

[16] Kaandorp, G. C., and Koole, G. Optimal outpatient appointment scheduling. *Health Care Management Science 10*, 3 (May 2007), 217–229.

[17] Kapamara, T., Sheibani, K., Haas, O. C. L., Reeves, C. R., and Petrovic, D. A review of scheduling problems in radiotherapy. *Proceedings of the eighteenth international conference on systems engineering (ICSE2006)* (2006), 201–207.

[18] Kutner, M., Nachtsheim, C., and Neter, J. *Applied Linear Statistical Models*. Irwin, 2004.

[19] Pinedo, M. L. *Scheduling: Theory, Algorithms, and Systems*, 4th ed. Springer Publishing Company, Incorporated, 2012.

[20] Press Ganey. Keep me waiting: Medical practice wait times and patient satisfaction. Tech. rep., July 2009.

[21] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[22] Rice, J. *Mathematical Statistics and Data Analysis*. No. p. 3 in Advanced series. Cengage Learning, 2006.

[23] Salkind, N. J., Ed. *Encyclopedia of Measurement and Statistics.* SAGE Publications, Inc., 2007.

[24] Sammut, C., and Webb, G. I., Eds. *Encyclopedia of Machine Learning.* Springer, 2010.

[25] Vanden Bosch, P., and Dietz, D. Minimizing expected waiting in a medical appointment system. *IIE Transactions 32*, 9 (2000), 841–848.

[26] Wickham, H. *ggplot2: elegant graphics for data analysis.* Springer New York, 2009.

[27] Zonderland, M. E. *Appointment Planning in Outpatient Clinics and Diagnostic Facilities.* Boston, MA: Springer US, 2014.

Appendix A

X-akseli data preprocessing summary

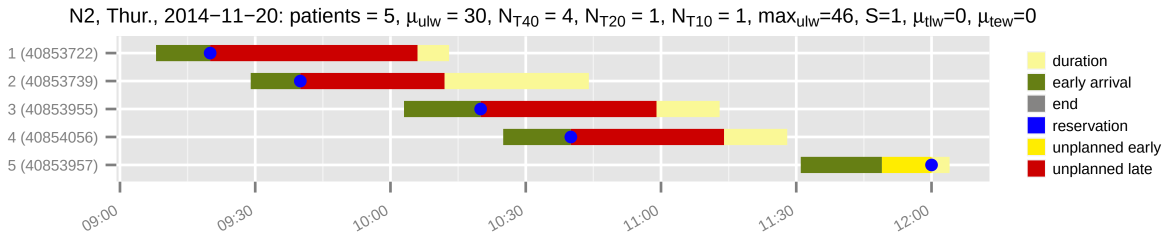
Problem	Solution	Data	Comments
Privacy	Replace personal details with dummy details	Patient data	This is required by law
Multi-waiting	Associate the waiting times to the correct appointments	Waiting time	A patient can be waiting for several appointments at once, this is a feature of the X-akseli system
Very long reservations	Remove	Duration of a reservation	Above 6 hours is unrealistic
Patient missing	Replace waiting period with missing	Waiting time	If a patient is called in many times they are deemed missing
Paused treatment	Treatments went to waiting status when paused, this is replaced with the paused status.	Waiting time	A treatment could be paused if a patient must wait after taking some medicine among other reasons.
Rejected call-in	Replace call-in status with rejected call-in status	Waiting time and duration	A call-in can be rejected, this should not be used as a start time for a reservation, so a new status is used.
Cancelled call-in	Rename status of in-treatment-room with called-to-treatment-room for cancelled call-ins	Waiting time and duration	
Called back after paused	New status for in-treatment-room-from-paused	Waiting time and duration	
Missing call-in	Replace with end time minus mean appointment duration	Waiting time and duration	All replaced data is flagged and omitted from statistical analysis; used in the Gantt charts
Identify resource groups	Use filters to find particular groups	Resources	Identify diagnostic examination, surgery appointments and surgery resource groups

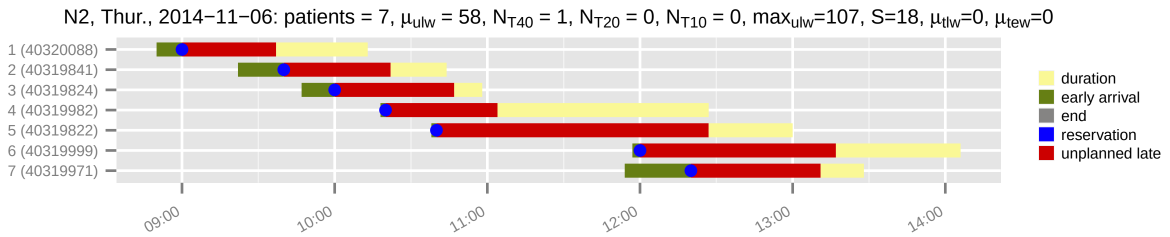
Table A.1: X-akseli data preprocessing summary

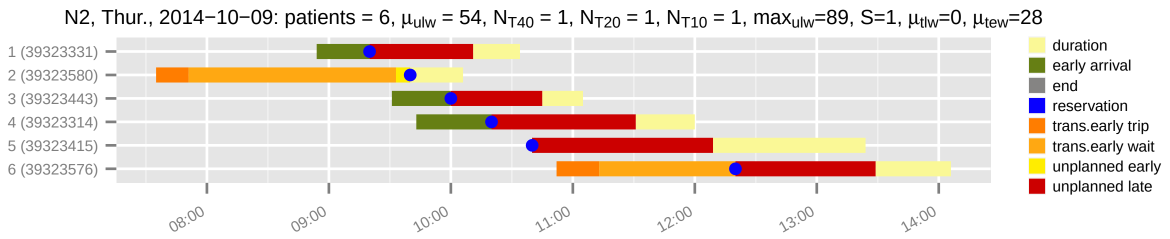
Appendix B

Session visualisation

Visualisations for the past four most recent sessions for the highest ranked weekly sessions that were a focus of this study. See Section 3.4 for a description of the waiting categories used.







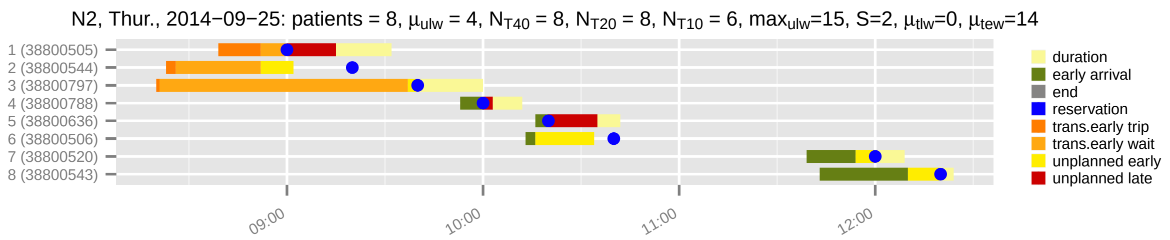


Figure B.1: Thursday's session for resource N2

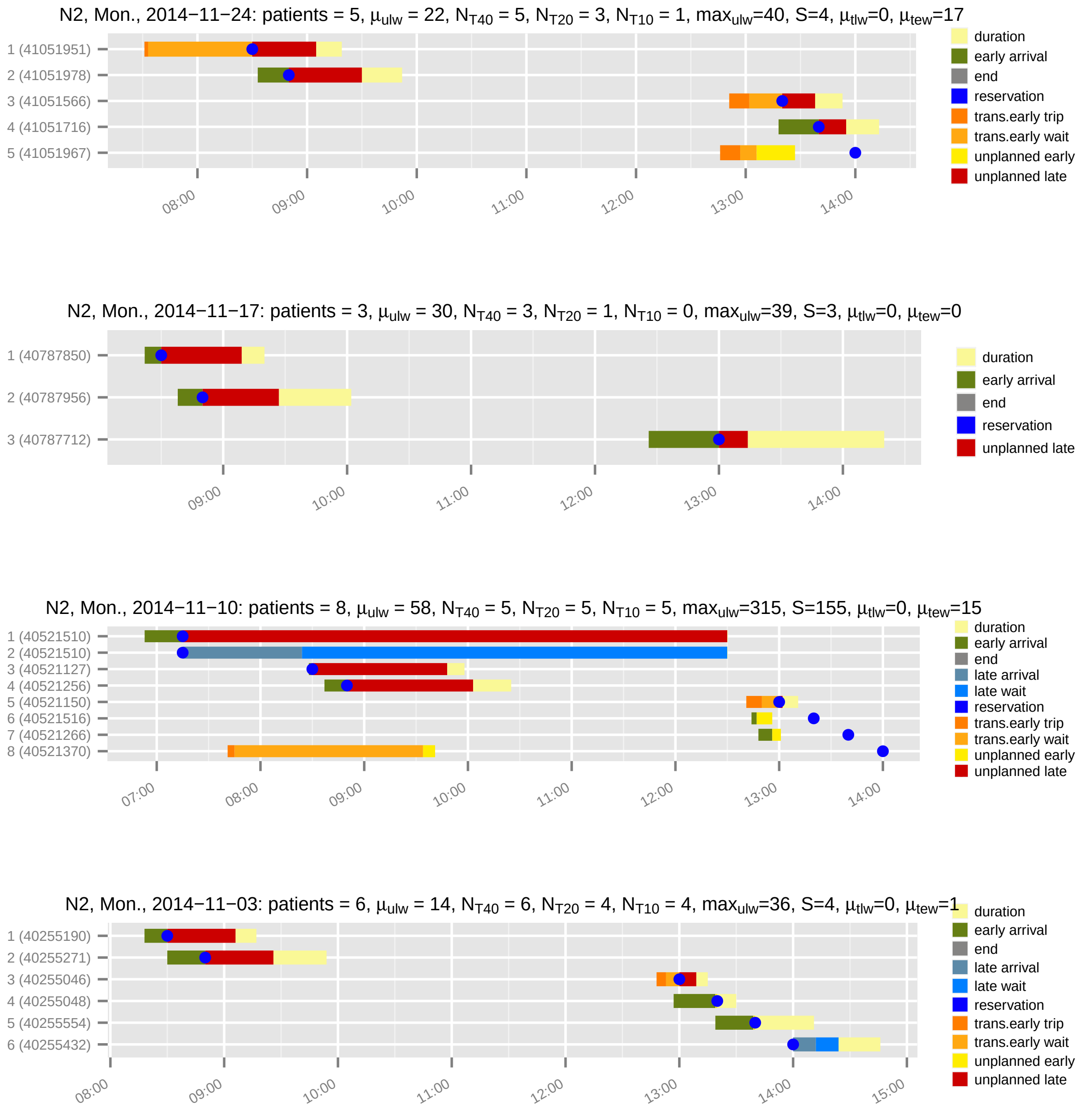


Figure B.2: Monday's session for resource N2

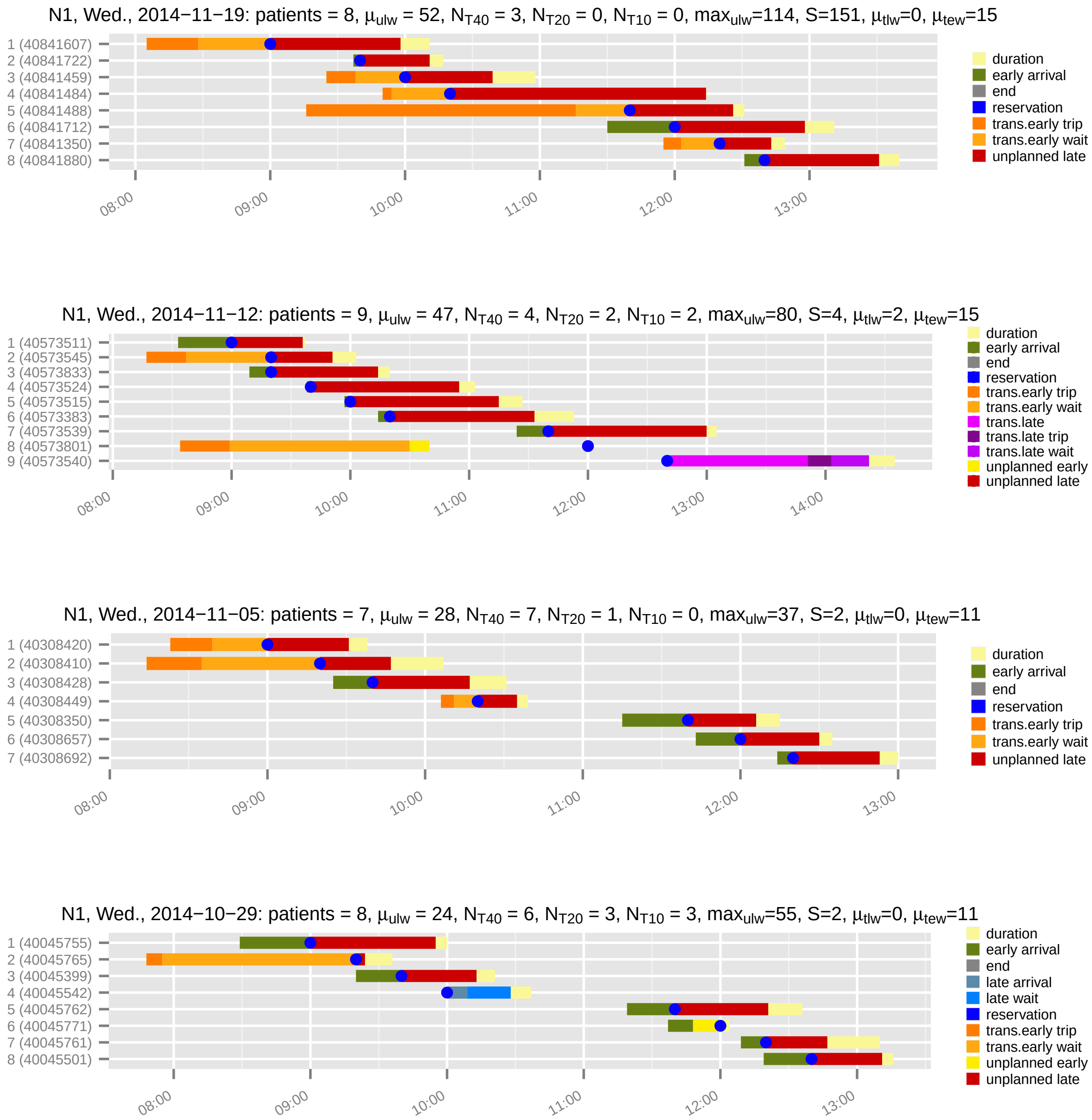
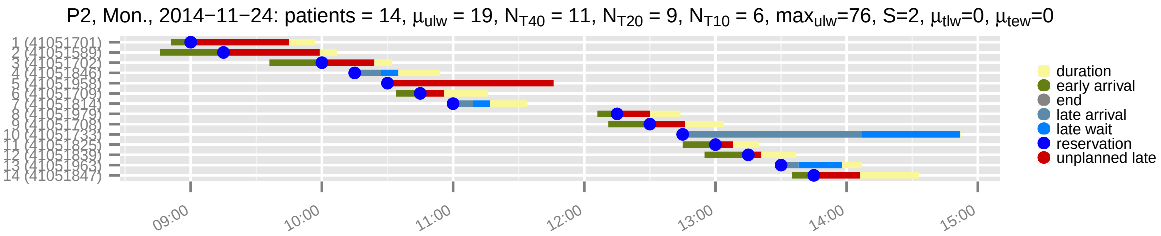
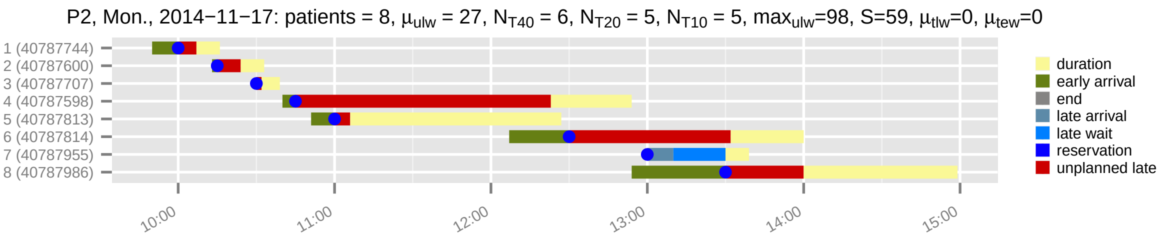
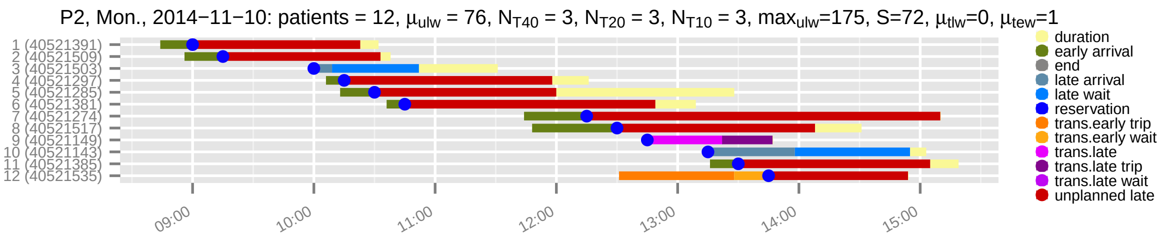


