Vrije Universiteit Amsterdam

Master Thesis

Business Analytics

# Hospital Bed Occupancy Prediction

Developing and Implementing a predictive analytics decision support tool to relate Operation Room usage to bed occupancy

*Author*
T.L. Bos

*Supervisors*
Prof. dr. G.M. Koole (VU)
R. Hindriks (VU)
M. Verschoor (ChipSoft B.V.)
K. Ruijter (ChipSoft B.V.)

November 3, 2017

# Preface

This thesis is written to obtain the Master's degree in Business Analytics at the Factulty of Science of the Vrije Universiteit (Amsterdam). The research was performed in the form of an internship with a total duration of six months at ChipSoft B.V. During the internship period, I studied the influence the Operation Room usage has on the clinical wards in hospitals and developed a decision support tool which assists hospitals in making capacity related decisions. The result of this research is to be further developed and optimized in order to finally implement it in ChipSoft's health care information software.

I would like to thank prof. dr. Ger Koole (Vrije Universiteit) for supervising this research project, as well as Rikkert Hindriks (Vrije Universiteit) for acting as second reader of this thesis. Besides the university supervisors, I would like to thank Matthijs Verschoor and Karen Ruijter from ChipSoft B.V. for allowing me to perform this research at their department and assisting me in making difficult decisions during this project.
Finally, I would like to explicitly thank the two hospitals, which will remain anonymous, for providing me with the required data to perform the research and provide me with feedback regarding the research's outcome.

Laurens Bos, October 2017

# Contents

# Chapter 1

# Introduction

Within hospital health care, most of the costs and revenues appear in and around the Operation Rooms (ORs) [15, 3]. Therefore, optimizing the inpatient care process is highly prioritized by hospitals (as it often includes surgeries in an OR), but might also be challenging due to conflicting priorities and the availability of the required resources such as doctors, ORs and expensive equipment.
An important part of optimizing the inpatient care process is to manage the available capacity, for example by creating collaboration between the hospital's ORs and the hospital's clinical wards. Patients arrive on the clinical wards to receive or to rehabilitate from surgery on a scheduled basis because of the use of a Master Surgical Schedule by the hospital and on a non-scheduled basis because of emergencies. The Master Surgical Schedule divides the OR capacity among the different specialties that require usage of the OR and results in a roster for patients to receive surgery. So, an important relation exists between the used Master Surgical Schedule and the clinical wards receiving the patients because of it [4, 9]. However, the development of such a Master Surgical Schedule usually does not incorporate the relationship certain scheduling decisions have on the clinical wards [2, 20].

In its current state, ChipSoft's health care information software does not include tools for hospitals to incorporate the relationship between the ORs and the clinical wards in the Master Surgical Scheduling development process. This is considered a lack by ChipSoft and they therefore strive to implement these tools in their software in the near future.
To do so, these (and other capacity related) tools are to be developed within the "Capacity Management" project so ChipSoft's software can act as a decision support tool for capacity related problems within hospitals in general. This research internship is part of this project and its main goal is to design and implement a model that extracts and exposes the relationship between the ORs and the hospital's clinical wards.

This research is structured as follows: this chapter gives an preliminary overview of the performed work, including the company, the department and the internship process followed to perform this research. The second chapter describes the clinical process and the context of this research. Chapter 3 describes the problem statement, including the predefined deliverables and chapter 4 states the model used. The datasets definitions and the data analysis is described in chapter 5. The implementation of the model, which describes an example tool developed to use the model in practice, is stated in chapter 6.
Chapter 7 states the results obtained by the model's implementation and its usage, whereas chapter 8 discusses the work performed in this research and states the interpretation of the results. The conclusion finalizes this report in chapter 9, followed by the appendices and the bibliography.

## 1.1 ChipSoft B.V.

In the year 1986, surgeon Gerrit Mulder realizes that instead of spending most of his time on improving patients' health, he spends most of his time on administrative tasks. His son Hans Mulder decided to design a computer program that performed the surgeon's administrative tasks automatically, which saved his father a lot of time. When the program turned out to be useful to his father, Gerrit and Hans Mulder founded ChipSoft B.V. [6]
Gerrit Mulder's colleagues in the health care field got word of the software program that automated administrative tasks and were also interested in ChipSoft's first product. This growth of clientele also expanded ChipSoft's software goals, as other divisions of the health care business (besides surgeons) were also interested in automatizing certain administrative processes, such as appointment scheduling. Multiple different subsystems were developed to support health care divisions like ORs, radiology departments and emergency departments.
The collection of subsystem requests kept growing over time, which resulted in one complete and integrated software information system in 1994, called CS-EZIS.Net (ChipSoft - Elektronisch Zorg Informatie Systeem).

As of today, the software does not only assist in administrative tasks, but also assists in many logistic tasks, such as medicine prescriptions and OR scheduling. There are currently over 58 health care institutes using some form of ChipSoft's software, ranging from independent clinics to academic hospitals. ChipSoft offers a fully integrated solution based on state-of-the-art Microsoft technology and many of ChipSoft's customers have been rewarded with a 'stage 6' EMRAM by HIMSS Analytics Europe (HAE), which classifies a hospital's progress on EMR (Electronic Medical Record) versus other health care organizations around Europe and across the world [10]. An important factor of ChipSoft's success is its ISO 13485 certification. This certificate ensures that the complete software solution is CE marked as a class IIb device in compliance with the Medical Device Directive 93/42/EEC [11].

The latest version of the software (called HiX) contains many different subsystems that are all integrated into one information system. The included components are shown in Figure 1.1.


Figure 1.1: HiX Components

**Datawarehouse Department**

This research was performed at the Datawarehouse Department of ChipSoft B.V. The department mainly focuses on the subjects "Statistics and reporting" and "Decision support and intelligence" from Figure 1.1. Approximately 30 employees work every day to improve and develop new content to enrich the Datawarehouse module that HiX offers, which has completely been renewed during the past year in an intensive development period. The newly designed Datawarehouse module uses up-to-date techniques and contains new content to satisfy hospitals in their information needs.

## 1.2 Internship Research Outline

This research project was divided into multiple steps and this chapter is dedicated to explain its process.
Figure 1.2 states general steps included in the research process, of which some will be clarified in this chapter and others will be individually discussed in different chapters.

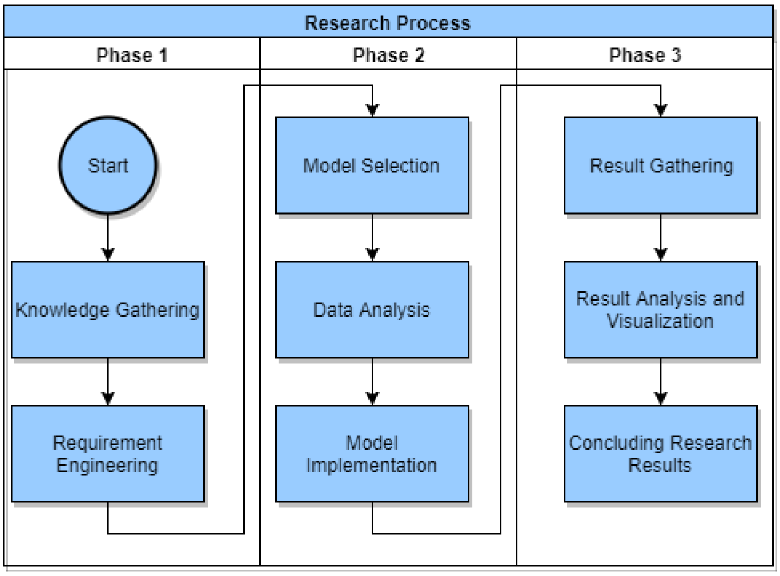



Figure 1.2: Research Process

As Figure 1.2 shows, the steps included in this research can be divided into three different phases. Each phase has its own main goal: the first phase's goal is to obtain knowledge and requirements (the input), the second phase's goal is to process the obtained knowledge and requirements into a suitable model and its implementation (the process) and the third phase's goal is to interpret the designed implementation and its results (the output).

As shown in Figure 1.2, the research started off with knowledge gathering and requirement engineering in the first phase. This was needed to align theoretical research with a practical application within ChipSoft's software packages, as well as aligning the research with existing health care processes in hospitals. As a result of this first step, the research's preliminary goals were translated into inputs (requirements, knowledge) and outputs (deliverables, result formats).
Afterwards, in phase 2, a model was selected for the predictive part of the research that suited the research's purpose and allowed the requested output requirements to be obtained. Next, the required datasets were obtained from ChipSoft's customers (hospitals) and analyzed using extensive data analysis. Once the required data analysis was performed, the model was programmed into an application in the model implementation step.
In phase 3, the model's implementation was used to gather, analyze, visualize and describe the results to conclude this research. The research steps in phase 1 will be briefly explained in this chapter, whereas the steps in phase 2 and 3 are described in different chapters.

### 1.2.1 Phase 1: Knowledge Gathering

Before the research framework was drawn, extensive knowledge needed to be gathered regarding the research's context (health care processes) and the capacity management project in general. Also, since the research should align with ChipSoft's software and the capacity management project, not only the general health care processes needed to be understood, but also the way ChipSoft has implemented these processes in its software. This resulted in a list of knowledge to be gathered that included (among others):

- The general inpatient care process and ChipSoft's implementation of this care process
- General OR Scheduling protocols used by ChipSoft
- The data structures used by ChipSoft to save the data

To obtain this knowledge, different methods were used, such as: interviewing colleagues from different departments, studying literature and previous research. Also, the two years of work experience already gained at ChipSoft was an important source of knowledge.

### 1.2.2 Phase 1: Requirement Engineering

This research evolves around analyzing the downstream effect the ORs have on the occupancy on the clinical wards, but before any model could be applied or any data could be analyzed, the input and output requirements for this research were discussed and clarified with all corresponding parties. Besides studying literature, many discussions and meetings took place in the requirement engineering process. The involved parties included:

- Vrije Universiteit supervisor
- Vrije Universiteit Exam committee
- ChipSoft's Datawarehouse management team
- ChipSoft's Inpatient Care Process consultants
- ChipSoft's Capacity Management consultants

Each of the involved parties listed above was responsible for a certain part of the requirements. The VU (Vrije Universiteit) mainly checked the research on scientific value and whether or not the research was considered within reach, whereas the ChipSoft Datawarehouse management team mainly verified the research's added practical value to the company and how this added value could be obtained (output requirements). ChipSoft's Inpatient Care Process and Capacity Management consultants assisted in obtaining the care process and the capacity related requirements respectively.

### 1.2.3 Phase 2

Once the requirements, health care process knowledge and ChipSoft's software knowledge was obtained, a suitable model was selected to perform the occupancy prediction part of this research. The selected model is discussed in chapter 4. The Data Analysis step and the Model Implementation step are discussed in chapters 5.1 and 6 respectively.

### 1.2.4 Phase 3

The final phase of this research consists of multiple steps regarding the different results obtained from the implementation of the model. These steps are explained in chapters 7 and 9.

# Chapter 2

# Process description

This section is devoted to giving insights in the general clinical care process as well as describing ChipSoft's implementation of this process in their software. The clinical care process is in general defined as the period between a patient's admission to the hospital and a patient's discharge from the hospital. This period can include many different care services, such as surgeries, screenings and observations.

## 2.1 Admission Priorities

A lot of different reasons (diagnosis) exist why a patient requires a hospital admission and its corresponding care process might be significantly different per reason and per patient. For hospitals, a big distinction between the required care process for a patient can be made by labeling an admission with a so-called admission priority. The admission priority denotes the urgency of an admission, which is registered using a binary indicator "emergency yes/no". Emergency admissions have priority over elective (scheduled, non-emergency) admissions and therefore follow a different care process. The main difference in care processes between the emergency and the elective admissions is the horizon in which the care services need to be provided. Elective patients receive their care on a scheduled basis, which can range from a few days until months between the date of scheduling and the actual admission date, whereas emergency patients need to be served as soon as possible and therefore are also often called the non-scheduled patient stream. Instead of using the term 'priority' for the admission, 'origination' is also used to denote the distinction between elective and emergency admissions.

## 2.2 Admission Categories

Another indicator to distinguish admissions is the admission category. The admission category denotes in general what kind of care/cure is offered to a patient and also where the services are offered. Currently, four different admission categories exist within ChipSoft's software: the inpatient, the day treatment, the outpatient and the observation category. Each of the categories has a different default care process, which will be briefly described below. The default care process for the inpatient admission category will be described in more detail in section 2.3 as it will be the main focus of this research.

**Inpatient Category**
The inpatient category contains the admissions that are considered clinical (receiving some form of cure services) and occupy a hospital bed for at least one night. It is possible that these admissions receive a surgery in one of the hospital's ORs and are usually admitted in one of

the hospital's clinical wards.
**Day Treatment Category**
The day treatment category contains the admissions that are considered clinical and possibly receive a surgery in one of the hospital's ORs. These patients are admitted and discharged on the same day and therefore never occupy a bed during the night. The day treatment patients are usually admitted in one of the hospital's day treatment wards.
**Outpatient Category**
The outpatient category contains patients that are neither clinical nor receive any surgery in one of the hospital's ORs. Patients in the outpatient category are treated outside of the clinical departments of a hospital.
**Observation Category**
The observation category contains only patients that are admitted for observation purposes, as its name reveals. Patients in this category will not receive any treatment, but will only be observed for a certain period of time.

An important note to make is that each admission can only be assigned one category and one admission priority, if some event arises that makes the admission belong to a different category, the hospital needs to change the admission's category.
In the next sections, the default inpatient care process is described for the inpatient category

## 2.3 Inpatient Care Process

Many of the hospital's health care services require patients to remain in the hospital for a certain period. The process from the moment of admittance of a patient in the hospital until the moment of discharge from the hospital is called the inpatient care process (for inpatient admissions). In general, patients that will receive a surgery which requires a rehabilitation period with at least one night will follow such an inpatient care process. Patients receiving surgery and rehabilitating for at least one night are the main subject of this research and this section will describe the inpatient care process in parts and ends with a summarizing process picture.

### 2.3.1 Preoperative period

For elective patients, the admission to the hospital is the result of a scheduled surgery. This means the admission usually starts on the day of surgery or a few days before and starts the inpatient care process. The period before surgery, called the preoperative period, might differ in length between admissions and surgeries based on many different patient and situational aspects. For instance, some surgeries require preoperative screenings to be done, which increases the duration of the preoperative period.
For emergency patients, the admission and the surgery are not chronologically related, as the admission and the surgery are both unscheduled. Emergency patients need to be admitted as soon as possible and if they require a surgery, that surgery will also be performed as soon as possible. However, when compared to the elective patients, the priority might not lie in giving surgery to the patient, but in stabilizing and observation first. This means that the unstructured time horizon and the unscheduled characteristics of emergency patients might result in a different pre-operative length of stay distribution when compared to the elective patients. This possible difference in distribution is addressed in the data-analysis section of this research.

### 2.3.2 Postoperative period

After surgery, most patients require a rehabilitation period to recover before being discharged from the hospital. Besides rehabilitation, some surgeries require the patients to stay in the hospital for observation reasons after performing the surgery (such as transplants). The length of the postoperative period is assumed to be less affected by the distinction between elective and non-elective patients, but mostly depends on patient aspects and the intensity of the surgery. This assumption will be tested in the data-analysis section of this research.

### 2.3.3 Length of Stay

When combining the preoperative, the perioperative (the surgical time) and the postoperative period, the length of stay of a patient is obtained. This length of stay denotes the full duration of a patient's admission in the hospital. The length of stay is often abbreviated as LoS and denoted in hours or days.

### 2.3.4 Clinical wards

Hospital beds are generally divided among the different hospital wards. The wards on which clinical patients (having admission categories 'day treatment' or 'inpatient') are hospitalized are called the clinical wards. These clinical wards differ from each other on aspects like specialty (types of patients it serves), size, severity (ICU, medium care) and age of the patients. Also, many hospitals use specific clinical wards with adjusted working hours for the day treatment patients, as they will not remain hospitalized during the nights.
Preferably, the patient's pre- and postoperative periods take place on the same clinical ward and therefore the bed assigned for the preoperative period will remain reserved for the postoperative period during surgery. Although preferred, it is not always possible for a patient to remain on the same clinical ward for the full admission duration. In many cases, patients need to be transferred to different wards (from/to ICU, different specialty etc.) during their admission.
Hospitals using ChipSoft's software trace a patients whereabouts in the hospital and register the possible changes happening in a patient's placement (using admission periods). This tracing is not only necessary for logistic reasons, it also has financial consequences as the financial compensation for nursing days can differ per ward (a nursing day on the ICU is more expensive than a nursing day on a medium care ward).

Figure 2.1 gives a general overview of an admission process for inpatient admissions with their respective origination (elective or non-elective), their admission divided into multiple admission periods as well as their destination at the moment of discharge.

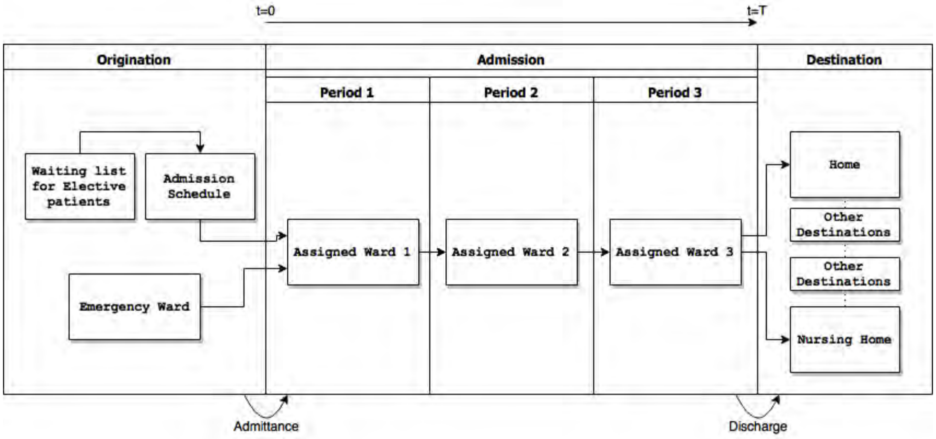



Figure 2.1: General Admission Process

## 2.4 Surgical Scheduling

This section describes the standard procedure offered in ChipSoft's software to support hospitals in scheduling surgeries and their corresponding admissions, meaning it is applicable for elective patients only. Surgeries are scheduled based on the availability of hospital resources (ORs, surgeons, etc.) and their corresponding admissions are the result of the scheduled surgeries and therefore scheduled afterward.

### 2.4.1 Master Surgical Schedule

Each year, hospitals formulate strategic goals, usually represented by OR time, for each of the different aggregate levels (such as specialties) that use the ORs. The OR time denotes the total amount of time the aggregate level has an OR at its disposal. This (yearly) total amount of OR time for a certain aggregate level is transformed into a repeating weekly or two-weekly tactical schedule that divides the total available OR time among all the aggregate levels. The distribution of the amount of OR time for a certain aggregate level in the tactical schedule is done with respect to the hospital's strategic goals. This tactical schedule is called the Master Surgical Schedule (MSS). There are many different ways to design a suitable MSS and make a decision on which aggregate level the OR time is distributed [7].
For example, consider a weekly Master Surgical Schedule representing three specialties as the aggregate level: Cardiology, Orthopaedics and Urology. If Cardiology requires 200 hours of OR time per year, Orthopaedics requires 400 hours and Urology requires 300 total hours per year, the MSS should represent these proportions accordingly. So, a fair distribution would be 200/52, 400/52 and 300/52 hours of weekly OR time for Cardiology, Orthopaedics and Urology respectively.

### 2.4.2 Sessions

Within ChipSoft's software, the periodic amount of OR time used for each aggregate level, of which specialty is the most common, is divided into multiple surgical time slots, called sessions. For instance, OR $i, i \in \{1...I\}$, is operational between 08:00-16:00 and should provide the specialties Cardiology, Orthopaedics and Urology with 3, 3 and 2 hours of surgical time respectively on day $q, q \in \{1, 2, .., Q\}$. This would result in one 3-hour session for Cardiology, one 3-hour session for Orthopaedics and one 2-hour session for Gynecology on day $q$. By structuring the surgeries from the different specialties into one session of sequenced surgeries, there is no potential OR time lost on changing personnel and equipment between the different specialties.

After the MSS is finalized and sessions are designed and allocated to the specialties accordingly, the MSS is applied to a certain time horizon so surgeries can be scheduled by assigning them to a certain session on a certain date. When a surgery is scheduled, the required admission process for the patient is also scheduled and the patient is notified regarding the date set for the surgery. For example, consider a weekly repeating MSS that allocates surgical specialties $SS_i, i \in 1, 2..., 7$ to sessions into a tactical surgical schedule using 5 ORs. An example for such an allocation in a tactical surgical schedule is shown in Figure 2.2.

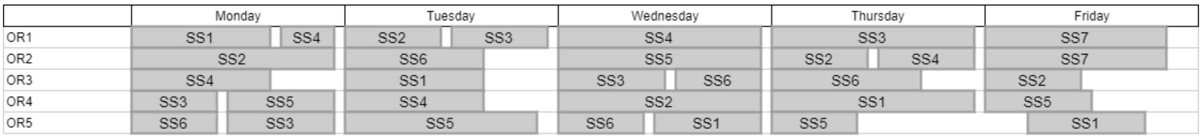



Figure 2.2: Example weekly tactical surgical schedule for 7 specialties and 5 ORs

# Chapter 3

# Problem Statement

The hospital health care is currently in need of efficiency improvements because of different reasons such as market changes, labor shortages, different costing calculation introduced in the Netherlands, an increase in health expenditures and long waiting lists [8, 21, 14].
An important and essential subject of efficiency improvement is nurse capacity management as it accounts for a large part of the hospital's budgets [22]. By preventing overstaffing, hospitals can achieve cost efficiency in their nurse capacity management and reduce unnecessary use of their budgets. However, appropriate staffing levels are required to provide the quality of care hospitals strive to provide and understaffing the clinical wards might result in a decrease of the quality of care [12, 16]. Obviously, another important reason to prevent understaffing is the increased pressure on the nurses in the understaffed wards, which might lead to stress and burnouts among the nurses [1].
To be able to adjust the staffing capacity on the care demand, which can be expressed as the number of hospitalized patients, hospitals need to be able to predict and estimate the care demand. An important influence on the number of hospitalized patients are the scheduled and unscheduled surgeries performed in the hospital. So, as stated in the introduction, there exists an important relation between the hospitals performed surgeries and the number of patients occupying a bed on the clinical wards [2, 20, 18].

