Stochastic Master Surgery Scheduling

Thomas Reiten Bovim[a,*], Marielle Christiansen[a], Anders N. Gullhav[a], Troels Martin Range[a,c], Lars Hellemo[b]

[a] *Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*
[b] *SINTEF Industry, Operations Research, NO-7034 Trondheim, Norway*
[c] *Hospital of South West Jutland and Department of Regional Health Research: Center of South West Jutland, University of Southern Denmark, Finsensgade 35, Esbjerg DK-6700, Denmark*

Abstract

The aim of the Master Surgery Scheduling Problem (MSSP) is to schedule the medical specialties to the different operating rooms available, such that surgeries may be performed efficiently. We consider a MSSP where elective and emergency patients can be treated in the same operating rooms. In addition to elective-dedicated operating room slots, flexible operating room slots are introduced to handle the fluctuating demand of emergency patients.

To solve the MSSP, we propose a simulation-optimization approach consisting of a two-stage stochastic optimization model and a discrete-event simulation model. For the two-stage stochastic optimization model, uncertain arrivals of emergency patients are represented by discrete scenarios. The discrete-event simulation model is developed to address uncertainty related to the surgery duration and the length of stay at the hospital, and to test the Master Surgery Schedule (MSS) developed by the optimization model in a stochastic operational-level environment. In addition, the simulation model is used to generate scenarios for the optimization model.

We present some general advice for surgery scheduling based on testing the optimization model in a numerical study. The simulation-optimization approach is applied to a case study from a hospital department that treats both elective and emergency patients. The optimized MSS outperforms the manually generated MSS, both in terms of emergency waiting time for surgery, and emergency interruptions to the flow of electives.

Keywords: OR in health services, Master Surgery Scheduling, Stochastic programming, Discrete-event simulation

*Corresponding author. Tel.: +47 95859178

Email addresses: thomas.bovim@ntnu.no (Thomas Reiten Bovim), mc@ntnu.no (Marielle Christiansen), anders.gullhav@ntnu.no (Anders N. Gullhav), Troels.Martin.Range@rsyd.dk (Troels Martin Range), Lars.Hellemo@sintef.no (Lars Hellemo)

1. Introduction

Demographic changes in Norway and many other countries are increasing the need for hospital services in the years to come. One of the major activities at a hospital is providing surgery to patients. Freeman et al. [9] state that 60-70% of all patients admitted to a hospital require some surgical intervention, and Essen et al. [6] state that surgical costs account for approximately 40% of the total hospital costs and that surgeries generate around 67% of hospital revenues. Developing ways to schedule surgery efficiently is key to proper utilization of scarce hospital resources, and a necessity to be able to treat more patients.

Patients are commonly divided into two groups: elective and emergency patients. Elective patients are not experiencing a medical emergency, and their surgery can be scheduled in advance to suit the availability of the surgeon and the patient. Emergency patients, on the other hand, may require surgery within hours or up to a few days. A triage system is often applied to further divide the emergency patients according to the urgency of their condition. Another classification of patients refers to whether the patient may leave the hospital following surgery or not.

One of the major issues within surgery scheduling is how to best balance efficiency and responsiveness when conducting surgeries for scheduled electives and high-priority emergencies. If the OR capacity is shared between electives and emergencies, emergencies can create disruptions to the handling of scheduled surgeries, implying longer elective waiting times, costly resource overtime, cancellations and rescheduling. If some of the OR capacity is dedicated to respond to emergencies and avoid disruptions of electives, there will be times when the dedicated capacity is not utilized as no emergencies are present [7].

Surgery planning may be divided into three decision stages [13]. At the strategic level, decisions on localization and dimension of the number and size of the operating rooms (ORs) are made. At the tactical level, a Master Surgery Schedule (MSS) is developed to schedule different specialties to the accessible ORs through the week. Finally, at the operational level, the individual patients are scheduled to the ORs covered by the respective specialty.

The main purpose of this paper is to provide tactical decision support for managers of departments that provide surgery to both elective and emergency patients. More specifically, we consider both elective and emergency patients in the Master Surgery Scheduling Problem (MSSP). The goal is to schedule the medical subspecialties to time slots in the ORs such that we ensure a sufficiently high throughput of elective patients while maintaining a high responsiveness for emergency patients. Two types of OR time slots are scheduled: elective slots and flexible slots that are primarily intended to handle sub-urgent emergency patients.

To perform the scheduling, we propose a simulation-optimization approach consisting of a two-stage stochastic optimization model and a discrete-event simulation model (see Figure 1). The optimization model generates an MSS, while the simulation model is used to evaluate the MSS and to generate new input scenarios for the optimization model. This allows us to generate a new MSS based on the scheduling rules that are applied in the simulation model and the MSS generated in the previous iteration. This is helpful to avoid using historic data that is dependent on the MSS and the scheduling regime that was present when the data was generated.

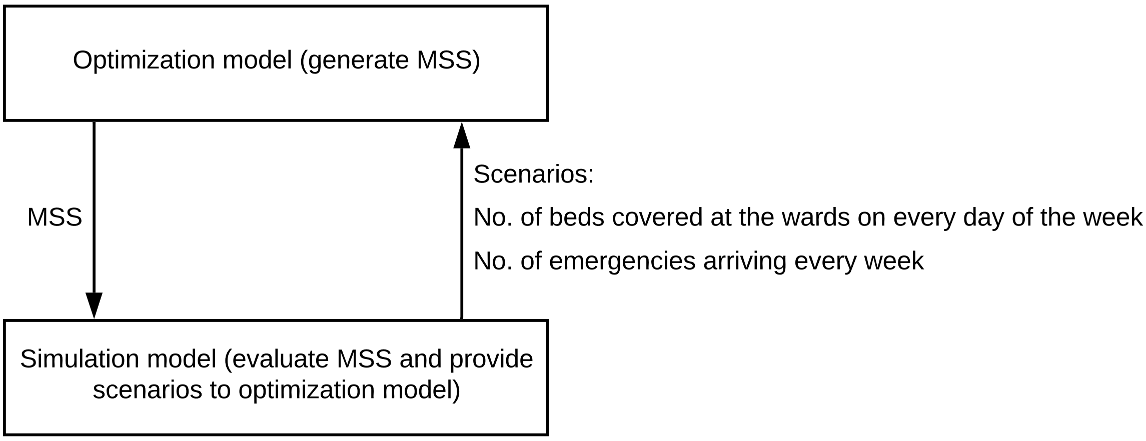

Figure 1: Illustration of our simulation-optimization approach. The optimization model generates an MSS, while the simulation model is used to evaluate the tactical schedule in a dynamic environment, and provides feedback to the optimization model in terms of scenarios.

The rest of the paper is outlined as follows: Section 2 presents relevant literature. The MSSP is introduced in Section 3. In Section 4, we present the approach for solving the MSSP. Following this, in Section 5, we present the results from a computational study where we investigate the value of applying a stochastic model formulation, managerial insights and a case study from the orthopaedic department at St. Olav's hospital in Trondheim, Norway. Finally, we conclude the paper and suggest topics for further research in Section 6.

2. Literature review

To provide context, we present relevant literature both on surgery planning and methods for solving the planning problems. First, we present an overview of decision levels within surgery planning.

Hans et al. [11] propose a holistic planning and control framework for a health care provider, which consists of four managerial areas, combined with a hierarchical decomposition of decision-making levels. Figure 2 illustrates the framework and provides examples of planning and control functions for each combination. The MSSP considers the managerial area of resource capacity planning. Furthermore, the MSS is a cyclical block schedule that is repeated for several months, implying that we consider the tactical planning level.

Within the field of OR planning, the majority of publications have considered only the elective patients [5]. The literature on the MSSP is no exception, and most authors argue that the emergency patients are handled with dedicated resources. However, some authors like Freeman et al. [9], Lamiri et al. [17], Razmi et al. [23] and Adan et al. [1] include emergency patients.

OR capacity is commonly divided into time blocks when solving the MSSP. Many authors consider surgery slots of equal length, while others, like Mannino et al. [20] include surgery blocks of different lengths. Testi and Tànfani [24], Mannino et al. [20] and Adan et al. [1] consider overtime work at the ORs.

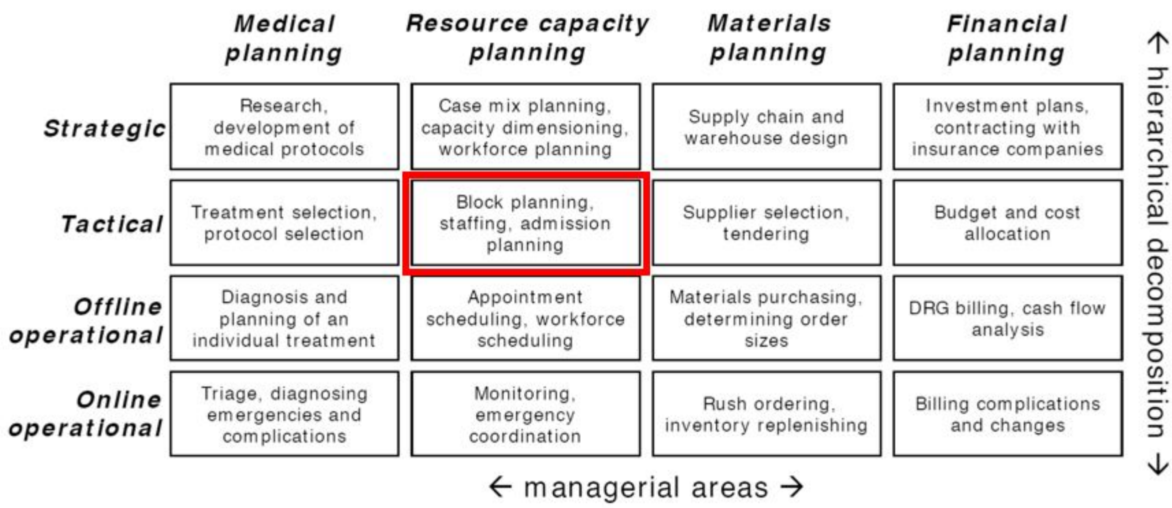

Figure 2: Framework for health care planning and control [11]

Testi and Tànfani [24] and Mannino et al. [20] minimize the overtime in the objective function, while Adan et al. [1] impose hard constraints on the overtime allowed for each OR. Koppka et al. [15] limit the total OR opening hours available through the week, but allows for the model to decide how the opening hours should be distributed over the different ORs.

The wards are frequently included. However, some authors, like Testi et al. [25], Li et al. [18] and Mannino et al. [20] disregard the wards in their model under the assumption that the access to beds is not imposing a bottleneck on the efficient flow of patients. A few authors, like Li et al. [18], Fügener et al. [10] and Adan et al. [1], include the intensive care unit (ICU) when handling the MSSP. The latter are also among the few that explicitly include the nurses, by imposing restrictions on the amount of nursing hours available at the ICU.

The literature on the MSSP presents numerous objective functions. Testi and Tànfani [24], Testi et al. [25] and Penn et al. [22] include aspects of welfare into their objective functions. Testi and Tànfani [24] propose an objective function that minimizes loss of welfare among the patients, and the latter two maximize surgeon preferences. A variety of objective functions regarding bed capacity is proposed in the literature. Oostrum et al. [21] aim to minimize both the number of ORs used and the maximum demand for hospital beds during the planning cycle. Ma and Demeulemeester [19] minimize the total bed deficit, the maximum daily spare bed volume and the maximum variance of the bed occupancy.

The inclusion of multiple criteria objective functions is used by several authors. In Beliën et al. [2], the objective function contains three parts: minimization of the total peak mean and variance bed occupancy, minimization of surgeons of the same specialty performing surgery in different rooms and minimization of surgeons not being scheduled to the same room on the same day every week of the planning horizon. Li et al. [18] aim at minimizing the number of patients not being scheduled, minimizing the underutilization of OR time, minimizing the maximum expected number of patients in the recovery unit and minimizing the expected range of patients in the recovery unit.