Figure B.3: Wednesday's session for resource N1







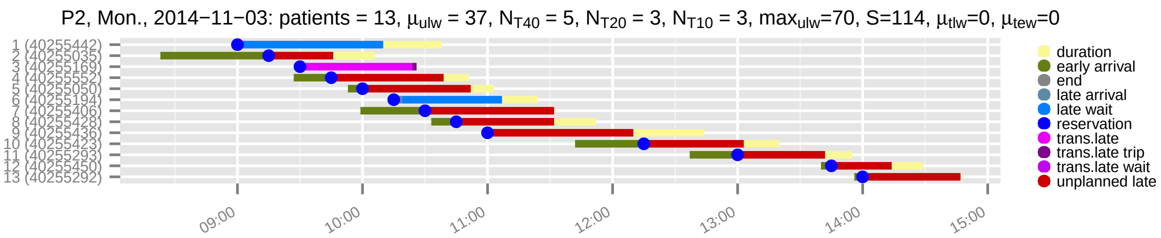


Figure B.4: Monday's session for resource P2

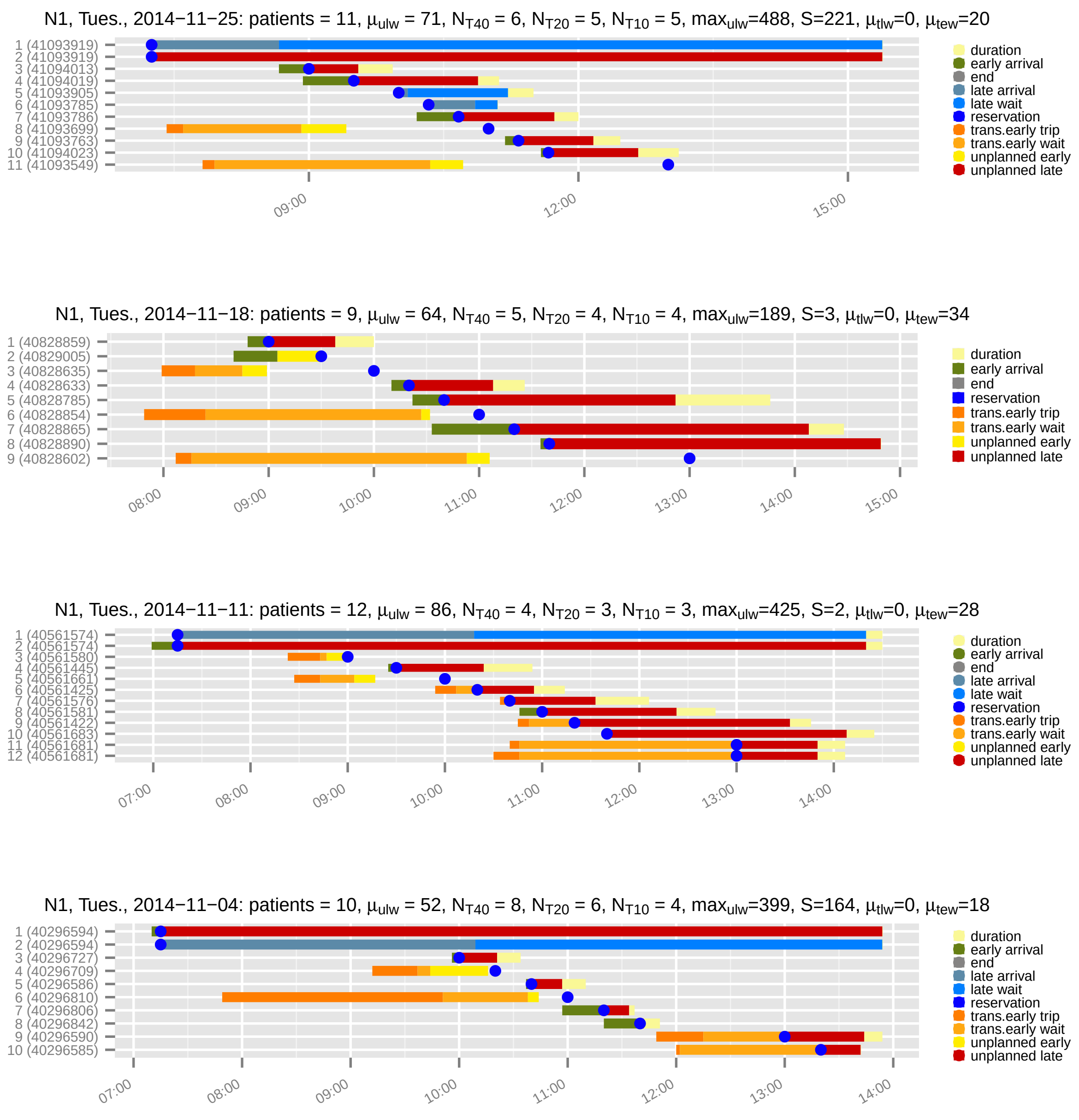


Figure B.5: Tuesday's session for resource N1

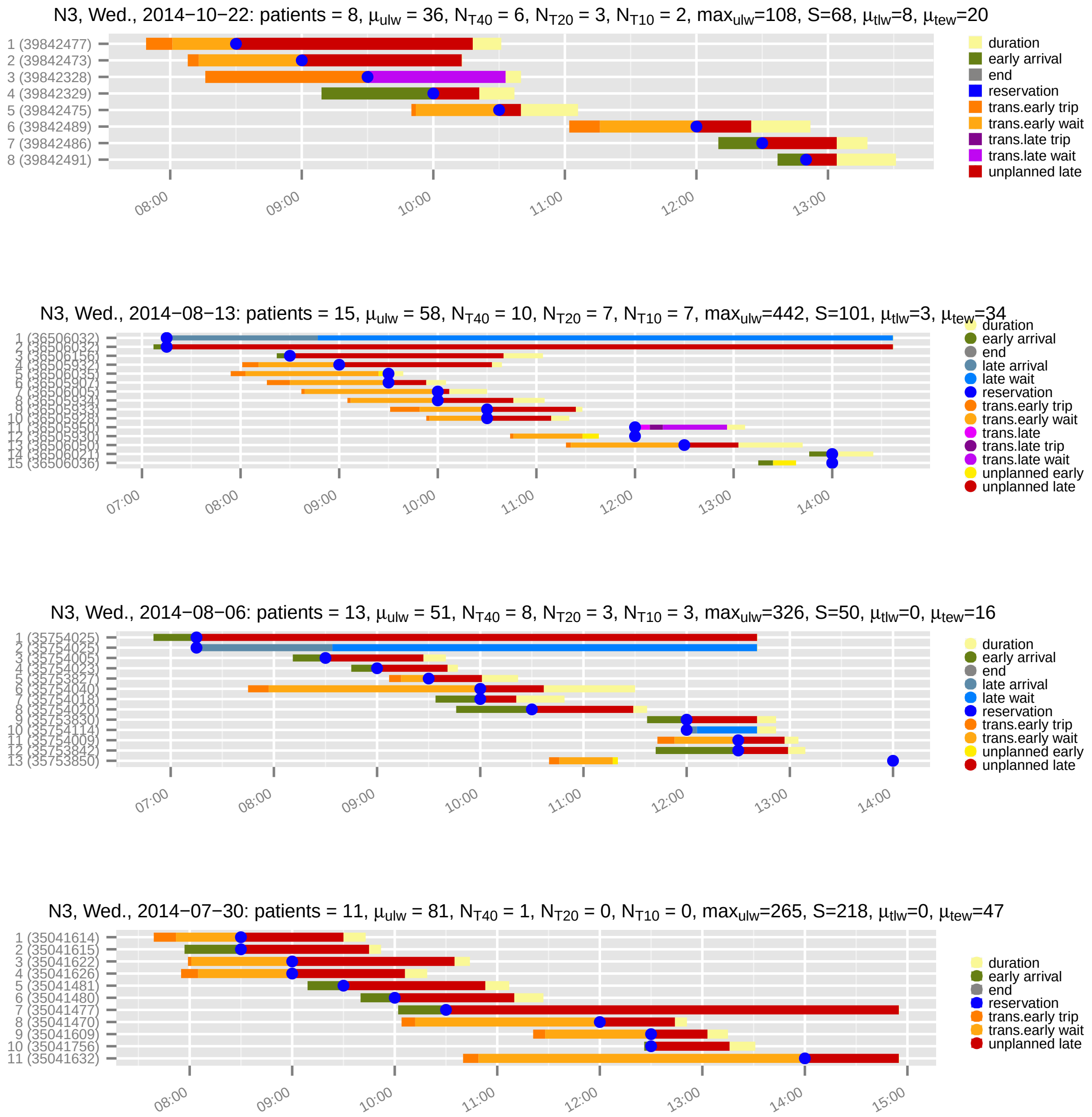


Figure B.6: Wednesday's session for resource N3

Appendix C

Clinic analysis plots

C.1 Duration

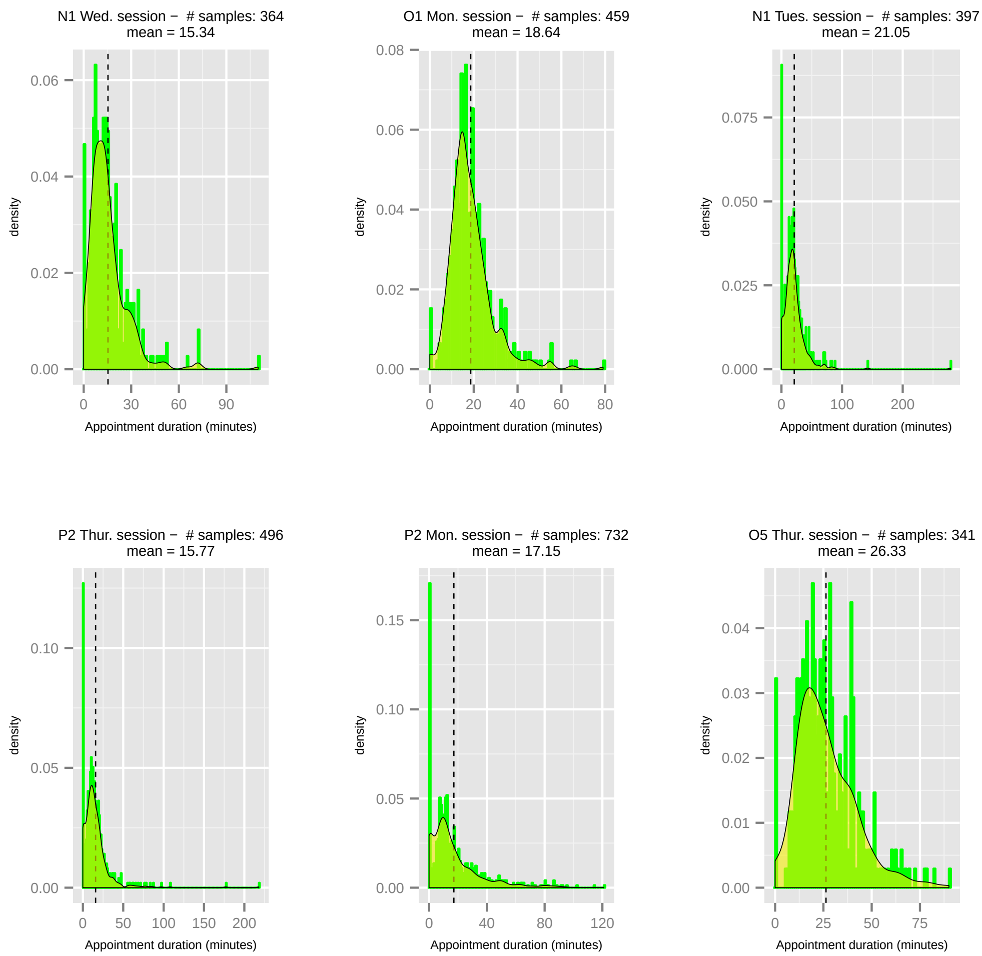


Figure C.1: Histograms of duration lengths with density estimation for the top-six ranking weekly sessions.

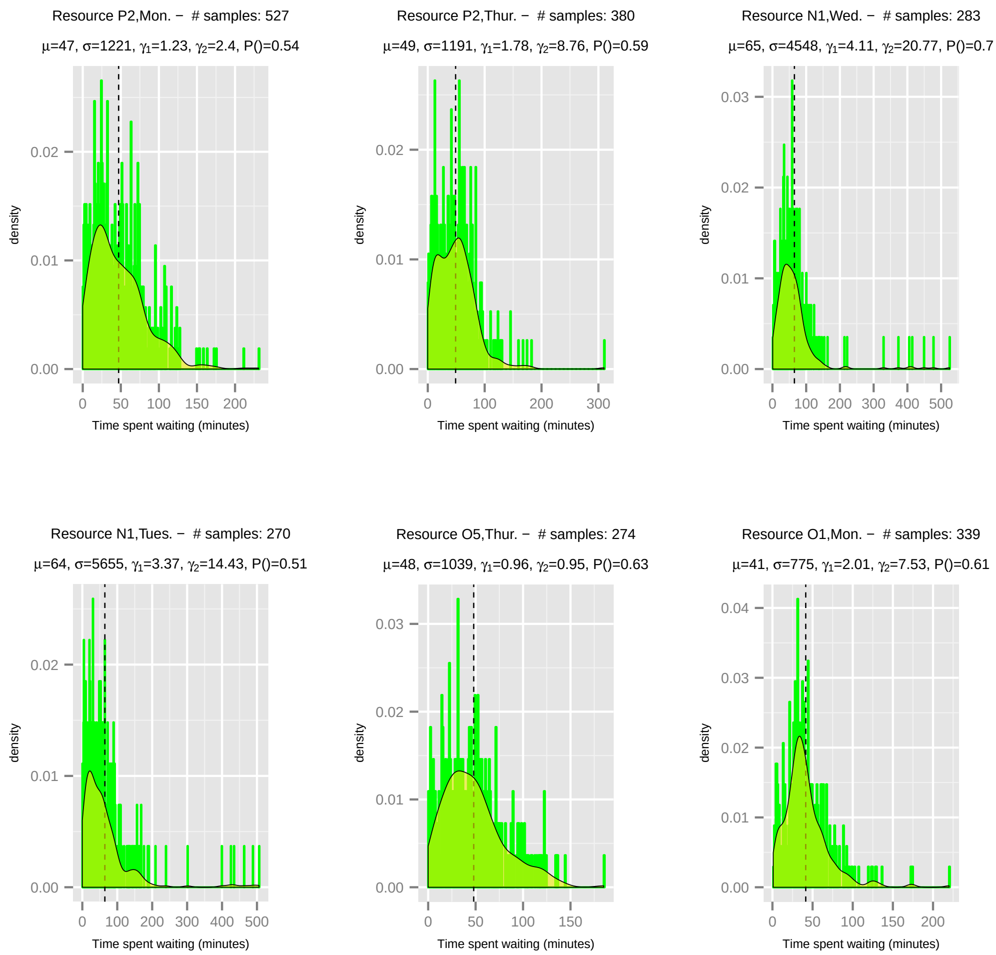


Figure C.2: Histograms of unplanned-late waiting with density estimation for the top-six ranking weekly sessions. $P() > 30$ is the percentage of the waiting times above thirty minutes.

C.2 Early transition

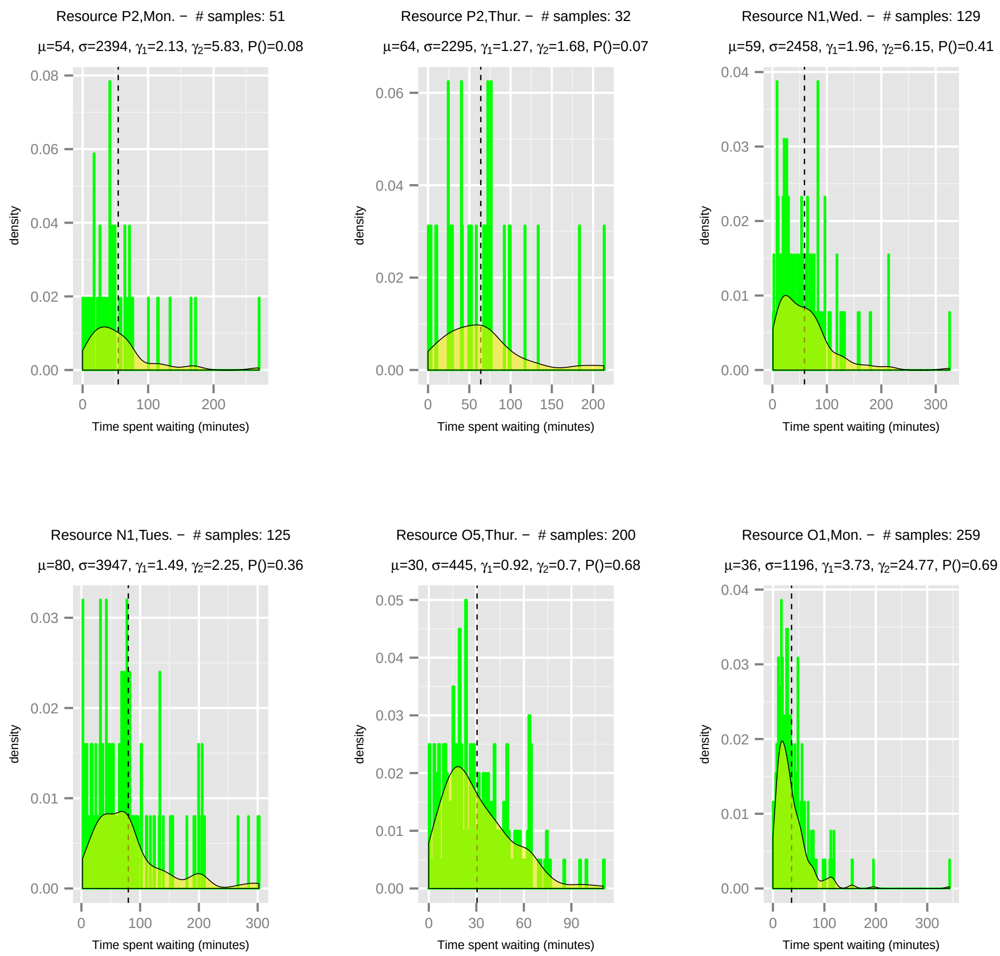


Figure C.3: Histograms of transition-early arrival (prior to the reservation time in minutes) with density estimation for the top-six ranking weekly sessions. P() is the percentage of early-transitions accounted for amongst all early arrivals.

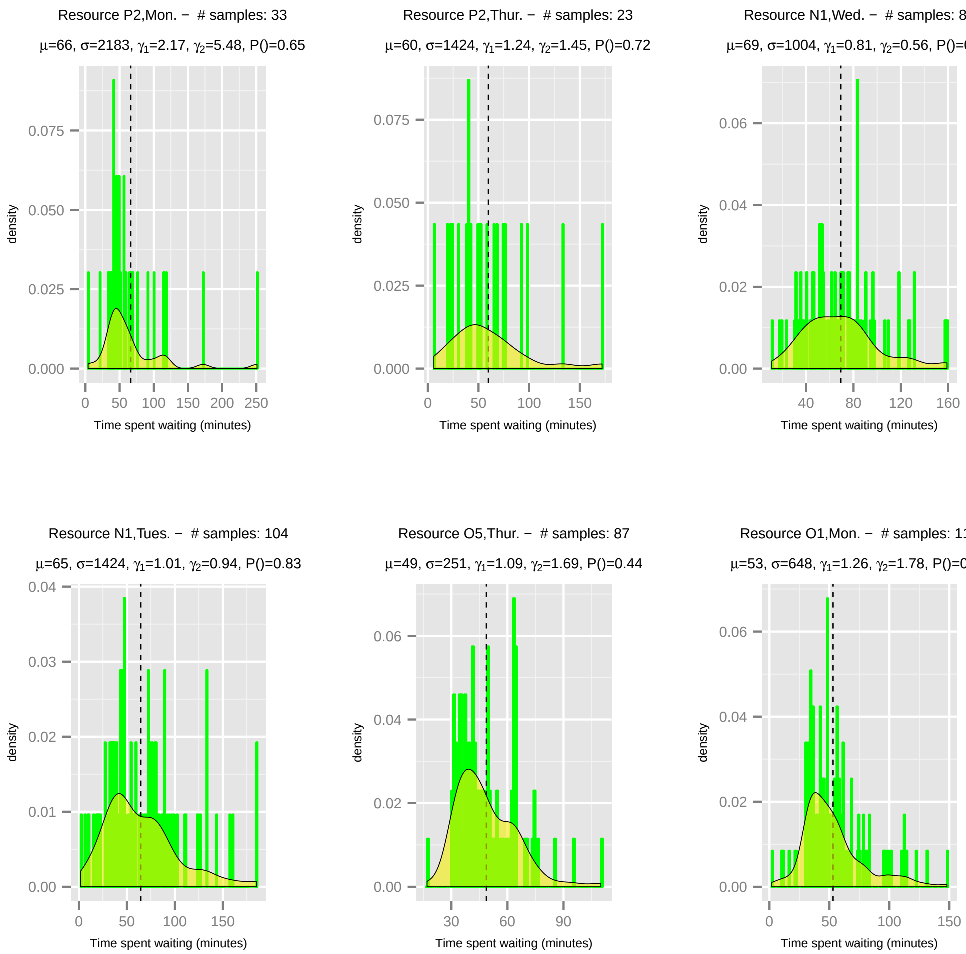


Figure C.4: Histograms of transition-early waiting with density estimation for the top-six ranking weekly sessions.