This relation between ORs and the clinical wards receiving the patients after or before surgery is non-negligible when developing a Master Surgical Schedule (MSS), because the MSS determines which specialty performs surgery when and therefore determines the patient stream into the clinical wards because of the ORs. This leads to the research question of this internship:

> "Is it possible to design and implement a predictive analysis model based on the downstream relationship between the ORs and the clinical wards to support hospitals in making nursing capacity related decisions?"

## 3.1 Deliverables

The deliverables of this research include a framework that allows hospitals to analyze the downstream relationship between the ORs and the clinical ward occupancies in a realistic environment with emergency and elective admissions. To do so, the framework should include the extraction of important occupancy related performance indicators as well as include a model to predict ward occupancies based on the MSS. The result of the research is to be given in a report, which also includes the Datawarehouse possibilities and requirements for ChipSoft to use the framework.

# Chapter 4

# Predictive Model

In this section, the model used to predict the occupancy at the hospital's wards is described in general, its implementation and application will be described in chapter 6. The basis of the predictive model is built by considering a timeline horizon as shown in Figure 4.1.

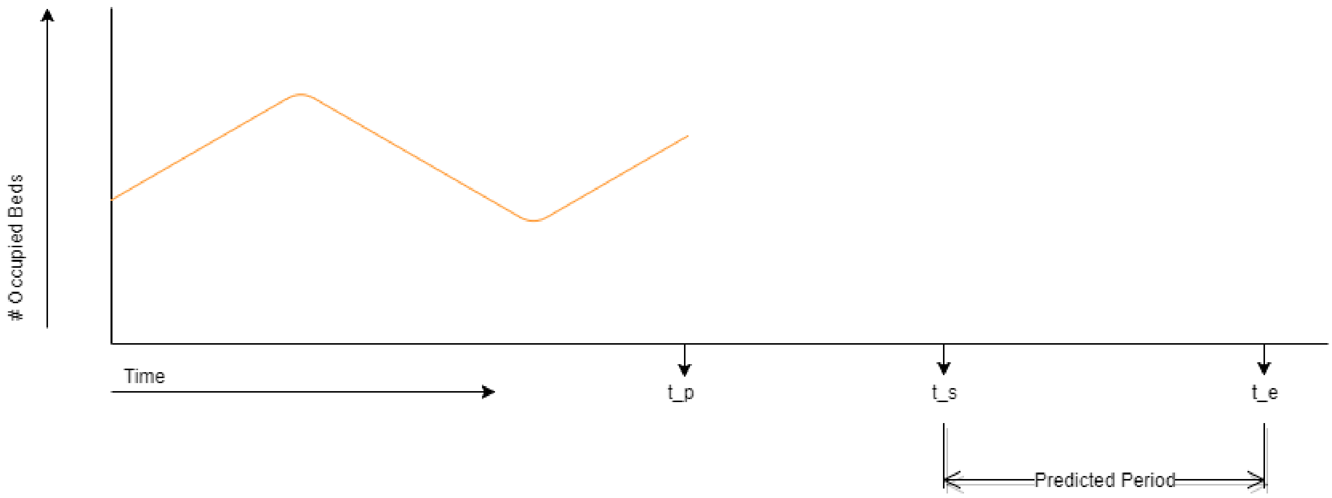



Figure 4.1: Prediction Timeline

In Figure 4.1, the orange line represents some number of occupied beds (vertical axis) at a certain time (horizontal axis). The moment of predicting is important, as it is assumed that all information regarding occupied beds is known when the prediction is made and the time difference between the moment of predicting and the predicted period is of influence to the predicted values. This moment of predicting is called $t_p$ and, as Figure 4.1 shows, the predicted period starts on $t_s$ and ends on $t_e$.
The predicted number of occupied beds at a certain time $t_n$, $n \in (s, s+1, ..., e)$, is assumed to be the result of three different sources with respect to the moment of predicting, $t_p$:

1. The remainder of the patients hospitalized at the moment of prediction ($t_p$).

2. The arriving elective patients as a result of the MSS sessions between $t_p$ and $t_n$.

3. The arriving emergency patients between $t_p$ and $t_n$.

For the first patient source, the probability that a patient hospitalized at $t_p$ is still hospitalized at $t_n$ depends on the patient's remaining length of stay, which obviously is uncertain and therefore estimated. This first source of patients is referred to as the 'known' patient group, as they are already present at the time of predicting ($t_p$) and the number of occupied beds at $t_p$ by

this first group is therefore known. Also, since the patients in this group are already present, the number of occupied beds as a result of this first group will only decrease over time from $t_p$ onward.
For the second patient source, the probability that a patient occupies a bed at $t_n$ depends on two things: firstly, the patient needs to be admitted because of an MSS session between $t_p$ and $t_n$ and secondly, the patient needs to be still occupying a bed at time $t_n$. This second group is referred to as the 'semi-known' group, as the hospital knows that some number of patients will arrive due to the execution of surgical sessions from the MSS between $t_p$ and $t_n$, but do not know the exact number of patients as a result of the performed or to be performed MSS sessions.
The third source of patients, the emergency arrivals, occupy a bed at time $t_n$ with a probability that depends on two things: firstly, an emergency admission needs to happen between $t_p$ and $t_n$ and secondly, the admitted emergency patient needs to be still occupying a bed at time $t_n$. Although hospitals expect some number of emergency arrivals in certain time intervals, this patient source is considered the 'unknown' group as they are completely unscheduled.

Each of the three sources is responsible for a certain fraction of the total number of occupied beds at the predicted time $t_n$. The next sections describe the model's prediction method for each of the different sources.

## 4.1 Patient Source 1: Remaining Patients

It is possible that patients hospitalized at $t_p$ remain hospitalized until $t_n$. This means they are still occupying a bed at $t_n$ and are therefore responsible for a fraction of the total number of occupied beds at $t_n$.
Whether or not a patient remains hospitalized between $t_p$ and $t_n$ depends on the patient's remaining length of stay and, which is obviously not fully certain.
In order to estimate the number of remaining present patients at the predicted time $t_n$, the present patients at time $t_p$ are separated into two groups: the present emergency patients and the present elective patients, both having their own length of stay characteristics calculated. These length of stay characteristics are used to calculate the expected remaining length of stay for the patients in the first patient source. The estimated remaining number of patients that are still present at time $t_n$ is therefore calculated differently for both groups and described in the following sections.

### 4.1.1 Remaining Elective Patients

The elective patients from patient source 1 received surgery in one of the ORs based on a MSS session and are in the postsurgical phase of their admission. Because of the repetitive use of the sessions in a MSS, it is possible to analyze the historically realized length of stay characteristics for patients that received surgery in each of the MSS sessions.
These length of stay characteristics of the MSS sessions are used to estimate the remaining length of stay for the elective patients present at $t_p$ resulting from the corresponding sessions. Denote $A_p = \{\text{number of elective patients present at } t_p\}$, who are the result of $S$ different MSS sessions performed before $t_p$. Then, for each session $s$, $s \in (1, 2, ..., S)$, the number of present patients at $t_p$ as a result of that session is denoted by $A_{s,p}$.
The probability that a present patient from session $s$ is still hospitalized at $t_n$ is calculated using the session's length of stay characteristics. To do so, denote:

$$L_s = \text{Length of stay of a patient from session } s$$

So, since it is known that the patient is present at $t_p$ and therefore arrived at some time $t_a$, $t_a \leq t_p$, the probability of a patient present at $t_p$ to be still hospitalized at $t_n$ is a conditional probability denoted as:

$$P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a) \\ \text{, where } t_a \leq t_p \tag{4.1}$$

In order to calculate the probability shown in (4.1) for patients resulting from session $s$, the empirical cumulative distribution function of the length of stay ($L_s$) for session $s$ is used. This empirical cumulative distribution function for the length of stay is obtained for each session $s$ and this cumulative distribution function is used to express the expected length of stay of patient's resulting from this session $s$.
Namely, given a cumulative probability distribution $F_{L_s} = P(L_s < l)$, its conditional probability $P(L_s < l|L_s < y)$, $l \leq y$, can be rewritten as:

$$\frac{P(L_s < l)}{P(L_s < y)} \tag{4.2}$$

The required probability for patients from session $s$ to be still hospitalized at $t_n$, given they were present at $t_p$, can be deducted from (4.2). However, instead of calculating the conditional probability $F_{L_s} = P(L_s < t_n - t_a|L_s < t_p - t_a)$, the probability $P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a)$ is calculated which is denoted as:

$$\frac{P(L_s \geq t_n - t_a)}{P(L \geq t_p - t_a)} = \frac{1 - P(L_s < t_n - t_a)}{1 - P(L_s < t_p - t_a} \tag{4.3}$$

The value of (4.3) is calculated by dividing the number of historical occurrences where $L_s >= t_n - t_a$ by the number of occurrences where $L_s >= t_p - t_a$:

$$\frac{\sum_{c=1}^{C_s} \mathbb{1}_{\{L_c \geq t_n - t_a\}}}{\sum_{c=1}^{C_s} \mathbb{1}_{\{L_c \geq t_p - t_a\}}} \tag{4.4}$$

Where $C_s$ represents the number of patients that received surgery in session $s$ before $t_p$ and $L_c$ the length of stay of patient $c$.
The calculation of (4.4) can also be expressed graphically, which is done in Figure 4.2. The probability from (4.4) can be interpreted as the number of historical patients with a length of stay in the striped part of the grey area, divided by the total number of historical patients with a length of stay in the grey area. In Figure 4.2 for example, the fraction of people in the grey area equals approximately 20%, of which approximately 7% lies within the striped part of the grey area. This would result in a value for (4.4) of $\frac{0.07}{0.20} = 0.35$.

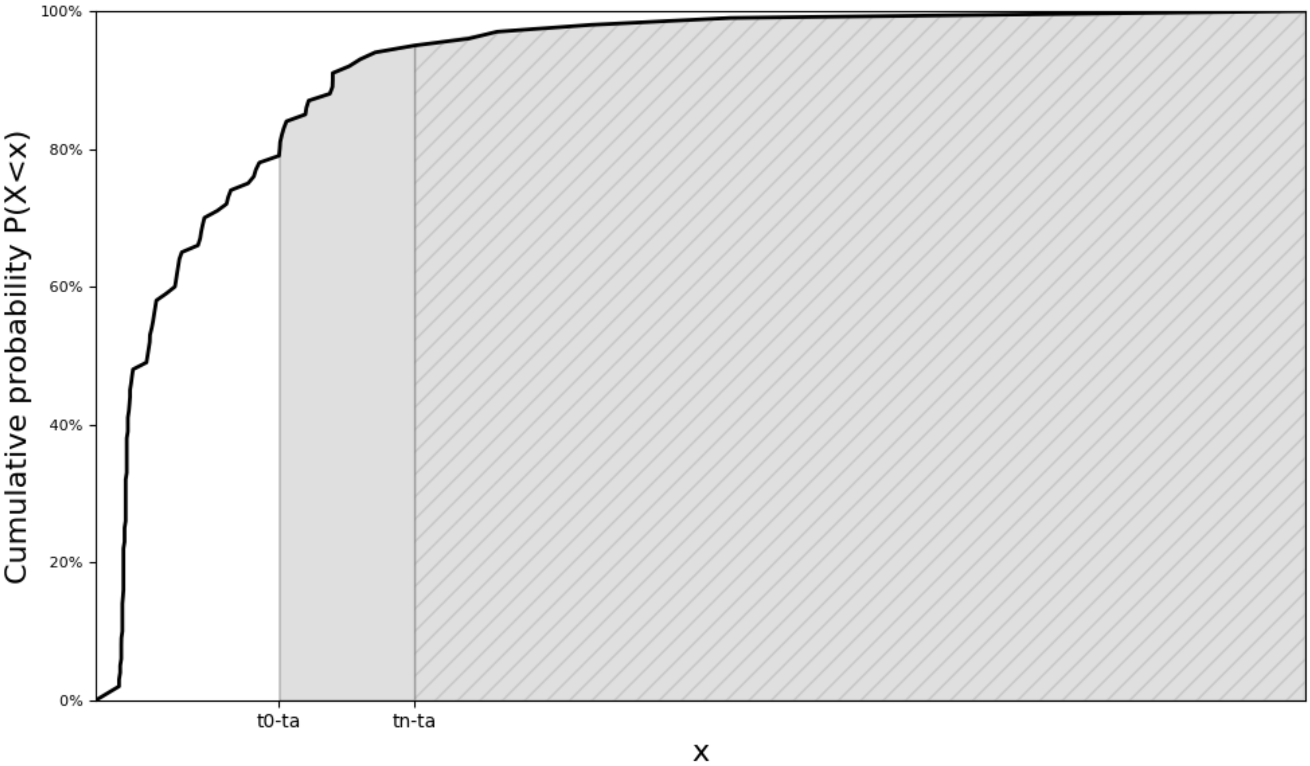



Figure 4.2: Cumulative Distribution Function

By using the probability of (4.4), it is possible to estimate the number of patients present at $t_n$, given they are present at $t_p$, by calculating:

$$\sum_{s=1}^{S_p} \sum_{a=1}^{A_{s,p}} P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a) \tag{4.5}$$

Where $S_p$ denotes the total number of different sessions the patients present at $t_p$ result from, $A_{s,p}$ the total number of present patients at $t_p$ as a result of session $s$ and $P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a)$ the probability stated in (4.4).

### 4.1.2 Remaining Emergency Patients

The calculation of the expected remaining emergency patients is also done based on the length of stay characteristics of the patients. However, instead of calculating the length of stay characteristics for a session, the length of stay characteristics are calculated for the different time intervals the patients arrived in.
For each possible time interval $i \in I$, which means a certain repetitive time horizon is divided into $I$ time intervals, the median length of stay is calculated for patients that arrived in the interval. For example, $I$ could represent each day of the week ($I = 7$). This median length of stay for time interval $i$ is assumed to deterministic for emergency arrivals that occur in future intervals of $i$. Denote $E_p$ the number of emergency patients present at $t_p$, of which each patient arrived in one of the $I$ time intervals, therefore $E_{i,p}$ denotes the total number of emergency patients present at $t_p$ as a result of time interval $i$.
Since each patient in $E_{i,p}$ arrived before $t_p$ and the length of stay for interval $i$ is assumed to be deterministic, the remaining length of stay of the patient is the difference between the median length of stay for emergency patients that arrived in time interval $i$ and the patient's current length of stay $t_p - t_a$.

For each patient in $E_{i,p}$, the probability of still being hospitalized at $t_n$ is defined as: 1 if the patients remaining length of stay at $t_p$ is bigger than $t_n - t_p$ and 0 elsewhere. Therefore, the expected number of remaining present emergency patients at $t_n$ is denoted by:

$$\sum_{i=1}^{I_p} \sum_{a=1}^{E_{i,p}} \mathbb{1}_{\{L_i^r \geq t_n - t_p\}} \tag{4.6}$$

Where $I_p$ denotes the total amount of time intervals before $t_p$ of which patients are still present, $E_{i,p}$ the total number of present emergency patients at $t_p$ as a result of time interval $i$, $L_i^r$ the remaining length of stay for patients that arrived in time interval $i$.

### 4.1.3 Combined Remaining Patients

The total predicted number of patients still hospitalized at $t_s$, given they were hospitalized at $t_p$, is obtained by combining (4.4) and (4.6):

$$\sum_{s=1}^{S_p} \sum_{a=1}^{A_{s,p}} P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a) + \sum_{i=1}^{I_p} \sum_{e=1}^{E_{i,p}} \mathbb{1}_{\{L_i^r \geq t_n - t_p\}} \tag{4.7}$$

Where:

- $S_p$ denotes the total number of different sessions of which patients are present at $t_p$
- $A_{s,p}$ denotes the total number of elective patients present at $t_p$ as a result of session $s$
- $P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a)$ the probability calculated as stated in formula 4.4 for session $s$
- $I_p$ denotes the total number of different time intervals of which patients are still present at $t_p$
- $E_{i,p}$ denotes the total number emergency patients present at $t_p$ that have arrived in time interval $i$
- $L_i^r$ denotes the remaining length of stay at $t_p$ for emergency patients that arrived in time interval $i$

## 4.2 Patient Source 2: Elective Patients

The patient included in patient source 2 are separated into two different groups: the group of patients that is present at $t_n$ because of preoperative reasons and the group that is present because of postoperative reasons.
The method of prediction for both patient groups is, besides some different input parameters, equal. This section will describe the method for the postoperative patient group and states the changes to be made to obtain the calculation for the preoperative patient group.

### 4.2.1 Prediction Postoperative Patients

The postoperative patient group from patient source 2 includes patients that are present at $t_n$, but not present at $t_p$. These patients received some form of surgery as a result of an MSS session between $t_p$ and $t_n$ (because they are present at $t_n$ for postoperative reasons).
It is assumed that the hospital does not know the exact number of patients that will receive surgery in a session between $t_p$ and $t_n$, let alone know whether or not the patients will still be rehabilitating at $t_n$. The prediction of the rehabilitating number of patients at $t_n$ is done by analyzing the currently active Master Surgical Schedule and its corresponding session characteristics.
To do so, denote $S_{n-}$ the total number of sessions that will be performed between $t_p$ and $t_n$. Then for each session $s \in (1, 2, ..., S_{n-})$ a certain number of $N$ elective patients will receive surgery and enter one of the hospital's clinical wards for rehabilitation afterward.
Considering the fact that the prediction is made at $t_p$ and $N$ is unknown for the sessions at that time, each possible $N$ is expected to occur in a session $s$ with a probability based on the historical analysis of the number of surgeries performed session $s$. Namely, the probability for $s$ to contain $n$ patients is denoted by:

$$P_s(N = n) = \frac{\#\text{ performed sessions } s \text{ having } n \text{ patients}}{\#\text{ performed sessions } s} \tag{4.8}$$

Each patient $x$, $x \in (1, 2, .., n)$ for a session $s$, is still hospitalized for rehabilitation at $t_n$ with a probability based on the empirical cumulative distribution function of the postsurgical length of stay for session $s$. This means:

$$P_s(L^{post} \geq t_n - t_{se}) = 1 - P_s(L^{post} < t_n - t_{se}) \tag{4.9}$$

Where $t_{se}$ denotes the end time of session $s$ and $L^{post}$ the postsurgical length of stay for patients resulting from session $s$. Probability (4.9) can be graphically described as the probability of $L$ to reach the grey part of Figure 4.3.

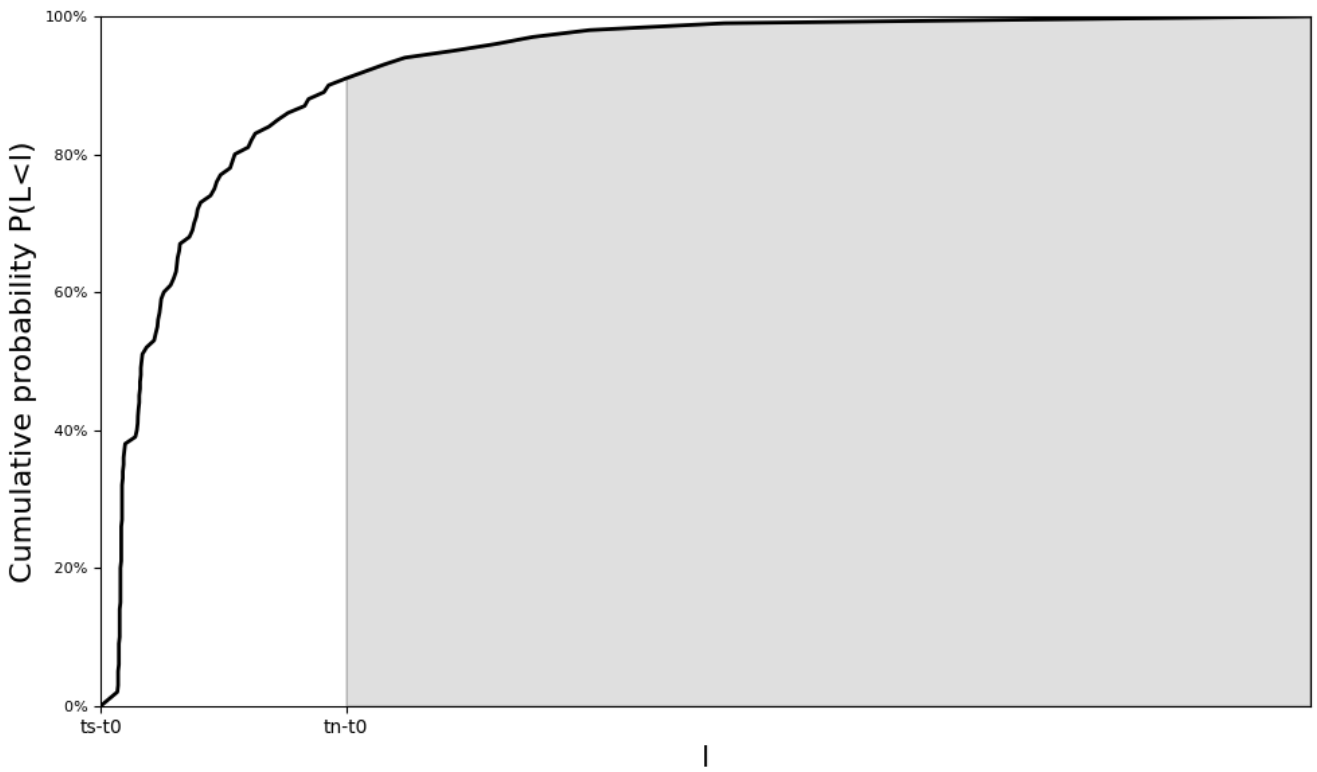



Figure 4.3: Cumulative Distribution Function $P(L < l)$ and $t_p = t_0$

After obtaining the probability for a patient from session $s$ to be still hospitalized at $t_n$, this probability needs to be translated to a probability that describes the chance of $k$ patients to be still hospitalized at $t_n$ as a result of this specific session $s$. To do so, assume that session $s$ contains $n$ patients and each of those patients has a probability as denoted in formula (4.9) to be still present at $t_n$, the probability of $k$ patients to be still present at $t_n$ as a result of this session $s$ is then described by:

$$P_s(K = k) = \binom{n}{k}(1 - P_s(L^{post} < t_n - t_{se})^k P_s(L^{post} < t_n - t_{se})^{n-k} \tag{4.10}$$

Combining (4.10) with (4.8), which describes the probability of session $s$ containing $n$ patients, for all sessions in $S_{n-}$ results in a formula that calculates the expected number of elective patients present at $t_n$ for postsurgical reasons:

$$\sum_{s=1}^{S_{n-}} \left( \sum_{n=1}^{n_{max}} P_s(N = n) \sum_{k=1}^{n} k \binom{n}{k}(1 - P_s(L^{post} < t_n - t_{se})^k P_s(L^{post} < t_n - t_{se})^{n-k} \right) \tag{4.11}$$

Where:

- $S_{n-}$ denotes the collection of sessions to be performed between $t_p$ and $t_n$
- $n_{max}$ denotes the maximum number of surgeries ever performed in session $s$
- $P_s(N = n)$ denotes the probability of session $s$ to contain $n$ patients (surgeries) as stated in formula 4.8
- $1 - P_s(L^{post} < t_n - t_{se})$ denotes the probability that the postsurgical length of stay of a patient from session $s$ is longer than $t_n - t_{se}$, as stated in formula (4.9)

### 4.2.2 Prediction Preoperative Patients

The method described in section 4.2.1 can be used for the prediction of the number of preoperative patients present at one of the clinical wards at $t_n$ with minimal adjustments. Instead of calculating the probability of a patient to be still present at $t_n$ because of a rehabilitation period after receiving surgery in a session, the probability of already being present at $t_n$ for a session occurring after $t_n$ is calculated. To do so, an empirical cumulative distribution of the preoperative length of stay is extracted from the historical data for every session $s$.
Also, the selection of sessions $S_{n-}$ that will occur between $t_p$ and $t_n$ is changed to the selection of sessions $S_{n+}$ occurring after $t_n$ but before $t_e$ ($t_e$ being the end of the predicted period). applying these changes to (4.11) results in a formula to calculate the expected number of preoperative patients in one of the hospital's clinical wards at $t_n$:

$$\sum_{s=1}^{S_{n+}} \left( \sum_{n=1}^{n_{max}} P_s(N=n) \sum_{k=1}^{n} k \binom{n}{k} (1 - P_s(L^{pre} < t_{se} - t_n)^k P_s(L^{pre} < t_{se} - t_n)^{n-k} \right) \tag{4.12}$$

Where $S_{n+}$ denotes the sessions that will occur after $t_n$ but before $t_{se}$ ($t_n <= t_{se} <= t_e$) and $P(L^{pre} < t_{se} - t_n)$ the probability of the pre-operative length of stay of a patient for session $s$ to be smaller than $t_{se} - t_n$, comparable with the probability stated in formula (4.9).