Several authors include aspects of uncertainty when handling the MSSP. Oostrum et al. [21] include a probability distribution for running into overtime in an OR as a function of the number of surgeries scheduled to that OR. Koppka et al. [15] consider the probability of running into overtime in the ORs that depends on the combination of patients that are scheduled for the OR. Adan et al. [1], Ma and Demeulemeester [19] and Li et al. [18] include probability distributions to account for uncertainty in the patient's length of stay (LOS) following surgery. Fügener et al. [10] calculate the distribution of patients resting in both the wards and the ICU resulting from a cyclical MSS. These distributions are used in the objective function to minimize the fixed costs, the overcapacity costs, and the staffing costs in both the wards and the ICU when generating the MSS.

According to Higle [12], stochastic programming is a technique which is well suited when some of the data elements are difficult to predict or estimate. A major framework within stochastic programming is two-stage recourse modeling. A two-stage recourse model consists of a first-stage problem and a second-stage (recourse) problem. The first-stage decisions are determined before knowing the outcome of the stochastic parameters, while the second-stage decisions are made after observing the realization of the stochastic parameters. The goal when applying a two-stage stochastic modelling approach is to identify a first stage solution that performs well in expectation, taking all possible realizations of the stochastic parameters into account. When approximating the continuous probability distributions of the stochastic parameters with discrete scenarios, extra care has to be taken to ensure stability of the solution (see e.g. Kall and Wallace [14]).

Some authors propose two-stage stochastic models at the tactical level within surgery planning. Koppka et al. [15] develop a two-stage stochastic model to deal with the varying number of elective patients that require surgery during the planning horizon. However, the authors do not include a recourse option. Kumar et al. [16] present a method that is inspired by the two-stage stochastic method with recourse. Uncertainty is incorporated by using various LOS scenario realizations, and non-anticipation (see Higle [12]) is imposed by constraining the model to schedule patients in the same order as their position in the queue. However, in contrast to the traditional two-stage stochastic models, the scenario realizations in this model are chronologically sequential and not parallel. This allows for the model formulation to be deterministic.

Figueira and Almada-Lobo [8] state that there are three major streams of simulation optimization research : Solution Evaluation (SE), Solution Generation (SG), and Analytical Model Enhancement (AME) approaches. Within these three streams several methods are available. The methods are categorized based on the interaction between simulation and optimization, and on the search algorithm design.

The SE approaches consist of developing a comprehensive simulation model to represent the system and use that model to evaluate performance of various solutions. The results of the simulation model is used to guide the search for new, better solutions. The SG approaches are used when optimization models can be formulated and solved, and their solutions simulated in order to compute realistic values of the variables. The purpose of simulation here is not to verify the advantage of one solution over another, but to compute realistic values for some variables and hence to be part of the whole solution generation procedure. In AME approaches, the optimization model is enhanced by the simulation results, for example by providing better estimates of parameters applied in the optimization model. One of the methods used in the AME approaches is called Stochastic Programming Deterministic Equivalent (SPDE), and here the simulation model is used to generate scenarios for the stochastic optimization model.

Several authors use the SG approach when formulating and solving surgery scheduling problems. Freeman et al. [9], Ma and Demeulemeester [19] and Testi et al. [25] use simulation models to evaluate the tactical schedule produced by the optimization model in an operational setting. This allows them to explore the realistic values of the decision variables when more uncertainty is included. Adan et al. [1] and Cappanera et al. [4] use discrete-event simulation to investigate different operational scheduling policies after having first generated a tactical surgery plan. By this they reveal realistic values of the tactical decision variables for different scheduling policies.

In accordance with the AME approach, Lamiri et al. [17] use Monte Carlo simulations and a sample average approximation (SAA) to solve a stochastic optimization problem with uncertainty related to emergency arrivals. Also Ma and Demeulemeester [19] apply the AME approach, as the simulated, operational level results are used to alter the parameter for the total bed capacity in the optimization model.

We propose a simulation-optimization approach to solve the MSSP. According to the AME approach (SPDE method), we use a discrete-event simulation model when generating the scenarios for our two-stage stochastic optimization model. We also use the SG approach, as we use the simulation model to evaluate the MSS generated by the optimization model. We introduce different levels of urgency for the emergency patients, and we allow for rescheduling of elective patients to provide capacity for the emergency patients in periods of excessive emergency patient loading. This allows us to investigate a major trade-off faced by the management at many surgical departments, namely the number of electives scheduled for surgery versus the amount of elective rescheduling needed in order to provide surgery for emergency patients.

3. Problem description for the MSSP

The MSSP is described in three steps. First, we provide the MSSP with a focus on elective patients. Then, we expand the problem by including emergency patients, and finally we present the concept of flexible slots that are used to handle the flow of emergency patients in periods of high emergency demand.

3.1. The elective MSSP

In this MSSP, the aim is to generate a cyclic MSS where the medical subspecialties are scheduled to the available ORs throughout the planning cycle (typically one week). A set of elective patient categories exist that share diagnostic similarities. The patient categories are either in- or outpatients. The surgeons performing the surgeries are divided into different medical subspecialties. Surgeons of a given subspecialty may perform surgery in several patient categories, but each patient category may only receive surgery from surgeons of one subspecialty.

The department has a given number of ORs where surgeries are performed. The ORs are heterogeneous and each OR may only accommodate certain subspecialties. The opening hours of the OR are divided into time slots, and each slot can be scheduled to one subspecialty. The number of slots allowed to schedule for a subspecialty on a given day, and during the cycle is limited. For an OR to be available for scheduling, an anaesthesia resource must be scheduled for the OR. The number of ORs that may be covered by an anesthesia resource each day is limited. Each patient category has an expected surgery duration, and the number of patients that can be scheduled for surgery in an OR is limited by the slot capacity scheduled for a suitable subspecialty.

There are several heterogeneous wards available, where inpatients rest following their surgery. In each ward a given number of beds can be staffed each day. An upper bound on the total number of beds that can be staffed each day. Within the total capacity, we can distribute the number of staffed beds that are available in each ward on a given day. A staffed bed is assigned to each inpatient entering the hospital on the day of arrival, and this bed is occupied by that patient throughout the stay. This scenario is not always true in real life, but it is a fair assumption from a tactical planning perspective.

The subspecialties are scheduled to the ORs according to the cyclical, fixed period MSS. The target throughput of elective patients to be scheduled for each patient category for the cycle is known. For all patient categories, a given share of the target throughput must be scheduled for surgery in each cycle. This minimum throughput should be set such that the average service rate is incrementally higher than the average arrival rate for each patient category, such that we maintain a stable waiting list of elective patients. Because elective patients are (in most cases) scheduled well in advance, short periods of peaks in elective patient demand can be smoothed out by the hospital planners. For that reason, we argue that planning according to the average demand is sufficient at a tactical level.

3.2. Considering emergency patients in the MSSP

The flow of emergency patients may cause elective cancellations and rescheduling in periods when the emergency OR resources are insufficient to handle the emergency demand for surgery. Therefore, the emergency patients should be considered in the MSSP. The emergency patients represent different urgency categories, and they are grouped according to three emergency scheduling regimes:

- The sub-urgent (SU) emergency scheduling regime applies to the least urgent emergencies. In periods when the emergency OR capacity becomes insufficient to handle all emergency patients, the least urgent emergency patients are scheduled for the elective ORs to free emergency OR capacity.
- The urgent (U) emergency scheduling regime applies to the urgent emergency patients. If no patients from the SU scheduling regime are present, and there is idle capacity in the elective ORs, the urgent emergency patients are scheduled for the elective ORs to free capacity for the most urgent emergency patients in the emergency ORs.
- The critically urgent (CU) emergency scheduling regime applies to the most urgent emergency patients. These patients are always scheduled for the emergency ORs.

Scheduling emergency patients to the elective surgery slots may cause elective cancellations. However, if there is excess capacity in an elective OR slot, after all elective patients have been scheduled, emergency

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
OR-1	Hand	Hand	Hand	Hand	Hand	Closed	Closed
OR-2	Back	Flexible	Back	Foot	Closed	Closed	Closed
OR-3	Foot	Prothesis	Prosthesis	Prosthesis	Flexible	Closed	Closed
OR-4	Flexible	Foot	Foot	Prosthesis	Arthroscopic	Closed	Closed

Figure 3: An example of a one-week MSS for four single-slot ORs, where both elective-dedicated and flexible slots are scheduled.

patients might be scheduled to the slot without causing cancellations (given that the idle capacity exceeds the planned surgery duration of the emergency patient).

In periods with a bed shortage in some of the wards, it is possible to let patients rest in wards dedicated to other patient groups. If the number of emergency patients requiring a bed increases, and no more scheduled beds are available, elective inpatients are cancelled to provide additional bed capacity. If the number of emergencies needing beds exceeds the total number of scheduled beds available, more beds need to be staffed. The random arrivals of emergencies cause both the emergency demand for surgery and the number of emergencies occupying beds in the wards each day to differ among cycles.

In periods of high emergency demand, SU emergency patients are displaced from the emergency ORs due to lower priority. These patients are called the excess demand of SU emergencies, and all of these should receive surgery in the elective ORs within the cycle. To handle the SU patients, we schedule flexible slots that are reserved for these patients. In periods of low demand for SU surgeries, the flexible slots can be used for scheduling of U emergency patients. Not all ORs may be accessible for scheduling of flexible slots. Figure 3 provides an example of an MSS including both elective-dedicated and flexible slots.

3.3. Objectives

Three objectives are relevant for the MSSP. The first is to maximize the number of elective patients scheduled for surgery. Secondly, we aim to minimize both the number of elective cancellations, and the number of patients resting in wards not designated for them.

4. Simulation-optimization approach

In this section, we present the simulation-optimization approach used to solve the MSSP. The optimization model is formulated as a two-stage stochastic optimization problem. Scenarios are used to represent the probability distributions of the stochastic parameters, which are the number of emergency patients resting in each ward during each day in the cycle, and the excess demand of SU emergency patients. Figure

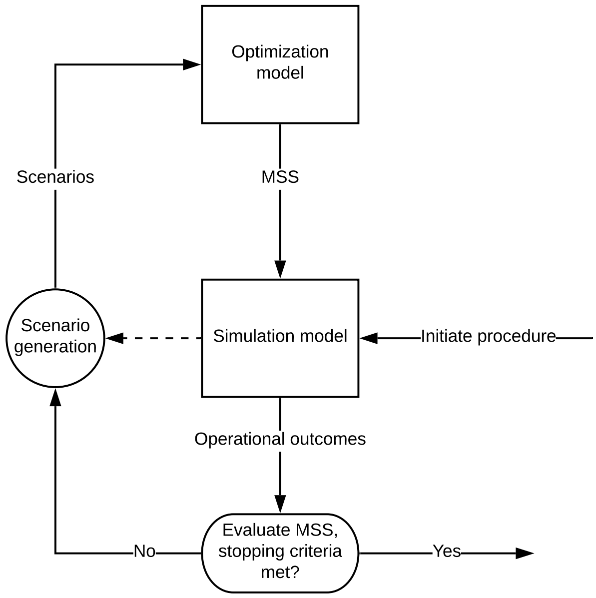

Figure 4: The set-up of the simulation-optimization approach, including both the optimization model and the simulation model.

4 illustrates the main components of the simulation-optimization approach, and their interactions. The simulation-optimization process consists of the following steps:

1. To initiate the procedure, the simulation model is run for an arbitrary MSS to generate input data for the scenarios applied in the optimization model.
2. The scenarios generated from the simulation model is used as input for the optimization model, and the optimization model generates a new MSS based on these.
3. The new MSS is then implemented in the simulation model, and new scenarios are generated. Relevant measures, such as patient waiting time for surgery and the number of elective cancellations are stored.
4. The procedure (2.-3.) is repeated until a stopping criterion is met, and is then terminated.