C.3 Late transition

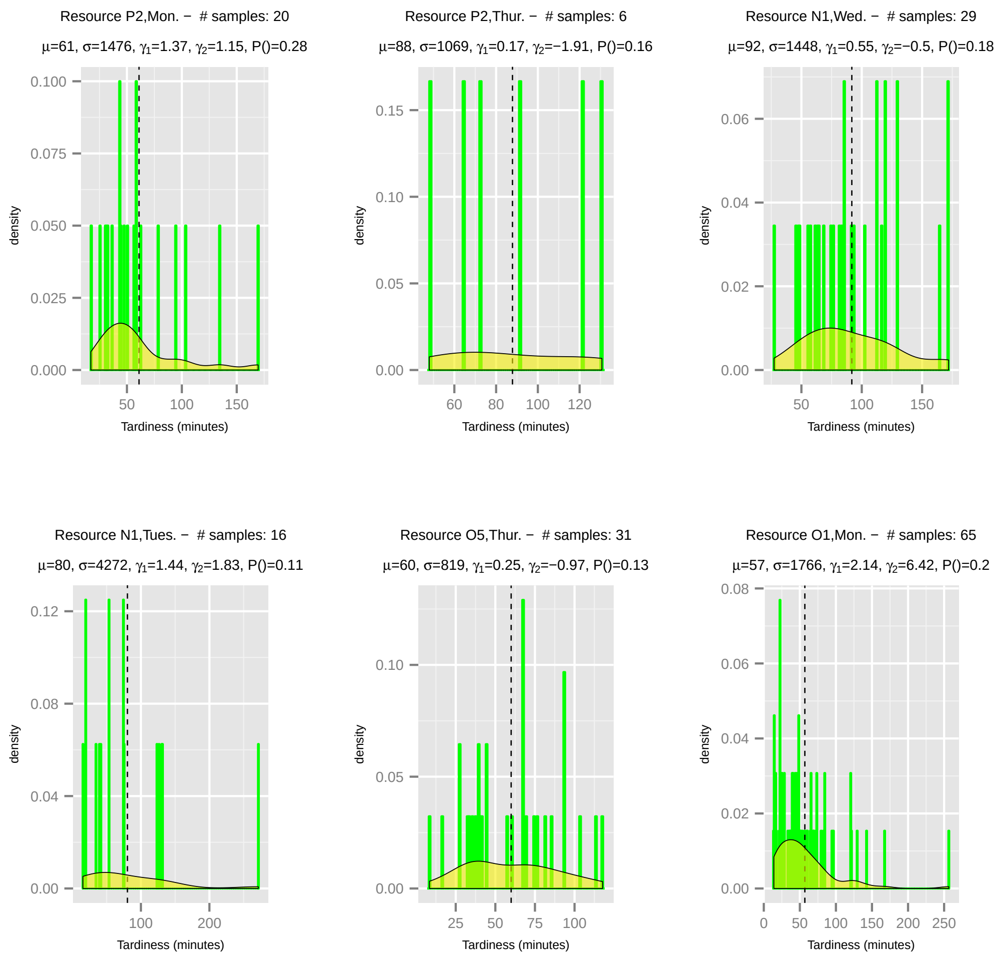


Figure C.5: Histograms of transition-late tardiness with density estimation for the top-six ranking weekly sessions. $P()$ is the percentage of transitions arrivals that occur late.

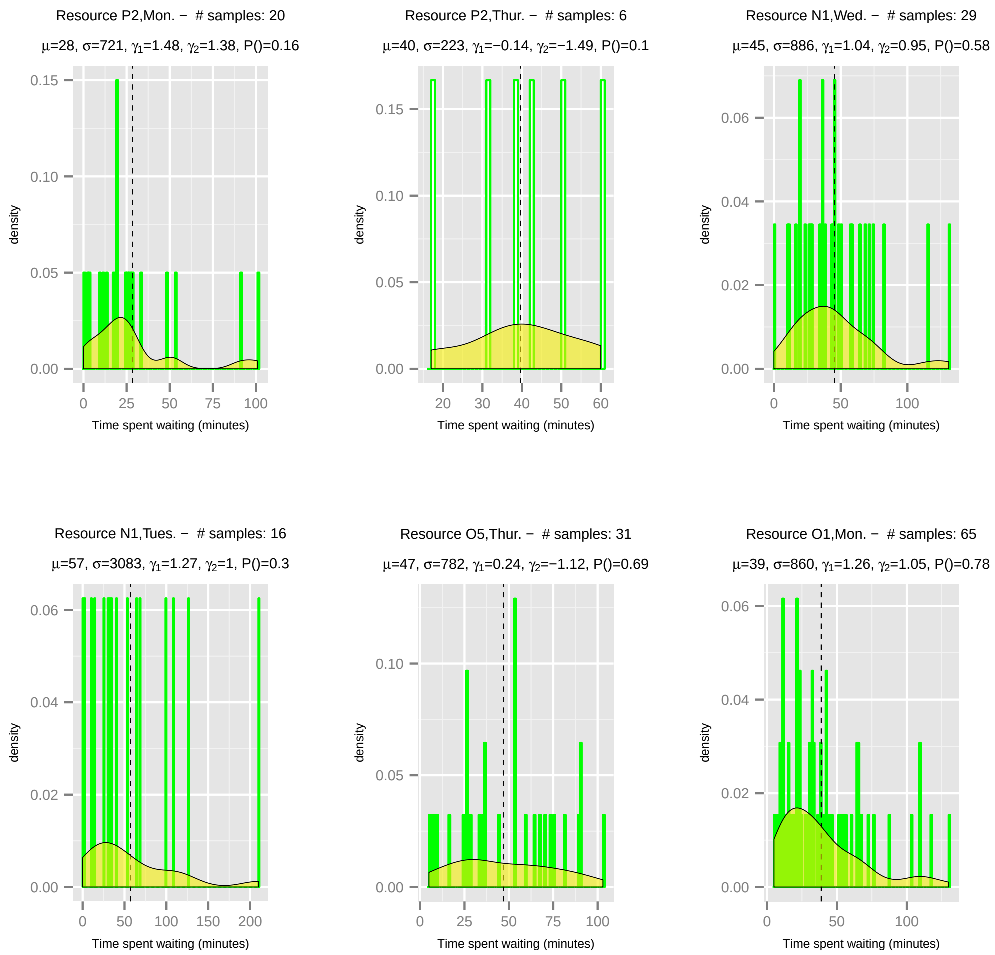


Figure C.6: Histograms of transition-late waiting with density estimation for the top-six ranking weekly sessions. $P()$ is the percentage of late arrivals that are due to a transition as oppose to the patient arriving late to the hospital.

C.4 Late start

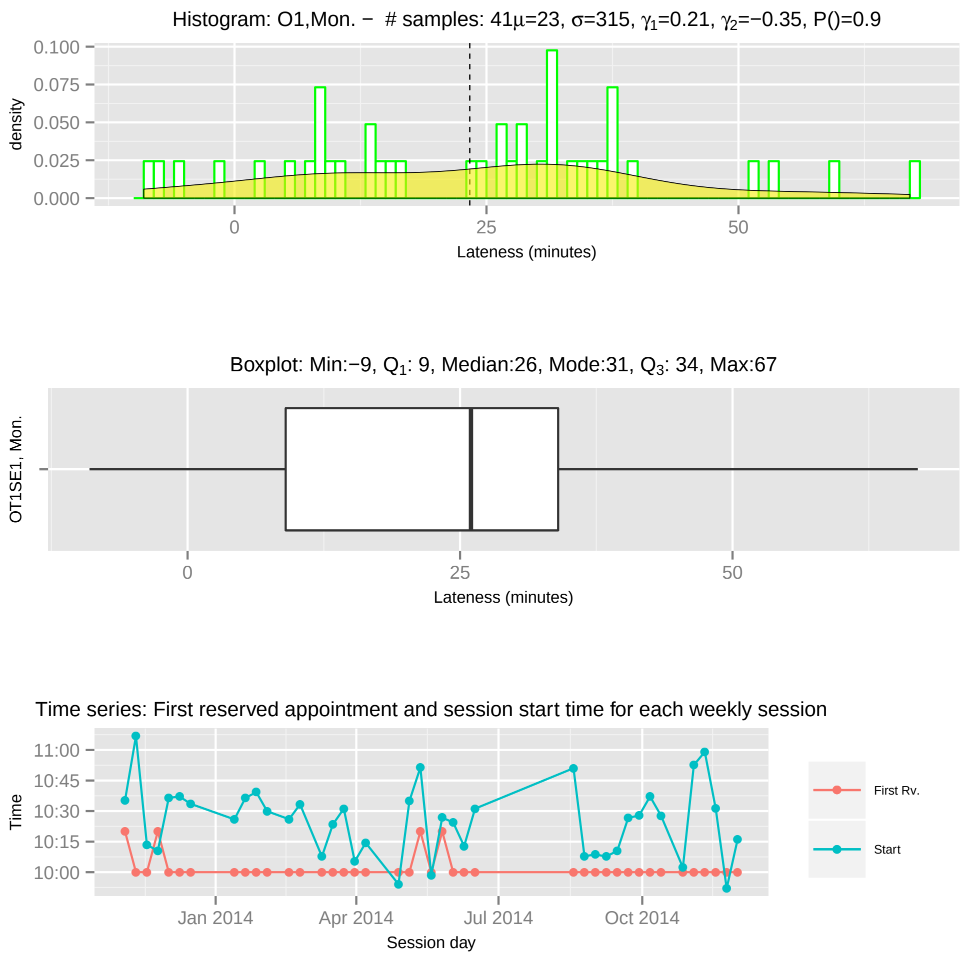


Figure C.7: Histogram, boxplot and time-series plots of the late starts for the top-six ranking weekly sessions. $P()$ is the percentage of session starts that were late.

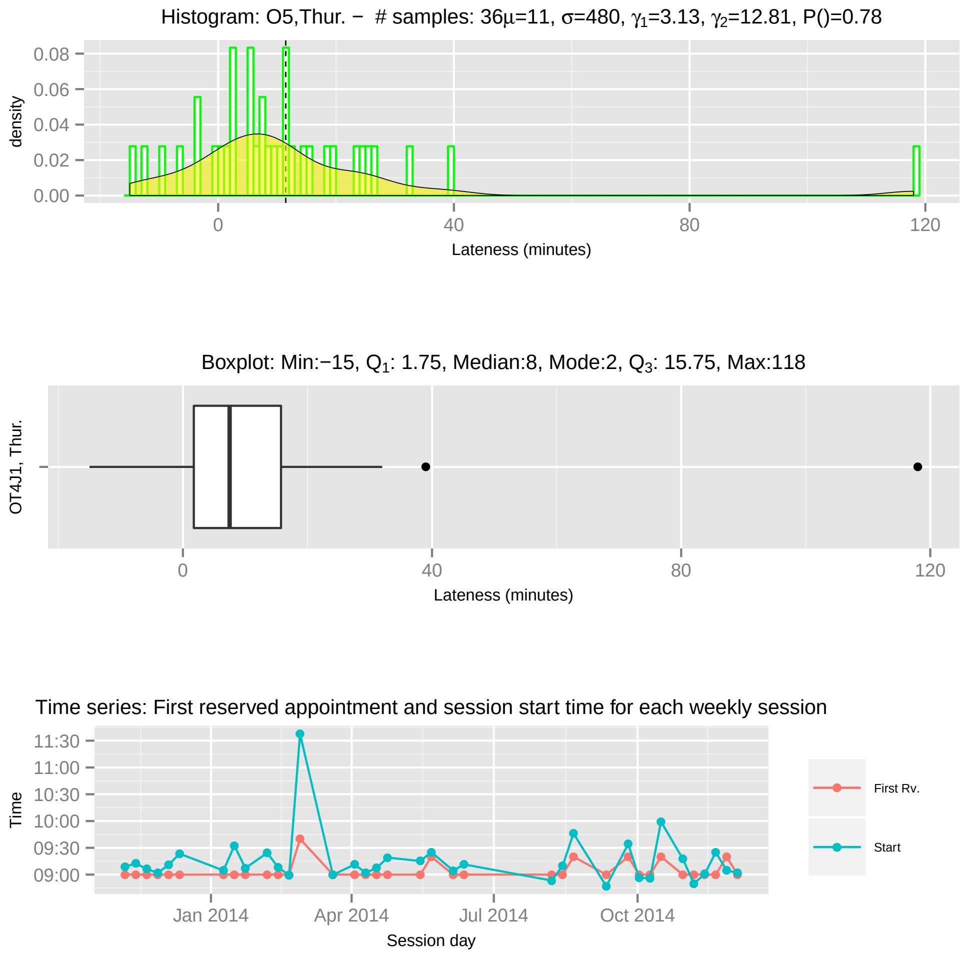
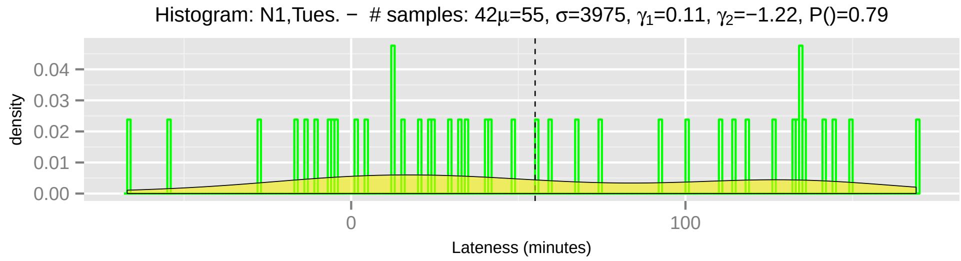
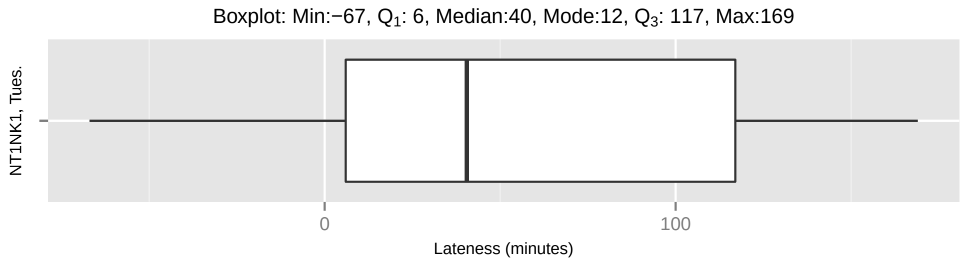


Figure C.8: Histogram, boxplot and time-series plots of the late starts for the top-six ranking weekly sessions. $P()$ is the percentage of session starts that were late.





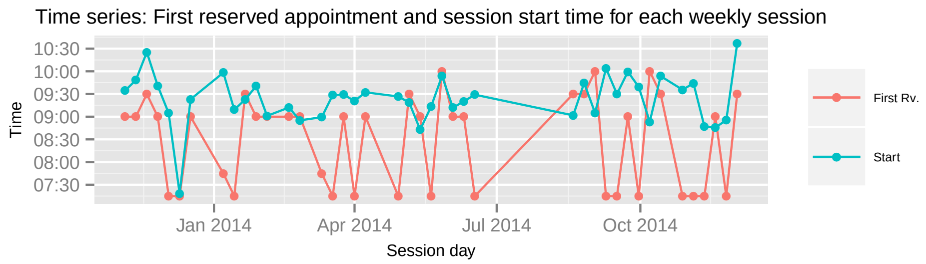


Figure C.9: Histogram, boxplot and time-series plots of the late starts for the top-six ranking weekly sessions. $P()$ is the percentage of session starts that were late.