The sum of (4.11) and (4.12) will represent the expected number of elective patients present at $t_n$ as a result of the MSS.

## 4.3 Patient Source 3: Emergency Patients

The method used to calculate the expected number of emergency patients present at $t_n$, given they are not present at $t_p$, is based upon a Poisson arrival process. This arrival process is historically shown to suit the emergency arrivals process in a hospital well and will be assumed applicable [23].
In this research, a time-dependent arrival rate $\lambda(i)$ is used for each time interval $i$. Also, the length of stay for emergency patients is time-dependent, meaning emergency arrivals can have a different expected length of stay based on the time interval of arrival.
This research considers the expected length of stay for emergency patients from a certain time interval $i$ to be the length a hospital should reserve a bed for the patient. Therefore, the length of stay is assumed to be deterministic for the practical purposes of this research. Considering the Poisson arrival process and the deterministic length of stay, the number of emergency patients present at time $t_n$, given they were not present at $t_p$, can be described by:

$$\sum_{i=0}^{I_{n-}} \lambda(i) \mathbb{1}_{\{L_i \geq t_n - t_i\}} \tag{4.13}$$

Where $I_{n-}$ denotes the number of time intervals between $t_p$ and $t_n$, $\lambda(i)$ the arrival rate of the corresponding time interval and $t_i$ the start time of time interval i. The indicator $\mathbb{1}_{\{L_t >= t_i\}}$ determines whether or not the arrivals in time interval $i$, having a deterministic length of stay $L_i$, are still present at $t_n$ (by having a $L_i$ larger than $t_n - t_i$).

## 4.4 Combined Patient Sources

Denote $X(t_n)$ the number of present patients at time $t_n$. Then, when considering a time horizon in which a prediction for the number of present patients at $t_n$ is made at $t_p$, the formulas (4.7), (4.11), (4.12) and (4.13) can be combined into one model to predict $X(t_n)$:

$$
\begin{aligned}
X(t_n) = & \sum_{s=1}^{S_p} \sum_{a=1}^{A_{s,p}} P(L_s \geq t_n - t_a \mid L_s \geq t_p - t_a) + \sum_{i=1}^{I_p} \sum_{e=1}^{E_{i,p}} \mathbb{1}_{\{L_i^r \geq t_n - t_p\}} \\
& + \sum_{s=1}^{S_{n-}} \left( \sum_{n=1}^{n_{max}} P_s(N=n) \sum_{k=1}^{n} k \binom{n}{k} (1 - P_s(L^{post} < t_n - t_{se})^k P_s(L^{post} < t_n - t_{se})^{n-k} \right) \\
& + \sum_{s=1}^{S_{n+}} \left( \sum_{n=1}^{n_{max}} P_s(N=n) \sum_{k=1}^{n} k \binom{n}{k} (1 - P_s(L^{pre} < t_{se} - t_n)^k P_s(L^{pre} < t_{se} - t_n)^{n-k} \right) \\
& + \sum_{i=0}^{I_{n-}} \lambda(i) \mathbb{1}_{\{L_i \geq t_n - t_i\}}
\end{aligned}
\tag{4.14}
$$

This model contains all three patient sources with their calculations as defined in formula's: (4.7), (4.11), (4.12) and (4.13).

**Confidence Interval**

The formula stated in (4.14) calculates the expected number of occupied beds at time $t_n$. Because of $X(t_n)$ being an expectation, it is very likely that the actual measured value is not exactly equal to the value predicted for $X(t_n)$. Therefore, a confidence interval around $X(t_n)$ is simulated with a higher probability of containing the actual measured value compared to the probability of the actual measured value being exactly equal to $X(t_n)$. The implementation of this simulation method is further clarified in section 6.1.2.

# Chapter 5

# Data

## 5.1 Data Selection

The data required to apply the model is extracted for two different hospitals, which use ChipSoft's software and therefore have the data stored in their databases. This chapter describes the used fields from the extracted datasets and states the performed data analysis, which includes the extraction of the input parameters for the model.
From the hospitals' databases, the production data regarding surgeries and their corresponding admissions as well as the data regarding the Master Surgical Schedules is extracted into two different datasets. The first dataset contains the surgery registrations with the required details and the second dataset contains the performed sessions of the MSS used over time. The next subsections describe both datasets in more detail.

### 5.1.1 Surgery dataset

The surgery dataset contains the required details regarding surgeries and their corresponding admissions registered in the database since 2014. It counts 12,959 surgeries in total and contains, besides the directly extracted data fields from the database, some details that are calculated for this research's purpose. Table 5.1 shows the directly extracted fields as well as the calculated fields that were used from the dataset. The column 'calculated or extracted' denotes whether a field is directly extracted from the database or calculated based on the extracted fields. A complete list of fields included in the dataset can be found in appendix A.

| Surgery Dataset | | |
|---|---|---|
| Field | Calculated\Extracted | Description |
| SESSIE | Extracted | Session name |
| PLANNR | Extracted | Unique admission number |
| START | Extracted | Start time of the session |
| STOP | Extracted | End time of the session |
| OPERATIENR | Extracted | Unique surgery number |
| DATUM | Extracted | Date of the session |
| admissionDateTime | Calculated | Combines date and time of admission |
| surgeryDateTime | Calculated | Combines date and time of surgery |
| dischargeDateTime | Calculated | Combines date and time of discharge |
| STATUS | Extracted | Admission status |
| SPECIALISM | Extracted | Admission specialty |
| SPOED | Extracted | Admission priority |
| SESSIENR | Extracted | Unique session number |
| AFDELING | Extracted | Admission ward |

| | | |
|---|---|---|
| STATUS | Extracted | Surgery status |
| ANNUDAT | Extracted | Admission cancelling date |
| CATEGORIE | Extracted | Admission main category |
| CATEGORIE | Extracted | Admission sub-category |
| losPreSurgery | Calculated | Length of Stay before surgery (hours) |
| losPostSurgery | Calculated | Length of Stay after surgery (hours) |
| LoS | Calculated | Total length of stay (hours) |

Table 5.1: Admission Dataset

In the surgery dataset, the field 'OPERATIENR' contains a unique key for each different surgery in the dataset, meaning 12,959 different values for the 'OPERATIENR' field exist in the dataset.
The calculated fields 'losPreSurgery' and 'losPostSurgery' are calculated differently for surgeries with a registered 'SESSIE' (which denotes the session in which the patient received surgery) field and surgeries without a registered 'SESSIE' field. For surgeries with a registered session, the length of stay before surgery ('losPreSurgery') is calculated by measuring the time difference between the admission datetime and the end of the session in which the patient received surgery. For surgeries without a registered session, this field denotes the time difference between the admission datetime and the actual surgery datetime, since no session end datetime exists. The same difference holds for the field 'losPostSurgery', which denotes the length of stay of a patient after receiving surgery; for surgeries with a session this field denotes the time difference between the end of the session and the discharge datetime, whereas for surgeries without a registered session it denotes the time difference between surgery and the discharge datetime. The calculated field 'LoS' is the total time difference between the admission datetime and the discharge datetime.

### 5.1.2 Roster Dataset

The roster dataset contains data regarding the performed MSS sessions since 2014. It contains 5,596 performed sessions with their corresponding details. Table 5.2 states the fields included in the roster dataset. All fields included in the roster dataset are directly extracted from the database.

| Roster Dataset | |
|---|---|
| Field | Description |
| weekday | Day of session date |
| SESSIE | Session name |
| DATUM | Date of the session |
| dayInMonth | Day number in month of session date |
| Year | Year of session date |
| Month | Month number of session date |
| START | Start time of the session (in minutes after 00:00) |
| STOP | End time of the session (in minutes after 00:00) |
| SESSIENR | Unique session number |
| duration | Length of session (in minutes) |

Table 5.2: Roster Dataset

In the roster dataset, the field 'SESSIENR' contains a unique key for each execution of a session on a certain date, meaning 5,596 different values for the field 'SESSIENR' exist.

## 5.2 Data Analysis

To allow occupancy forecasting as described in the Problem Statement from section 3, data analysis has to be performed to obtain the model's input parameters. Besides the extraction of the input parameters, this section also describes the data exploration performed to gain some feeling regarding the data. In the first subsection, this data exploration is described by calculating basic session and admission characteristics, whereas the second subsection describes the extraction of the model's input parameters.
The data analysis shown in this section is performed on the datasets extracted from hospital A's database.

### 5.2.1 Session and Admission Characteristics

By analyzing the performed sessions and admissions contained in the datasets, some first insights can be obtained regarding the hospital's usage of sessions and their corresponding admissions, this section describes the analysis of important session and admission characteristics used in this research.

**Session Usage**
The roster dataset contains a total of 4,592 performed sessions with at least one surgery (which counts for a patient) since 2014. These 4,592 performed sessions consist of 86 unique sessions (sessions are repeated in an MSS), of which the most frequently used session is performed 610 times.
The surgery dataset contains 12,959 surgeries and it is important to know the number of surgeries performed in certain sessions, as it is assumed that every surgery results in a patient entering one of the hospital's clinical wards for rehabilitation. Although most of the surgeries are registered to a performed session, 1,828 surgeries lack a registered session and thus cannot be included in the session usage analysis. The remaining 11,131 surgeries are divided among the 86 unique sessions.
The possible number of surgeries performed in an individual session depends, among others, on the total duration of that session and the duration of the surgeries, therefore the average number of surgeries performed in a session is not a good statistic when comparing the 86 unique sessions. However, to be able to compare the sessions based on performed surgeries, the number of surgeries performed per hour is calculated for the 86 unique sessions and Figure 5.1 shows the distribution of this surgeries-per-hour calculation.

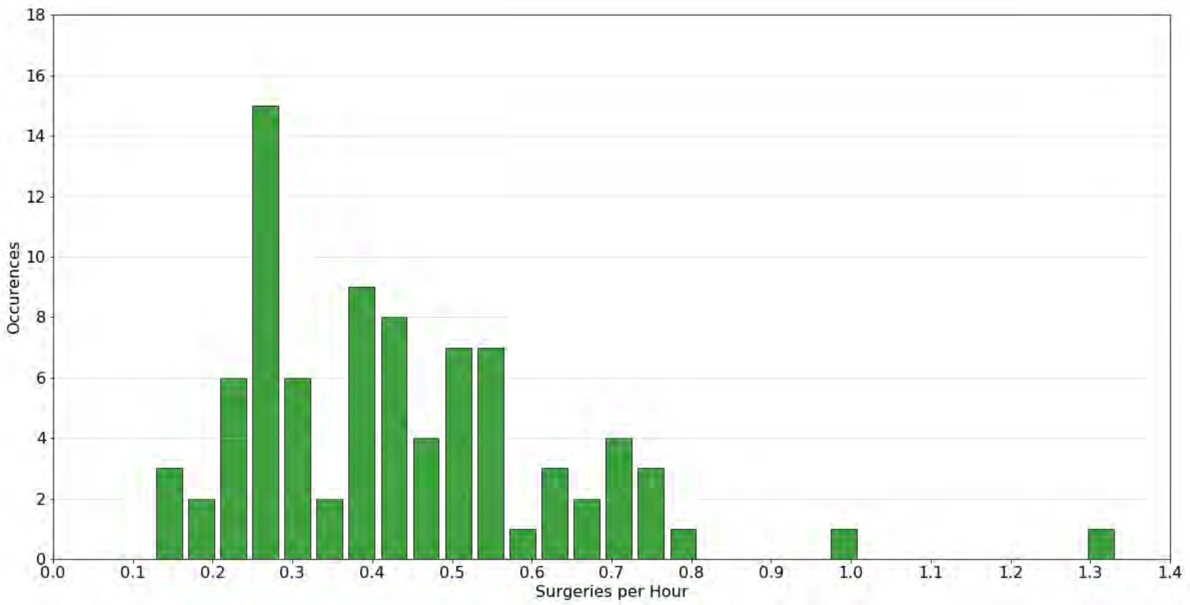



Figure 5.1: Histogram: average number of surgeries per hour among the 86 unique sessions

The histogram from Figure 5.1 counts the number of unique sessions having a certain average amount of surgeries per hour. It becomes clear that most sessions perform roughly between 0.15 and 0.8 surgeries per hour on average. When only analyzing sessions that have been performed for at least 20 times since 2014 (which accounts for $\approx 94\%$ of all performed sessions divided among 27 unique sessions) the average number of surgeries performed in the sessions lies between 0.15 and 0.7, without any exceptions.

**Length of Stay**

Besides the number of patients as a result of the sessions, the length of stay of the patients is also an important characteristic as it denotes the duration a clinical bed will be occupied by the patient. The length of stay for patients is calculated as described in section 5.1.1.

When analyzing all 12,959 surgeries and their corresponding admissions, the mean length of stay is 106 hours with a minimum of 0 and a maximum of 2,248 hours. The median of the length of stay is equal to 51, which is less than half of the mean length of stay and the length of stay is therefore expected to be a nonsymmetrical distribution with a positive skew.

The distribution of the length of stay is shown in the upper graph of the histogram shown in Figure 5.2. The graph at the bottom of Figure 5.2 shows the length of stay distribution for the surgeries with a length of stay below the 0.95-quantile. Both the upper and the lower histograms also show their corresponding means and medians.

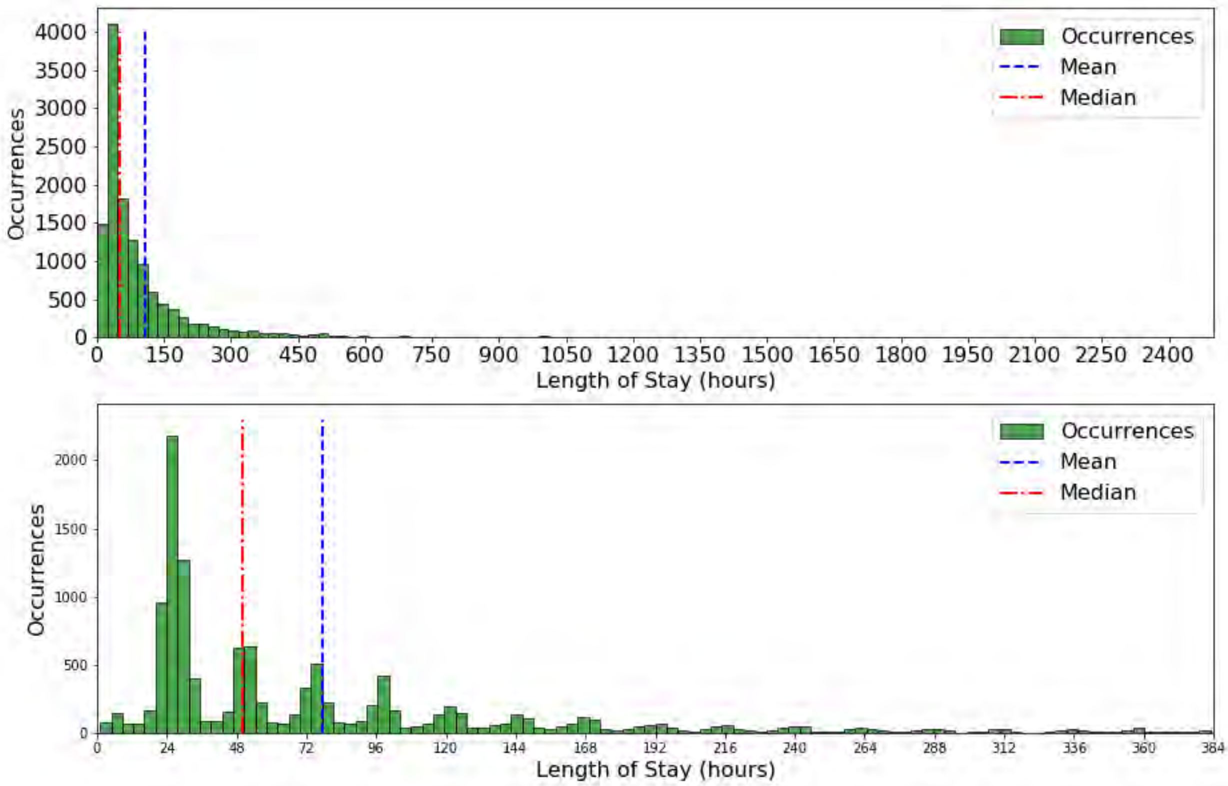



Figure 5.2: distribution of the length of stay for admissions corresponding to all sessions (upper) and admissions corresponding to sessions with at least 20 occurrences (lower)

The expectation of a positively skewed (and thus nonsymmetric) length of stay distribution is endorsed by Figure 5.2. It clearly shows a big difference in means and medians, as well as the right tails of the occurrences. Also, the bottom picture in figure 5.3 show a high number of occurrences around every 24 hours and a low amount of occurrences in between. A possible explanation for this is the limited availability (very few in nighttime) of surgeons to discharge a patient or the usage of "discharge-rounds" at certain times during the day. In these "discharge-rounds", surgeons walk through the clinical wards to check if patients are ready to be discharged from the hospital. These "discharge-rounds" are usually done at specific times during the working days on a daily repetitive basis.

**Length of Stay related to Sessions**

To allow modeling of patients entering the clinical wards as a result of performing certain sessions, length of stay characteristics are required on a session based level.

Between the 86 unique sessions performed since 2014, the mean of the length of stay ranges from 9 hours to 628 hours, whereas the median of the length of stay ranges between 9 hours and 521 hours. When, however, only taking the sessions into account that have been performed 20 times or more since 2014, the mean and median of the length of stay ranges from 27 to 259 and 23 to 153 hours respectively. The means and medians of the lengths of stay from patients as a result of these (20 times or more performed) sessions are shown in Figure 5.3.

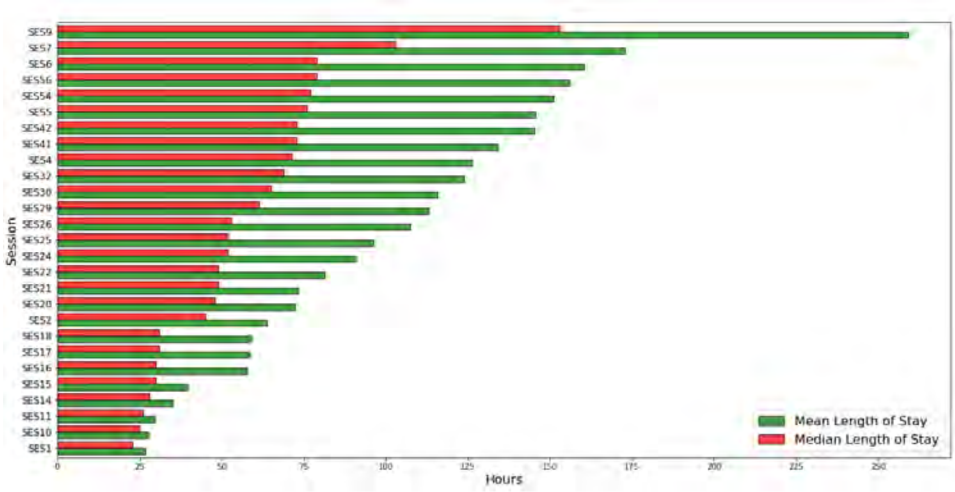



Figure 5.3: Means and Medians of lengths of stay for sessions performed at least 20 times since 2014

Figure 5.3 clearly shows a big difference between the mean and the median for the stated sessions. Therefore, when modeling the expected length of stay for patients as a result of performing a certain session, using its average (mean) length of stay or its median length of stay makes a significant difference in the prediction process and is sometimes considered wrong [17]. For this reason, the model used in this research uses the empirical cumulative distribution functions of the sessions as described in chapter 4.