The simulation model takes the MSS generated by the optimization model as input, and evaluates it in an operational environment. In the optimization model, the LOS and surgery duration are deterministic parameters. In the simulation model, these values are drawn from empirical probability distributions for each patient. By generating scenarios from the simulation output, we are able to generate input to the optimization model that is dependent on the scheduling regime applied in the simulation model and the MSS at the previous iteration.

In Sections 4.1 and 4.2, we present the optimization and simulation models respectively. In Section 4.3, we discuss the scenario generation, and in Section 4.4, we describe the stopping criterion that terminates the simulation-optimization procedure.

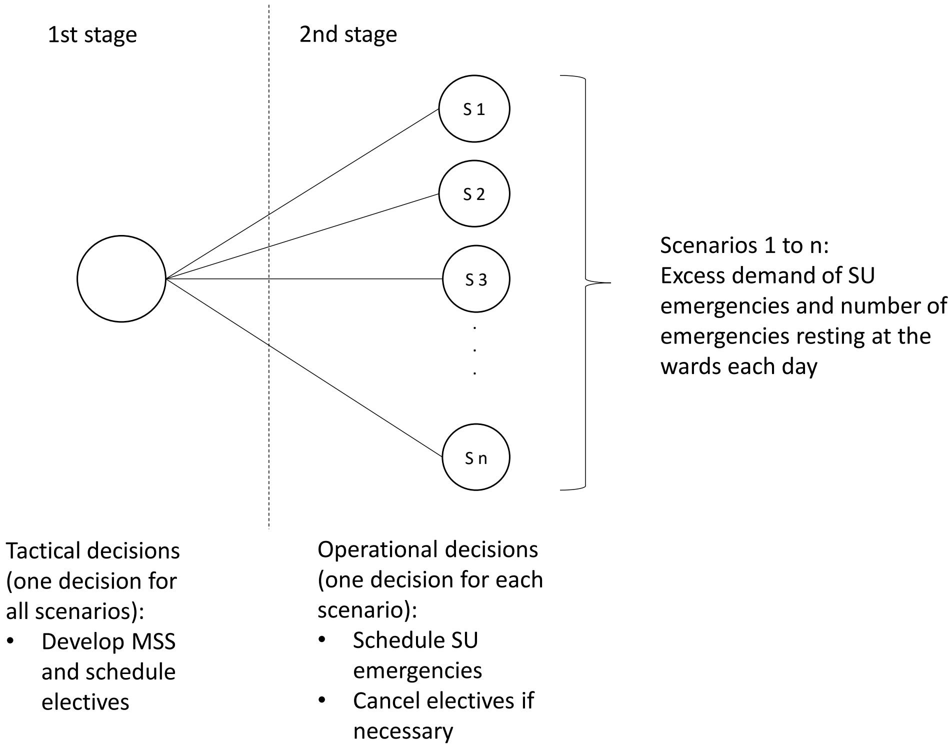

Figure 5: Illustration of the two-stage decision model

4.1. The optimization model

We must decide on the MSS before knowing the exact number of SU emergencies that require surgery in every cycle, or the number of emergencies occupying beds in the wards each day. Therefore, a two-stage modelling framework is suitable for the optimization model, as illustrated in Figure 5. The first-stage decisions are tactical-level decisions, and are as follows:

- Assign an anaesthesia resource to the ORs in use.
- Schedule the available OR slots as either flexible or elective, and assign a subspecialty responsible to each slot.
- Schedule the elective patients for surgery in the elective slots.
- Decide on the number of staffed beds in each ward on every day throughout the cycle.

The second stage decisions are operational-level decisions that are made within each cycle:

- Schedule the excess demand of SU patients to the flexible and elective slots (if no more flexible slots are available).
- Cancel elective surgeries if necessary.

- Send inpatients to the wards and let patients rest in wards not designated for them, if necessary due to shortage of bed capacity.
- Staff more beds if necessary to handle all the emergency inpatients.
- If excess capacity is available in the flexible slots, schedule emergency patients belonging to the U emergency scheduling regime to these slots.

The uncertainty included in the optimization model is the excess demand of SU patients that require surgery in the cycle, and the number of emergencies resting in the wards each day through the cycle. These are represented by scenarios, and each scenario contains a complete realization of both aspects for one cycle.

In the following, sets are indicated by calligraphic letters, parameters are given by uppercase letters, and variables are in lowercase Latin or Greek letters. Appendix A provides a detailed overview of all indices, sets, parameters and variables used in the model formulation.

4.1.1. The first-stage constraints

The total number of slots available for scheduling is given by constraint (1). The variable n_{jkd} represents the number of flexible slots scheduled for subspecialty j in OR k on day d, while y_{jkd} does the same for the elective slots. The parameter M^{CYCLE} represents the number of slots available through the cycle. Constraints (2) and (3) allocate OR slots to the different subspecialties. Parameter N^{D}_{jd} gives the number of OR slots available to subspecialty j on day d, while N_j states the number of OR slots available to subspecialty j in the cycle.

$$\sum_{j\in\mathcal{J}}\sum_{d\in\mathcal{D}}\left(\sum_{k\in\mathcal{K}_j\cap\mathcal{K}^F} n_{jkd} + \sum_{k\in\mathcal{K}_j} y_{jkd}\right) \leq M^{CYCLE} \tag{1}$$

$$\sum_{k\in\mathcal{K}_j\cap\mathcal{K}^F} n_{jkd} + \sum_{k\in\mathcal{K}_j} y_{jkd} \leq N^{D}_{jd} \qquad j\in\mathcal{J}, \quad d\in\mathcal{D} \tag{2}$$

$$\sum_{d\in\mathcal{D}}\left(\sum_{k\in\mathcal{K}_j\cap\mathcal{K}^F} n_{jkd} + \sum_{k\in\mathcal{K}_j} y_{jkd}\right) \leq N_j \qquad j\in\mathcal{J} \tag{3}$$

Scheduling according to the target level for the elective patient categories is ensured by constraints (4) and (5). The variable x_{ikd} represents the number of elective patients belonging to elective patient category i that are scheduled for OR k on day d, while v_i gives the number of patients belonging to a given elective patient category that are scheduled above the minimum level. The target level of electives belonging to patient category i is given by T_i, and this parameter imposes an upper limit on the number of elective patients belonging to a patient category that may be scheduled for surgery through the cycle. The parameters V_i represents the minimum share of patients belonging to patient category i that should be scheduled for surgery through the cycle.

$$\sum_{k\in\mathcal{K}}\sum_{d\in\mathcal{D}} x_{ikd} \leq T_i \qquad i\in\mathcal{I}^{EL} \tag{4}$$

$$\sum_{k \in \mathscr{K}} \sum_{d \in \mathscr{D}} x_{ikd} - v_i = \lceil T_i V_i \rceil \qquad i \in \mathscr{I}^{EL} \tag{5}$$

Parameter S_i represents the surgery duration of a patient belonging to patient category i, while B_{kd} is the slot length in OR k on day d. Constraints (6) ensure that the scheduled surgery duration at OR k does not exceed the time available for surgery at that OR on day d.

$$\sum_{i \in \mathscr{I}_j^J} S_i x_{ikd} \leq B_{kd} y_{jkd} \qquad j \in \mathscr{J}, \quad k \in \mathscr{K}_j, \quad d \in \mathscr{D} \tag{6}$$

The anaesthesia resources are modelled in constraints (7) and (8). Binary variable α_{kd}^A indicates whether OR k is covered by an anaesthesia resource on day d, while parameter M_d^A represents the number of ORs that may be covered by an anesthesia resource on day d. The parameter M_{kd}^{OR} represents the maximum number of slots that can be scheduled for OR k on day d. In constraints (7) the number of ORs covered by an anaesthesia resource on a given day is restricted by the total amount of anaesthesia resources available on that day. Constraints (8) ensure that no more slots, elective or flexible, may be scheduled for an OR on a day than the total number of slots available at the OR on that day. In order to schedule subspecialties to the OR, an anaesthesia resource has to cover the OR.

$$\sum_{k \in \mathscr{K}} \alpha_{kd}^A \leq M_d^A \qquad d \in \mathscr{D} \tag{7}$$

$$\sum_{j \in \mathscr{J}} (n_{jk'd} + y_{jkd}) \leq M_{kd}^{OR} \alpha_{kd}^A \qquad k' \in \mathscr{K}^F, \qquad k \in \mathscr{K}, \quad d \in \mathscr{D} \tag{8}$$

The first-stage bed constraints are given by (9) and (10).The number of staffed beds in ward w on day d is given by the variable a_{wd}, while parameters A_w^{MAX} and A_d represent the maximum number of beds available in ward w per day, and the maximum number of staffed beds available on day d, respectively. In constraints (9), we require that the number of staffed beds in a ward on a given day cannot exceed the maximum number of beds available in the ward. Constraints (10) ensure that we cannot staff more beds in total on a given day than the total number of staffed beds available on that day.

$$a_{wd} \leq A_w^{MAX} \qquad d \in \mathscr{D}, \quad w \in \mathscr{W} \tag{9}$$

$$\sum_{w \in \mathscr{W}} a_{wd} \leq A_d \qquad d \in \mathscr{D} \tag{10}$$

4.1.2. The second-stage constraints

To model the scheduling of emergency patients and cancellations of elective patients, constraints (11) to (14) are applied. Each constraint and variable belongs exclusively to a scenario $s \in \mathscr{S}$. Each scenario is defined by the stochastic parameters T_{is}^{SU} and U_{wds}^{EM}. T_{is}^{SU} represents the number of SU emergencies from a given patient category that should be scheduled in the cycle, and U_{wds}^{EM} represents the number of emergencies covering beds at the different wards every day. In Section 4.3, we describe how the random parameters are generated.

The variables e_{ijkds} and e^{EL}_{ijkds} are the number of emergency patients that belong to patient category i scheduled for subspecialty j in OR k on day d. The first variable represents the number of emergencies scheduled to flexible slots, while the latter is the number of SU emergencies scheduled to the elective slots. Furthermore, the variable x^C_{ikds} gives the number of elective cancellations of elective patient category i in OR k on day d. Constraints (11) ensure that all SU emergencies are scheduled for surgery during the cycle. Constraints (12) restrict the number of emergency patients scheduled for the flexible slots in an OR by the surgery duration of the patients. In constraints (13) the number of electives and SU emergencies scheduled for an OR on a given day are restricted by the slot time scheduled for subspecialties able to handle the patients in that OR on that day. Finally, constraints (14) state that we cannot cancel patients who are not scheduled.

$$\sum_{j\in\mathscr{J}}\sum_{d\in\mathscr{D}}\left(\sum_{k'\in\mathscr{K}_j\cap\mathscr{K}^F} e_{ijk'ds}+\sum_{k\in\mathscr{K}} e^{EL}_{ijkds}\right)=T^{SU}_{is} \qquad i\in\mathscr{I}^{SU},\quad s\in\mathscr{S} \tag{11}$$

$$\sum_{i\in\mathscr{I}^{UJ}_j\cup\mathscr{I}^{SUJ}_j} S_i e_{ijkds}\le B_{kd}n_{jkd} \qquad j\in\mathscr{J},\quad k\in\mathscr{K}_j\cap\mathscr{K}^F,\quad d\in\mathscr{D},\quad s\in\mathscr{S} \tag{12}$$

$$\sum_{i\in\mathscr{I}^J_j} S_i(x_{ikd}-x^C_{ikds})+\sum_{i\in\mathscr{I}^{SUJ}_j} S_i e^{EL}_{ijkds}\le B_{kd}y_{jkd} \qquad j\in\mathscr{J},\quad k\in\mathscr{K}_j,\quad d\in\mathscr{D},\quad s\in\mathscr{S} \tag{13}$$

$$x^C_{ikds}\le x_{ikd} \qquad i\in\mathscr{I}^{EL},\quad k\in\mathscr{K},\quad d\in\mathscr{D},\quad s\in\mathscr{S} \tag{14}$$