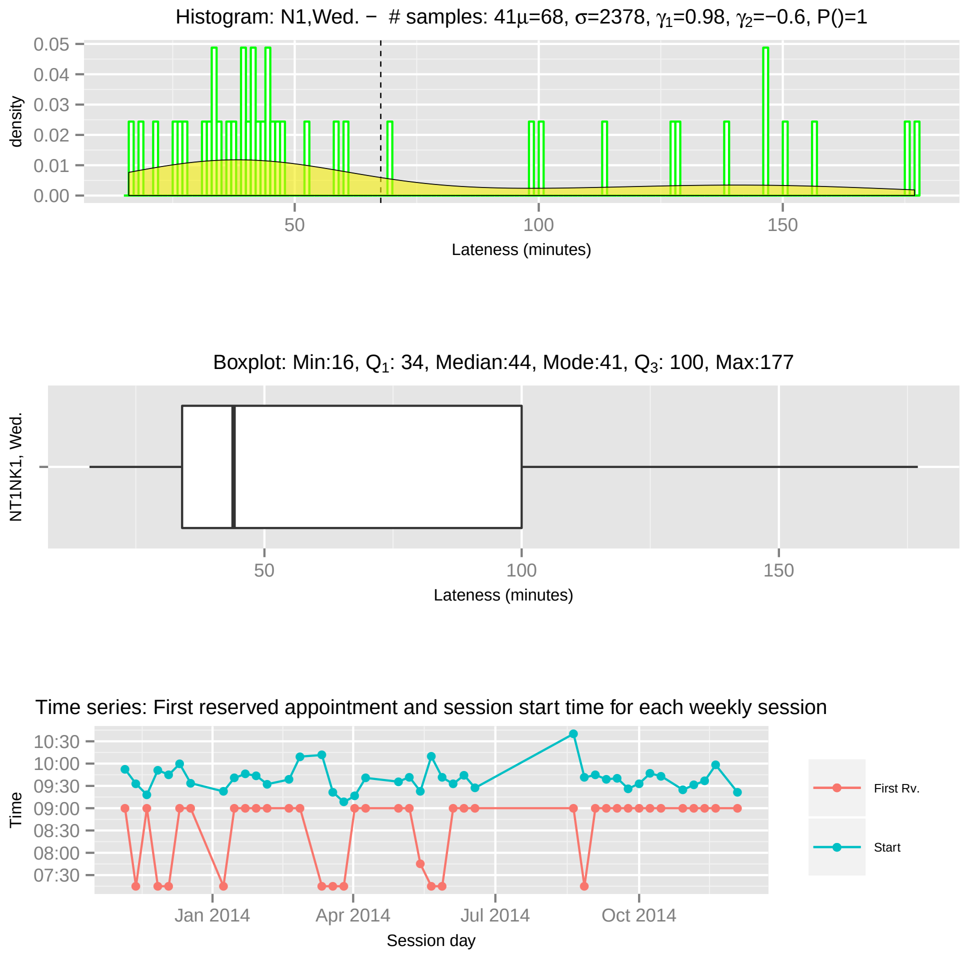
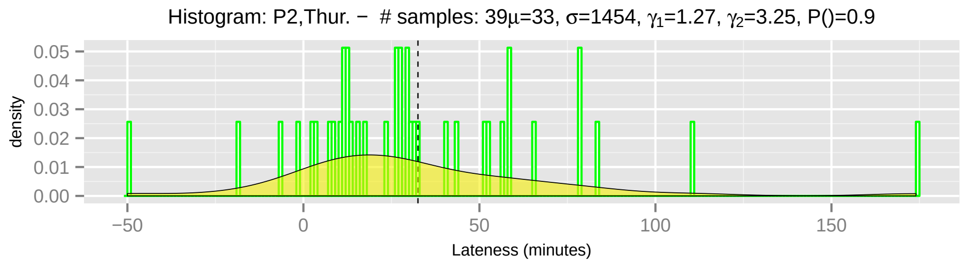
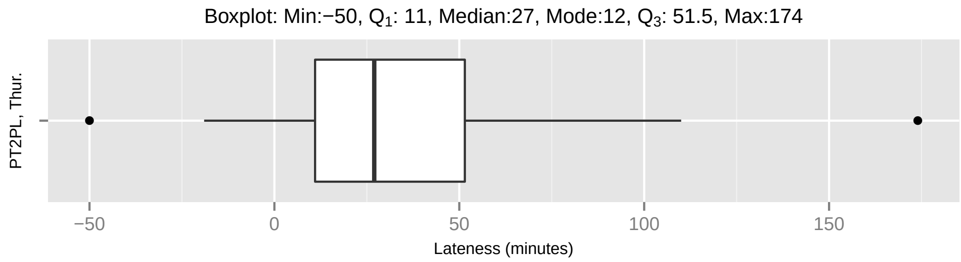


Figure C.10: Histogram, boxplot and time-series plots of the late starts for the top-six ranking weekly sessions. $P()$ is the percentage of session starts that were late.





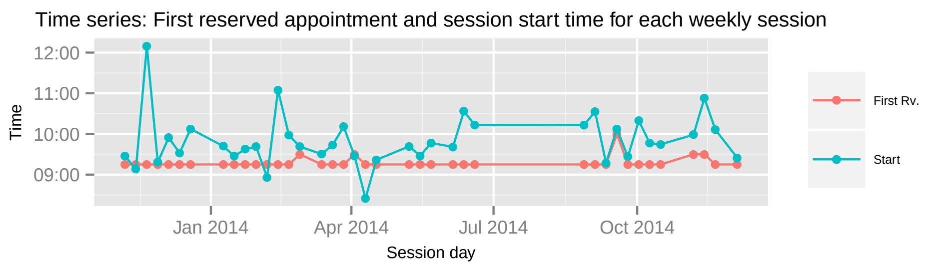


Figure C.11: Histogram, boxplot and time-series plots of the late starts for the top-six ranking weekly sessions. $P()$ is the percentage of session starts that were late.

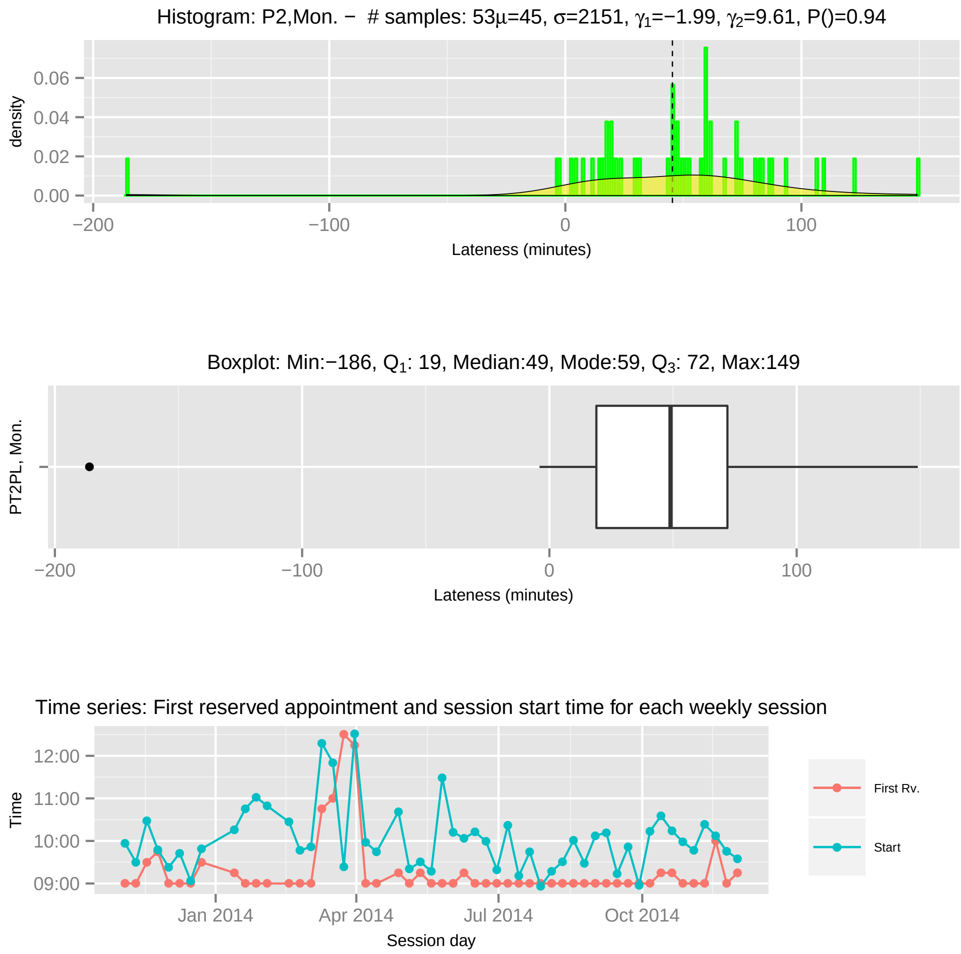


Figure C.12: Histogram, boxplot and time-series plots of the late starts for the top-six ranking weekly sessions. $P()$ is the percentage of session starts that were late.

C.5 Patient punctuality

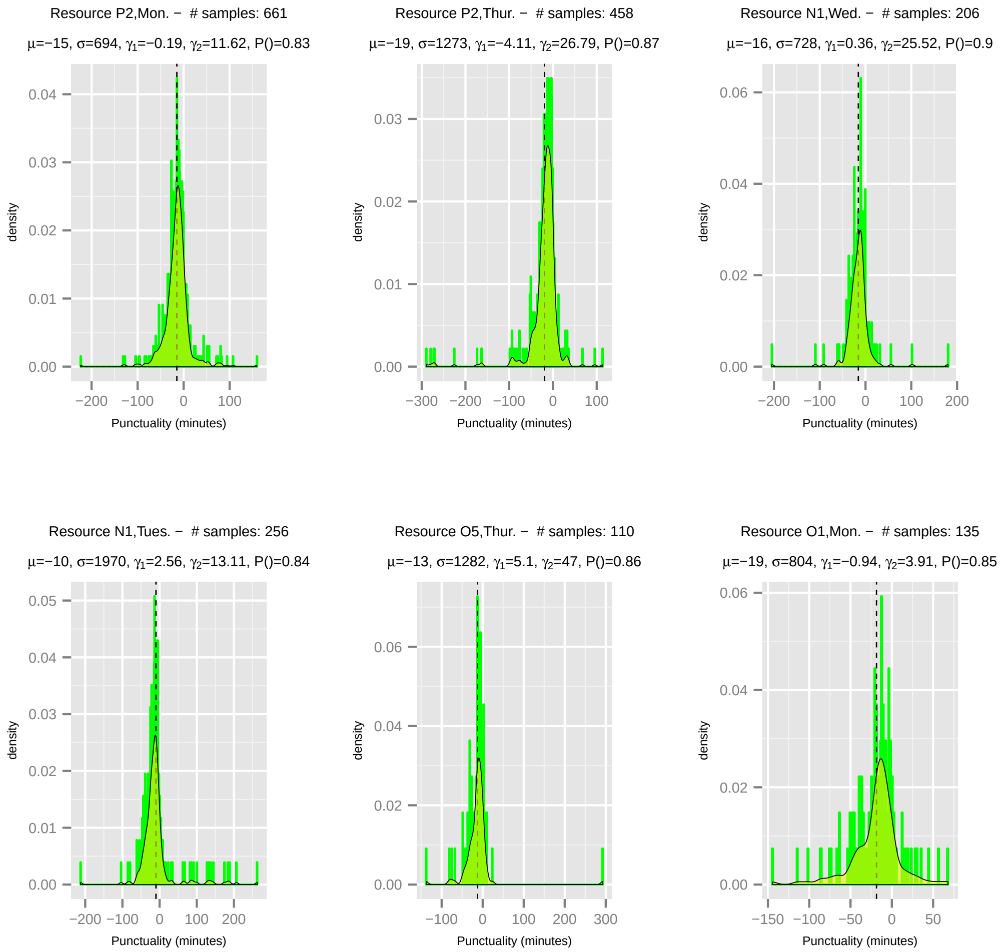


Figure C.13: Histogram with density estimation plot of the patient punctuality for the top-six ranking weekly sessions. $P()$ is the percentage of patients that arrived prior to the reservation time.

Appendix D

Duration distribution-fit plots

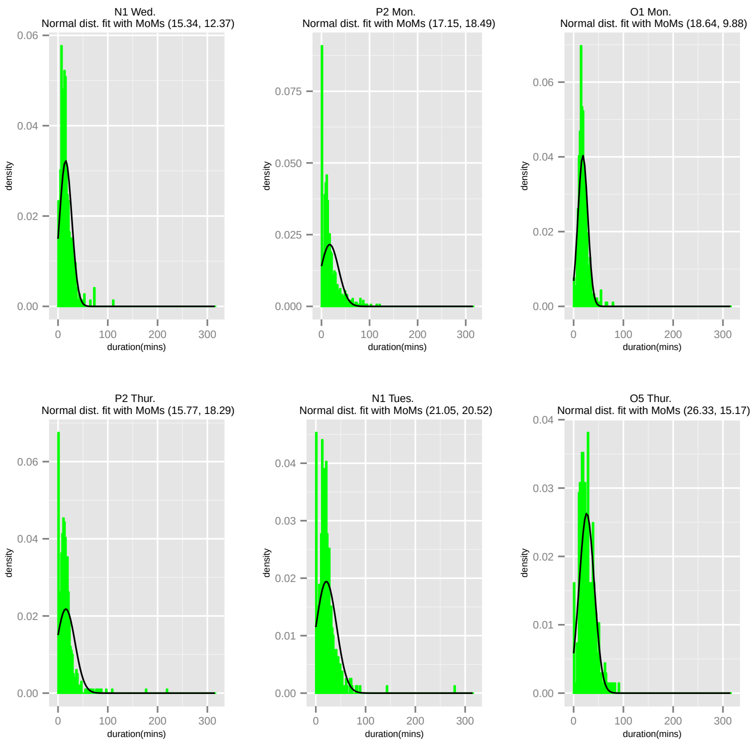


Figure D.1: Fitting the appointment duration with a Gaussian distribution using method of moments. Numbers in brackets are the first moment and the standard deviation of the data.

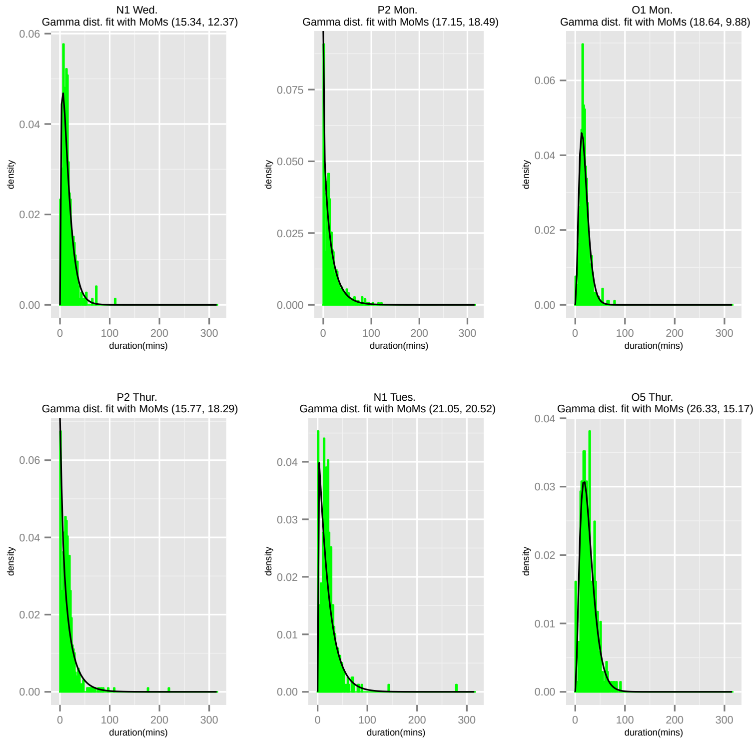


Figure D.2: Fitting the appointment duration with a Gamma distribution using method of moments. Numbers in brackets are the first moment and the standard deviation of the data.

Appendix E

Extracted features for duration prediction

feature	data type	details
age	integer	patient's age
gender	categorical	patient's gender (male, female or unknown)
time	time	time of the reservation
hour of day	integer	hour of day of the reservation booking
surgery past month?	boolean	has patient had surgery in the past month?
surgery past 2-3 months?	boolean	has patient had surgery in the past 2-3 months?
surgery past 3-6 months?	boolean	has patient had surgery in the past 3-6 months?
surgery past 6-12 months?	boolean	has patient had surgery in the past 6-12 months?
surgery past 12-24 months?	boolean	has patient had surgery in the past 12-24 months?
app. past month?	boolean	has patient had a clinic appointment in the past month?
app. past 2-3 months?	boolean	has patient had a clinic appointment in the past 2-3 months?
app. past 3-6 months?	boolean	has patient had a clinic appointment in the past 3-6 months?
app. past 6-12 months?	boolean	has patient had a clinic appointment in the past 6-12 months?
app. past 12-24 months?	boolean	has patient had a clinic appointment in the past 12-24 months?
previous same-day reservation?	boolean	has patient got a same-day reservation prior to the appointment?
previous same-day resource	categorical	where was the patient's previous reservation ("NONE" if none)
diagnostic same day?	boolean	does the patient have a previous same-day diagnostic examination?
diagnostic past 1-7 days?	boolean	does patient have an examination 1-7 days before the appointment?
diagnostic past 8-30 days?	boolean	does patient have an examination 8-30 days before the appointment?
previous surgery weeks	integer	weeks since the patient's previous surgery (-1 if none)
previous surgery appointment weeks	integer	weeks since the patient's previous appointment (-1 if none)

Table E.1: Features extracted for predicting the appointment duration. The abstraction of previous events is to aid the learning to categorise appointments.