**Length of Stay related to Priorities**

As discussed in sections 2.3.1 and 2.3.2, the priority of an admission (elective/emergency) might influence its pre- and postsurgical length of stay because of differences in the care process between the different priorities. For both the presurgical and the postsurgical length of stay an empirical cumulative distribution is extracted from the dataset for each of the priorities. Also, the distribution of the lengths of stay are shown for the admissions within the lower 0.95-quantile is calculated. These empirical cumulative distributions of the lengths of stay are shown in Figure 5.4.

Figure 5.4 clearly shows a big difference in the shapes of the distributions shown in the top left and bottom left graphs, which represent the presurgical length of stay distributions for the emergency and the elective patients. Although the top right and bottom right graphs also show some difference between the values in the distributions shown, their shapes are more or less equal. Therefore, the expected difference stated in sections 2.3.1 and 2.3.2 regarding the presurgical length of stay distribution between the emergency and the elective patients is emphasized by Figure 5.4.

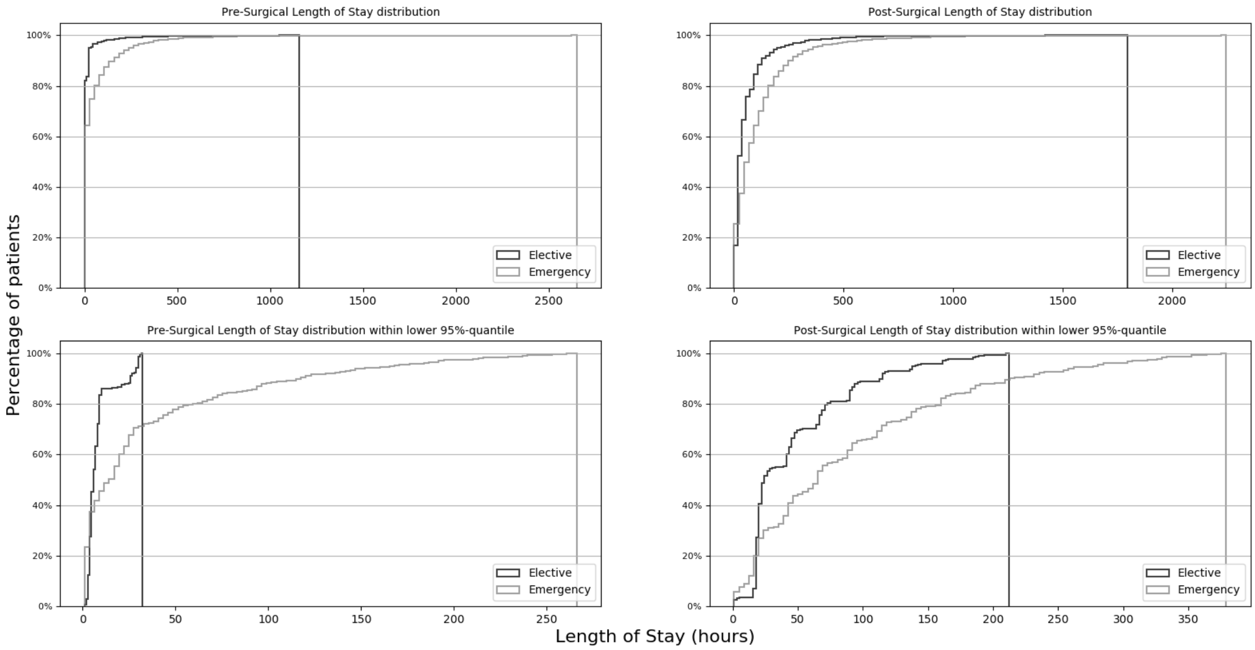



Figure 5.4: Presurgical Length of Stay distribution (cumulative) for: emergency/elective patients (top left) and emergency/elective patients in the lower 0.95-quantile (bottom left). Postsurgical Length of Stay distribution (cumulative) for: emergency/elective patients (top right) and emergency/elective patients in the lower 0.95-quantile (right)

## 5.2.2 Parameter Extraction

The model used in this research requires multiple input parameters related to the sessions (for elective patients) and time intervals (for emergency patients) to be able to calculate the expected number of occupied beds. This section describes the extraction of these parameters from the datasets.

**Empirical Distributions**

The model requires the empirical distribution of the length of stay for patients that enter the clinical wards as a result of performing a session. This empirical distribution is calculated for each of the unique sessions and Figure 5.5 shows an example of this empirical distribution of the lengths of stay for sessions 16 and 50. Also, since previous data analysis of the lengths of stay showed a heavy positively skewed distribution, the empirical distributions of the lengths of stay within the 0.95-quantile for both sessions are included in Figure 5.5 for comparison.

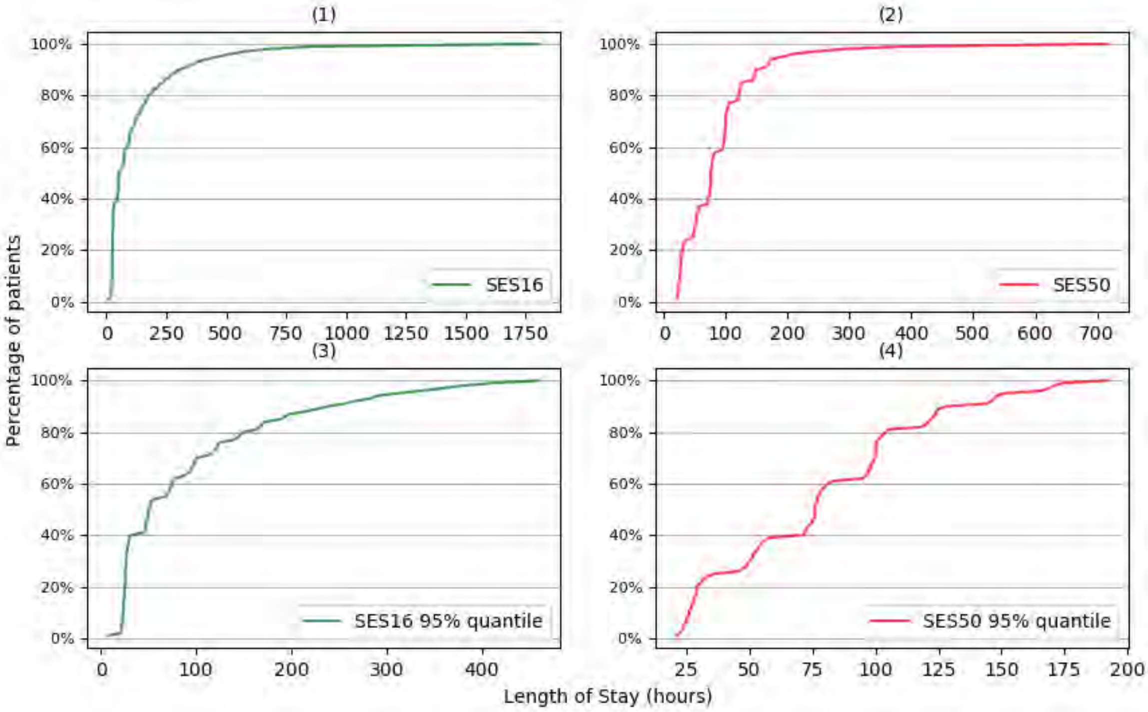



Figure 5.5: Empirical distribution of the length of stay for session 16 (1) and session 50 (2) and their corresponding empirical distributions of the length of stay within the 95%-quantile (3) and (4)

Figure 5.5 shows the lengths of stay distribution for the different sessions on the x-axis and the percentage of patients having a length of stay equal to or less than the x-axis value on the y-axis. The skewed characteristic of the length of stay is endorsed by graphs (1) and (2) from Figure 5.5 as both graphs clearly show a minority of the patients having a significantly longer length of stay, which means they also have a significant influence on the mean length of stay for that session.

Graphs (3) and (4) state the empirical length of stay for sessions 16 and 50 respectively when only surgeries with a length of stay within the 95%-quantile are taken into account. These graphs show a more equally distributed length of stay among the included surgeries.

**Emergency Arrivals**

To simulate the expected number of occupied beds as a result of emergency arrivals, the expected number of emergency arrivals in a certain time interval is required. The model used in this research uses a different expected number of emergency arrivals based on time intervals separated by weekday and the part of the day. The weekdays obviously range from Monday until Sunday whereas the part of the day is denoted:

$$PartofDay = \begin{cases} \text{"Morning"} & \text{if 06:00 <= Admission Time < 14:00} \\ \text{"Afternoon"} & \text{if 14:00 <= Admission Time < 22:00} \\ \text{"Night"} & \text{if 22:00 <= Admission Time < 06:00} \end{cases} \tag{5.1}$$

For each combination of the weekday and the part of the day, the average number of emergency arrivals is calculated based on the historical data. The result is shown in Table 5.3.

**Emergency Length of Stay**

Besides the number of emergency arrivals occurring in a certain time interval, the expected

| | | Part of Day | | |
|---|---|---|---|---|
| | | Morning | Afternoon | Night |
| **Weekday** | Monday | 1.29 | 1.71 | 0.45 |
| | Tuesday | 1.12 | 1.58 | 0.51 |
| | Wednesday | 1.09 | 1.43 | 0.42 |
| | Thursday | 0.95 | 1.27 | 0.45 |
| | Friday | 0.90 | 1.40 | 0.67 |
| | Saturday | 0.91 | 1.48 | 0.45 |
| | Sunday | 0.98 | 1.20 | 0.52 |

Table 5.3: Number of emergency arrivals per part of day per weekday

length of stay is also required to simulate the emergency occupation. The median length of stay for emergency patients is calculated for each time interval (comparable with the Emergency Arrivals) and can be found in table 5.4

| | | Part of Day | | |
|---|---|---|---|---|
| | | Morning | Afternoon | Night |
| **Weekday** | Monday | 67 | 70 | 62 |
| | Tuesday | 58 | 93 | 67 |
| | Wednesday | 71 | 65 | 81 |
| | Thursday | 43 | 84 | 111 |
| | Friday | 68 | 60 | 98 |
| | Saturday | 52 | 67 | 98 |
| | Sunday | 66 | 97 | 69 |

Table 5.4: Length of Stay (hours) for emergency arrivals per part of day per weekday

# Chapter 6

# Model Implementation

The goal of this research is not only to develop a model to analyze and predict bed occupancy as a result of elective and emergency patients receiving surgery in the hospital's ORs, but also address possibilities to implement such a predictive model in ChipSoft's software. The purpose of this chapter is twofold. Firstly, state the developed application that uses the research's model in a practical and easily applicable manner to obtain the simulation/prediction results. Secondly, describe the possibilities for ChipSoft to use the used predictive model and application to enrich their software package, which is done by describing the most important processes an implementation of this research would require.

## 6.1 Developed Application

Multiple actions performed in this research are programmed into an application using the Python programming language [19]. The application only requires the datasets as described in 5.1 as input and will the predict the number of occupied beds during and based on the period to predict. This subsection will describe the application's input, processes and outputs in general, as well as the assumptions or generalizations made to the model for practical reasons.

### 6.1.1 Input

Besides the datasets from section 5.1, the application requires some parameters to be set beforehand, namely:

- The moment of prediction ($t_p$) in datetime format, required as input for the model described in section 4.
- The starting moment of the predicted period ($t_s$) in datetime format, required as input for the model and to calculate the end moment of the predicted period ($t_e$).
- The length of the predicted period in days, will be added to $t_s$ to obtain $t_e$.
- The starting moment of the realization period ($t_{rs}$) in datetime format, determines the moment from which the realization (true measured values) will be calculated and compared to the predicted values.
- The start date for the data analysis ($t_{data}$) in datetime format, only data registered after this parameter will be analyzed in the data analysis process.

The input parameters listed above will result in a timeline as shown in Figure 6.1, on which the model is applied and simulation is performed.

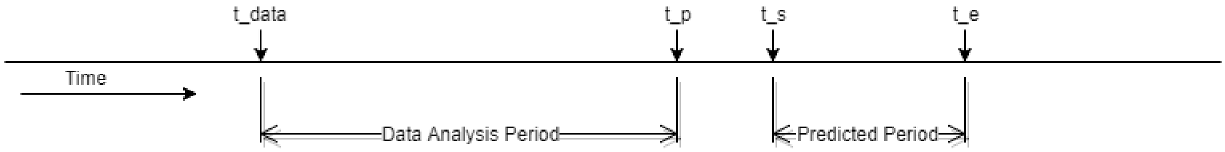



Figure 6.1: Prediction Timeline

The starting moment of the realization period ($t_{rs}$) is not shown in figure 6.1, as it is kept equal to the starting moment of the simulated period ($t_s$) throughout the rest of this research. Using the timeline shown in Figure 4.1, a series of processes is executed in the program in order to obtain a prediction of the number of occupied beds during the predicted period.

### 6.1.2 Processes

**Reading the MSS**

After the input parameters are set, the program starts with the first sub-process of the prediction process, namely reading and storing the Master Surgical Schedule used in the predicted period. To do so, the unique sessions used in the predicted period are extracted from the roster dataset (5.1.2) and stored in the program's memory, with their corresponding details such as session name, session date, session start time and the end time of the session.

An important note to make is that sessions can be slightly adjusted within the MSS on an incidental basis. For example, when a surgeon is only available until 12.00, whereas the original session in the MSS ends at 15.00, hospitals can adjust this planned session to the 12.00 end time without having to create a new session.
Since the (postsurgical) length of stay of elective patients is calculated based on the end times of their sessions and we assume the discharge moments during the day to be limited and not affected by the end times of the sessions, it is assumed that the end time of a session is of importance on the (postsurgical) length of stay of patients.

To take this possible postsurgical length of stay difference into account in the implementation, a generalization is made for the end time of the sessions: each session (unique by name) is stored twice, once with an end time classified as "Morning" (end time before 12.30) and once with end time classified as "Afternoon" (end time after 12.30), this way the distinction in end time is taken into account in the calculation, but small changes in the end time should not result in a different session stored. Further calculated characteristics are stored for both the "Morning" version and the "Afternoon" version of a unique session.

**Data Analysis for Session Characteristics**

Once all required sessions are stored in the program's memory, the sessions used in the predicted period need their corresponding characteristics to be calculated based on data analysis performed on the 'Data Analysis Period' part of the surgery dataset (section 5.1.1). Since the prediction will be made for the predicted period and the prediction is done at time $t_p$, it is obvious that data registered after $t_p$ is not allowed to be used in the data analysis, as it did not exist when the prediction was made at $t_p$. Also, the extraction of the parameters is only done based on the elective patients included in the surgery dataset, as it will only be used to predict the elective patients occupying a bed within the predicted period.
From this 'Data Analysis Period' subset of the surgery dataset, the empirical cumulative dis-

tribution for both the preoperative and the postoperative period is extracted for each session in twofold: once for the session with a "Morning" classified end time and once for the "Afternoon" classified end time. So, for each session, four different cumulative distributions are stored: preoperative length of stay for sessions with a "Morning" end time, post-operative length of stay for sessions with a "Morning" end time, preoperative length of stay for sessions with an "Afternoon" end time and postoperative length of stay for sessions with an "Afternoon" end time. Again, it is important that this extraction of the length of stay characteristics for the pre- and postoperative periods for a session is only based on the elective surgeries, because the length of stay distributions for both periods are significantly different between the elective and emergency patients (the preoperative periods in particular, see section 5.2.1).

Besides the lengths of stay distributions, the model described in chapter 4 also requires the number of operations in the sessions to be extracted from the dataset. This number of operations in a session is assumed to be equal to the number of patients entering the hospital to receive surgery in the session at some point. Obviously, since the length of a session also influences the number of surgeries performed, the number of surgeries distribution is calculated for each session for the "Morning" end time version as well as the "Afternoon" end time version.

**Data Analysis for Emergency Characteristics**

In order for the program to use the model and include emergency arrival simulation, the model requires the emergency parameters to be estimated based on historical data. The required emergency parameters are calculated and stored per unique time interval. In this research, the intervals are implemented as stated in section 5.2.2: three different parts of the day for each different weekday, resulting in 21 different time intervals. For each of the time intervals, the average number of arrived emergency patients and their corresponding median length of stay is calculated and stored.

**Prediction Process**

Once all the required parameters for the model described in chapter 4 have been extracted and stored in the program, formula 4.14 can be implemented and calculated by parts.
Instead of predicting a single moment $t_n$ as described in the model, the program predicts multiple $t_n$ values in one run. In fact, each hour between the start moment of the predicted period ($t_s$) and the end moment of the predicted period ($t_e$, obtained by adding the number of predicted days from the parameter input to $t_s$), is considered a value for $t_n$. For example, when the start moment of the predicted period is 01-10-2015 00:00 and the program is set to predict 2 days (48 hours), the end moment of the predicted period will be 03-10-2015 00:00 and $t_n$ will be set to each different hour between those dates.
Then, for each value of $t_n$, the value for formula 4.14 is calculated in multiple steps, which will be described shortly.

First, all the active admissions (elective and emergency) at $t_p$ are extracted from the surgery dataset and separated into a list of elective admissions and a list of emergency admissions. The predicted number of remaining elective patients present at $t_n$ is calculated by using formula 4.5, with $S_p$ being the collection of different sessions (different based on name and "Morning" or "Afternoon" end time) active elective admissions at $t_p$ result from and $A_{s,p}$ the collection of active elective patients at $t_p$ for session $s \in (1, 2, ..., S_p)$. The fraction from formula 4.5 can be calculated by using the cumulative distribution functions stored in the program for each different session $s$. The obtained value for formula 4.5 is stored in the program. The remaining emergency admissions based at $t_n$, given the obtained list of active emergency admissions at $t_p$, is calculated using formula 4.6. The value used for $I_p$ is the collection of different time intervals

the emergency arrivals originate from and $E_{i,p}$ is the collection of active admissions at $t_p$ for time interval $i$.
Once both the values for formula 4.5 and 4.6 have been obtained, they are stored as the number of remaining present patients at the predicted time $t_n$.

Next, the value for formulas 4.11 and 4.12 are calculated using the stored characteristics for the sessions to be executed between $t_p$ and $t_n$ ($S^{sk}_{n-}$) for 4.11 and the stored characteristics for the sessions to be executed between $t_n$ and $t_e$ ($S^{sk}_{n+}$) for formula 4.12. Both the summation over the possible number of surgeries for the different sessions and the parameters for the binomial distribution part can be deducted from the stored values obtained in the data analysis part of the program. The values for 4.11 and 4.12 together form the predicted number of elective patients present at $t_n$.

Finally, the predicted number of emergency patients present at $t_n$ is calculated using formula 4.13. With $I_{n-}$ being the different time intervals occurring between $t_p$ and $t_n$ and $\lambda(i)$ being the expected number of arriving emergency patients for interval $i$, as stored in the program as a result of the data analysis steps.

The combined values of the three steps above represent the value for formula 4.14 for the predicted moment $t_n$. Once these steps have been performed for each $t_n$, a prediction of the number of occupied beds is obtained for every hour in the predicted period, $X(t_n), t_n \in (t_s, t_{s+1}, ..., t_e)$.

**Simulation Process**
Instead of only calculation the predicted value $X(t_n)$, the program simulates confidence intervals around each value for $X(t_n)$ within the predicted period based on formulas 4.5, 4.11, 4.12 and 4.13. This section will describe the simulation performed for each of the formulas.
The simulation for the remaining elective patients (formula 4.5) is performed as:

1. Each patient in $A_{s,p}$ is still present at $t_n$ with a probability $p$ as deducted from the stored session characteristics.

2. Simulate 10,000 random samples for each patient $a$ in $A_{s,p}$, with probability $p_a$ to be present, resulting in $(\hat{a}_1, ..., \hat{a}_{10000})$ values for patient $a$ (with $\hat{a}_i$ is 1 with probability $p_a$ and 0 with probability $1 - p_a$).

3. The expected simulated number of present patients based on $A_{s,p}$ is then given by vector $\vec{a}$: $\sum_{a=1}^{A_{s,p}} a_i$ for $i \in (1, 2, ..., 10000)$. The obtained vector $\vec{a}$ is stored in the program.

To simulate prediction values for the collection of sessions $S^{sk}_{n-}$ (formula 4.11), the following steps are followed:

1. Draw 100 samples from the number of surgeries distribution for session $s \in S^{sk}_{n-}$ and store as $(x_1, x_2, ..., x_{100})$.

2. For each $x_i \in (x_1, x_2, ..., x_{100})$, draw 100 samples from the binomial distribution, using the parameters $n = x_i$ and $p = 1 - P(L^{post}_s < t_n - t_{se})$, the probability for a patient resulting from session $s$ to have a postsurgical length of stay longer than $t_n - t_{se}$ (with $t_{se}$ being the end time for session $s$). This probability can be extracted from the cumulative distribution functions obtained in the data analysis part of the program. The results of this second simulation are stored for each $x_i$, to obtain $x_{i,j} \in (x_{1,1}, x_{1,2}, ..., x_{1,100}, x_{2,1}, x_{2,2}, ..., x_{2,100}, ..., x_{100,100})$. So, a collection of 10,000 simulations is obtained for each session $s$ and stored as vector $\vec{x}_s$.

3. The combined simulation for the elective postsurgical patients present at $t_n$ is represented by the vector $\vec{x} : \sum_{s=1}^{S_{n-}^{sk}} \vec{x}_s$ for each $i, j \in (1, 2, ..., 100)$. Meaning the value in vector $\vec{x}$ on coordinate (i,j) is the sum of the corresponding values on the same coordinate in the individual session vectors $\vec{x}_s$.