The second-stage bed constraints are given by (15) to (17). Variables u_{iwds} and u^{SU}_{iwds} represent the number of electives and SU-emergencies from patient category i that cover beds in ward w on day d. Variable $b_{ww'ds}$ represents the number of beds occupied in ward w' by patients belonging to ward w on day d in scenario s, and the variable β_{wds} represents the number of additional beds staffed in ward w on day d. The two first bed-constraints count the number of elective and SU emergency patients who rest in the different wards each day, while the last ones ensure that the total bed capacity is respected.

$$\sum_{k\in\mathscr{K}}\sum_{d'=1}^{E_{id}}(x_{ik(d-d'+1)}-x^C_{ik(d-d'+1)s})\le u_{iwds} \qquad w\in\mathscr{W},\quad i\in\mathscr{I}^W_w,\quad d\in\mathscr{D},\quad s\in\mathscr{S} \tag{15}$$

$$\sum_{j\in\mathscr{J}}\sum_{d'=1}^{E^{SU}_{id}}\left(\sum_{k'\in\mathscr{K}_j\cap\mathscr{K}^F} e_{ijk'(d-d'+1)s}+\sum_{k\in\mathscr{K}} e^{EL}_{ijk(d-d'+1)s}\right)\le u^{SU}_{iwds} \qquad w\in\mathscr{W},\quad i\in\mathscr{I}^{SUW}_w,\quad d\in\mathscr{D},\quad s\in\mathscr{S} \tag{16}$$

$$\sum_{i \in \mathscr{I}_w^W} u_{iwds} + \sum_{i \in \mathscr{I}_w^{SUW}} u_{iwds}^{SU} + \sum_{w' \in \mathscr{W} | w' \neq w} b_{w'wds} - \sum_{w' \in \mathscr{W} | w' \neq w} b_{ww'ds} \leq a_{wd} + \beta_{wds} - U_{wds}^{EM} \quad w \in \mathscr{W}, \quad d \in \mathscr{D}, \quad s \in \mathscr{S} \tag{17}$$

When implementing constraints (15) and (16), we link the last day of the cycle to the first day of the cycle. In the implementation, we handle non-positive values in the expression $(d - d' + 1)$ by adding the number of days in the cycle. For a weekly cycle (7 days), day 0 maps to day 7, day -1 maps to day 6, and so on. This way of modelling is possible because the MSS is cyclic, and it allows us to only model the days of the cycle, while at the same time making sure that the bed usage of the whole length of stay is accounted for.

4.1.3. The objective function

The objective function is given by (18). Parameter R_i^{EL} represents the gain obtained by scheduling more patients belonging to elective patient category i than the lower limit, while P_s gives the probability of ending up in scenario s. Furthermore, C_i^C represents the penalty of cancelling elective patients belonging to category i, while $C_{ww'}^W$ yields the penalty of assigning a patient belonging to ward w to ward w'. C^{SU} indicates the penalty of scheduling SU emergencies for elective ORs, and C^β is the penalty related to staffing more beds than scheduled. Finally, R^U is the growth obtained when scheduling U emergencies to the flexible slots. The objective function maximizes the gains from scheduling more elective patients than the lower limit in the first stage. In the second stage, we minimize the penalty of cancelling electives, providing beds for patients in wards not originally intended for them, scheduling SU patients to elective slots and staffing more beds than scheduled on the wards. In addition, we maximize the amount of urgent emergency patients scheduled for surgery in the second stage.

$$\begin{aligned} max \sum_{i \in \mathscr{I}^{EL}} R_i^{EL} v_i - \sum_{s \in \mathscr{S}} P_s \Bigg[& \sum_{i \in \mathscr{I}^{EL}} \sum_{k \in \mathscr{K}} \sum_{d \in \mathscr{D}} C_i^C x_{ikds}^C + \sum_{w \in \mathscr{W}} \sum_{w' \in \mathscr{W}} \sum_{d \in \mathscr{D}} C_{ww'}^W b_{ww'ds} + \\ \sum_{i \in \mathscr{I}^{SU}} \sum_{j \in \mathscr{J}} \sum_{k \in \mathscr{K}} \sum_{d \in \mathscr{D}} C^{SU} e_{ijkds}^{EL} & + \sum_{w \in \mathscr{W}} \sum_{d \in \mathscr{D}} C^\beta \beta_{wds} - \sum_{i \in \mathscr{I}^U} \sum_{j \in \mathscr{J}} \sum_{k \in \mathscr{K}} \sum_{d \in \mathscr{D}} R^U e_{ijkds}^U \Bigg] \end{aligned} \tag{18}$$

4.2. The discrete-event simulation model

The discrete-event simulation model encompasses all elective and emergency patient categories, and it performs the scheduling of these patients either to the elective ORs governed by the MSS or to the emergency ORs. The system is modelled as a queuing network where the emergency patients are the customers waiting in line, and the ORs and the wards are the servers. There are four queues of emergency patients: the CU, the U, the SU outpatient and the SU inpatient queue. The elective patients are also treated by the servers, but they are not waiting in line as they are scheduled for specific time slots and arrive just before surgery. Table 1 lists the activities, states and events.

Figure 6 illustrates the flow of patients in the model. The emergency patients arrive with an exponentially distributed inter-arrival time. The expected inter-arrival time, $1/\lambda_{i^{EM},\tau}$, is dependent both on the

Table 1: The activities, states and events of the system considered in the simulation model

Activities	States	Events
Preop. stay in wards for em. inpatients	No. of em. inpatients resting at each ward, preop.	Arrival of em. patients
Preop. stay at home for em. outpatients	No. of em. outpatients waiting at home, preop.	
Postop. stay at the wards for inpatients	No. of patients resting at each ward, postop.	Inpatient leaving the ward postop.
Transportation of em. patients from the ward to the OR		
The surgery (including cleaning of the OR)	Whether an OR is idle or busy	Completion of a surgery

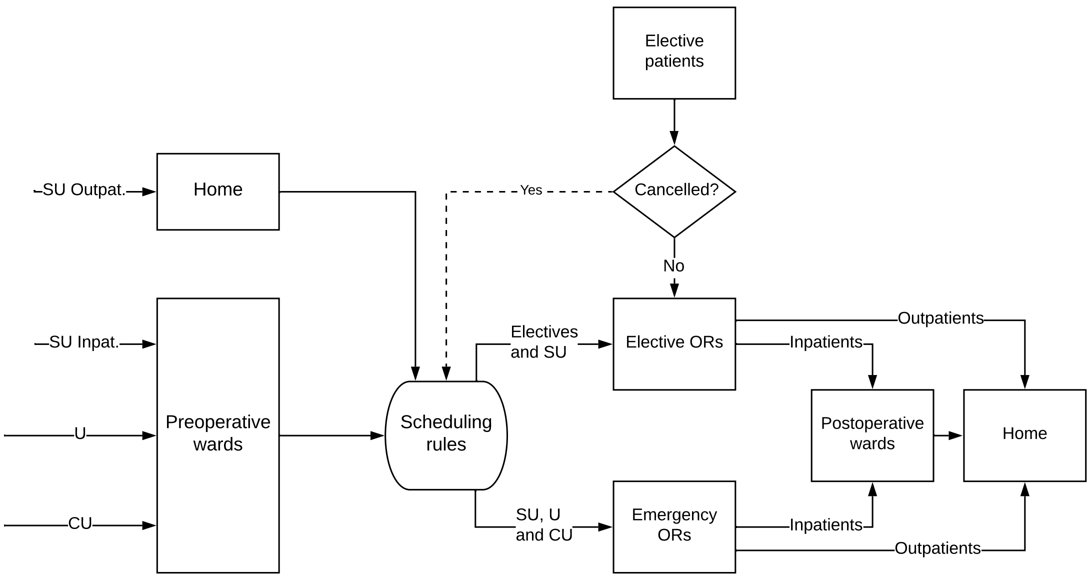

Figure 6: A flow chart describing the flow of patients in the simulation model.

emergency patient category, i^{EM}, and on the time, τ, of the day. Both the U, the CU and the SU inpatients rest in the preoperative wards while waiting in line for surgery. The SU outpatients are sent home to wait. If the number of SU emergencies waiting in line for the emergency ORs is above a threshold value, the SU patients are scheduled for surgery in a flexible surgery slot, preferably within the deadline for surgery. If a flexible surgery slot is not available within a given number of days, elective patients are cancelled in order to provide capacity for the SU emergency patients.

For the emergency ORs, scheduling rules are used to determine which patient should enter next. The first patient in line in the U and CU queues, and the first patient in the SU in- and outpatient queues who have not been scheduled for flexible slots are the first to be chosen for surgery. The scheduling rules applied to choose among the candidates may vary depending on the case department. For the elective ORs, the next patient is the next scheduled patient. This may be either an elective patient or a SU emergency patient. The elective patients that are cancelled are rescheduled to a flexible slot some days ahead. All elective patients are assumed to show up at the scheduled surgery time.

The surgery duration is random, with empirical probability distributions for each patient category. The scheduling of emergency patients is based on the expected surgery duration. We can not schedule for overtime, but overtime may occur as a result of the actual surgery duration. Following each surgery, the OR must be cleaned, implying that the room is unavailable for some time following surgery.

After surgery, the patients are either sent to the ward or directly home. The preoperative waiting time for emergency patients is calculated as the time from arrival to the system until surgery. The postoperative length of stay is calculated as the time interval from leaving the OR to leaving the postoperative ward. The postoperative length of stay is random, and drawn from empirical probability distributions for each patient category.

Unlike in the optimization model, the wards in the simulation model are assumed to have infinite capacity, implying that no rescheduling is done as a result of the wards being overloaded. The unlimited ward capacity is chosen to allow for scheduling of more beds in periods when many emergencies arrive. This is the normal protocol at many hospitals, as the number of physical beds exceeds the number of staffed beds. Scheduling more beds is not penalized in the simulation model, but rather in the optimization model. All emergency inpatients return to the same ward where they were resting at prior to surgery.

4.3. The scenario generation procedure

Each scenario contains data from one cycle in the simulation output. Recall that a cycle corresponds to the number of days considered in the optimization model. In each iteration of the simulation-optimization approach illustrated in Figure 4, a number of simulation replications, N^{REP}, are produced, and for each replication, a given number of separated cycles, N^{CYC}, are sampled as the scenarios. In total, this yields $N^{REP} \cdot N^{CYC}$ scenarios in each iteration. By letting each scenario consist of data gathered from a number of consecutive days from the simulation output, dependency is ensured between the days in each scenario. Furthermore, to keep the scenarios reasonably independent of each other, the selected cycles are separated by a number of cycles. All simulation replications start with an empty system, so to ensure that the scenarios

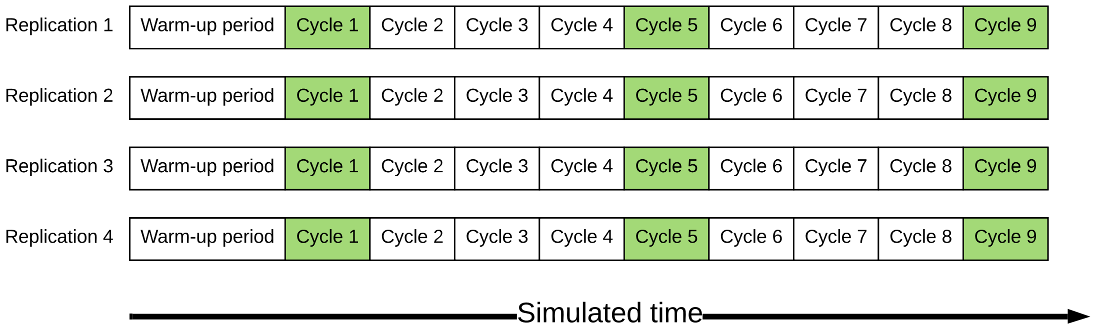

Figure 7: An example of how scenarios are chosen in each iteration of the simulation-optimization procedure. Here, the number of simulation replications, $N^{REP} = 4$, and the number of cycles chosen from each replication, $N^{CYC} = 3$, providing a total of 12 scenarios. Note that in each simulation replication there is a warm-up period, and that each cycle is separated by a number of cycles (3 in this example) to ensure reasonable independence between the cycles.

are not affected by this, a warm-up period is implemented in each replication. In Figure, 7 an example of the scenario generation procedure is illustrated. Here, $N^{REP} = 4$, and $N^{CYC} = 3$, providing a total of 12 scenarios.