Appendix F

Linear Regression statistical results

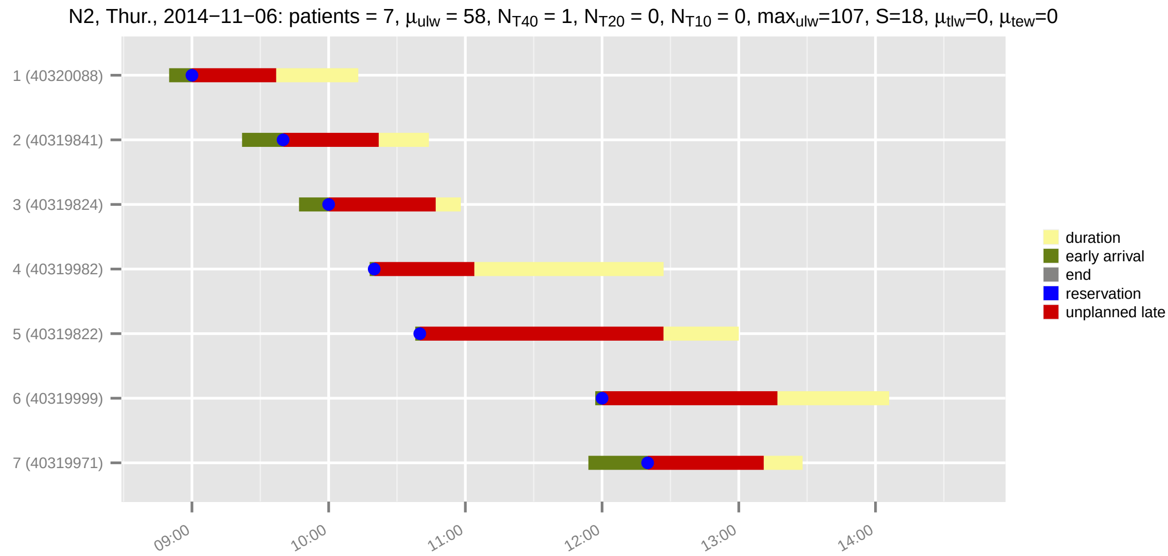
Table F.1:

	Dependent variable:
	rv_duration
patient_age	−0.005 (0.097)
gendermale	5.406* (2.945)
past_yr_mean	
surgery_past_month	
surgery_past_1_2	−1.532 (4.225)
surgery_past_2_3	3.590 (5.702)
surgery_past_3_6	20.912** (7.939)
surgery_past_6_12	0.775 (5.727)
surgery_app_past_month	1.432 (7.183)
surgery_app_past_1_2_months	−1.458 (5.037)
surgery_past_2_3_months	0.373 (4.683)
surgery_past_3_6_months	4.813 (3.944)
surgery_past_6_12_months	−2.462 (3.420)
prev_same_day_rv	−1.003 (6.393)
prev_same_day_rsN115	−0.123 (7.156)
prev_same_day_rsNONE	
prev_same_day_rsNT1PS	−3.676 (12.773)
prev_same_day_rsRMR3T	35.542** (14.161)
prev_same_day_rsRMRI7	6.958 (10.570)
prev_same_day_rsRNAT11	−4.833 (12.459)
prev_same_day_rsRNAT12	28.311** (13.539)
prev_same_day_rsRNAT13	18.042 (12.126)
diag_same_day	−7.522 (5.924)
diag_1_7_past	0.837 (4.982)
diag_8_30_past	1.667 (7.896)
rv_time_numeric	0.0005* (0.0003)
Constant	−655,006.300* (388,806.900)
Observations	74
R^2	0.443
Adjusted R^2	0.187
Residual Std. Error	10.562 (df = 50)
F Statistic	1.728* (df = 23; 50)
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix G

Optimization visualisations

G.1 Duration prediction chaining



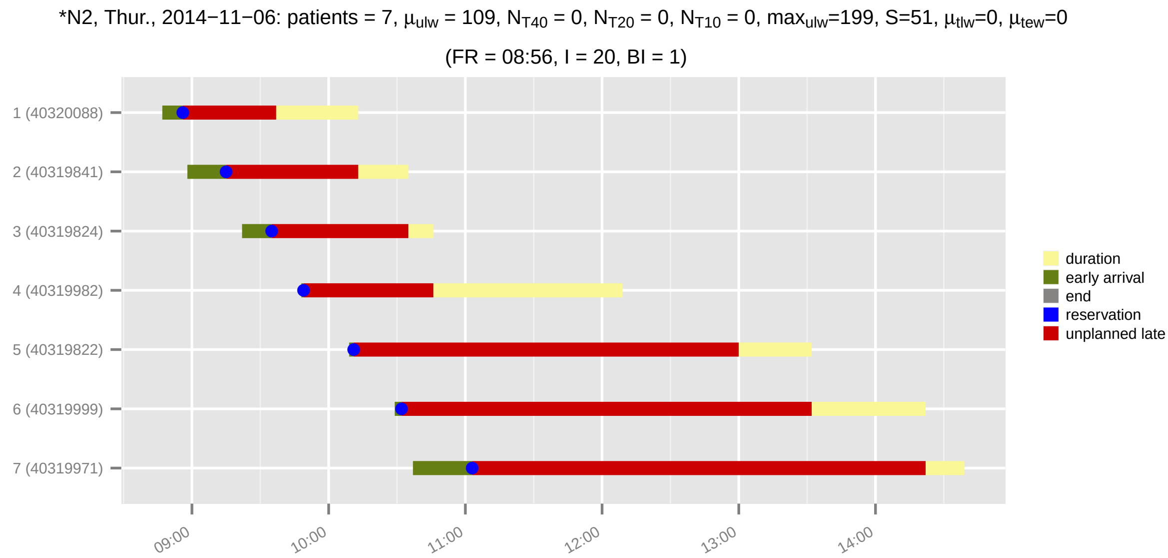


Figure G.1: Thursday's session for resource N2 optimised with prediction chaining

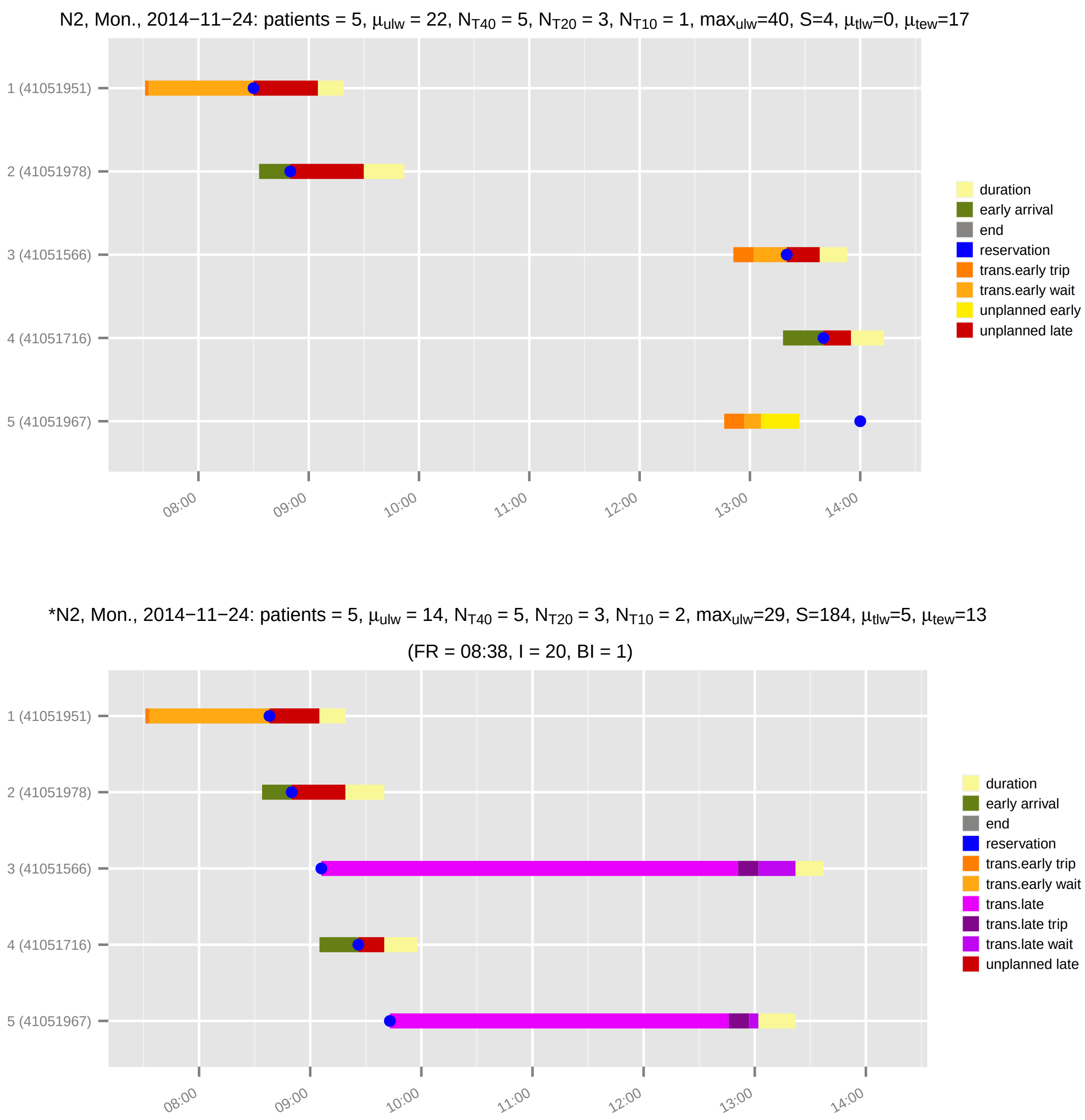
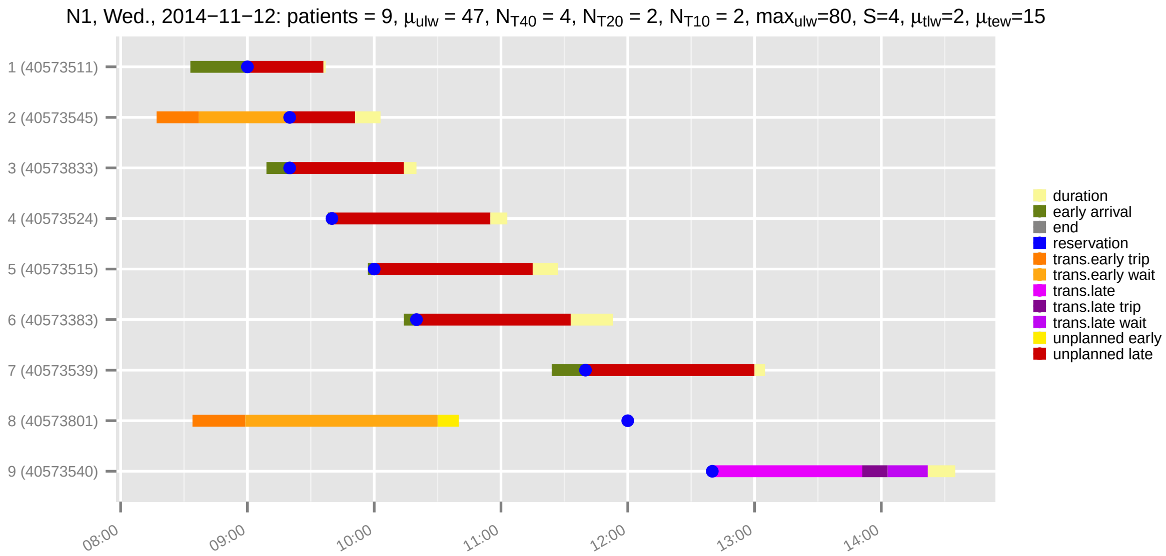


Figure G.2: Monday's session for resource N2 optimised with prediction chaining



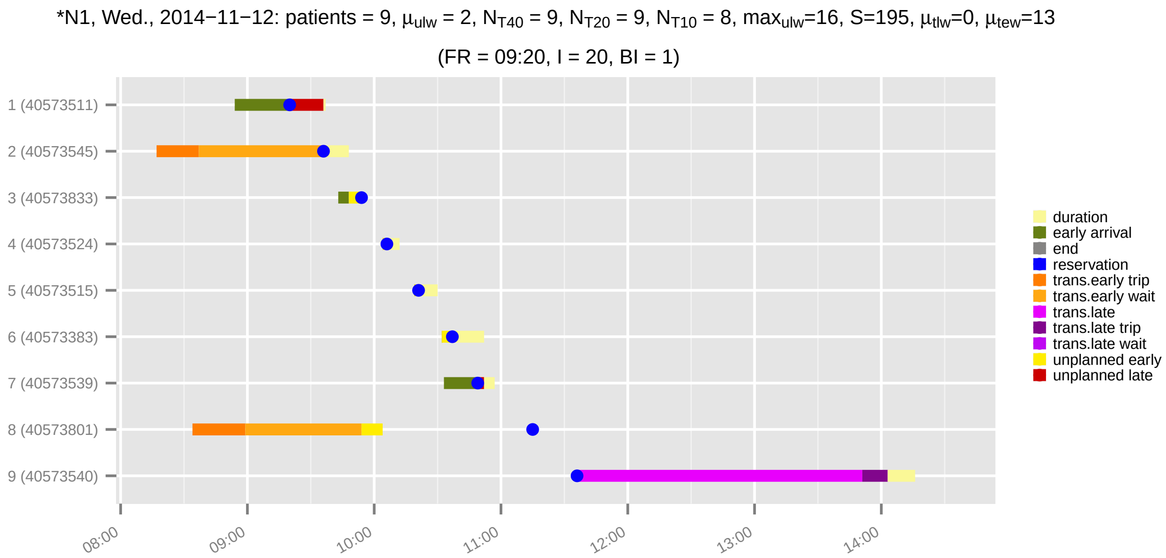
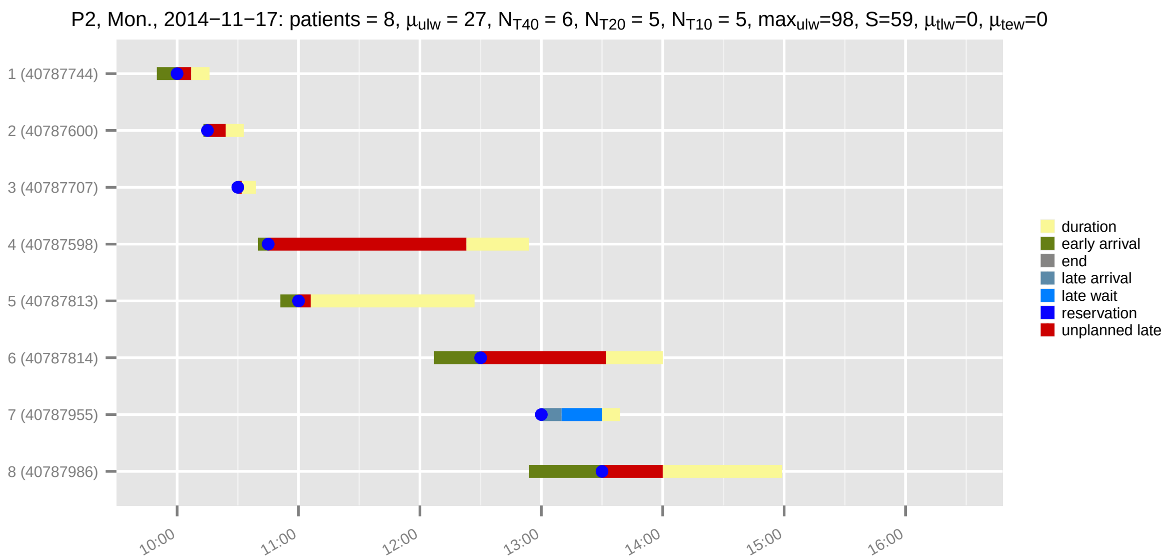


Figure G.3: Wednesday's session for resource N1 optimised with prediction chaining



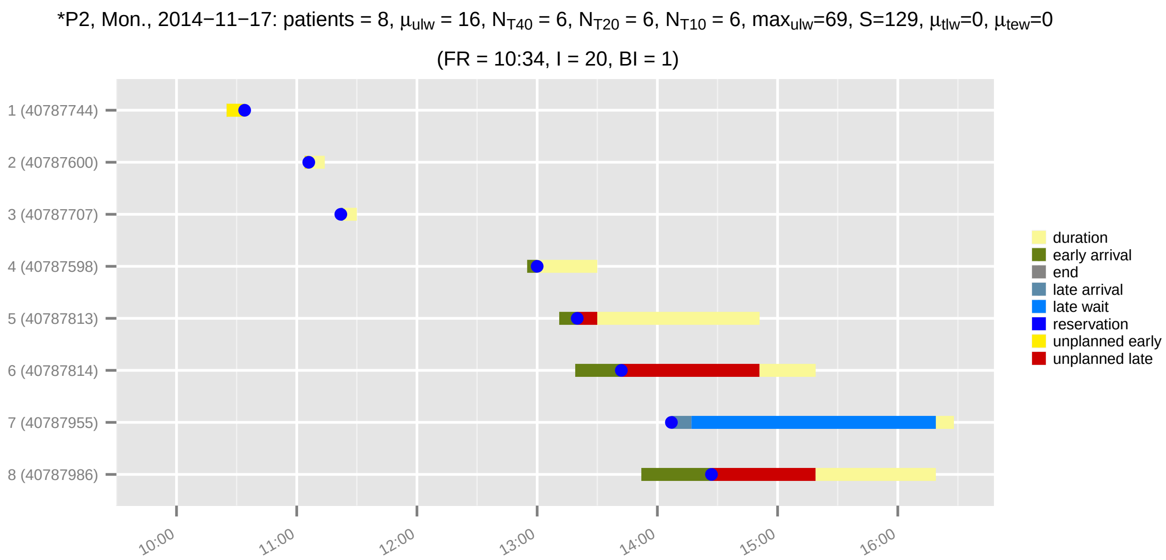
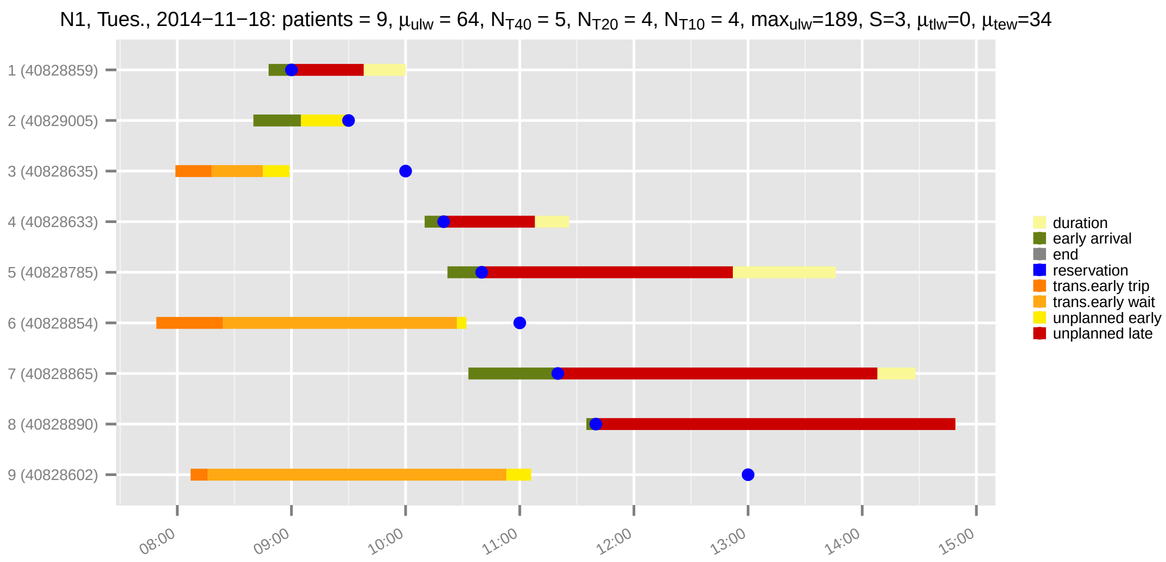