The same method of simulation is performed for the collection of sessions $S_{n+}^{sk}$, with small adjustments to calculate the elective presurgical instead of the postsurgical patients present at $t_n$. So, for the collection of sessions $S_{n+}^{sk}$:

1. Draw 100 samples from the number of surgeries distribution for session $s \in S_{n+}^{sk}$ and store as $(y_1, y_2, ..., y_{100})$.

2. For each $y_i \in (y_1, y_2, ..., y_{100})$, draw 100 samples from the binomial distribution, using the parameters $n = y_i$ and $p = 1 - P(L_s^{pre} < t_n - t_{se})$. This probability can also be extracted from the cumulative distribution functions obtained in the data analysis part of the program. The results of this second simulation are stored for each $y_i$, to obtain $y_{i,j} \in (y_{1,1}, y_{1,2}, ..., y_{1,100}, y_{2,1}, y_{2,2}, ..., y_{2,100}, ..., y_{100,100})$. So, a collection of 10,000 simulations is obtained for each session $s$ and stored as $\vec{y}_s$.

3. The combined simulation for the elective pre-surgical patients present at $t_n$ is represented by the vector $\vec{y} : \sum_{s=1}^{S_{n+}^{sk}} \vec{y}_s$ for each $i, j \in (1, 2, ..., 100)$.

A different method is used to simulate the predicted number of emergency patients present at $t_p$. Namely, for the collection of intervals between $t_p$ and $t_n$, $I_{n-}$:

1. Since a deterministic length of stay is used for the different intervals $i$, the value of $\lambda_n^I = \sum_{i=0}^{I_{n-}} \lambda(i) \mathbb{1}_{\{L_i >= t_n - t_i\}}$ (with $t_i$ being the start moment of interval $i$), can be used as a Poisson input parameter for simulation at time $t_n$. The parameter $\lambda(i)$ can be extracted from the program as it was calculated in the data analysis part, the same holds for the length of stay for interval $i$, $L_i$.

2. Use $\lambda_n^I$ to simulate 10,000 samples of a random Poisson distribution with parameter $\lambda_n^I$. These results are saved in a vector with 10,000 values $\vec{z} = (z_{1,1}, z_{1,2}, ..., z_{100,100})$

After obtaining the four simulation results $\vec{a}, \vec{x}, \vec{y}$ and $\vec{z}$ for the predicted moment $t_n$, they are combined to obtain 10,000 simulated values for $X(t_n)$: $\vec{X}(t_n) = a_{i,j} + x_{i,j} + y_{i,j} + z_{i,j}$ for $i, j \in (1, 2, ..., 100)$. The values in $\vec{X}(t_n)$ are sorted and percentiles can be extracted to act as the confidence lower or upper bounds, with a default extraction of the 0.05-quantile as the lower interval bound and the 0.95-quantile as the upper interval bound.

### 6.1.3 Output

Once all the predicted and simulated values are obtained, the prediction is scored versus the real measured values, which can be obtained from the surgery dataset. To score the prediction done for the predicted period the precision statistic is used, which is defined as the fraction of real measured values being smaller than the upper quantile extracted from the simulation and higher than the lower quantile extracted. For example, when predicting 48 different values (hours) within the predicted period and the real measured occupation value lies within the lower and upper bound 40 times, the precision would be $40/48 \approx 0.83$.
Besides the scoring result of the prediction, the program also visualizes the prediction made versus the actual measured values for the predicted period. An example of this output is shown in Figure 6.2.

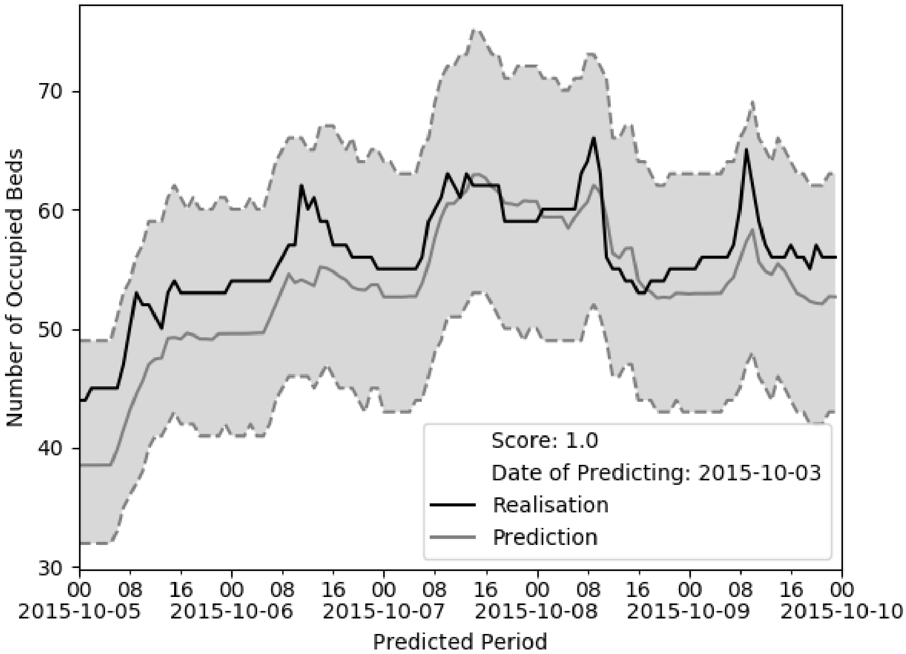



Figure 6.2: Example of Result Visualization

The horizontal axis of Figure 6.2 states the predicted moment ($t_n$) within the predicted period with time labels every 8 hours, whereas the vertical axis states the number of occupied beds for the predicted moment. The grey line represents the predicted value by formula 4.14 and the grey dotted lines represent the upper and lower quantiles extracted from the simulated values ($\vec{X}(t_n)$) as explained in the previous section.
The figure also shows the program's input parameters in the legend, which in this example figure were set to: the date of predicting ($t_p$)=2015-10-03, the predicted period ($t_s$ to $t_s$) = 05-10-2015 00:00 until 10-10-2015 00:00, meaning the prediction was done 2 days in advance of the start of the predicted period. The precision-score of this example is shown below the date of predicting; 1.0, meaning all realized values for $t_n$ lie within the upper and lower boundaries of the predicted confidence interval.

# Chapter 7

# Results

This chapter describes the results obtained by running the program for different parameter settings as well as the results obtained when using the program to analyze "what-if" scenarios. For each analysis, the results of the program are presented visually and by using the precision scoring function, both described in section 6. Since the datasets were extracted from two different hospitals, the results are visually shown for hospital A only, whereas the results of the precision function are shown for both hospitals A and B.
The first part of this chapter will show the program's performance (results) based on different parameter settings, the second part describes the results for different "what-if" scenarios.

## 7.1 Program Results

This section describes multiple program outcomes for different parameter settings. Among these parameter settings, the moment of predicting, the predicted period and the different patient sources included are varied and their results are stated.

### 7.1.1 Varying $t_p$

When varying the moment of predicting, $t_p$, while maintaining the same predicted period, differences might occur in the prediction due to the time difference between $t_p$ and the predicted period. This section describes the results obtained while varying the moment of predicting, while the other input parameters as described in section 6.1.1 were given the values as stated in table 7.1. In the table, the column 'Type' denotes the parameter type ('static' means the parameter did not change during the simulations and 'variable' means different results were gathered by varying the parameter's value).

| **Description** | **Parameter** | **Value** | **Type** |
|---|---|---|---|
| Moment of prediction | $t_p$ | *N/A* | Variable |
| Start moment of predicted period | $t_s$ | 2015-10-5 00:00 | Static |
| End moment of predicted period | $t_e$ | 2015-10-10 00:00 | Static |
| Data analysis period start date | $t_{data}$ | 2014-01-01 00:00 | Static |
| Simulation Lower Quantile | $\alpha$ | *N/A* | Variable |
| Simulation Upper Quantile | $\beta$ | *N/A* | Variable |

Table 7.1: Parameters used in simulation

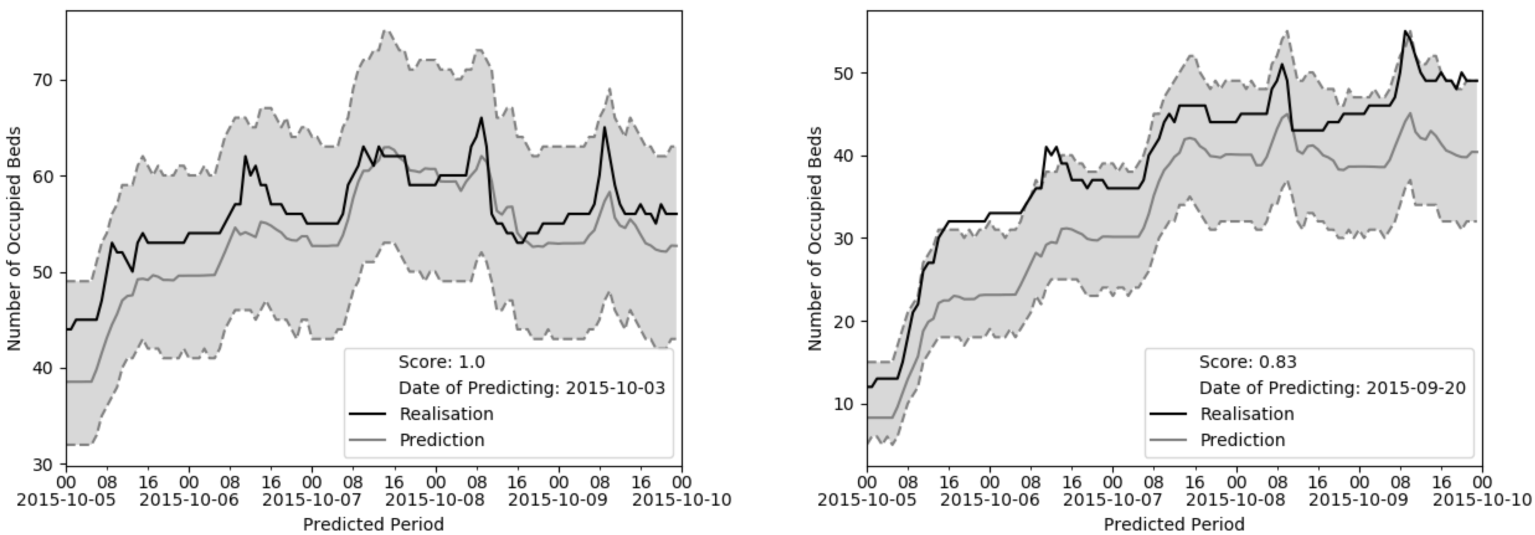
A simulation was run with the parameter setup as stated in table 7.1 and by varying the parameters with type 'Variable'. The results of the simulation for different $t_p$ and $[\alpha;\beta]$ values are shown for hospital A and B (B between parenthesis) in table 7.2.

| | | $[\alpha; \beta]$ | |
|---|---|---|---|
| | $[0.050; 0.950]$ | $[0.075; 0.925]$ | $[0.100; 0.900]$ |
| $t_p$ 2015-10-05 00:00 | 1.00 (1.00) | 1.00 (1.00) | 0.99 (0.97) |
| 2015-10-03 00:00 | 1.00 (0.93) | 1.00 (0.87) | 1.00 (0.86) |
| 2015-10-01 00:00 | 1.00 (0.88) | 1.00 (0.76) | 1.00 (0.64) |
| 2015-09-29 00:00 | 1.00 (1.00) | 0.98 (0.97) | 0.96 (0.86) |
| 2015-09-27 00:00 | 1.00 (1.00) | 0.97 (1.00) | 0.97 (1.00) |
| 2015-09-25 00:00 | 0.99 (1.00) | 0.97 (1.00) | 0.91 (0.94) |
| 2015-09-20 00:00 | 0.83 (0.88) | 0.76 (0.79) | 0.63 (0.76) |

Table 7.2: Precision score based on parameters from table 7.1 for hospital A and B (B between parenthesis)

The scores in table 7.2 for hospital A decrease for all $[\alpha; \beta]$ when the time difference between the moment of prediction, $t_p$, and the start of the predicted period, $t_s$, increases. For hospital B, the decreasing of the scores because of increasing the time difference between $t_p$ and $t_s$ is much less evident, if existing at all.

As stated in the beginning of this chapter, the results of the simulations are also obtained, as described in section 6.1.3. For two different parameters $t_p$, while having $[\alpha; \beta] = [0.050; 0.950]$, these graphs are shown in Figure 7.1. The visual results of the prediction done for hospital B for the same two parameter settings are shown in Appendix B.1.




(a) result for $t_p$=2015-10-03

(b) result for $t_p$=2015-09-20

Figure 7.1: Prediction results for hospital A based on parameter input as stated in table 7.1 and different values for $t_p$.

Figure (a) from Figure 7.1 state a precision score of 1.00 with a prediction performed 2 days before the start of the predicted period. Figure (b) states a precision score of 0.83 with a prediction performed 15 days before the start of the predicted period. A difference is observed between the number of occupied beds at $t_s$ between figures (a) an (b), whereas both predictions predict a value below the realized number of occupied beds at the start and the end of the predicted period.

### 7.1.2 Varying the Predicted Period

Instead of simulating different results for the same predicted period as done in previous section, this section states the program's performance for different predicted periods, as this might result in different prediction scores and visual results.
This simulation, the input parameters were set to the values as shown in table 7.3.

| Description | Parameter | Value | Type |
|---|---|---|---|
| Moment of prediction | $t_p$ | 4 days before $t_s$ | Variable |
| Start moment of predicted period | $t_s$ | $N/A$ | Variable |
| End moment of predicted period | $t_e$ | 5 days after $t_s$ | Variable |
| Data analysis period start date | $t_{data}$ | 2014-01-01 00:00 | Static |
| Simulation Lower Quantile | $\alpha$ | 0.05 | Static |
| Simulation Upper Quantile | $\beta$ | 0.95 | Static |

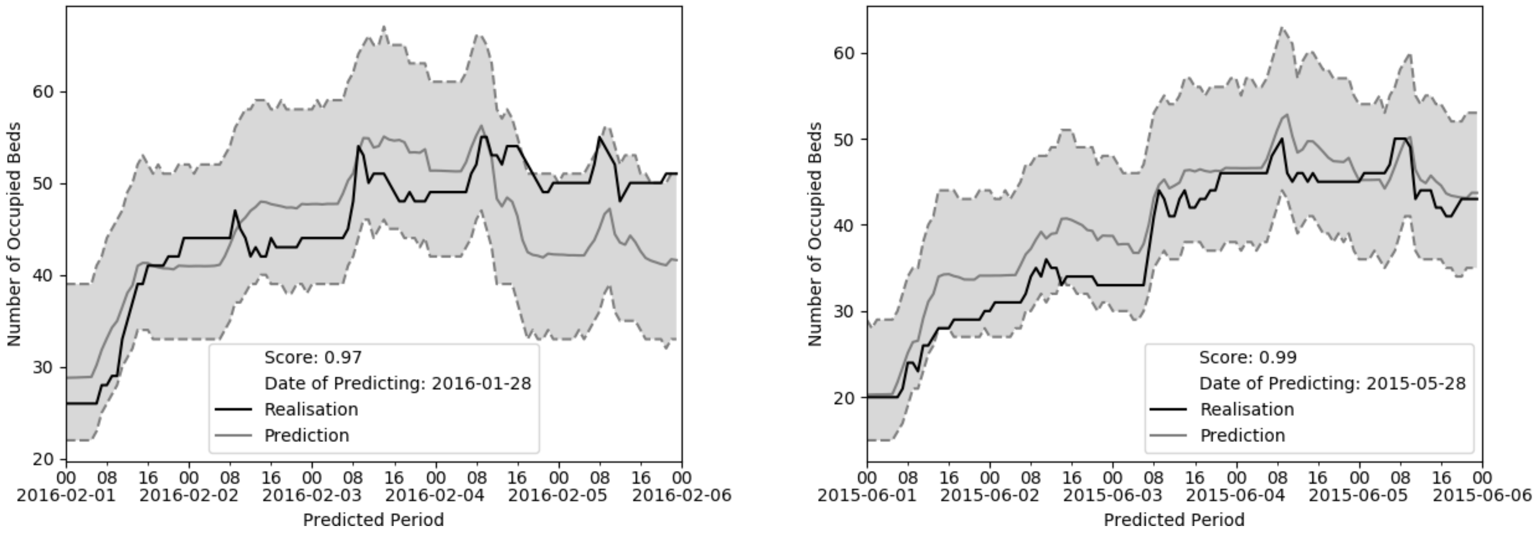
Table 7.3: Parameters used in simulation

The scores as a result of the simulation based on the input parameters from table 7.3 are shown in table 7.4 for hospital A, the scores of hospital B are shown between the parenthesis.

| | | Score |
|---|---|---|
| $t_s$ | 2016-10-03 00:00 | 0.79 (0.97) |
| | 2016-02-01 00:00 | 0.97 (0.95) |
| | 2015-11-23 00:00 | 0.94 (0.84) |
| | 2015-10-21 00:00 | 0.84 (0.51) |
| | 2015-08-10 00:00 | 0.97 (1.00) |
| | 2015-06-01 00:00 | 0.99 (0.98) |
| | 2015-03-23 00:00 | 0.71 (0.98) |

Table 7.4: Precision Scores for Hospital A (and B between parenthesis) for different predicted periods and using the parameter as stated in table 7.3

The precision scores for hospital A, as shown in table 7.4, range between 0.71 and 0.99 for the different predicted periods. For hospital B, the scores range from 0.51 and 1.00.
The visual results for hospital A are shown in Figure 7.2 for hospital A, whereas the visual results for hospital B for the same predicted period is shown in Appendix B.2. Figure 7.2 states a precision score of 0.97 and 0.99 for the predicted periods in graphs (a) and (b) respectively.




(a) result for $t_s$=2016-02-01 (b) result for $t_p$=2015-06-01

Figure 7.2: Prediction results for hospital A based on parameter input as stated in 7.3 and two different predicted periods $t_s$.

### 7.1.3 Varying Patient Sources

As described in section 4, the prediction method used in this research uses three different patient sources. Since the prediction formula used for each of the patient sources is different, the program's performance is expected to differ based on the sources included in the prediction. This section presents the obtained simulation results while varying the patient sources and the moment of prediction, $t_p$. The different patient sources are described in section 4, whereas the static input parameters that are not changed during the simulation are the same as stated in table 7.1, with values 0.05 for $\alpha$ and 0.95 for $\beta$.

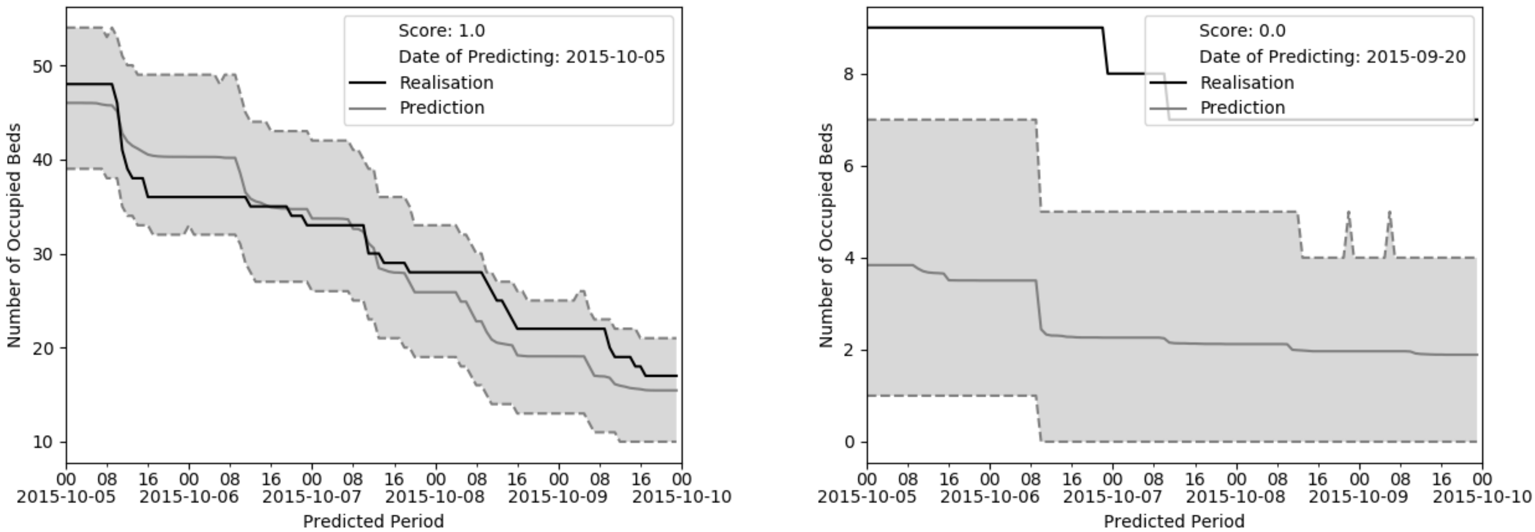
#### Results for Patient Source 1

When only taking patient source 1 (the remaining patients) into account, the predicted values for $t_n$ in the predicted period are only based on formula 4.7. The results for patient source 1 are shown in table 7.5 for different values of $t_p$ for hospital A and hospital B (between the parenthesis).

| | | **Score** |
|---|---|---|
| $t_p$ | 2015-10-05 00:00 | 1.00 (1.00) |
| | 2015-10-03 00:00 | 1.00 (1.00) |
| | 2015-10-01 00:00 | 1.00 (1.00) |
| | 2015-09-29 00:00 | 0.99 (1.00) |
| | 2015-09-27 00:00 | 1.00 (1.00) |
| | 2015-09-25 00:00 | 1.00 (1.00) |
| | 2015-09-20 00:00 | 0.00 (0.55) |

Table 7.5: Results for patient source 1 while varying $t_p$ for hospital A and B (B between the parenthesis)

Table 7.5 shows that the precision scores for patient source 1 range between 0.00 and 1.00 for hospital A and between 0.55 and 1.00 for hospital B. Both hospitals having their lowest score occurring when the prediction was done 15 days before the start of the predicted period ($t_s$). The visual results for two different values of $t_p$ are also shown graphically in Figure 7.3 for hospital A, the visual results for hospital B for the same input parameters are shown in Appendix B.3.