Initially, each scenario contains the number of SU emergency patients that are scheduled for the elective ORs within the cycle and the number of emergency patients resting in each ward every day. Because the SU emergency patients are scheduled in the optimization model, they may be scheduled for other days compared to what the simulation output shows. Because of this, they may end up covering beds on different days as well. Therefore we subtract the postoperative LOS for the SU patients when generating the scenarios. However, we do not subtract the preoperative LOS for the SU patients when generating scenarios. The reason for this is that the optimization model does not explicitly account for the preoperative LOS. Hence, if we had subtracted the preoperative LOS, we would have underestimated the bed loading through the cycle when generating the scenarios.

Equations (19) describe how the number of SU emergency patients that require surgery at the elective ORs within the cycle is obtained. SU_{id} is the number of SU emergency patients from patient category i that entered on day d in the cycle and were scheduled for the elective ORs. The set D_s^S represents the days in the drawn cycle that is used to provide scenario s. Equations (20) show how we calculate the number of emergency patients resting in the different wards each day of the cycle. E_{wd} state the number of emergency patients resting in ward w on day d, while $E_{wd}^{SU_{post}}$ is the number of SU emergencies that rest in ward w on day d following their surgery.

$$T_{is}^{SU} = \sum_{d \in \mathscr{D}_s^S} SU_{id} \qquad i \in \mathscr{I}^{SU}, \quad s \in \mathscr{S} \tag{19}$$

$$U_{wds}^{EM} = E_{wd} - E_{wd}^{SU_{post}} \qquad w \in \mathscr{W}, \quad s \in \mathscr{S}, \quad d \in \mathscr{D}_s^S \tag{20}$$

4.4. *Stopping criterion*

The levels of detail in the optimization model and the simulation model differ in their construction. This presents a challenge in terms of the convergence and stability of the solution, as the difference in representation typically leads to different objective values, even for identical MSS. While we do not prove that the model presented is guaranteed to converge, in practice the solution stabilizes after a few iterations of the algorithm illustrated in Figure 4.

For the test cases and case study in this paper, we have chosen to use the number of flexible slots scheduled through the cycle as the stopping criterion. At a tactical decision level, determining the share of flexible slots is of high clinical interest for planners at the hospital. If additional stability is required by the user, adding inter-day stabilization would probably be the most useful requirement, but would take more computation time to reach a stable solution.

5. Computational study

In the computational study, the optimization model is first run for a set of test instances, providing results that can give both technical and managerial insights. Then we perform a case study, where we apply the simulation-optimization approach on a case department to develop an MSS that handles a fluctuating demand of emergency patients. In all instances, a cycle is set to be one week.

5.1. *Implementation and setup of study*

The optimization model is implemented in the Mosel language and is solved in Xpress 8.3. The simulation model is built in MATLAB. In each simulation-optimization iteration, 20 simulation replications are made ($N^{REP} = 20$), each representing a period of half a year (after half a year warm-up). From each simulation replication, five scenarios are selected ($N^{CYC} = 5$) yielding a total of 100 scenarios. All scenarios are replaced from one iteration to the next.

To ensure independence between the scenarios, we would choose $N^{REP} = 100$ and $N^{CYC} = 1$. However, because of the relatively long warm-up period we decided to use several cycles from each simulation replication to save computational time. The drawn cycles are separated by four weeks to ensure a reasonably degree of independency.

For the test instances, three levels of emergency patient loading are implemented: low, medium and high (EL, EM and EH, respectively). For each of the emergency loading cases, two sets of target throughput of electives - low and high (TL and TH, respectively) - are applied, and for each of these targets we test three bed capacities, resulting in 18 different instances. Because the number of patients is so variable in the instances, we apply four different bed capacities (W1, W2, W3 and W4) to provide three levels of bed capacity for each of the three emergency loading cases. Applying the lowest bed capacity (W1) to the high- and medium emergency (EH and EM) loading cases will yield an unrealistically low bed capacity, while the complete opposite will yield a very large bed capacity.

Note that for the experiments on these test instances, the focus is on the optimization model. The simulation model is only run once for each of the emergency loading levels to generate scenarios for the optimization model. Appendix B illustrates the values assigned to the input parameters of the optimization model in the test instances and the case department instance.

5.2. The value of the stochastic solution

We will use the value of the stochastic solution (VSS), as described by Birge [3], to evaluate our model. The VSS measures the value of applying a stochastic, rather than a deterministic, model to solve the problem at hand. It can be interpreted as valuing the flexibility that the stochastic solution provides that is not present in the deterministic solution. The VSS is calculated by first constructing the mean value problem (MVP). In the MVP, the stochastic parameters take their expected value, and the stochastic model is solved deterministically. The values obtained for the first-stage variables when solving the MVP are applied as input parameters for the stochastic model, and we solve the second-stage problems (one problem for each scenario) of the stochastic model. This leaves us with an optimal objective function value referred to as the mean value solution (MVS). The VSS is calculated as the difference between the stochastic solution (SS) obtained from solving the stochastic model (equations (1)-(18)) and the MVS.

Neither the MVPs nor the stochastic problems are solved to optimality within three hours, so the true VSS may not be calculated. However, the LP gaps are small, and we can provide quite narrow intervals for the VSS. The upper limit is calculated as the difference between the objective value of the upper bound obtained from solving the stochastic model and the MVS. The lower limit is calculated as the difference between the objective value of the IP-solution from solving the stochastic model and the MVS.

When solving the MVP, the values of the first-stage variables (that are fed to the MVS) are the same regardless of the costs associated with the random parameters. Therefore, by increasing these costs, the VSS increases. Furthermore, for this specific problem, the number of flexible slots increases as the costs associated with the random parameters increase. Scheduling flexible slots can be regarded as investing in an insurance against periods of high emergency loading.

In Table 2 we present the number of flexible slots scheduled in both the SS and the MVS, when the cost of cancelling a patient is set to be the same as the revenue from scheduling the patient. As we see, the SSs provide fewer flexible slots than the MVS. This is counterintuitive. The reason for this is that the stochastic model also accommodates scenarios where few emergency patients arrive; the need for flexible slots in these scenarios is less, and the capacity can be utilized to treat more elective patients if they are scheduled in the first stage. As a consequence, more elective patients are scheduled and more electives receive surgery (despite more cancellations in general) in the stochastic solutions.

Table 3 includes results from solving both the stochastic model and its deterministic counterpart, when the cost of cancelling elective patients is set to be higher than the revenue from scheduling the same patients. In addition to showing the SS, the MVS and the VSS, the table includes the number of flexible slots scheduled, the mean number of elective cancellations and the mean number of electives treated in solving both the stochastic and the deterministic model for all the instances.

Table 2: The number of flexible slots scheduled in the SS and the MVS when the cost of cancelling an elective patient is the same as the revenue from scheduling the same patient.

Instance	Flexible slots (SS)	Flexible slots (MVS)
EL-TL-W1	9	9
EL-TL-W2	9	9
EL-TL-W3	9	9
EL-TH-W1	2	2
EL-TH-W2	1	3
EL-TH-W3	0	3
MN-TL-W2	9	9
MN-TL-W3	9	9
MN-TL-W4	9	9
MN-TH-W2	2	5
MN-TH-W3	3	4
MN-TH-W4	2	5
EH-TL-W2	10	10
EH-TL-W3	9	9
EH-TL-W4	10	9
EH-TH-W2	5	8
EH-TH-W3	8	9
EH-TH-W4	7	9

Comparing the instances, the VSS is often higher when the bed capacity is scarce. This has to do with the fact that we penalize rescheduling of beds in the second stage. Proper scheduling of beds in the first stage can decrease the number of elective patients resting in wards not intended for them in the second stage. The benefits of this scheduling are typically higher when the bed capacity is scarce.

5.3. *Managerial insight*

We have destilled some insights from the test instances. Table 4 provides the results from running the 18 instances for two levels of cost related to cancelling elective patients. For the low cost cases, the cost of cancelling an elective patient is the same as the revenue from scheduling the same patient. The table includes the number of flexible slots scheduled, the number of electives scheduled for surgery, the mean share of SU surgeries performed in flexible slots, the mean number of electives cancelled, the mean number of electives treated and the mean number of elective inpatients resting in inconvenient wards.

Figures 8 and 9 illustrate how the number of flexible slots scheduled in the first stage depends on the available OR capacity and the bed capacity, respectively. The number of flexible slots increases when the OR capacity is high - in other words, when the target level of elective patients is low. The reason for this is that more OR capacity can be made flexible without having to sacrifice the scheduling of elective patients, and that we assume there is always an U patient who will fill the idle flexible OR capacity.

When the cost of cancelling elective patients is high, the number of flexible slots decreases as the bed capacity increases. There are two reasons for this. First, the flexible slots may be utilized for either SU outpatients or U patients if bed capacity is scarce, while the SU inpatients may be moved to elective slots where they receive surgery without exceeding the bed capacity, providing a kind of option towards scarce bed capacity. Secondly, scheduling flexible slots reduces the number of electives scheduled, implying less demand for beds.

Table 3: The value of the stochastic solution. For both the stochastic solution (SS) and the mean value solution (MVS) we provide the number of flexible slots scheduled, the mean number of cancellations and the mean number of electives treated. Because neither the MVPs or the stochastic problems were solved to optimality, intervals are given for the SS and the VSS.

	Flexible slots		Cancellations		Electives treated		Objective function value		
Instance	SS	MVS	SS	MVS	SS	MVS	SS	MVS	VSS
EL-TL-W1	14	9	2.71	6.20	80.29	84.80	[-9.63, -9.04]	-25.74	[16.11, 16.70]
EL-TL-W2	9	9	0.20	1.55	90.80	89.45	[113.43, 114.96]	100.05	[13.38, 14.91]
EL-TL-W3	9	9	0	0	91.00	91.00	[115.51, 117.14]	115.26	[0.25, 1.88]
EL-TH-W1	8	2	3.29	9.03	93.71	100.97	[-21.35, -19.08]	-46.40	[25.05, 27.32]
EL-TH-W2	4	3	1.22	2.74	106.78	106.26	[107.90, 113.64]	96.24	[11.66, 17.40]
EL-TH-W3	3	3	0.82	0.80	108.18	108.20	[115.26, 120.30]	115.22	[0.04, 5.08]
EM-TL-W2	12	9	3.58	5.98	84.42	85.02	[-278.02, -276.04]	-296.47	[18.45, 20.43]
EM-TL-W3	9	9	0.21	0.57	90.79	90.43	[-19.25, -15.87]	-22.99	[3.74, 7.12]
EM-TL-W4	10	9	0.25	0.27	89.75	90.73	[108.62, 114.22]	109.99	[0, 4.23]
EM-TH-W2	7	5	4.39	7.68	95.61	98.32	[-297.10, -289.21]	-318.15	[21.05, 28.94]
EM-TH-W3	7	4	1.25	2.75	99.75	104.25	[-36.28, -27.83]	-42.76	[6.48, 14.93]
EM-TH-W4	7	5	1.23	2.20	99.77	103.80	[93.76, 102.45]	92.02	[1.74, 10.43]
EH-TL-W2	16	10	6.01	11.86	70.99	78.14	[-6732.91, -6731.50]	-6760.24	[27.33, 28.74]
EH-TL-W3	9	9	1.03	2.91	88.97	88.09	[-1764.45, -1757.14]	-1759.48	[0, 2.34]
EH-TL-W4	11	9	0.33	2.85	87.67	88.15	[-700.23, -691.45]	-719.14	[18.91, 27.69]
EH-TH-W2	12	8	6.65	12.85	79.35	85.15	[-6758.48, -6753.5]	-6786.54	[28.06, 33.04]
EH-TH-W3	11	9	1.20	4.83	89.80	91.17	[-1785.64, -1779.42]	-1813.86	[28.22, 34.44]
EH-TH-W4	11	9	0.62	2.32	90.38	93.68	[-719.23, -712.14]	-728.99	[9.76, 16.85]

Table 4: Output from solving the stochastic problem for the 18 instances. Low indicates that the cost of cancelling an elective patient is the same as the revenue from scheduling the same patient, while high represents a higher cost of cancelling electives. We include the number of flexible slots and the number of elective patients scheduled in the first stage. For the second stage we include the mean number of SU patients scheduled for flexible slots, the mean number of elective cancellations, the mean number of electives treated and the mean number of patients resting in wards not intended for them.