Figure G.4: Monday's session for resource P2 optimised with prediction chaining



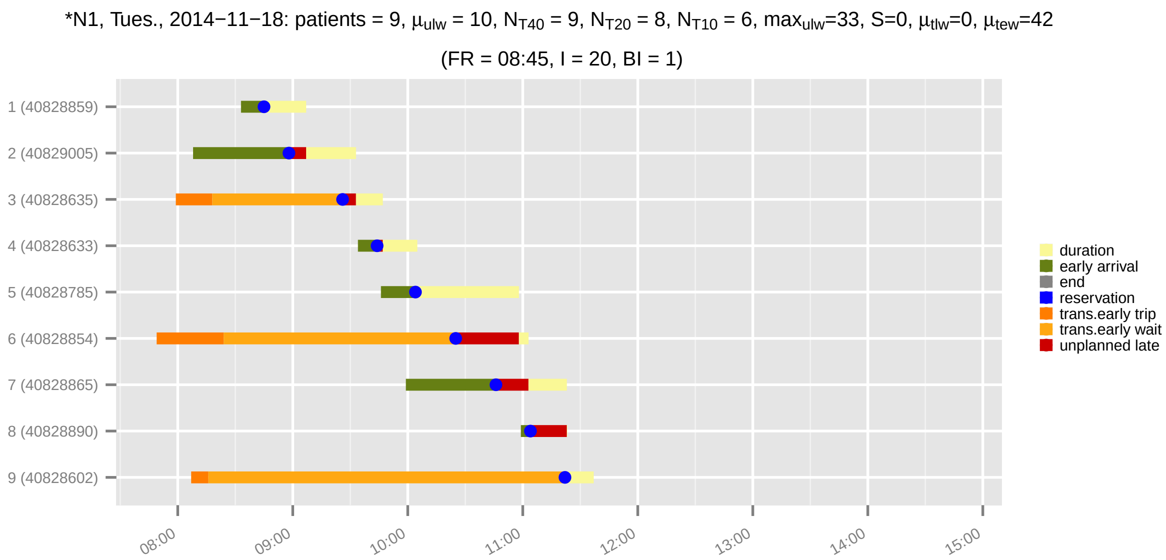
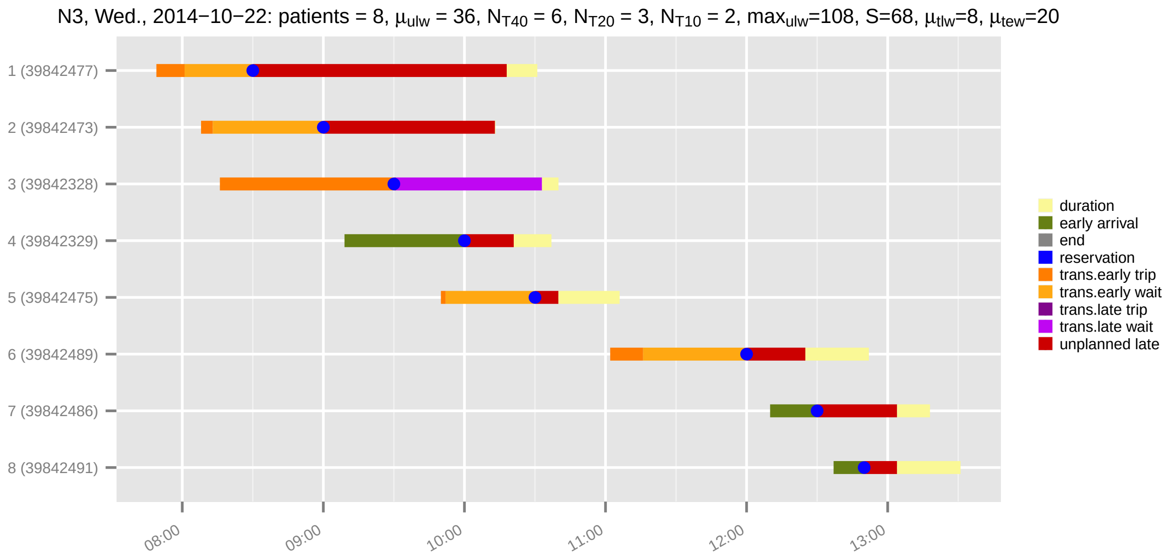


Figure G.5: Tuesday's session for resource N1 optimised with prediction chaining



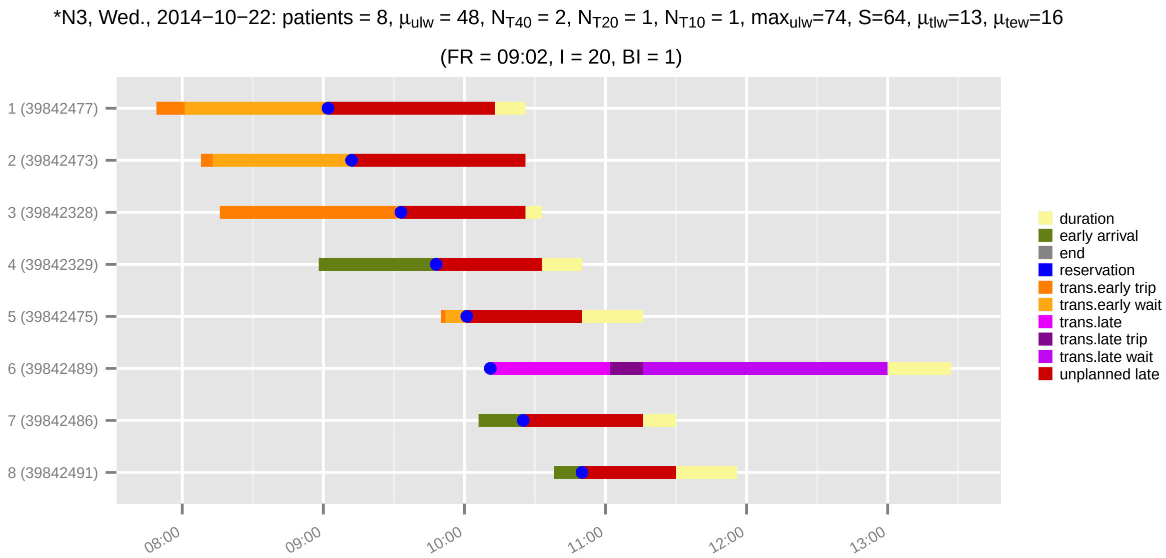
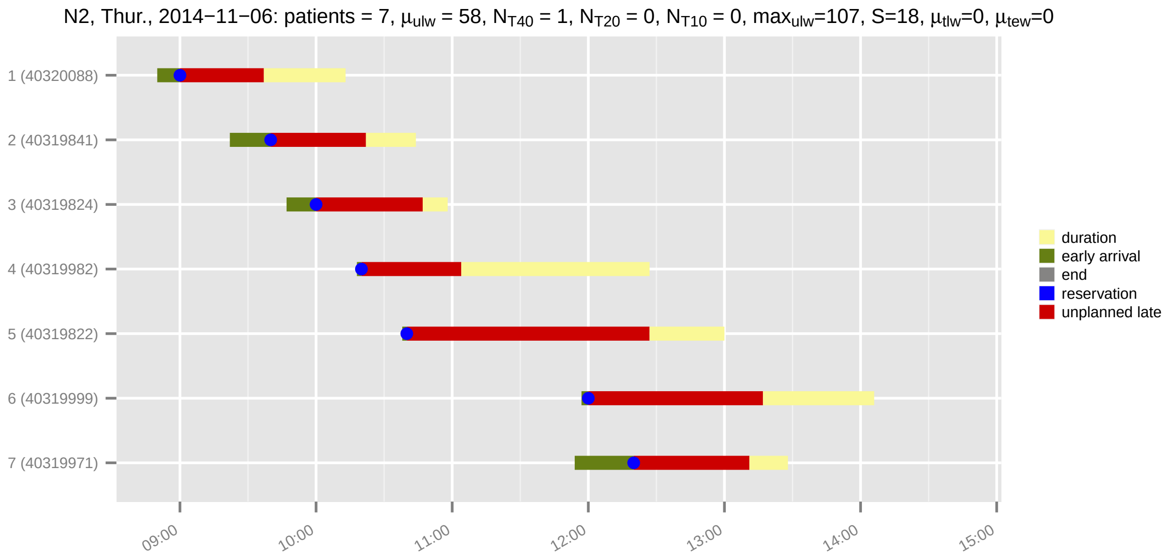


Figure G.6: Wednesday's session for resource N3 optimised with prediction chaining

G.2 Hill climbing optimisation: late-waiting only



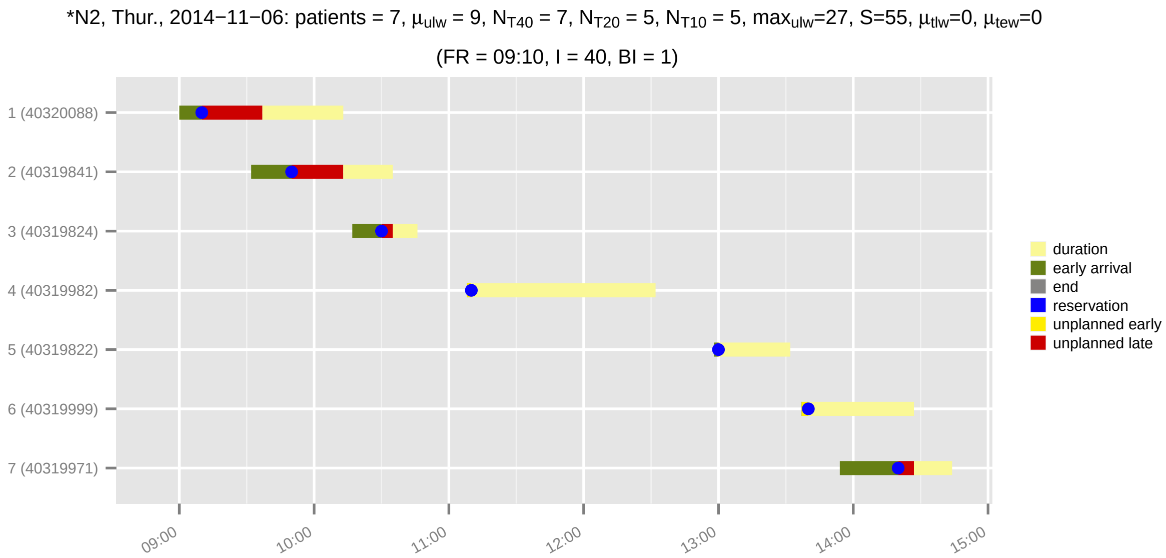
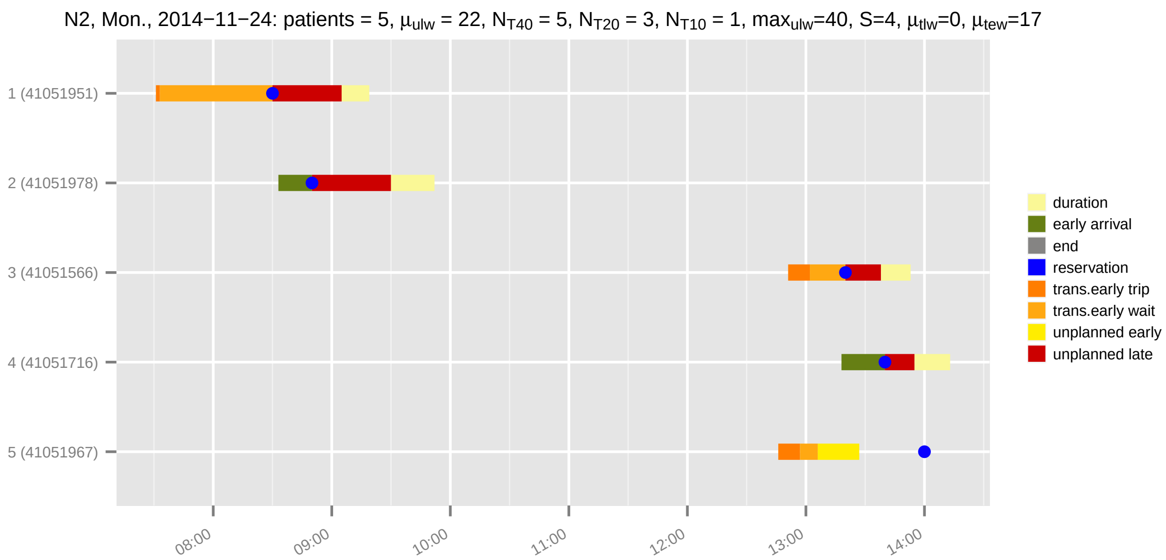


Figure G.7: Thursday's session for resource N2 optimised with hill climbing local search optmisation



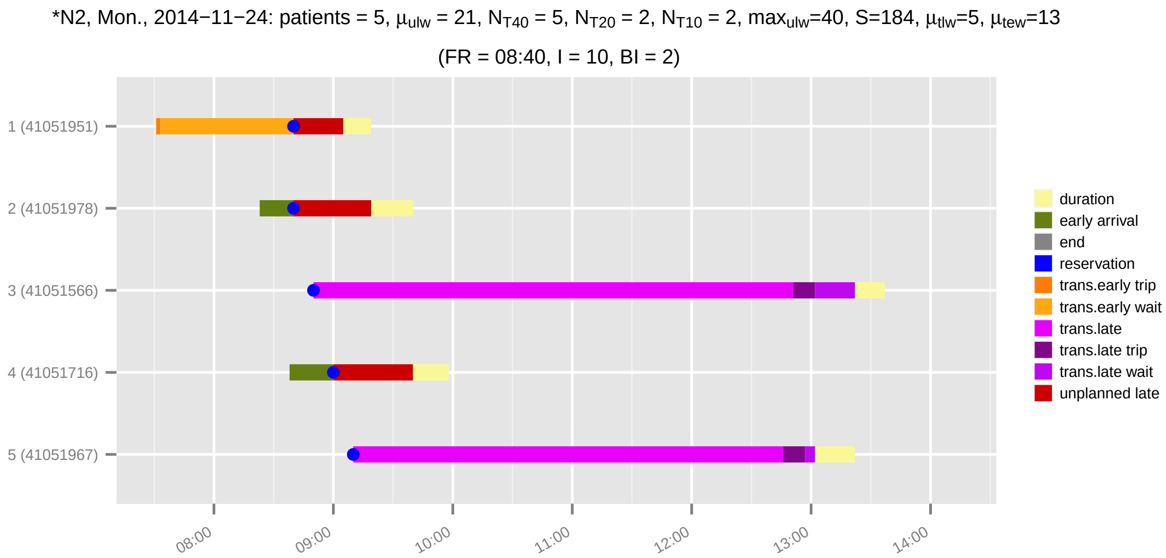
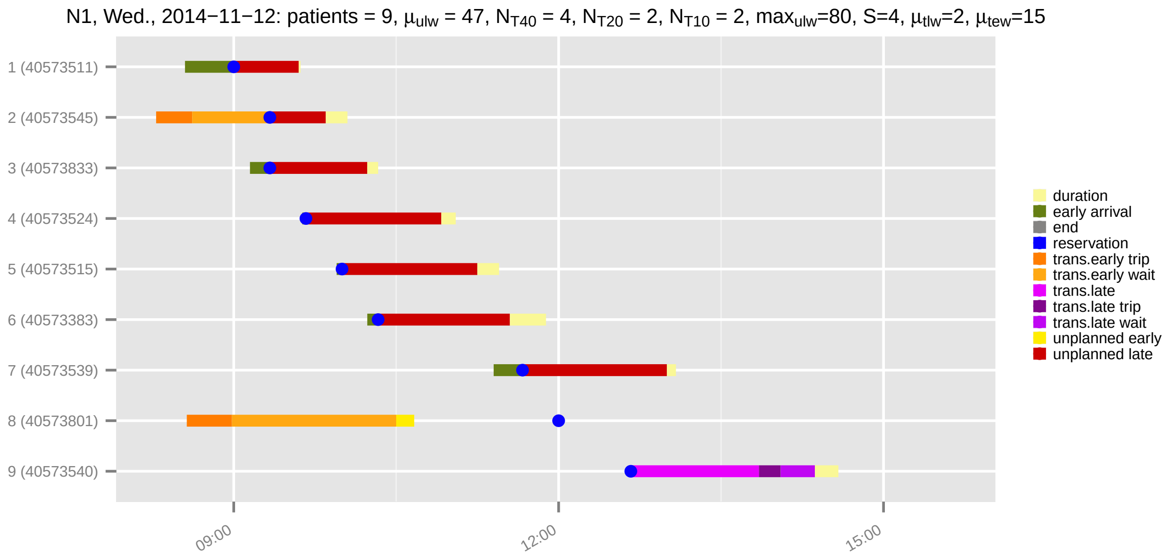


Figure G.8: Monday's session for resource N2 optimised with hill climbing local search optmisation



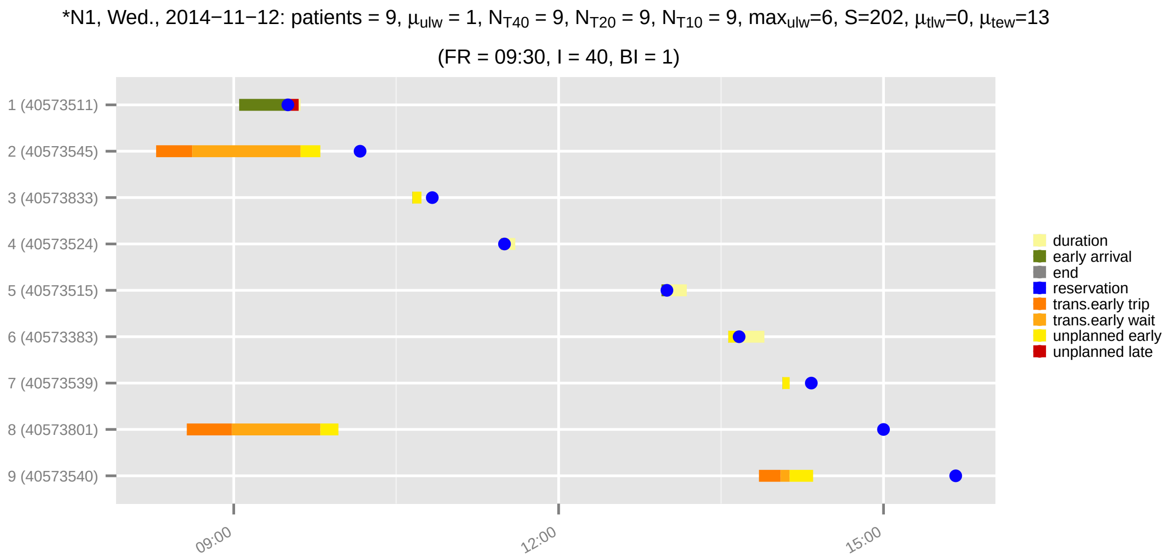
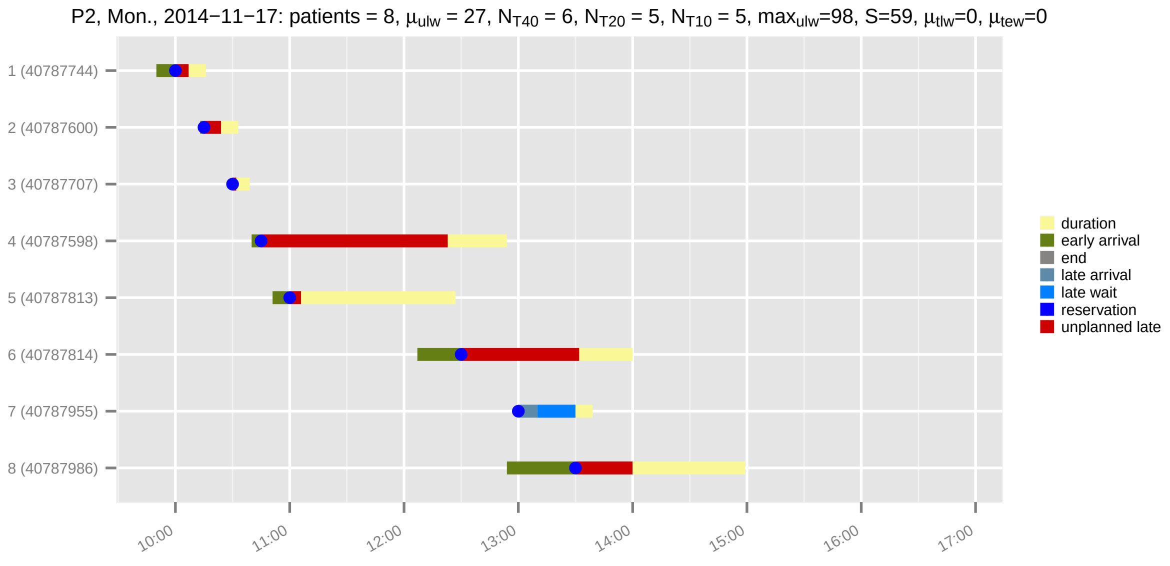