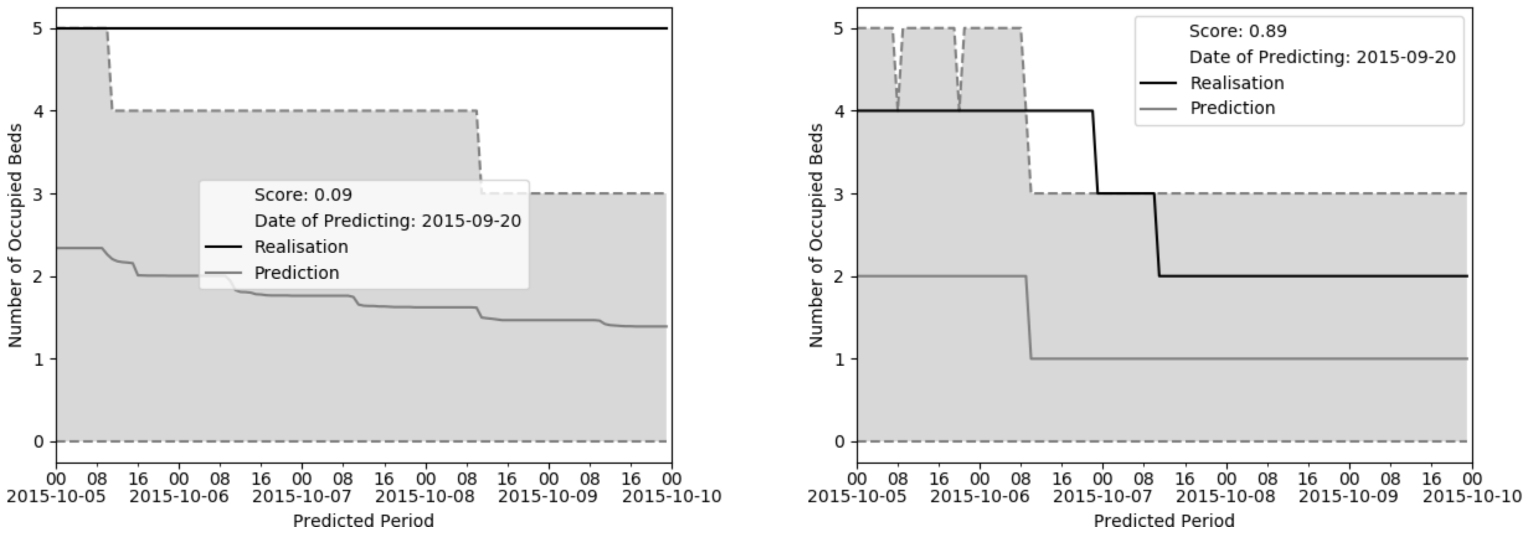

(a) result for $t_p$=2015-10-03

(b) result for $t_p$=2015-09-20

Figure 7.3: Prediction results based on patient source 1 and the parameter input as stated in 7.1 for hospital A, using different values for $t_p$.

Both graphs in Figure 7.3 show a decreasing number of occupied beds over time, with a score of 1.00 for graph (a) and a score of 0.00 for graph (b). The obtained confidence interval bounds in graph (a) at $t_s$ are $\approx 39$ for the lower bound and $\approx 55$ for the upper bound, whereas graph (b) states the values $\approx 3$ and $\approx 7$ for the lower and upper bounds at $t_s$ respectively.

Since patient source 1 consists of two individual calculations, one for the elective priority and one for the emergency priority, the results are also calculated for each priority individually. The visual result for the individual simulation is shown using the same input parameters and $t_p$=2015-09-20 in Figure 7.4 for hospital A, the figure obtained for hospital B using the same input is shown in Appendix B.4.




(a) elective priority only

(b) emergency priority only

Figure 7.4: Prediction results based on patient source 1 and parameter input as stated in 7.1 for hospital A.

Graph (a) from Figure 7.4 shows a constant non-changing line of 5 realized occupied number of beds as a result of the session remainders in the predicted period. The predicted value and its corresponding confidence interval decrease in value over time, resulting in a score of 0.09. Graph (b) represents the number of predicted/realized occupied beds based on the remaining emergency patients. Both the realized and the predicted values decrease over time in the prediction period, resulting in a score of 0.89.

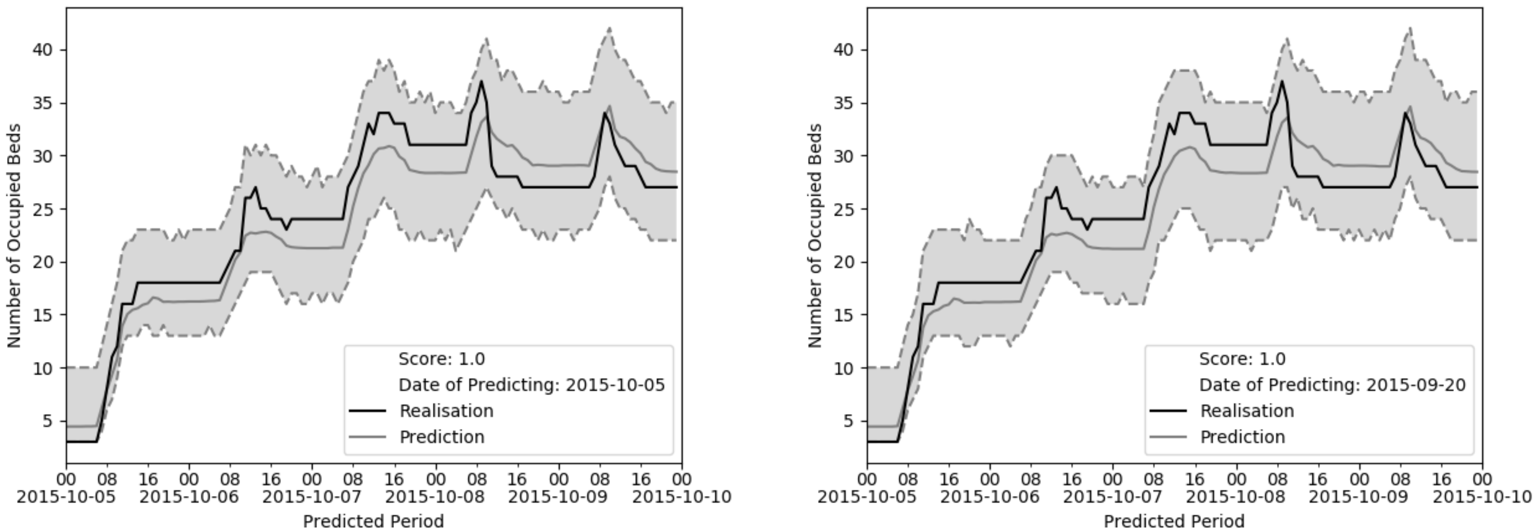
### Results for Patient Source 2

When only considering the elective patients that arrive in the future because of a MSS session, the prediction is only based upon formulas 4.11 and 4.12. The parameter settings used in this simulation is stated in table 7.1, using 0.05 and 0.95 for $\alpha$ and $\beta$ respectively. The precision score of these simulations are shown in table 7.6 for hospital A and B (B between the parenthesis).

| | | **Score** |
|---|---|---|
| $t_p$ | 2015-10-05 00:00 | 1.00 (0.83) |
| | 2015-10-03 00:00 | 1.00 (0.83) |
| | 2015-10-01 00:00 | 1.00 (0.66) |
| | 2015-09-29 00:00 | 1.00 (0.68) |
| | 2015-09-27 00:00 | 1.00 (0.68) |
| | 2015-09-25 00:00 | 1.00 (0.86) |
| | 2015-09-20 00:00 | 1.00 (0.86) |

Table 7.6: Results for patient source 1 only and varying $t_p$

Table 7.6 shows that the precision scores for hospital A are 1.00 for all the different values of $t_p$, whereas the scores for hospital B range between 0.66 and 0.86. The results are also shown graphically, using the same two values for $t_p$ as used in Figure 7.3, in Figure 7.5 for hospital A and in Appendix B.5 for hospital B.




(a) result for elective priority only

(b) result for emergency priority only

Figure 7.5: Prediction results based on parameter input as stated in 7.1 and different values for $t_p$.

Both graphs (a) and (b) from Figure 7.5 show the same values for the predicted and realized number of occupied beds for each $t_n$ in the predicted period. The simulated upper and lower bounds of the confidence interval does differ between graphs (a) and (b), however. Graph (a) as well as graph (b) obtain a precision score of 1.00.

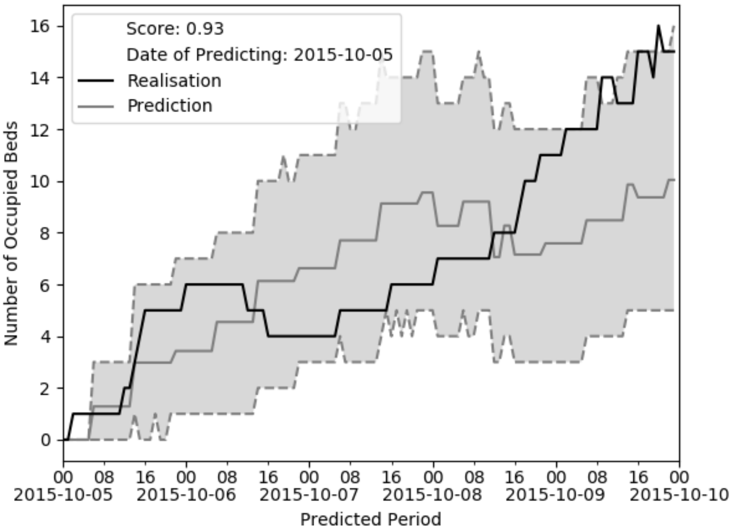
### Results for Patient Source 3

Patient source 3 only consists of emergency patients arriving in the future. The predicted values in this simulation are obtained based on formula 4.13 and parameter settings as stated in table 7.1, using $\alpha = 0.05$ and $\beta = 0.95$. The results for the simulations using variable $t_p$ is shown in table 7.7 for hospital A and B (B between the parenthesis).

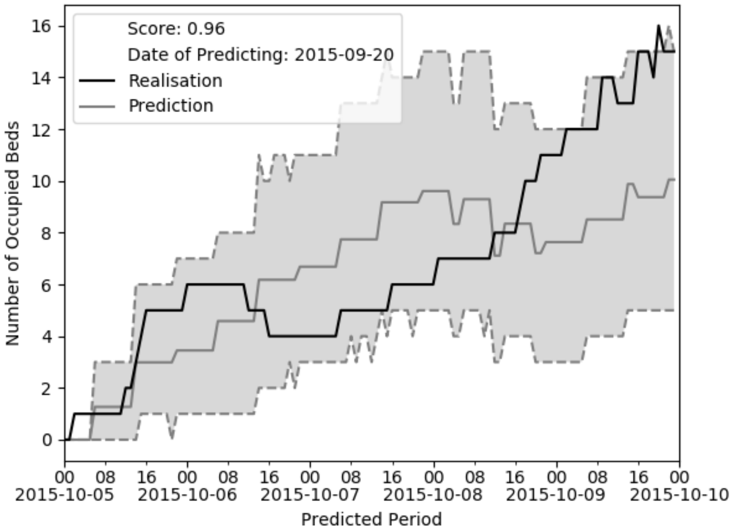
| | | **Score** |
|---|---|---|
| $t_p$ | 2015-10-05 00:00 | 0.93 (0.91) |
| | 2015-10-03 00:00 | 0.94 (0.91) |
| | 2015-10-01 00:00 | 0.94 (0.91) |
| | 2015-09-29 00:00 | 0.96 (0.91) |
| | 2015-09-27 00:00 | 0.94 (0.91) |
| | 2015-09-25 00:00 | 0.95 (0.91) |
| | 2015-09-20 00:00 | 0.96 (0.91) |

Table 7.7: Results for patient source 3 and varying $t_p$ for hospital A and B (B between the parenthesis)

For hospital A, the scores in table 7.7 range between 0.93 and 0.96, whereas the scores for hospital B are 0.91 for all the $t_p$ values. The results are also shown graphically, using the same two values for $t_p$ as used in Figure 7.5, in Figure 7.6 for hospital A and in Appendix B.6 for hospital B.




(a) result for $t_p$=2015-10-05 00:00




(b) result for $t_p$=2015-09-20 00:00

Figure 7.6: Prediction results based on parameter input as stated in 7.1 and different values for $t_p$.

The precision scores for graphs (a) and (b) from Figure 7.6 are 0.93 and 0.96 respectively. Both graphs show a prediction without clear confidence interval bounds at the start of the predicted period ($t_s$).

## 7.2 Result of what-if scenarios

An important goal of this research internship is to obtain a model and its implementation, to allow analysis of scenarios regarding different Master Surgical Schedules and session/patient characteristics. This section shows results of the program's usage in the following example scenarios:

- Swapping two days in the MSS
- Opening an extra OR
- Shortening the length of stay

For each of the scenarios listed above, a simulation is run based on the data of hospital A to obtain the results of the prediction before and after implementing the scenario. This results in a figure of the simulated week, containing both the situations as before and after the implementation of the scenario.

### 7.2.1 Swapping two days in the MSS

This section describes the result of the program when analyzing a difference in the Master Surgical Schedule used. Namely, it shows the result of the program obtained when swapping the sessions of the MSS on Monday and Tuesday in one week of the MSS. This scenario is run using the parameter setup as stated in table 7.8.

Within the simulated period, both the original MSS and the adjusted MSS by swapping the

| Description | Parameter | Value | Type |
|---|---|---|---|
| Moment of prediction | $t_p$ | 2016-03-07 00:00 | Static |
| Start moment of predicted period | $t_s$ | 2016-03-07 00:00 | Static |
| End moment of predicted period | $t_e$ | 2016-03-12 00:00 | Static |
| Data analysis period start date | $t_{data}$ | 2014-01-01 00:00 | Static |
| Simulation Lower Quantile | $\alpha$ | 0.05 | Static |
| Simulation Upper Quantile | $\beta$ | 0.95 | Static |

Table 7.8: Results for patient source 1 only and varying $t_p$

sessions on Monday and Tuesday are shown in table 7.9. The four sessions executed in the original roster on Monday are executed on Tuesday, whereas the nine sessions executed on Tuesday in the original roster are swapped to the Monday in the adjusted MSS.

| Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|
| SES16 | SES17 | SES16 | SES50 | SES8 |
| SES23 | SES66 | SES23 | SES23 | SES7 |
| SES19 | SES75 | SES76 | SES36 | SES16 |
| SES50 | SES44 | SES39 | SES19 | SES19 |
| | SES54 | SES50 | SES16 | SES24 |
| | SES32 | SES9 | SES8 | SES50 |
| | SES16 | | | SES79 |
| | SES7 | | | SES50 |
| | SES8 | | | |

$\longrightarrow$

| Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|
| SES17 | SES16 | SES16 | SES50 | SES8 |
| SES66 | SES23 | SES23 | SES23 | SES7 |
| SES75 | SES19 | SES76 | SES36 | SES16 |
| SES44 | SES50 | SES39 | SES19 | SES19 |
| SES54 | | SES50 | SES16 | SES24 |
| SES32 | | SES9 | SES8 | SES50 |
| SES16 | | | | SES79 |
| SES7 | | | | SES50 |
| SES8 | | | | |

Table 7.9: Original MSS (left) and the adjusted MSS (right) by swapping the sessions on Monday and Tuesday

Based on the MSS change stated in table 7.9, the results are obtained for the original MSS as well as the adjusted MSS and shown in one figure.

Figure 7.7 shows the result obtained for both schedules, using the parameter setting as stated in table 7.8.

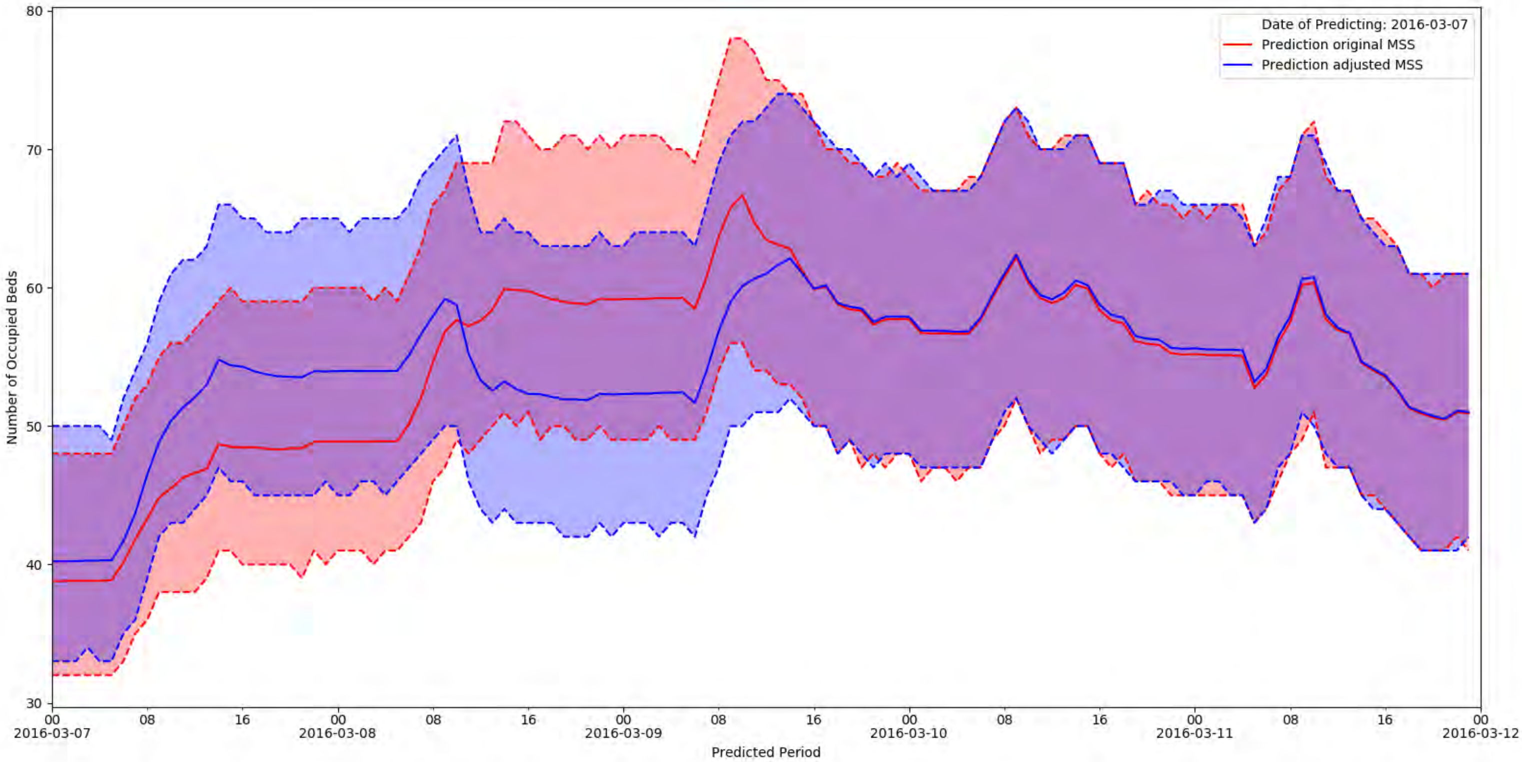



Figure 7.7: Result for the original as well as the adjusted MSS, having the Monday and Tuesday sessions swapped

Figure 7.7 shows the predicted number of occupied beds as well as the upper and lower confidence interval bounds for the original and the adjusted MSS. Using the adjusted MSS, the program predicts a higher number of occupied beds on Monday and Tuesday morning, whereas the original MSS predicts more occupied beds on Tuesday and Wednesday morning. The difference between the predicted values is much less significant after 2016-03-09 00:00.

### 7.2.2 Opening an extra OR

The second scenario simulated is the scenario in which an extra OR is opened and used in the MSS. Due to opening an extra OR, it is possible to schedule more surgeries, which obviously is expected to result in more patients entering the clinical wards for rehabilitation. The parameters used to run this simulation are stated in table 7.10.

| **Description** | **Parameter** | **Value** | **Type** |
|---|---|---|---|
| Moment of prediction | $t_p$ | 2016-04-11 00:00 | Static |
| Start moment of predicted period | $t_s$ | 2016-04-11 00:00 | Static |
| End moment of predicted period | $t_e$ | 2016-04-11 00:00 | Static |
| Data analysis period start date | $t_{data}$ | 2014-01-01 00:00 | Static |
| Simulation Lower Quantile | $\alpha$ | 0.05 | Static |
| Simulation Upper Quantile | $\beta$ | 0.95 | Static |

Table 7.10: Results for patient source 1 only and varying $t_p$

In this example scenario, the extra OR used in the Master Surgical Schedule is assigned an extra session to be performed each day. Both the original and the adjusted MSS are shown in table 7.11, in which the extra sessions to be performed in the adjusted MSS have a grey background color.

| **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** |
|---|---|---|---|---|
| SES16 | SES17 | SES16 | SES36 | SES7 |
| SES19 | SES66 | SES23 | SES23 | SES56 |
| SES23 | SES32 | SES76 | SES53 | SES16 |
| SES36 | SES44 | SES39 | SES19 | SES19 |
| SES47 | SES16 | SES50 | SES16 | SES24 |
| SES50 | SES75 | | SES9 | SES50 |
| SES7 | | | SES7 | SES79 |
| SES8 | | | SES2 | |
| | | | SES54 | |

$\longrightarrow$

| **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** |
|---|---|---|---|---|
| SES16 | SES17 | SES16 | SES36 | SES7 |
| SES19 | SES66 | SES23 | SES23 | SES56 |
| SES23 | SES32 | SES76 | SES53 | SES16 |
| SES36 | SES44 | SES39 | SES19 | SES19 |
| SES47 | SES16 | SES50 | SES16 | SES24 |
| SES50 | SES75 | SES39 | SES9 | SES50 |
| SES7 | SES16 | | SES7 | SES79 |
| SES8 | | | SES2 | SES19 |
| SES16 | | | SES54 | |
| | | | SES16 | |

Table 7.11: Original MSS (left) and the adjusted MSS (right) by using an extra or in the schedule. The grey cells in the right MSS are the extra performed sessions due to opening an extra OR

As table 7.11 shows, five extra sessions are performed in the predicted week when using the adjusted MSS. The result of the program's simulation for both Master Surgical Schedules are shown in Figure 7.8. It shows clearly a higher expected number of occupied beds during the predicted period when using the adjusted MSS.