	Flexible slots		Electives scheduled		SU pat. in flex.		Electives cancelled		Tot. el. treated		Pat. moved	
Instance	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
EL-TL-W1	9	14	91/91	83/91	2.36/2.40	2.40/2.40	5.08	2.71	85.92	80.29	5.76	4.17
EL-TL-W2	9	9	91/91	91/91	2.40/2.40	2.40/2.40	0.27	0.20	90.73	90.80	1.41	1.18
EL-TL-W3	9	9	91/91	91/91	2.40/2.40	2.40/2.40	0	0	91.00	91.00	0	0
EL-TH-W1	2	8	111/115	97/115	1.20/2.40	2.31/2.40	9.89	9.03	101.11	93.71	6.31	4.20
EL-TH-W2	1	4	112/115	108/115	0.60/2.40	1.90/2.40	3.18	2.74	108.82	106.78	2.73	1.93
EL-TH-W3	0	3	114/115	109/115	0/2.40	1.65/2.40	2.62	0.80	111.38	108.18	0.05	0.01
EM-TL-W2	9	12	91/91	88/91	4.80/5.24	5.20/5.24	4.52	3.58	86.48	84.42	6.05	5.42
EM-TL-W3	9	9	91/91	91/91	4.83/5.24	4.78/5.24	0.36	0.21	90.64	90.79	0.49	0.52
EM-TL-W4	9	10	91/91	90/91	4.82/5.24	4.99/5.24	0.47	0.25	90.53	89.75	0.21	0.20
EM-TH-W2	2	7	111/115	101/115	1.83/5.24	4.34/5.24	9.32	4.39	101.68	95.61	10.15	5.24
EM-TH-W3	3	7	109/115	101/115	2.40/5.24	4.37/5.24	3.92	1.25	105.08	99.75	0.43	0.44
EM-TH-W4	2	7	111/115	101/115	1.82/5.24	4.35/5.24	4.68	1.23	106.32	99.77	0.20	0.29
EH-TL-W2	10	16	90/91	77/91	8.86/9.88	9.88/9.88	11.65	6.01	78.35	70.99	9.37	7.39
EH-TL-W3	9	9	91/91	90/91	8.47/9.88	8.50/9.88	1.48	1.03	89.52	88.97	2.55	2.22
EH-TL-W4	10	11	89/91	88/91	9.02/9.88	9.42/9.88	1.01	0.33	87.99	87.67	0.57	0.62
EH-TH-W2	5	12	104/115	85/115	5.15/9.88	9.55/9.88	16.05	6.65	87.95	79.35	9.29	7.17
EH-TH-W3	8	11	97/115	101/115	7.60/9.88	9.41/9.88	3.63	1.20	93.37	89.80	2.12	2.78
EH-TH-W4	7	11	101/115	91/115	7.12/9.88	9.38/9.88	3.98	0.62	97.02	90.38	0.70	0.59

We do not find the same pattern for the low cost instances. In Table 4, we observe that going from instance EM-TH-W2 to EM-TH-W3 and from EH-TH-W2 to EH-TH-W3, the number of scheduled flexible slots increases. In the instances EM-TH-W2 and EH-TH-W2, bed capacity is very scarce, resulting in many elective cancellations. Cancelling elective inpatients because of the scarce bed capacity provides ample

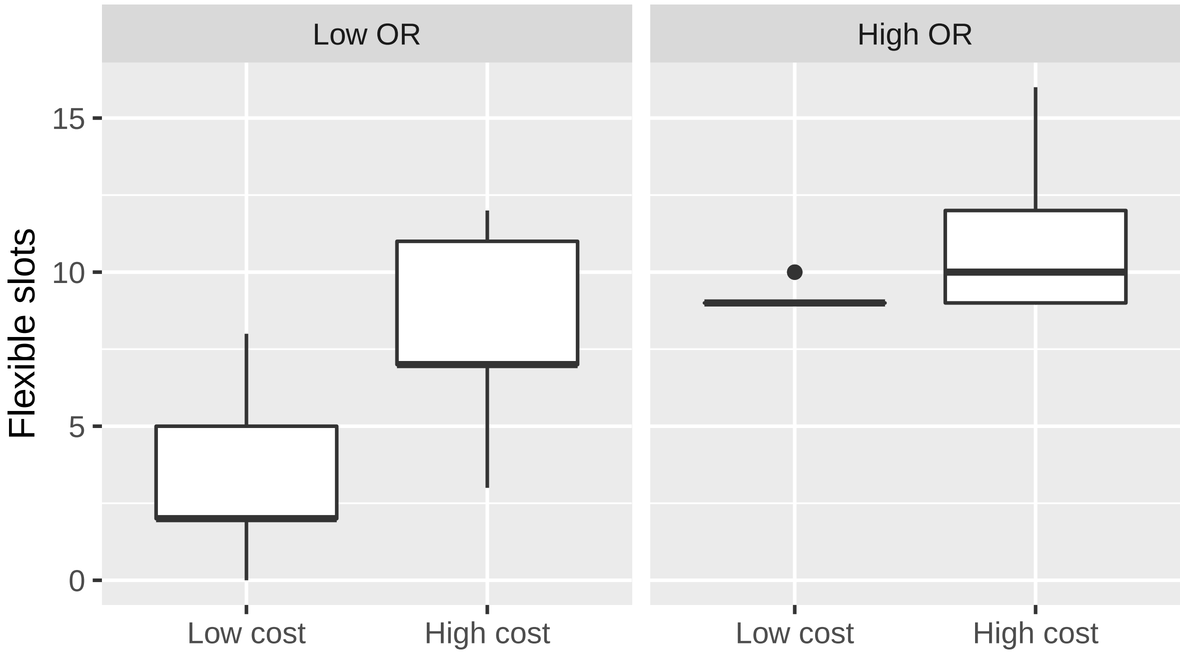

Figure 8: How the number of flexible slots scheduled in the first stage depends on the OR capacity when the cost of cancelling elective patients is either low or high.

spare elective OR capacity. This idle OR capacity may be used to schedule SU patients (if they are less demanding in terms of beds than the elective patients that were cancelled), decreasing the need for flexible slots. This happens because the penalty for cancelling elective patients is very low in these instances. For three of the low-target instances, the number of flexible slots increases as we go from the medium to the high bed capacities. This results from the high penalty associated with staffing more beds than scheduled to handle the peaks of emergency bed loading. The chances for avoiding additional staffing are greater for the instances with high bed capacity, and may be avoided if we schedule fewer electives in the first stage. Decreasing the penalty of staffing more beds than scheduled yields solutions that are more aggressive in terms of elective bed loading for the instances with high bed capacity.

Figure 10 illustrates the mean share of flexible slots scheduled for the three levels of emergency loading. As expected, the number of scheduled flexible slots increases as emergency loading increases. For the low emergency loading instances, the share of emergency patients is 2.6% for the TL instances and 2.0% for the TH instances. These numbers are 5.4% and 4.4%, and 9.8% and 7.9% for the medium and high emergency loading instances, respectively. For the TH-low cost instances, the mean share of flexible slots is quite similar to the share of emergency patients, implying that the emergency patients receive OR capacity according to the relative size of the group. For the TL-low cost instances, there is less competition for the OR capacity, and the emergency patients receive excessive OR capacity. For the TH-high cost instances, the emergency patients receive a much higher share of the OR capacity compared to the TH-low cost instances,

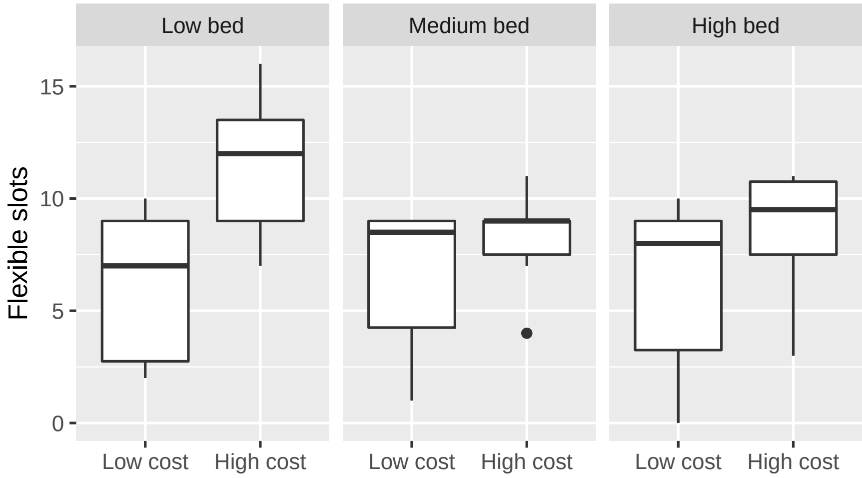

Figure 9: How the number of flexible slots scheduled in the first stage depends on the bed capacity when the cost of cancelling elective patients is either low or high.

indicating that they gain power in the competition for OR capacity. For the TL instances, the competition from electives is less, and therefore the difference is less between the two cost regimes. In general, if the penalty for cancelling a patient is at least the same as the revenue from scheduling the patient, the share of flexible slots scheduled should be higher than the share of emergency patients.

5.4. Case study

For the case study we consider the Master Surgery Scheduling Problem (MSSP) in the orthopaedic department at St. Olav's Hospital. St. Olav's Hospital is a university hospital located in Norway, and it is a relatively large hospital with approximately 1000 beds. The orthopaedic department performs approximately 7000 surgeries every year, and roughly 3000 of these are emergency surgeries. To perform these surgeries, the department has access to 7 elective and 3 emergency ORs. The case department faces many of the issues discussed in this paper.

In this section we apply the simulation-optimization approach to the case department. The aim is to provide an MSS that enables the department to handle fluctuating emergency patient demand and at the same time provide a sufficient throughput of elective patients. First, the scheduling rules applied in the simulation model to mimic the case department is presented. Secondly, the new MSS is introduced, and finally we compare outcomes from running both the new MSS and the MSS present at the case department today in the simulation model.

5.4.1. Scheduling rules and the flow of patients in the simulation model

CU patients will always have priority over the U and SU patients. U patients have priority over SU patients except when the next SU patient has exceeded the time limit more than the next U patient. Only

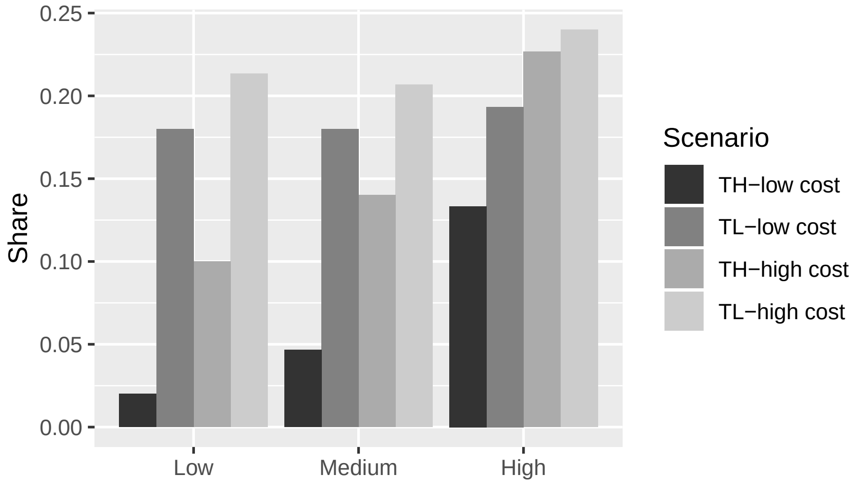

Figure 10: How the number of flexible slots scheduled in the first stage depends on the emergency loading when the cost of cancelling elective patients is either low or high.