Figure G.9: Wednesday's session for resource N1 optimised with hill climbing local search optmisation



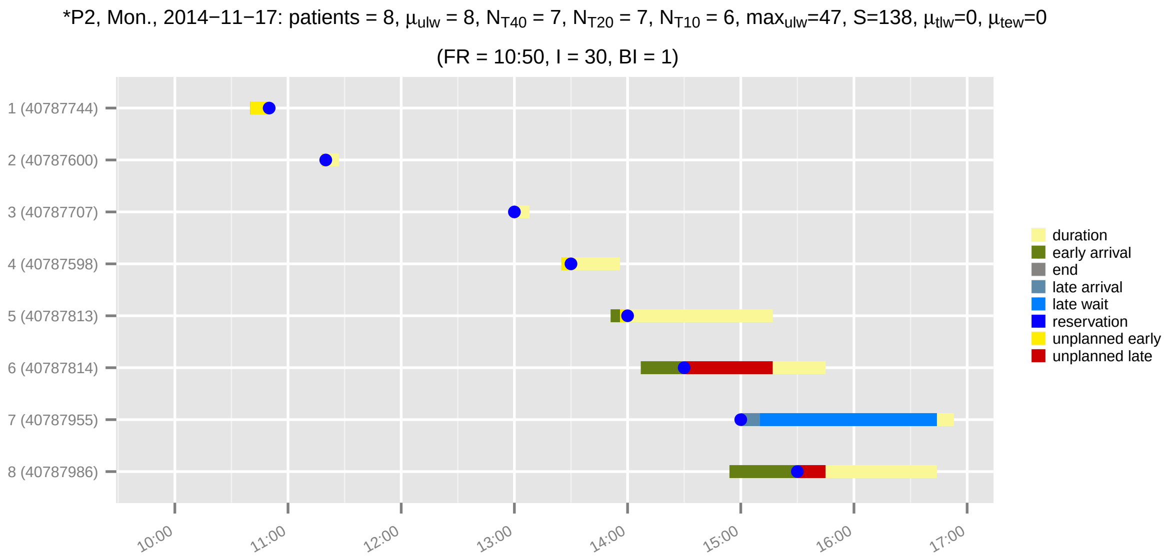
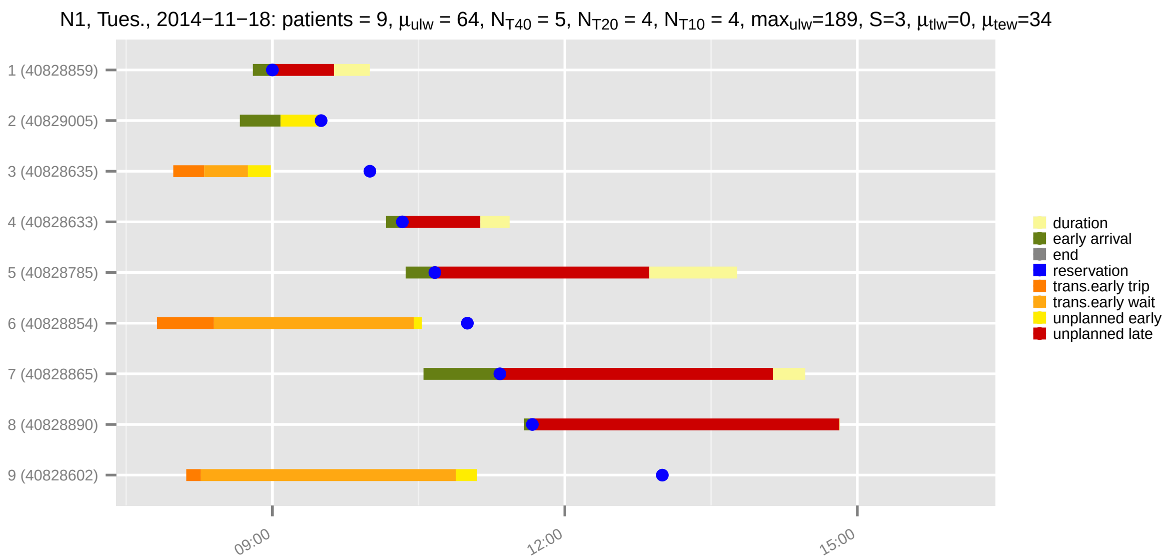


Figure G.10: Monday's session for resource P2 optimised with hill climbing local search optmisation



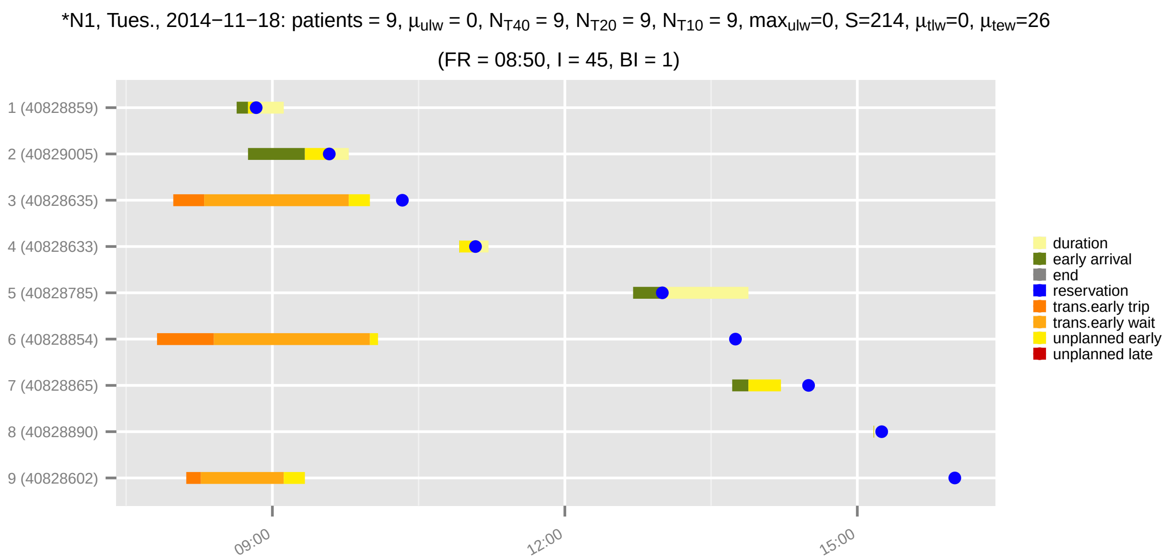
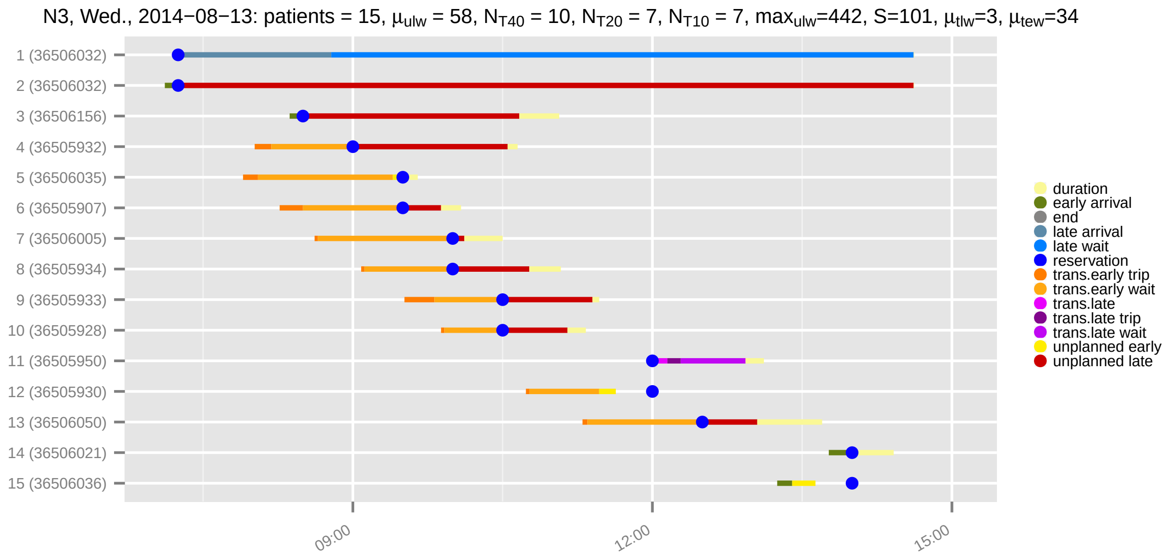


Figure G.11: Tuesday's session for resource N1 optimised with hill climbing local search optmisation



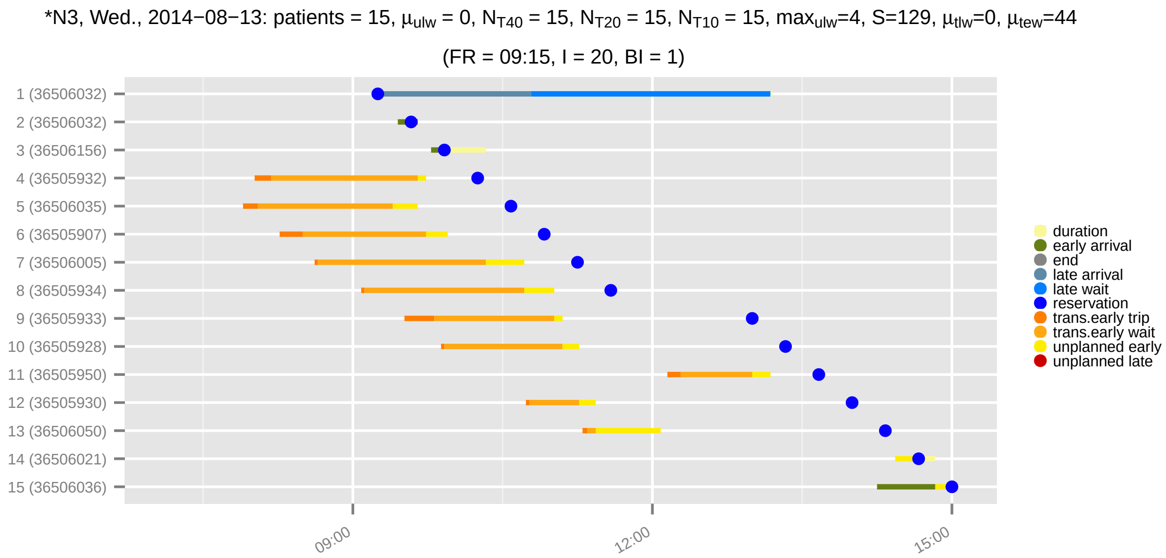
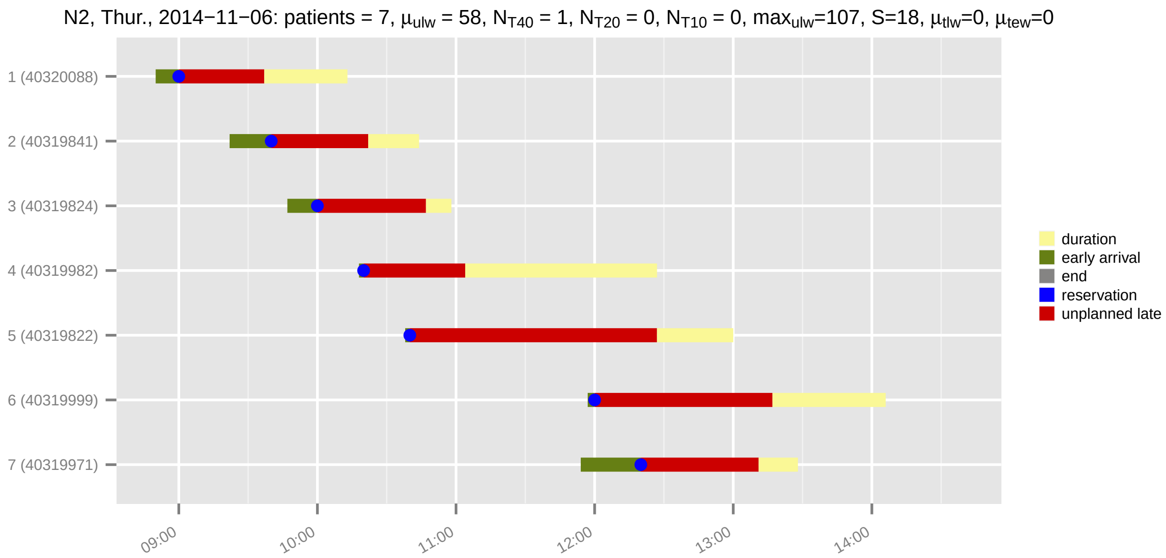


Figure G.12: Wednesday's session for resource N3 optimised with hill climbing local search optmisation

G.3 Hill climbing optimisation: multi-objective



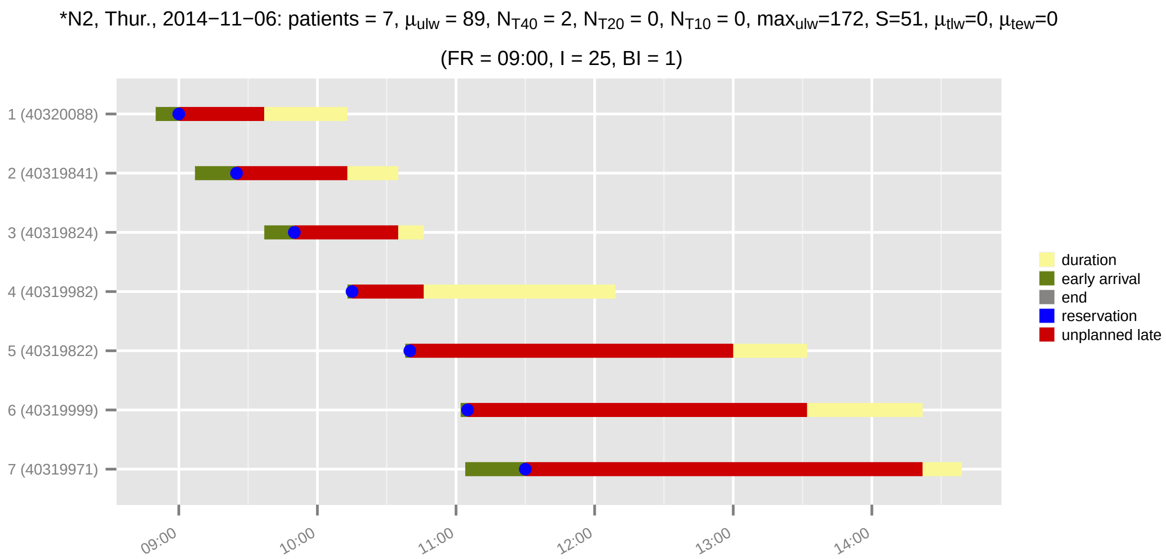


Figure G.13: Thursday's session for resource N2 optimised with hill climbing local search optmisation using a multi-objective cost function