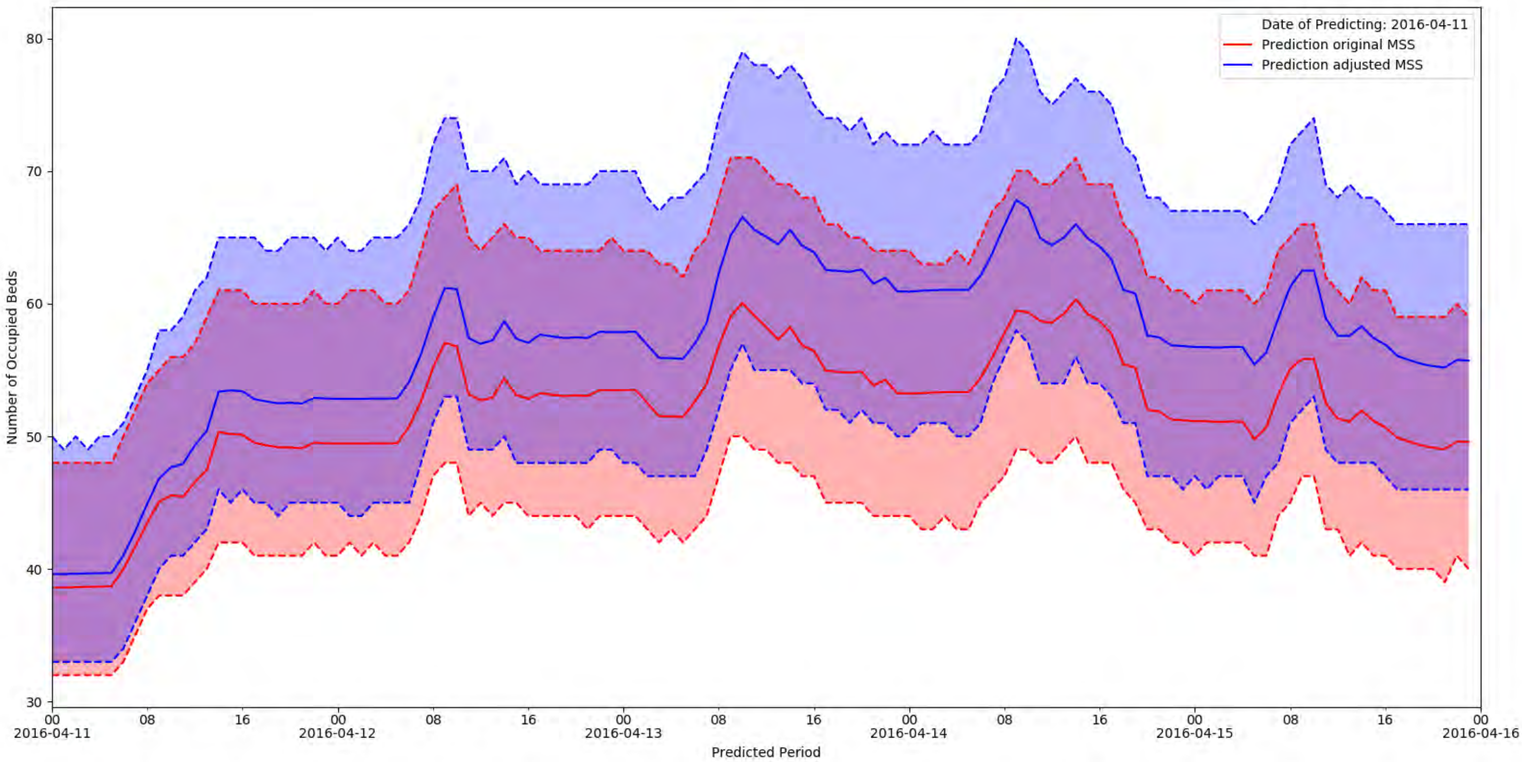



Figure 7.8: Result for original MSS and the edited MSS, having an extra OR available which is scheduled to perform one extra session each day

### 7.2.3 Shortening the Length of Stay

The final what-if scenario analyzed in this research concerns the length of stay of patients. Namely, this section shows the results of a scenario in which the length of stay of patients is decreased by 10%, compared to the original result. The parameter settings used in this scenario are the same as used in the first what-if scenario analysis, which is stated in table 7.8. Based on these parameter settings and the original length of stay as well as the adjusted length of stay, the program's simulation result is shown in Figure 7.9.

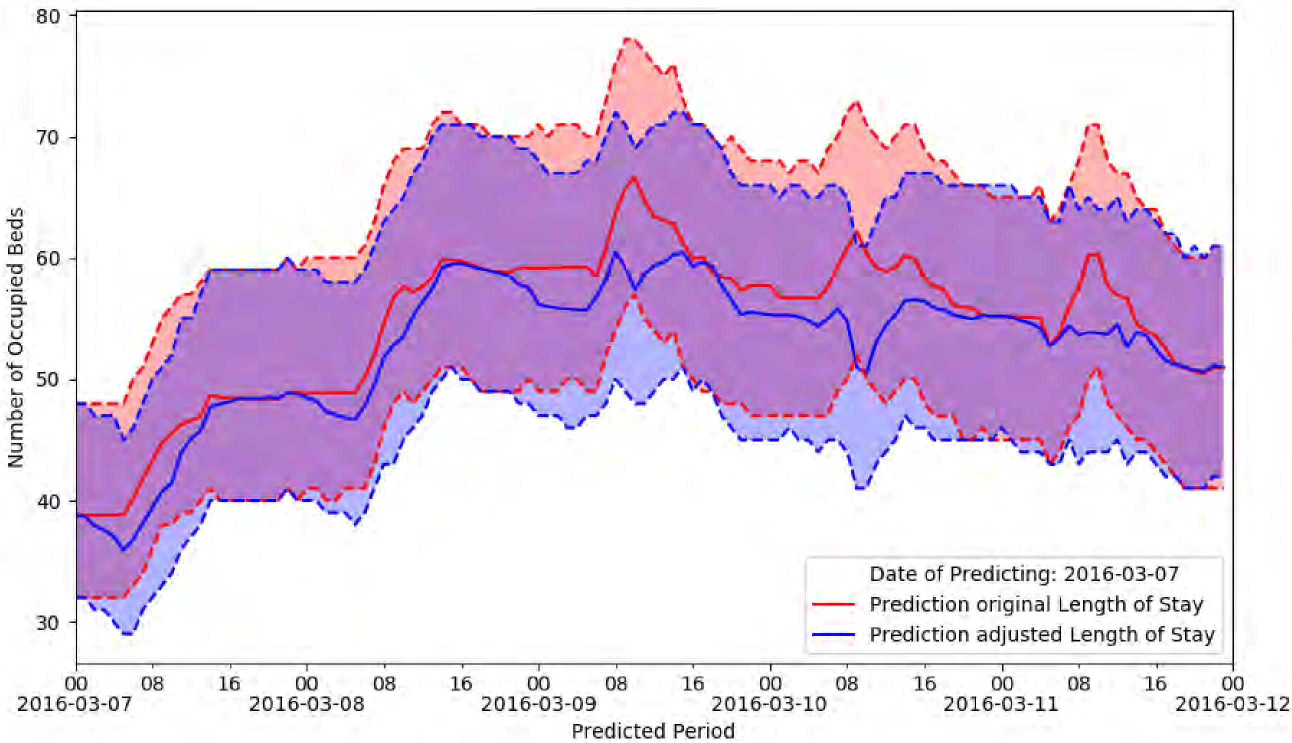



Figure 7.9: Result for original length of stay and the adjusted length of stay, which is decreased 10% compared to the original value

# Chapter 8

# Discussion

This section is dedicated to give an interpretation to the obtained results, discuss the performed work and address possible improvements to the model and its implementation.

## 8.1 Interpretation of Results

The interpretation of the results from this research internship is stated in twofold, first the performance of the designed program (and therefore the model) is discussed based on the first section of chapter 7, second, the program's output for the example what-if scenarios is discussed.

### 8.1.1 Program's Performance

Considering the results stated in section 7.1, the program's performance reacts differently on varying different input parameters, which will be addressed in this section. Also, possible clarifications of the showed behavior is stated for the predicted scores, but first, the general performance of the model and its implementation is discussed.

#### General Performance

The general performance interpretation of the model and its implementation (the developed program) can be drawn based on the results showed in section 7.1. By varying the predicted period and maintaining a full simulation (including all patient sources), 7 independent simulation/prediction results are obtained for each hospital. Of these 14 results, the lowest precision scores obtained are 79% for hospital A and 51% for hospital B. The lowest value for hospital B is obtained with $t_s$ being a Wednesday and $t_p$ being a Saturday, the fact that the program tries to predict a period starting in the middle of the week (and thus not including the Monday and Tuesday predicted admissions), might be of influence to the score (for hospital A the value was also low compared to the average score).

Looking at the shown graphs for hospital A and B (in the appendix) for the different predicted periods, neither the results for hospital A nor the results for hospital B indicate a structural higher or lower predicted value in comparison to the real measured value. In fact, graph (a) in Figure 7.2 starts off having a predicted value more or less equal to the real measured value and ends with a predicted value significantly lower than the real measured value. Graph (b), however, starts off with a predicted value being significantly higher than the real measured value, but ends the predicted period with values that are nearly equal to the real measured value. For hospital B, the graphs using different predicted periods also show a predominantly lower prediction values for one of the graphs and one predominantly higher prediction values. Therefore, although most precision scores are reasonably high, it is assumed that the actual prediction value of formula 4.14 does not meet the real measured value exactly very often. The simulation of the confidence interval is therefore a good solution to apply some flexibility to the prediction value by including possible situations to be obtained. An important reason of the difference between the predicted values by formula 4.14 and the real measured value is that the predicted value $X(t_n)$ by the prediction formula is a continuous value, whereas the real measured values are discrete (as are the simulated confidence bounds).

Furthermore, it is important to note the influence of the scoring function on the obtained results stated in table 7.4. The scoring function is defined as the ratio real measured values being included in the predicted confidence intervals, which is simulated based on two simulation-bound parameters: $\alpha$ and $\beta$. These two parameters denote the lower and upper bounds of the possible scenarios (described as a number of occupied beds) obtained through the simulation for the predicted moment $t_n$. This simulation is obviously based on the historical data, which means it is based on the historically occurred scenarios. Therefore, it is expected that, when predicting many different predicted periods and obtaining the precision-score, the obtained average precision-score is equal to the confidence interval width $\beta - \alpha$. This is obviously only the case under the assumption that the predicted period is comparable to the periods analyzed in the data analysis part of the program and thus included in the simulation process.

### Varying $t_p$

The results of the precision score obtained when varying the time of predicting, but maintaining the same predicted period, indicate a worse prediction when the time difference between the making of the prediction and the predicted period is large. Although this holds for hospital A specifically, hospital B also shows some lower predicted values for predictions made further in advance.
When looking at the graphs, the predictions made for hospital A and B (see appendix) start off at $t_s$ (the start moment of the predicted period) with a predicted value that is unequal to the real measured value. This indicates a miscalculation in the value obtained for patient source 1 (the remaining patients). Since formula 4.14, which calculates the predicted value for $t_n$, is only dependent on $t_p$ based on the remaining patients, the influence of $t_p$ on the prediction score can further be interpreted in section 7.1.3. Using the same predicted period, but only including patient source 1, the graph for $t_p$ being 2 days before $t_s$ clearly shows a reasonable prediction compared to the real measured values. However, when $t_p$ was set to 15 days in advance of $t_s$, the precision score obtained equals 0.00 and the graph shows that the error with which the predicted period starts ($\approx 5$ beds difference between prediction and realization) is never solved in the predicted period. This behavior is endorsed by the graphs of hospital B (see appendix), indicating that a relatively large difference will remain during the predicted period, although the absolute difference is only small.
When separating the predicted values for patient source 1 into priorities (emergency and elective), as done in Figure 7.4, it shows that for both hospital A and B the predicted value for the elective patients barely changes, whereas the predicted value for the emergency patients roughly follows the shape of the real measured values.

### Varying the patient sources

Besides the influence of $t_p$ on patient source 1, more interpretation can be made from the results with respect to the different patient sources. For both hospital A and B holds that the prediction for patient source 2 closely follows the real measured values, except between 2015-10-08 08:00 and 2015-10-09 08:00 for hospital B. This might be due to a different discharge policy in this specific predicted week, as the model predicts patients to remain longer in the

hospital compared to the realization. Besides the wrongly predicted moments for hospital B, the absolute difference between the prediction and the realization is $\approx 5$ beds at maximum for both hospitals, which can be interpreted as a good estimation for the elective patients included in patient source 2.
Patient source 3 uses a deterministic length of stay for each of the time intervals and a time-dependent arrival rate for the expected arrivals in the time intervals. The predictions (moreover the simulated confidence bounds) appear to include most of the realizations obtained in the predicted period, except for a mismatch in values between the simulated bounds and the realized values at the start and end of the predicted period ($t_s$ and $t_e$) for both Hospitals.

### 8.1.2 Results of What-If Scenarios

This research states the outcome of the designed program for three different what-if scenarios that might occur while designing a Master Surgical Schedule or making capacity related decisions. Although the result of the prediction can not be scored versus the real measured values (because these are obviously unknown for what-if scenarios), the result of the program before and after the what-if scenario has been implemented can be interpreted. For the first scenario (swapping Monday and Tuesday in the MSS for one week), the influence of this change becomes clear in the predicted number of occupied beds by the program. The prediction without the what-if scenario states lower values of predicted number of occupied beds for the Monday and a large increase in prediction on Tuesday, whereas the prediction with the what-if scenario states a higher value of prediction values on Monday and a smaller increase in the prediction on Tuesday. Apparently, the sessions performed on Tuesday in the original MSS result in more rehabilitating patients in the clinical wards than the sessions performed on Monday in the original MSS. At the end of the predicted week, no difference in before/after the what-if scenario implementation is visible. Results, as stated in this example, can act as a decision support tool for hospitals to decide if changes made to the MSS are expected to be desirable or not.
The second what-if scenario shows the difference in predicted number of occupied beds between the original MSS and an MSS with an extra OR, which has already comparable sessions scheduled on its extra OR. The result of the program for the predicted period clearly shows an increase in predicted number of occupied beds for the MSS with extra OR, as expected. Using this what-if scenario, hospitals are able to predict the influence of opening an extra OR, when it is used for performing already existing sessions (as showed in this example).

## 8.2 Discussion of Work, Possible Improvements and Further Research

#### Towards the model

The model is designed to analyze the influence of an MSS on the clinical wards based on a number of occupied beds. Because of this, the session characteristics required for the model are calculated based for each unique session. A unique session is defined unique based on session name and end time category, with end time category being "Morning" or "Afternoon". It is possible that clear differences exist between the included performed sessions to calculate the characteristics for the corresponding unique session they belong to. For example, both performed sessions of SES18 with end times 14.00 and 16.00 are included in the calculation of the characteristics for "SES18 - Afternoon", however, the 2-hour difference in length is obviously important for the number of performed surgeries (and thus patients). The model might be improved by using a different definition for a unique session to be used in the prediction model.

By performing further research towards the optimal definition of a "unique" session and corresponding characteristics this improvement could be realized.

Another possible improvement to the model is to change the way it includes the emergency patients. In its current form, the model includes emergency arrivals based on a time-dependent Poisson arrival process with a time-dependent deterministic length of stay. A lot of research has already been performed in analyzing the emergency arrivals in hospitals and their outcomes could be used in the model used in this research to increase its ability to simulate the reality regarding the emergency arrivals. Examples of different handling of the emergency patients are used in [5, 14, 13, 4].

As stated in the interpretation of the results section of this chapter, the model tends to struggle with the probability of a patient present at $t_p$ to be still present at the predicted moment $t_n$. To calculate the probability, the model uses the empirical cumulative distribution of the postsurgical length of stay, but sometimes this distribution does not result in an acceptable probability. For example, when a patient is the first patient ever from a session to reach a certain length of stay, the model's probability for the patient to remain hospitalized is zero, due to the fact that historically the patient's length of stay has not been reached a single time. This property of the model seems counter-intuitive in practice since usually a patient having a significant long length of stay already is more likely to remain hospitalized for another period. Research regarding this remaining length of stay and the implementation of its result could improve the model's prediction regarding this aspect.

The use of the empirical cumulative distribution function of the length of stay could also be improved in general. For example, when a session is only performed a few times in the historical data, the representative value of these historical occurrences can be questioned. Also, because of the lack of historical occurrences for some sessions, their corresponding probabilities for the current patient are less likely to describe the situation of the current patient. For example, when the historical lengths of stay corresponding to a session were 20, 30 and 130 hours, and a patient currently hospitalized for 40 hours would receive a probability of 1 to stay until 130 hours. Of course, since the MSS are applied for longer periods and on a repetitive basis, the usage of sessions with very few historical occurrences and patients is expected to only happen on an incidental basis.

**Towards the Developed Program**

The implementation of the model is done by developing a computer program that performs the prediction within a user-defined prediction period. By doing so, the practical value of the model can be tested and the required adjustments can be added in order to use the model in a practical manner. Although the obtained results from the simulation examples are reasonable and the difference between the original and the adjusted situations in the what-if scenarios are easily recognized using the developed program, improvements still exist to further optimize the model's usage.

Using the developed program, a result is obtained in which the expected number of occupied beds is stated for every hour during the predicted period. However, looking at the real measured values in the graphs, little to no changes occur in the number of occupied beds during approximately 20:00 o'clock and 06:00 o'clock the next day. This is possibly the case due to the lack of surgeons to discharge patients as well as the simple fact that discharging patients in that time frame is undesirable in general. Of course, it is still possible for emergency patients to arrive during these hours, so an increase in number of occupied beds during the stated time frame is visible from time to time. The decision to not discharge patients during the stated time slot is a decision made by the hospital, the model currently does not take a "non-discharge time frame" into account and the implementation of such a time frame could improve the model's implementation in reality.

Another improvement to the model would be the inclusion of the actual clinical wards of the hospital. Currently, the model predicts the total number of occupied beds as a result of the MSS and the emergency arrivals using the hospital's ORs. However, the distribution of this total number of patients among the different clinical wards is not stated. To predict the distribution of the patient among the clinical wards, more research is required towards the way the hospitals distribute the patients in different scenarios. It is unlikely and therefore probably wrong to assume that all patients from a certain session are hospitalized at the same clinical wards, therefore some form of allocation policy needs to be implemented to distribute patients among the clinical wards. An example research that addresses an allocation policy to obtain a predicted value for the clinical wards is [14].

# Chapter 9

# Conclusion

This chapter provides conclusions that can be drawn regarding the research question as well as the research internship in general.

Based on the previous chapters, the research question:

> "Is it possible to design and implement a predictive analysis model based on the downstream relationship between the ORs and the clinical wards to support hospitals in making nursing capacity related decisions?"

can be answered with yes. The performance of the model's implementation results in an average precision score of 88%, obtained by simulating 7 different predicted periods for two completely different hospitals, which is an acceptable result in answering the research question with "yes". Also, the computer program designed to apply the model is used in three example what-if scenarios and clearly states the expected differences between the situation before and after the what-if scenario is implemented, which proves that this research can be used to assist hospitals in making capacity related decisions, which on its turn endorses answering the research question with "yes".

During this internship research, a model is designed that assists hospitals in analyzing the downstream relationship between the ORs and the clinical wards. To do so, the model predicts the expected number of occupied beds at a certain predicted period based on multiple patient sources that require OR usage. Included in these patient sources are:

- The patient present in one of the clinical wards at the time of predicting
- The patients arriving in one of the clinical wards due to a scheduled OR session to be performed in the future
- The emergency patients that require OR usage in the future

The prediction of occupied beds based on the first patient source is done using the cumulative distribution function of the postsurgical length of stay, which is calculated for each unique session. The prediction of occupied beds based on the second patient source is also done using the cumulative distribution function of the postsurgical length of stay for the unique sessions, but extended with a cumulative distribution function of the presurgical length of stay. These cumulative distribution functions are then combined with the distribution of the number of patients in the corresponding sessions and a binomial distribution to predict the expected patients present at the predicted period. In order to calculate the predicted number of occupied beds based on the third patient source, a Poisson arrival process is used with a time-dependent arrival rate, as well as a time-dependent expected length of stay.

Afterwards, the model is implemented in a computer application using the Python programming language. This allows flexible predictions to be made based on user input, such as the predicted period and the moment of prediction. Instead of predicting the expected number of occupied beds at one point in time, the application calculates a prediction for each hour in a user-defined predicted period.
Besides a value for the predicted number of occupied beds in a time horizon, the computer program calculates a confidence interval for each predicted moment, based on a simulation it performs. The final step of the computer program is to calculate a precision score, which represents the fraction of actual measured values within the simulation confidence interval. The bounds of the confidence intervals are also part of changeable user input. As output, the computer program visualizes the predicted period with its predicted values, the actual measured values as well as the precision score obtained in the prediction.

The designed computer program is then used for two purposes, first, its performance is tested versus the real measured value using the precision score to obtain insights in its reliability and second, its output is used to analyze three different what-if scenario examples known to exist within hospitals in general.