CU patients can have surgery during the night, and only the U and CU patients are admitted to surgery on weekends.

All U and CU patients may be summoned for surgery immediately after arrival, while SU patients are scheduled in the morning after arrival. The SU inpatients are primarily lined up for the emergency ORs. If the number of SU inpatients waiting in the queue is above a threshold limit - in this case two - we start scheduling these patients for the flexible slots in the elective ORs. The SU outpatients have to return to the hospital on the day of surgery, so we want to make sure that they do not have their surgery postponed since this would involve too much traveling for the patients. Therefore, the SU outpatients are not lined uo for the emergency ORs, but are scheduled directly into the flexible slots. If no flexible slots are available within a given number of days - in this case eight - elective surgeries must be cancelled to provide capacity for the SU in- and outpatients.

All displaced elective patients will be rescheduled to a flexible slot some days ahead. Rescheduling a patient just before surgery is not preferable, so we introduce a limit on the number of days prior to surgery that rescheduling is not allowed.

5.4.2. Iterative outcomes and the optimization-based MSS

To create an MSS for the case department, we populate the optimization model with the data described in Appendix B. On request from the case department, no flexible slots may be scheduled for two of the elective ORs (OR-6 and OR-7), and no elective inpatients should be scheduled on Friday. To initialize the simulation-optimization approach, we apply the MSS present at the orthopaedic department today when running the simulation model in the first iteration. Because no flexible slots are available in the present MSS, all emergency patients are sent to the emergency ORs in the first iteration. However, if the SU emergencies do not receive surgery within a set time, they are rescheduled to the elective ORs, and there will be elective cancellations.

Tables 5 and 6 present the main outcomes from the optimization model and the simulation model, respectively. The optimality gap in Table 5 is calculated as the difference between the objective function of the best integer solution and the upper bound, divided by the upper bound. The MSS generated in the last iteration is illustrated in Appendix D.

Table 5: Outcomes from each iteration with the optimization model when generating the simulation-optimization-based MSS

	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5	Iter. 6	Iter. 7	Iter. 8
Obj. func. val.	-2566.701	-75.346	-140.293	-241.623	-44.047	-33.668	-1137.51	58.55
Upper bound	-2563.714	-71.017	-134.99	-230.729	-40.951	-29.512	-1132.74	61.21
Optimality gap	0.12%	5.75%	3.77%	4.51%	7.03%	12.34%	0.42%	4.35%
Flex. slots	12	12	11	14	12	12	12	12
El sched.	71/80	74/80	74/80	70/80	74/80	74/80	74/80	74/80
SU in flex. slots (mean)	4.41/4.53	6.47/6.70	6.29/6.58	6.66/6.67	7.26/ 7.60	5.79/6.04	5.91/6.01	7.04/7.16
El. canc. (mean)	6.95	2.35	3.63	1.94	2.02	2.18	3.45	1.58
U to receive surg. (mean)	11.22	8.03	6.48	7.94	7.57	9.53	8.76	8.24

Table 6: Outcomes from each iteration with the simulation model when generating the simulation-optimization-based MSS

	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5	Iter. 6	Iter. 7	Iter. 8
Cancellations per week	4.16	0	0	0	0	0	0	0
Mean waiting time (CU)	3.9 h	3.7 h	3.5 h	3.8 h	3.7 h	3.8 h	3.8 h	3.7 h
Mean waiting time (U)	31.3 h	20.9 h	20.5 h	20.7 h	18.0 h	20.7 h	21.0 h	18.8 h
Mean waiting time (SU)	3.81 d	1.80 d	1.80 d	1.83 d	1.73 d	1.86 d	1.87 h	1.78 d

The number of SU emergencies treated in flexible slots and the number of U emergencies treated reveal information about the flexible slot capacity. Almost all SU emergencies are treated in the flexible slots, and there is ample flexible capacity to treat U emergencies when no more SU emergencies are present. This indicates that the flexible slot capacity is good, and that most of the elective cancellations are due to the shortage of beds.

The simulated results include both the number of elective surgeries that are cancelled each week and the waiting time to receive surgery for the emergency patients. Scheduling flexible slots dramatically decreases the waiting time for both U and SU patients from iteration one to the subsequent iterations. Note that since the bed capacity is treated as unlimited in the simulation model, no electives are cancelled due to the shortage of beds. After the first iteration no electives are cancelled, indicating that the flexible OR capacity is sufficient.

5.4.3. Analysing the performance of the optimization-based MSS

Figure 11 illustrates the waiting time to receive surgery for U and SU emergencies when running both the MSS present in the case department today and the optimized MSS 20 times in the simulation model. The optimization-based MSS performs better on average, and the waiting times are less affected by fluctuations in emergency demand for surgery. Figure 12 illustrates the queue of SU emergencies at 08.00 through the simulated period when running both the MSS present in the case department today and the optimized MSS once in the simulation model. Note that the minimum number of SU patients waiting in queue for surgery

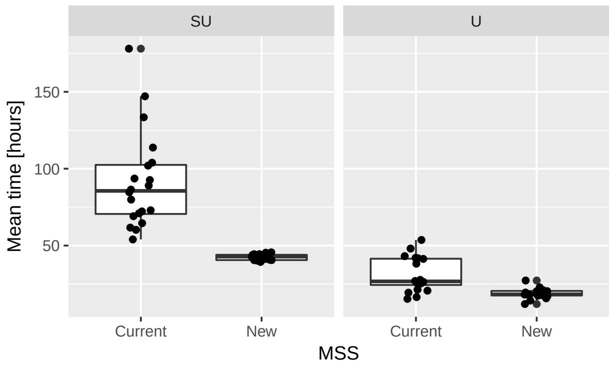

Figure 11: The simulation mean waiting time in hours for SU (left) and U (right) emergencies for the current and new optimization-based MSS, results from 20 simulations.

in the MSS present today is almost as large as the maximum number of SU patients waiting for surgery in the optimized MSS. Figure 13 shows the number of SU patients treated in both the flexible and elective slots each week from running the two MSSs. For the MSS present today, all SU emergencies are treated in elective slots (as no flexible slots are available), while for the optimized MSS all SU emergencies are treated in flexible slots. As a consequence, there are no elective interruptions when applying the new MSS.

6. Conclusion

The main purpose of this paper is to present a simulation-optimization approach for developing an MSS and to provide tactical decision support for the management in a department with both elective and emergency patients. Flexible slots are scheduled to the elective ORs to handle the fluctuating demand for emergency surgeries.

A two-stage stochastic optimization model is presented, where uncertainty related to emergency arrivals is included. Furthermore, a discrete-event simulation model is developed to include aspects of uncertainty related to the length of stay of patients following surgery and the surgery duration. The simulation model allows us to evaluate the MSS produced by the optimization model. Also, the simulation model provides scenarios for the optimization model, allowing the model to adapt to different MSSs and scheduling regimes.

The stochastic model outperforms its deterministic counterpart in terms of the Value of Stochastic Solution (VSS), and for realistic cancellation costs the number of flexible slots is higher in the stochastic solution. Furthermore, if the OR capacity is sufficient, or the ward capacity is scarce, a relatively large share of the ORs should be scheduled as flexible.

The simulation-optimization approach is applied to an orthopaedic department at a Norwegian hospital that treats both elective and emergency patients. With the optimized MSS, the emergency waiting time for surgery decreases, and it proves to be able to handle fluctuating emergency surgery demand with less interruptions to the flow of elective patients.

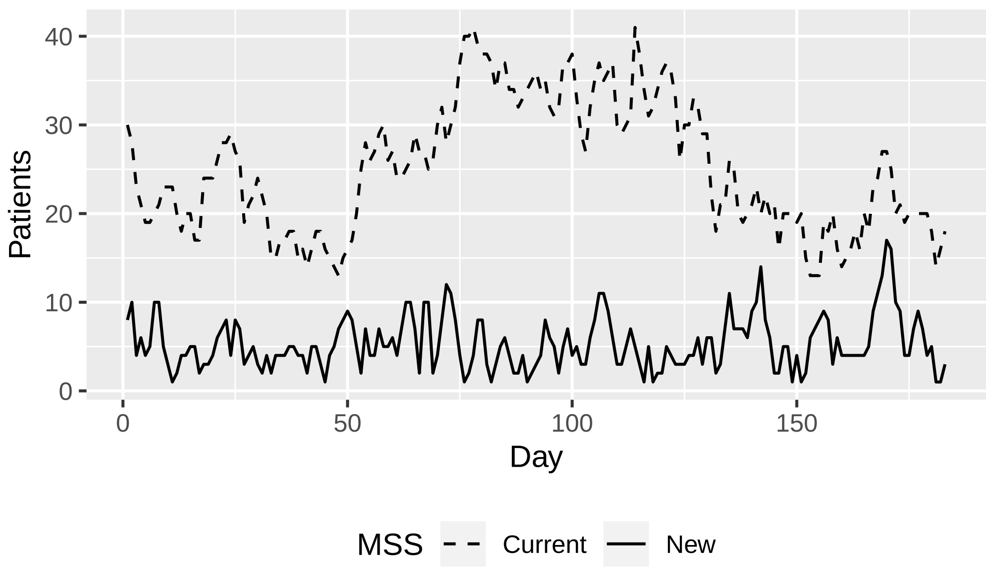

Figure 12: The graph illustrates the queue of SU emergencies as 08.00 through the simulated period for both the optimization-based MSS and the MSS present at the case department today.

7. Acknowledgements

We would like to thank the management in the Orthopaedic Department at St. Olav's Hospital, and in particular Head of Department, Vigleik Jessen, MD for good collaboration throughout the project. We are grateful for the comments and suggestions made by three anonymous reviewers, which helped to improve the paper.

References

[1] Adan, I., Bekkers, J., Dellaert, N., Jeunet, J., Vissers, J., 2011. Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. European Journal of Operational Research 213 (1), 290 – 308.

[2] Beliën, J., Demeulemeester, E., Cardoen, B., 2008. A decision support system for cyclic master surgery scheduling with multiple objectives. Journal of Scheduling 12 (2), 147.

[3] Birge, J. R., 1982. The value of the stochastic solution in stochastic linear programs with fixed recourse. Mathematical Programming 24 (1), 314–325.

[4] Cappanera, P., Visintin, F., Banditori, C., 2014. Comparing resource balancing criteria in master surgical scheduling: A combined optimisation-simulation approach. International Journal of Production Economics 158, 179 – 196.

[5] Cardoen, B., Demeulemeester, E., Beliën, J., 2010. Operating room planning and scheduling: A literature review. European Journal of Operational Research 201 (3), 921 – 932.

[6] Essen, J. V., Hans, E., Hurink, J., Oversberg, A., 2012. Minimizing the waiting time for emergency surgery. Operations Research for Health Care 1 (2), 34 – 44.

[7] Ferrand, Y. B., Magazine, M. J., Rao, U. S., 2014. Managing operating room efficiency and responsiveness for emergency and elective surgeries—a literature survey. IIE Transactions on Healthcare Systems Engineering 4 (1), 49–64.