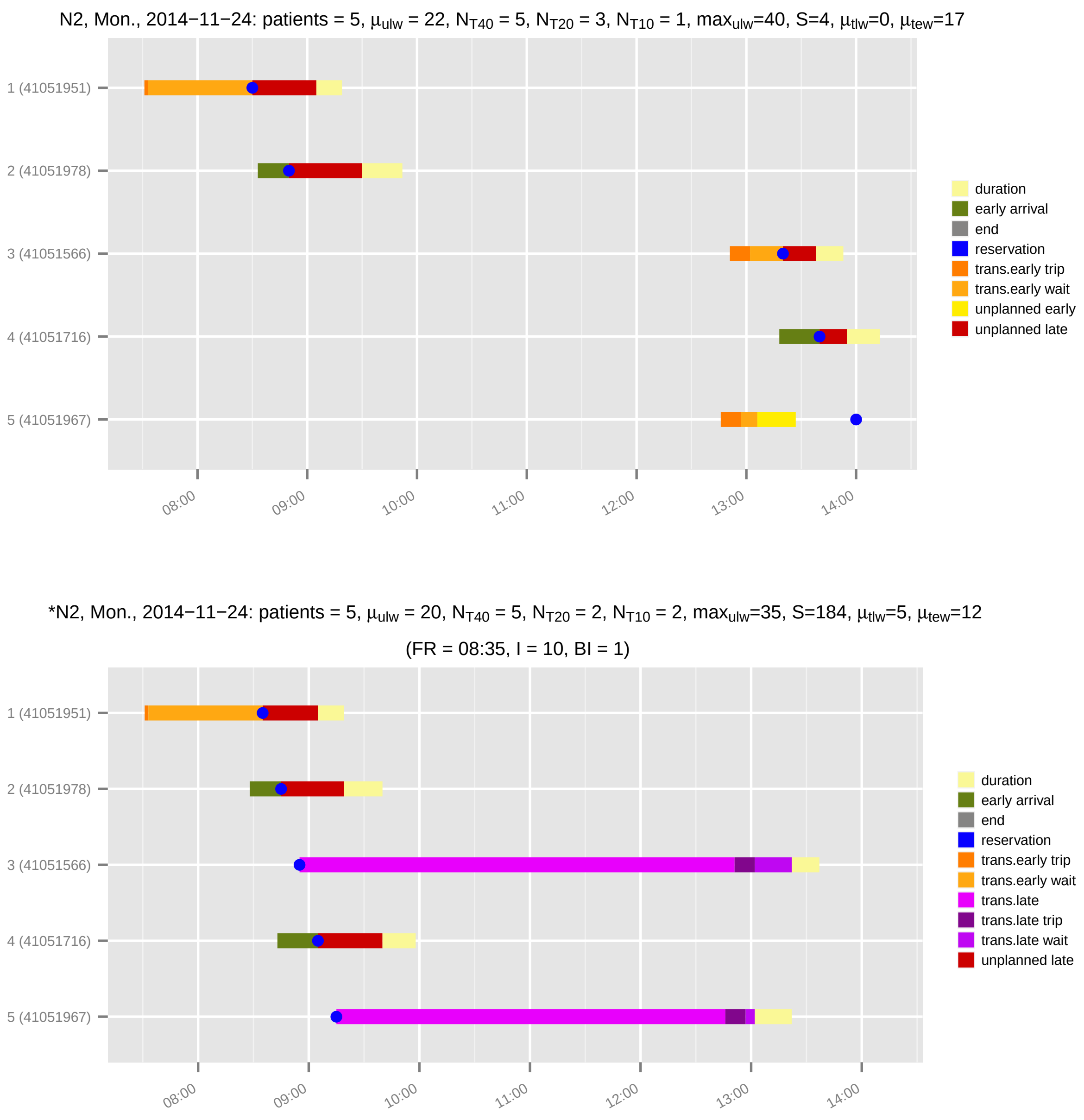
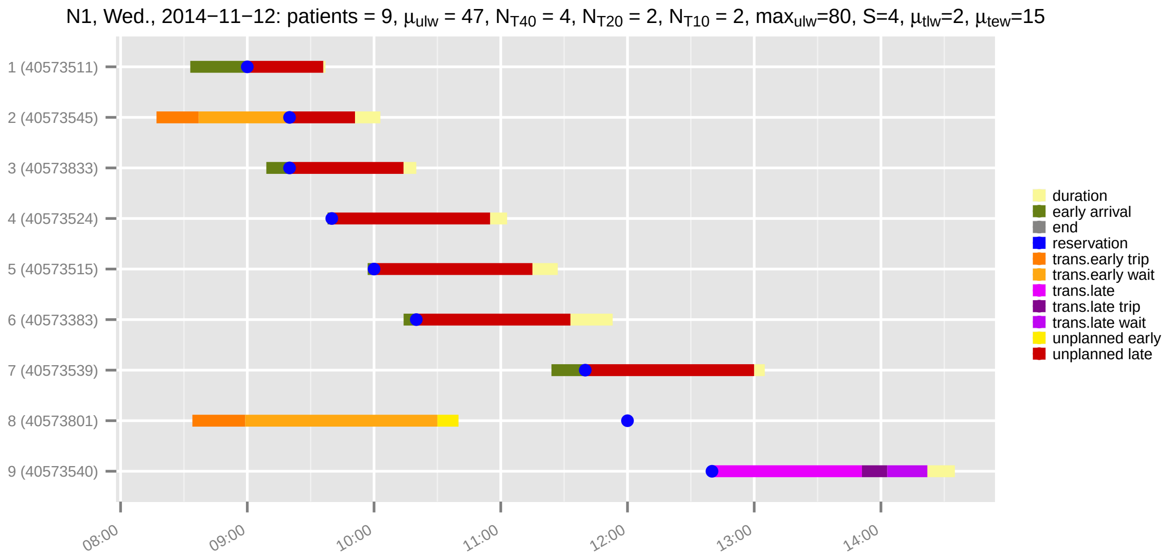


Figure G.14: Monday's session for resource N2 optimised with hill climbing local search optmisation using a multi-objective cost function



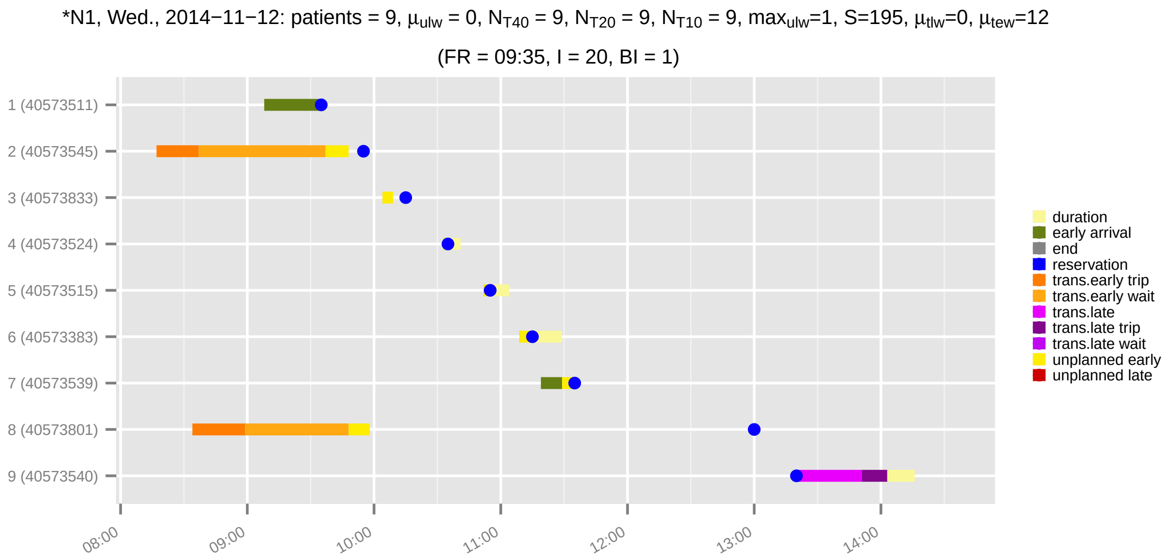
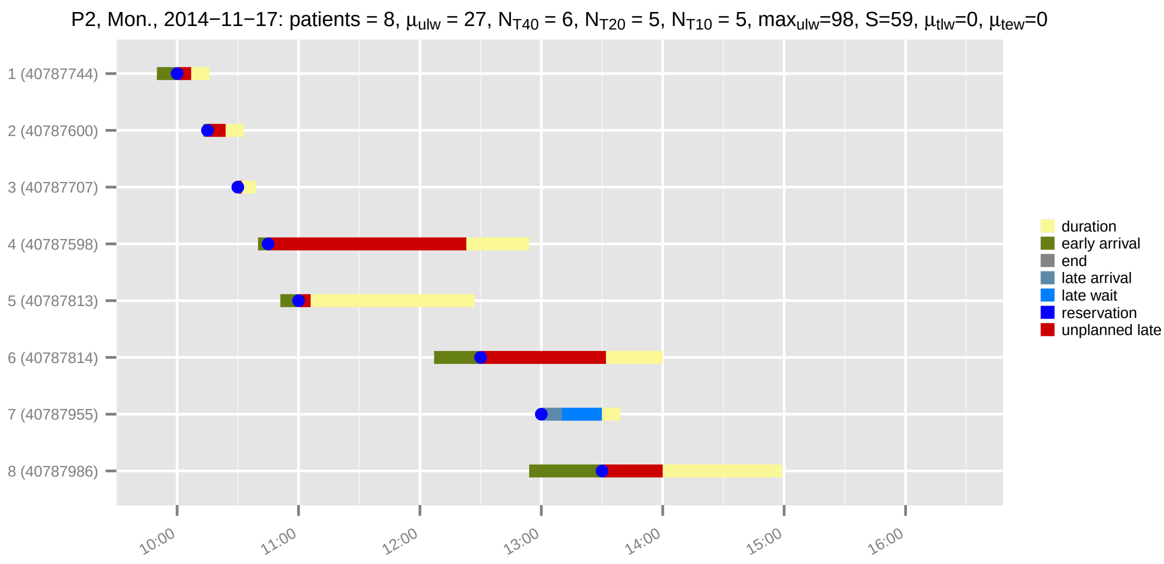


Figure G.15: Wednesday's session for resource N1 optimised with hill climbing local search optmisation using a multi-objective cost function



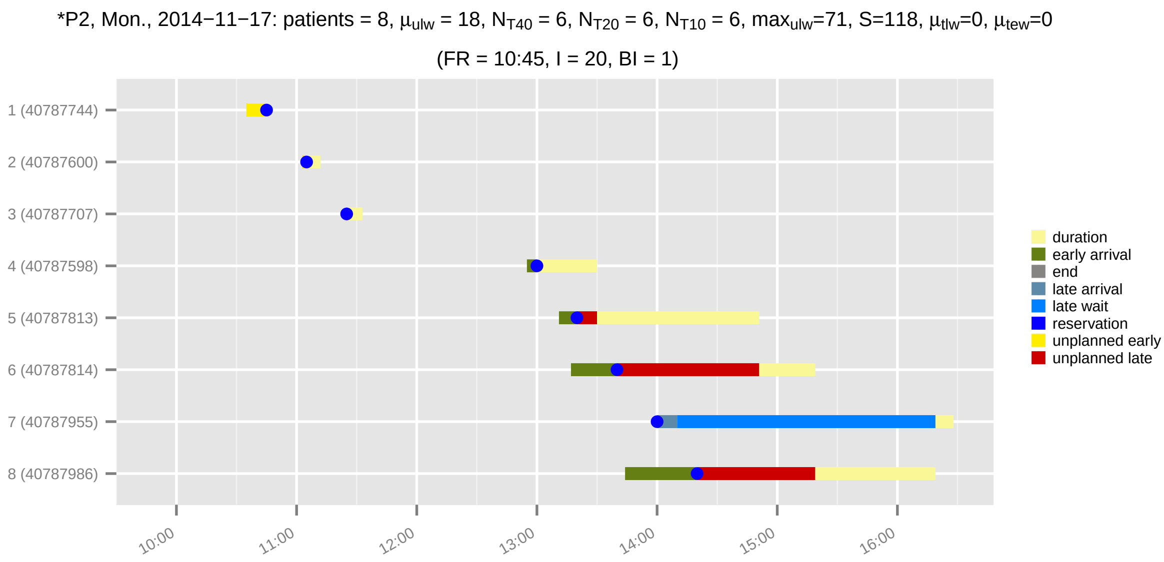
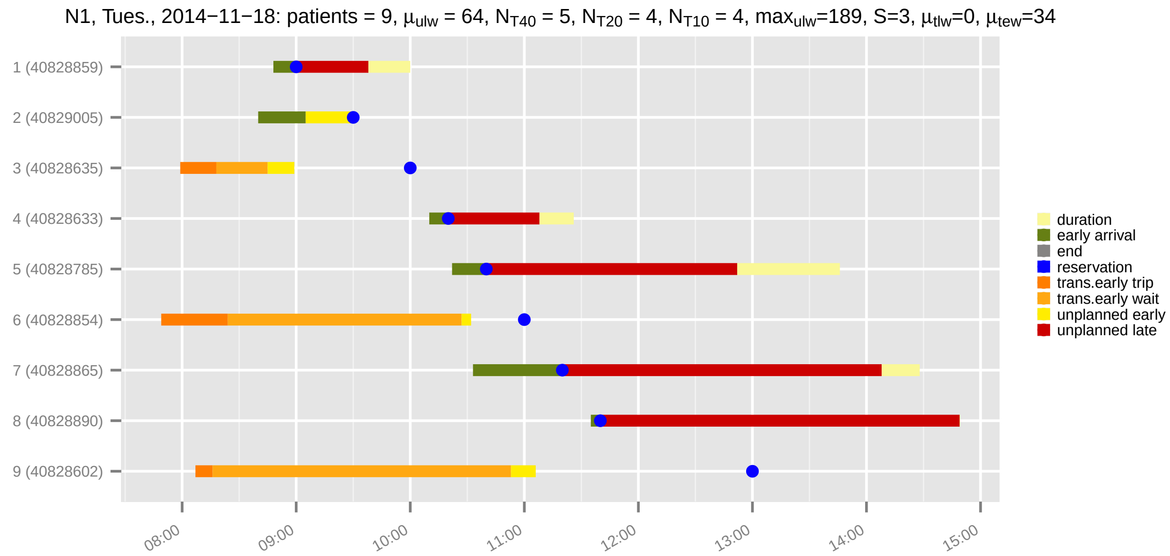


Figure G.16: Monday's session for resource P2 optimised with hill climbing local search optmisation using a multi-objective cost function



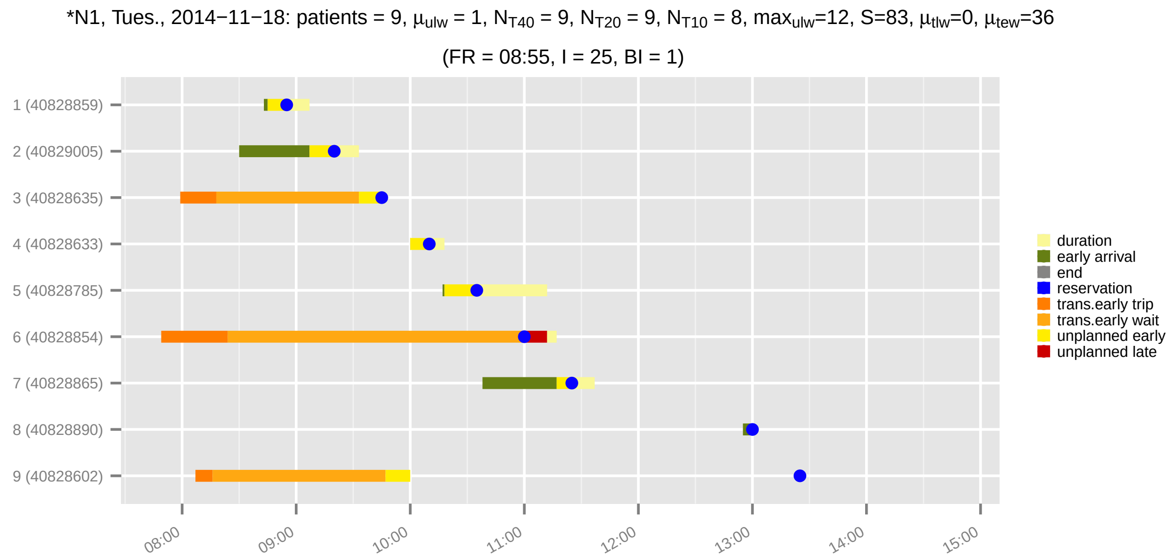
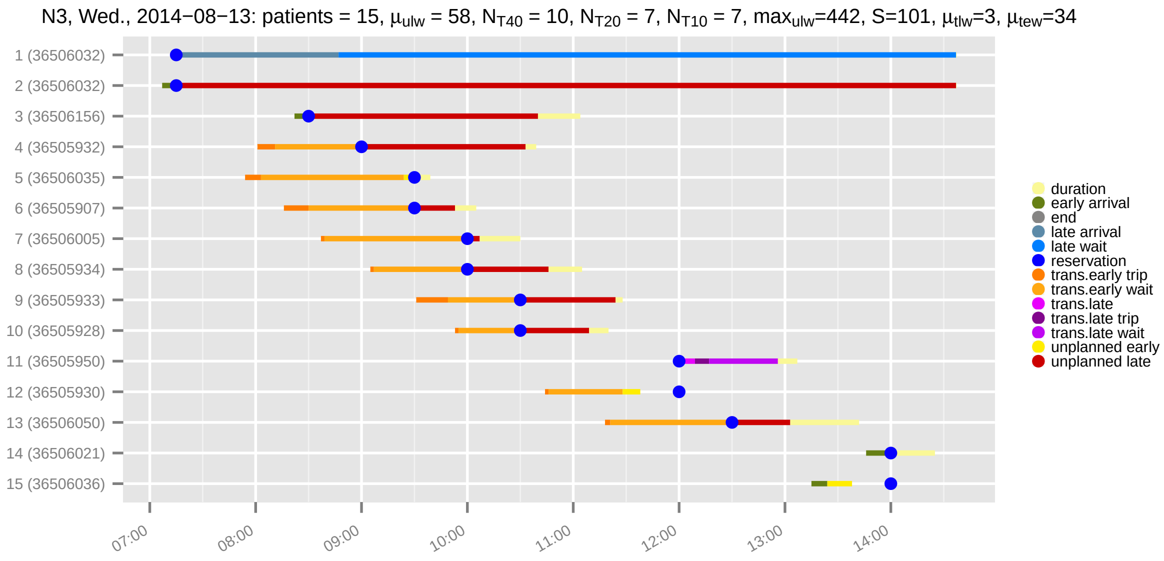


Figure G.17: Tuesday's session for resource N1 optimised with hill climbing local search optmisation using a multi-objective cost function



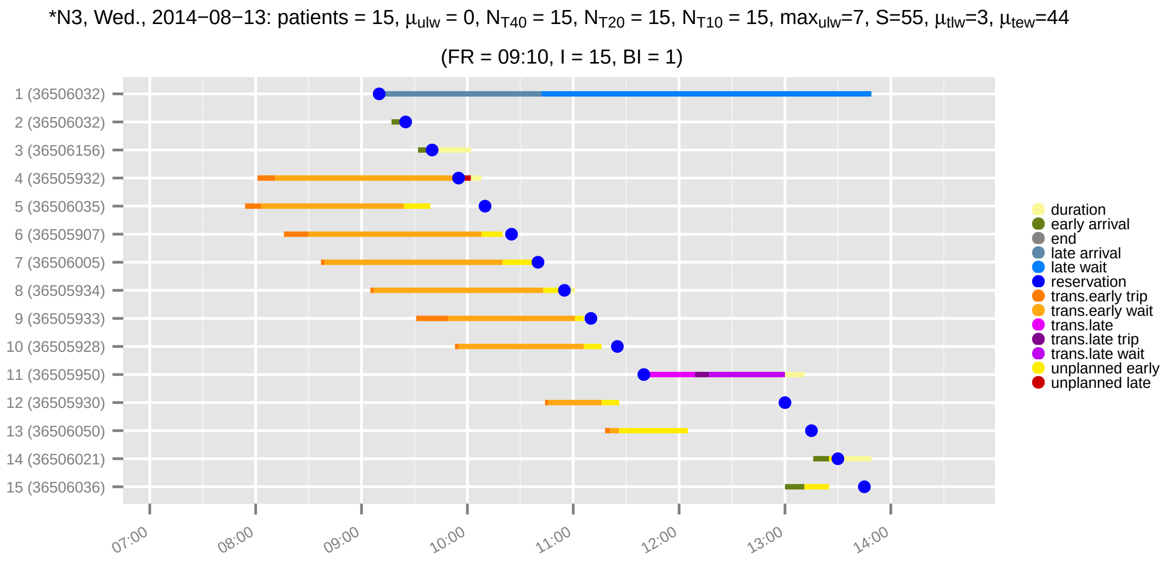


Figure G.18: Wednesday's session for resource N3 optimised with hill climbing local search optmisation using a multi-objective cost function