# Appendices

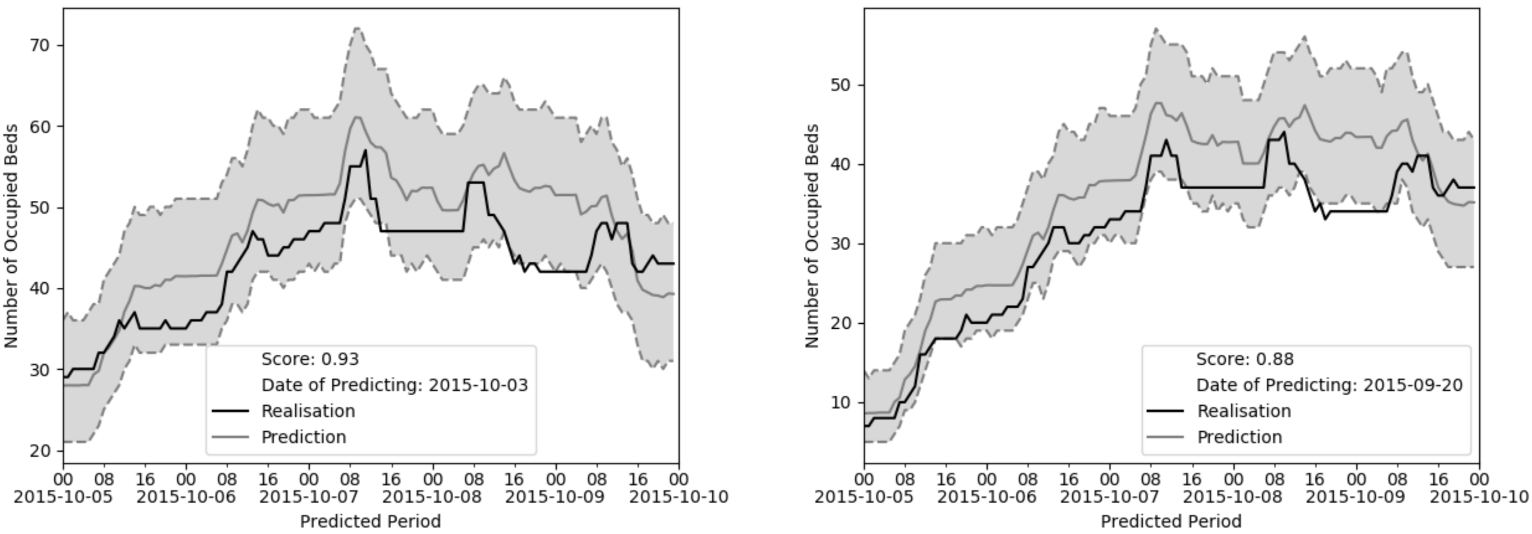
# Appendix A

# Admission datasets

| Admission Dataset | | |
|---|---|---|
| Field | Calculated\Extracted | Description |
| SESSIE | Extracted | Session name |
| OK | Extracted | OR name used in session |
| PLANNR | Extracted | Unique admission number |
| START | Extracted | Start time of the session |
| STOP | Extracted | End time of the session |
| OPERATIENR | Extracted | Unique surgery number |
| DATUM | Extracted | Date of the session |
| DEFOPNDAT | Extracted | Date of admittance |
| DEFOPNTIJD | Extracted | Time of admittance |
| admissionDateTime | Calculated | Combines date and time of admission |
| BT_OPERATI | Extracted | Time of surgery |
| OPERATIE_D | Extracted | Date of surgery |
| surgeryDateTime | Calculated | Combines date and time of surgery |
| losPreSurgery | Calculated | Length of Stay before surgery (hours) |
| DEFONTSLDAT | Extracted | Date of discharge |
| DEFONTSLTIJD | Extracted | Time of discharge |
| dischargeDateTime | Calculated | Combines date and time of discharge |
| losPostSurgery | Calculated | Length of Stay after surgery (hours) |
| LoS | Calculated | Total length of stay (hours) |
| STATUS | Extracted | Admission status |
| SPECIALISM | Extracted | Admission specialty |
| CODE | Extracted | Surgical intervention code |
| SPOED | Extracted | Admission priority |
| BEHCODE | Extracted | Admission treatment code |
| SESSIENR | Extracted | Unique session number |
| AFDELING | Extracted | Admission ward |
| ASASCORE | Extracted | Risk score of patient before surgery |
| OKID | Extracted | Unique surgical intervention number |
| InterventionAmount | Calculated | Amount of interventions during surgery |
| STATUS | Extracted | Surgery status |
| ANNUDAT | Extracted | Admission cancelling date |
| CATEGORIE | Extracted | Admission main category |
| CATEGORIE | Extracted | Admission sub-category |

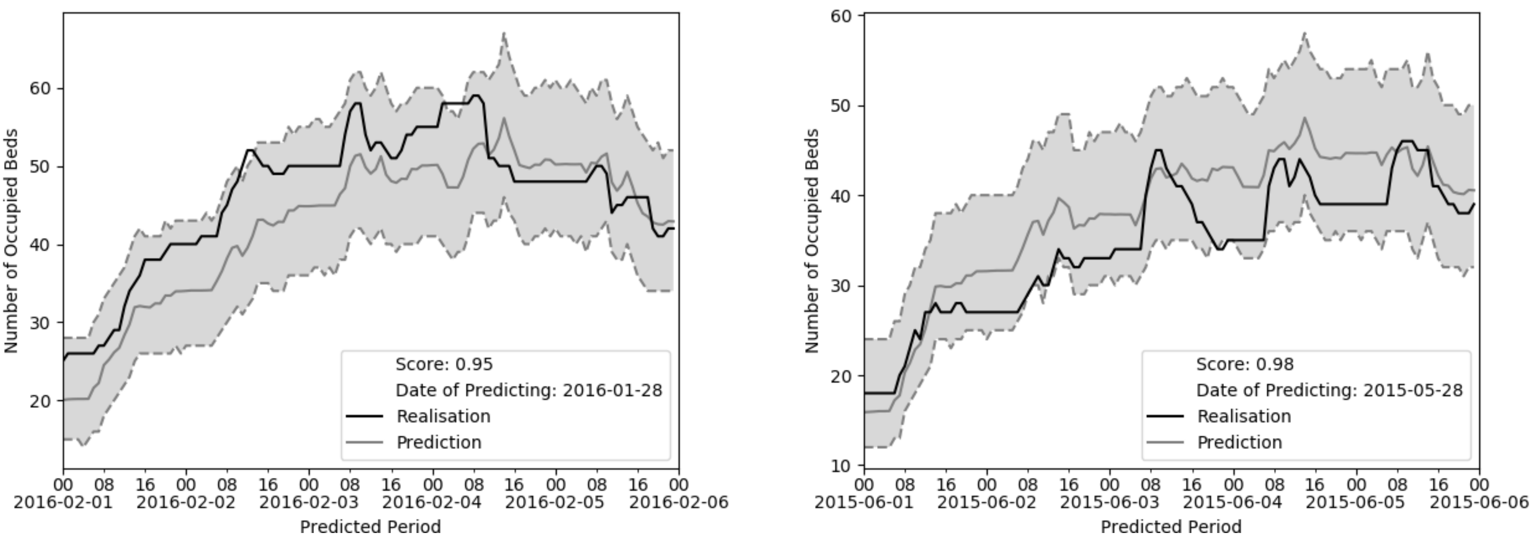
# Appendix B

# Visual Results Hospital B



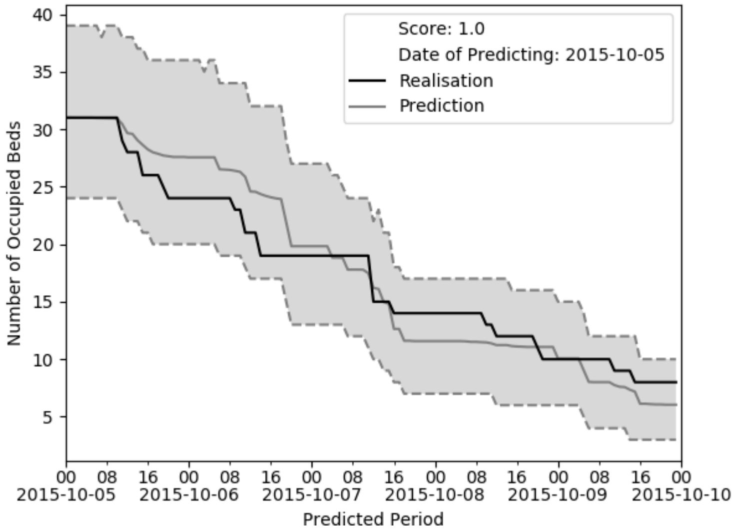

(a) result for $t_p$=2015-10-03 (b) result for $t_p$=2015-09-20

Figure B.1: Prediction results for hospital A based on parameter input as stated in 7.1 and different values for $t_p$.

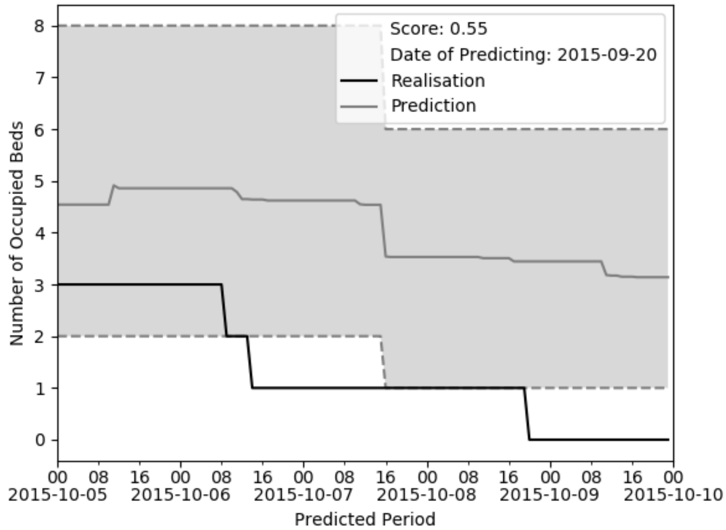



(a) result for $t_s$=2016-02-01 (b) result for $t_p$=2015-06-01

Figure B.2: Prediction results for hospital A based on parameter input as stated in 7.3 and two different predicted periods $t_s$.

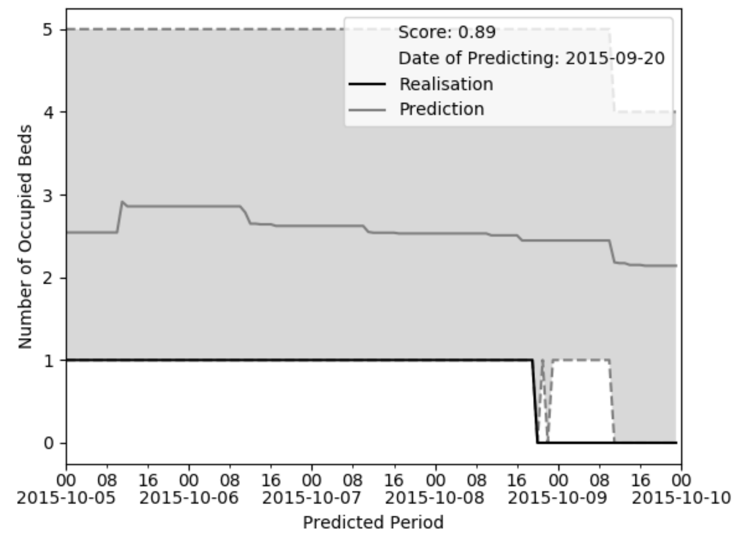



(a) result for $t_p$=2015-10-03

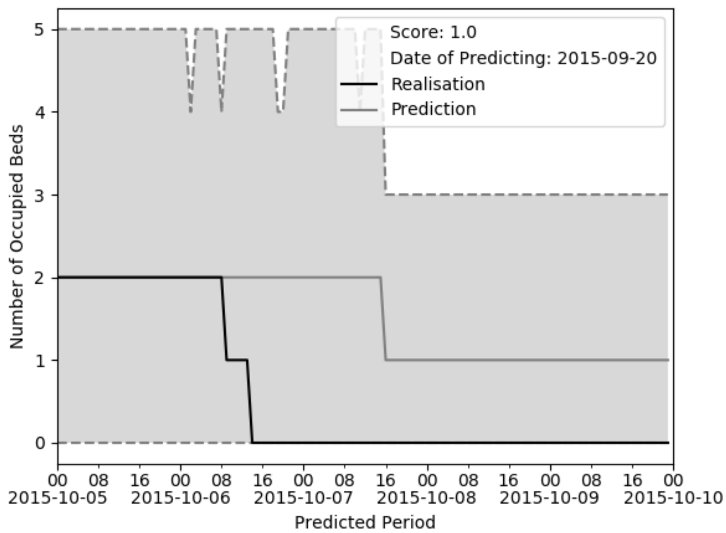



(b) result for $t_p$=2015-09-20

Figure B.3: Prediction results based on parameter input as stated in 7.1 and different values for $t_p$ for hospital B.

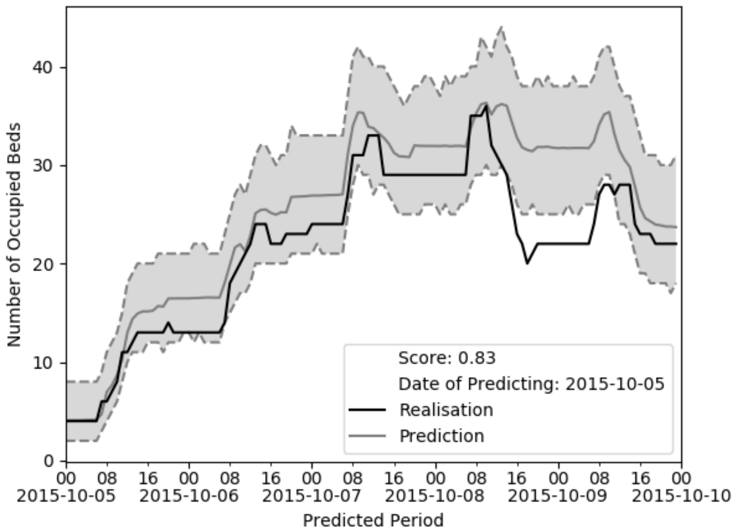



(a) result for elective priority only

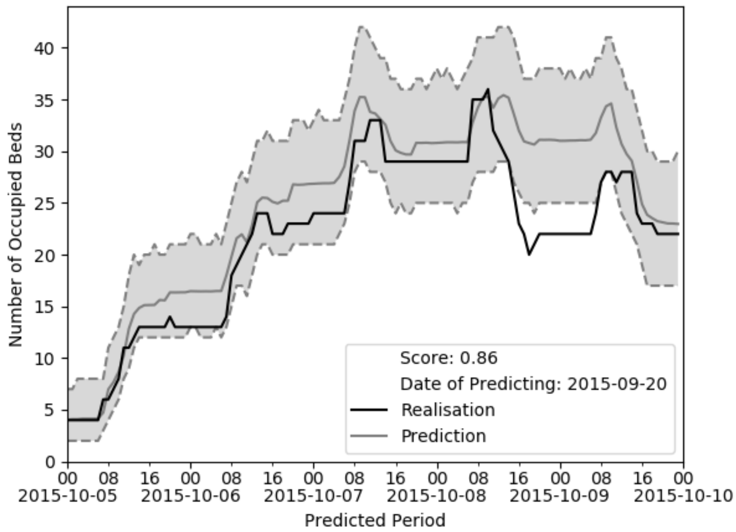



(b) result for emergency priority only

Figure B.4: Prediction results based on patient source 1 and parameter input as stated in 7.1 for hospital B, each graph using a different values for $t_p$.

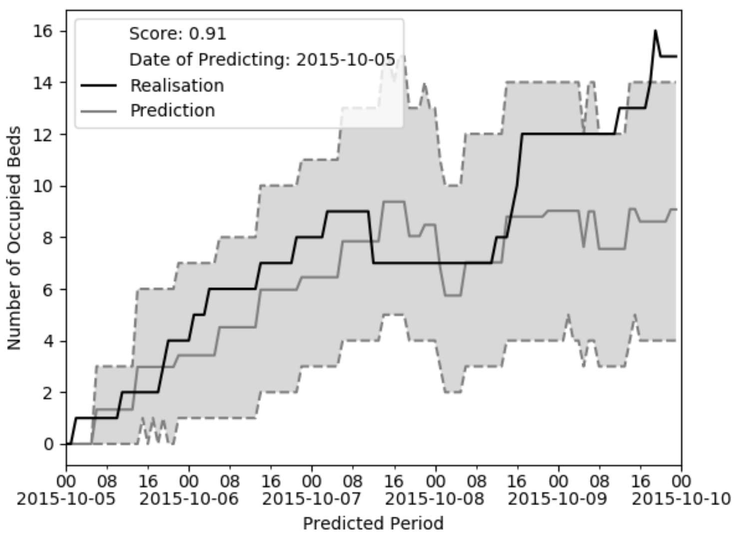



(a) result for elective priority only

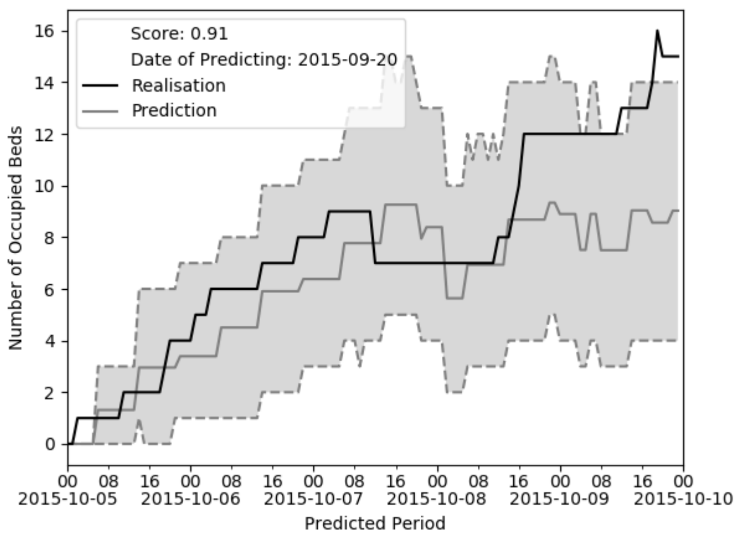



(b) result for emergency priority only

Figure B.5: Prediction results based on parameter input as stated in 7.1 and different values for $t_p$.




(a) result for elective priority only




(b) result for emergency priority only

Figure B.6: Prediction results based on parameter input as stated in 7.1 and different values for $t_p$ for hospital B.

# Bibliography


[1] L.H. Aiken et al. "Hospital nurse staffing and patient mortality, nurse burnout and job dissatisfaction". In: *Journal of the American Medical Association* 288.16 (2002), pp. 1987–1993.

[2] David Anderson et al. "Examining the discharge practices of surgeons at a large medical center". In: *Health Care Management Science* 14 (2011), pp. 338–347.

[3] Health Care Financial Management Association. "Achieving operating room efficiency through process integration, Technical Report". In: *HFMA* (2005).

[4] R Bekker and A.M. de Bruin. "Time-dependent analysis for refused admissions in clinical wards". In: *Annals of Operations Research* (2009). DOI: 10.1007/s10479-009-0570-z.

[5] Arnoud M. et al de Bruin. "Modeling the emergency cardiac in-patient flow: an application of queuing theory". In: *Health Care Management Science* 10 (2007), pp. 125–137. DOI: 10.1007/s10729-007-90009-8.

[6] ChipSoft B.V. *ChipSoft B.V.* URL: https://www.chipsoft.com/.

[7] B. Cardoen, E. Demeulemeester, and J. Beliën. "Operating room planning and scheduling: a literature review". In: *European Journal of Operational Research* 201 (2010), pp. 921–932. DOI: 10.1016/j.ejor.2009.04.011.

[8] Sylvia G. Elkhuizen et al. "Capacity management of nursing staff as a vehicle for organizational improvement". In: *BMC Health Services Research* 7 (2007). DOI: 10.1186/1472-6963-7-196.

[9] Andreas Fügener et al. "Master surgery scheduling with consideration of multiple downstream units". In: *European Journal of Operational Research* 239 (2014), pp. 227–236. DOI: 10.1016/j.ejor.2014.05.009.

[10] HIMSS. *EMRAM.* URL: http://www.himss.eu/healthcare-providers/emram.

[11] ISO. *ISO 13485.* URL: https://www.iso.org/standard/59752.html.

[12] R.L. Kane et al. "The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis". In: *Medical Care* 45.12 (2007), pp. 1195–1204.

[13] E. Kortbeek et al. "Flexible nurse staffing based on hourly bed census predictions". In: *International Journal Production Economics* 161 (2015), pp. 167–180. DOI: 10.1016/j.ijpe.2014.12.007.

[14] E. Kortbeek et al. "Integral resource capacity planning for inpatient care services based on hourly bed census predictions". In: *Journal of the Operational Research Society* 66.1061 (2015). DOI: 10.1057/jors.2014.67.

[15] A. Macario et al. "Where are the costs in perioperative care? Analysis of hospital costs and chargers for inpatient surgical care". In: *Anesthesiology* 83 (10) (1995), pp. 1138–1144.

[16] J. Needleman et al. “Nurse-staffing levels and the quality of care in hospitals”. In: *New England Journal of Medicine* 346.22 (2002), pp. 1715–1722.

[17] Marco Papi, Luca Pontecorvi, and Roberto Setola. “A new model for the length of stay of hospital patients”. In: *Health Care Management Science* 19 (2016), pp. 58–66. DOI: 10.1007/s10729-014-9288-9.

[18] C Price. “Applications of operations research models to problems in health care”. In: *Dissertation, University of Maryland* (2009).

[19] Python. *Python*. URL: https://www.python.org/.

[20] P. Vanberkel et al. “Accounting for inpatient wards when developing Master Surgical Schedules”. In: *Anesthesia* 112(6) (2011). DOI: 10.1213/ANE.0b013e3182159c2f.

[21] P. Vanberkel et al. “An exact approach for relating recovering surgical patient workload to the master surgical schedule”. In: *Journal of the Operational Research Society* 62(10):1851-1860 (2010).

[22] P.D. Wright et al. “Reexamining the nurse scheduling problem: staffing ratios and nursing shortages”. In: *Decision Sciences* 37.1 (2006), pp. 39–70.

[23] JP Young. “Stabilization of inpatient bed occupancy through control of admissions”. In: *Journal of the American Hospital Association* 39 (1965), pp. 41–48.