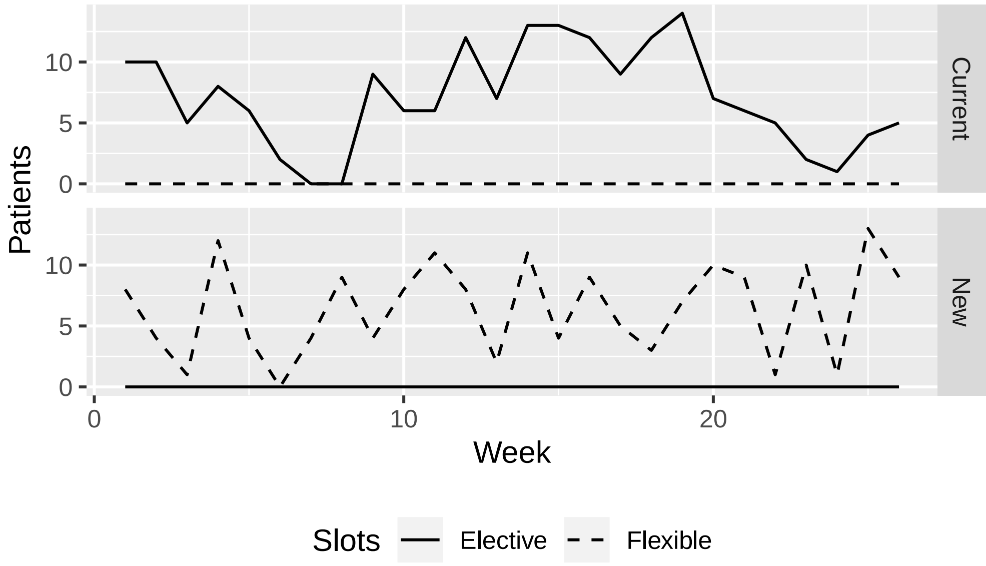

Figure 13: The two graphs illustrate the weekly number of SU emergencies treated in flexible and elective slots. The upper graph illustrates the MSS present today, and the lower one illustrates the optimization-based MSS.

[8] Figueira, G., Almada-Lobo, B., 2014. Hybrid simulation–optimization methods: A taxonomy and discussion. Simulation Modelling Practice and Theory 46, 118 – 134, simulation-Optimization of Complex Systems: Methods and Applications.

[9] Freeman, N., Zhao, M., Melouk, S., 2018. An iterative approach for case mix planning under uncertainty. Omega 76, 160 – 173.

[10] Fügener, A., Hans, E. W., Kolisch, R., Kortbeek, N., Vanberkel, P. T., 2014. Master surgery scheduling with consideration of multiple downstream units. European Journal of Operational Research 239 (1), 227 – 236.

[11] Hans, E. W., Houdenhoven, M. V., Hulshof, P. J. H., 2012. A Framework for Healthcare Planning and Control. Springer US, Boston, MA, pp. 303–320.

[12] Higle, J. L., 2005. Stochastic Programming: Optimization When Uncertainty Matters. Ch. Chapter 2, pp. 30–53.

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

[14] Kall, P., Wallace, S. W., 1994. Stochastic Programming. John Wiley Sons.

[15] Koppka, L., Wiesche, L., Schacht, M., Werners, B., 2018. Optimal distribution of operating hours over operating rooms using probabilities. European Journal of Operational Research 267 (3), 1156 – 1171.

[16] Kumar, A., Costa, A. M., Fackrell, M., Taylor, P. G., 2018. A sequential stochastic mixed integer programming model for tactical master surgery scheduling. European Journal of Operational Research 270 (2), 734 – 746.

[17] Lamiri, M., Xie, X., Dolgui, A., Grimaud, F., 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. European Journal of Operational Research 185 (3), 1026 – 1037.

[18] Li, X., Rafaliya, N., Baki, M. F., Chaouch, B. A., 2017. Scheduling elective surgeries: the tradeoff among bed capacity, waiting patients and operating room utilization using goal programming. Health Care Management Science 20 (1), 33–54.

[19] Ma, G., Demeulemeester, E., 2013. A multilevel integrative approach to hospital case mix and capacity planning. Computers Operations Research 40 (9), 2198 – 2207, operations research for health care delivery.

[20] Mannino, C., Nilssen, E. J., Nordlander, T. E., 2012. A pattern based, robust approach to cyclic master surgery scheduling. Journal of Scheduling 15 (5), 553–563.

[21] Oostrum, M. J. V., Houdenhoven, M. V., Hurink, J. L., Hans, E. W., Wullink, G., Kazemier, G., 2008. A master surgical

scheduling approach for cyclic scheduling in operating room departments. OR Spectrum 30 (2), 355–374.

[22] Penn, M. L., Potts, C. N., Harper, P. R., 2017. Multiple criteria mixed-integer programming for incorporating multiple factors into the development of master operating theatre timetables. European Journal of Operational Research 262 (1), 194 – 206.

[23] Razmi, J., Barati, M., Yousefi, M. S., Heydari, J., Jun 2015. A stochastic model for operating room planning under uncertainty and equipment capacity constraints. Journal of Industrial Engineering International 11 (2), 269–279.

[24] Testi, A., Tànfani, E., 2008. Tactical and operational decisions for operating room planning: Efficiency and welfare implications. Health Care Management Science 12 (4), 363.

[25] Testi, A., Tanfani, E., Torre, G., 2007. A three-phase approach for operating theatre schedules. Health Care Management Science 10 (2), 163–172.

Appendix A. The mathematical model

In Tables A.7, A.8, and A.9 all the notation used in the mathematical formulation is presented.

Appendix B. Input parameters for the optimization model

Here, we provide the values assigned to the input parameters in the optimization model for both the test instances used in Sections 5.2 and 5.3, and the large instance uses in Section 5.4. In Tables B.10 to B.12, and Tables B.16 and B.17, we present the values applied for the input parameters in the large instance. The number of ORs and wards in the large instance are similar to the case department, and the elective patient categories represent the main patient categories treated at the department. In Tables B.13 to B.15, and Tables B.16 and B.17, we provide the values applied for the test instances.

Appendix C. Input for the simulation model

In Tables C.18 and C.19 we provide the input values used in the simulation model when running the model in Section 5.4. In Table C.20 we provide the probability distributions that are used to model the stochastic processes in the simulation model. These distributions are based on historical data obtained from the case department.

Appendix D. The case study

Table D.21 provides the optimized MSS.

Table A.7: Sets used in the mathematical formulation

Set	Description	Indices
$\mathscr{D}$	Set of days in a cycle	$d \in \mathscr{D}$
$\mathscr{I}$	Set of patient categories	$i \in \mathscr{I}$
$\mathscr{J}$	Set of surgical subspecialties	$j \in \mathscr{J}$
$\mathscr{K}$	Set of ORs	$k \in \mathscr{K}$
$\mathscr{K}^F$	Set of ORs that are available for scheduling of flexible slots	$k \in \mathscr{K}^F \subseteq \mathscr{K}$
$\mathscr{W}$	Set of wards	$w \in \mathscr{W}$
$\mathscr{S}$	Set of scenarios	$s \in \mathscr{S}$
$\mathscr{I}^{EL}$	Set of elective patient categories	$i \in \mathscr{I}^{EL} \subseteq \mathscr{I}$
$\mathscr{I}^{IN}$	Set of elective patient categories that are inpatients	$i \in \mathscr{I}^{IN} \subseteq \mathscr{I}^{EL}$
$\mathscr{I}_j^J$	Set of elective patient categories that can be treated by subspecialty j	$i \in \mathscr{I}_j^J \subseteq \mathscr{I}^{EL}$
$\mathscr{I}_k^K$	Set of elective patient categories that can be scheduled to OR k	$i \in \mathscr{I}_k^K \subseteq \mathscr{I}^{EL}$
$\mathscr{I}_w^W$	Set of elective patient categories meant for ward w	$i \in \mathscr{I}_w^W \subseteq \mathscr{I}^{IN}$
$\mathscr{I}^{EM}$	Set of emergency patient categories	$i \in \mathscr{I}^{EM} \subseteq \mathscr{I}$
$\mathscr{I}^{SU}$	Set of emergency patient categories that belong to the SU emergency scheduling regime	$i \in \mathscr{I}^{SU} \subseteq \mathscr{I}^{EM}$
$\mathscr{I}^{U}$	Set of emergency patient categories that belong to the U emergency scheduling regime	$i \in \mathscr{I}^{U} \subseteq \mathscr{I}^{EM}$
$\mathscr{I}_j^{UJ}$	Set of U emergency patient categories that can be treated by subspecialty j	$i \in \mathscr{I}_j^{UJ} \subseteq \mathscr{I}^{U}$
$\mathscr{I}^{SUIN}$	Set of emergency patient categories that belong to the SU emergency scheduling regime that are inpatients	$i \in \mathscr{I}^{SUIN} \subseteq \mathscr{I}^{SU}$
$\mathscr{I}_j^{SUJ}$	Set of SU emergency patient categories that can be treated by subspecialty j	$i \in \mathscr{I}_j^{SUJ} \subseteq \mathscr{I}^{SU}$
$\mathscr{I}_w^{SUW}$	Set of SU emergency patient categories meant for ward w	$i \in \mathscr{I}_w^{SUW} \subseteq \mathscr{I}^{SUIN}$
$\mathscr{K}_j$	Set of ORs that can be managed by surgeons with subspecialty j	$k \in \mathscr{K}_j \subseteq \mathscr{K}$

Table A.8: Parameters used in the mathematical formulation

Parameter	Description
A_w^{MAX}	Maximum amount of beds available on ward w
A_d	Number of staffed beds available on day d
B_{kd}	Time (minutes) available for surgery in one slot in operation room k at day d
C_i^C	Penalty for cancelling an elective patient of category i
C^{SU}	Penalty for scheduling an SU patient to an elective surgery slot
$C_{ww'}^W$	Penalty for putting a patient belonging to ward w in ward w'
C^β	Penalty for staffing more beds than scheduled
E_{id}	Expected length of stay (days) of patient category i scheduled for surgery on day d
E_{id}^{SU}	Expected length of stay for SU emergency patient of category i scheduled on day d
M_{kd}^{OR}	Maximum number of slots that can be reserved at an OR k on day d within the opening hours of the OR
M_d^A	Maximum number of ORs that can be covered by anaesthesia staff on day d
M^{CYCLE}	Total amount of slots available through one cycle
N_j	Maximum surgeon capacity (slots) of subspecialty j in one cycle
N_{jd}^D	Maximum surgeon capacity (slots) of subspecialty j at day d
R_i^{EL}	Reward for scheduling more patients from patient category i than the lower limit
P_s	Probability of ending up in scenario s
R^U	Reward for scheduling a U emergency patient to a flexible slot
S_i	Expected surgery duration of a patient in elective patient category i
T_i	Target throughput in number of elective patients belonging to elective patient category i
T_{is}^{SU}	Excess demand in number of SU emergency patients belonging to SU patient category i
U_{wds}^{EM}	Amount of emergency patients resting in ward w on day d in scenario s
V_i	At least, a share of V_i of the patients belonging to patient category i should be scheduled for surgery

Table A.9: Variables used in the mathematical formulation

Variable	Description
a_{wd}	Number of staffed beds in ward w on day d
n_{jkd}	Number of slots scheduled as flexible for subspecialty j in OR k on day d
v_i	Number of elective patients from patient category i scheduled above the lower limit
x_{ikd}	Number of elective patients of patient category i scheduled to an elective slot in OR k on day d
y_{jkd}	Number of elective slots scheduled for subspecialty j in OR k on day d
α_{kd}^A	Indicates whether OR k is covered by anaesthesia staff on day d or not
$b_{ww'ds}$	Number of beds occupied in ward w' by patients belonging to ward w on day d and scenario s
e_{ijkds}	Number of SU emergency patients of category i scheduled to subspecialty j in a flexible surgery slot in OR k on day d in scenario s
e_{ijkds}^{EL}	Number of SU emergency patients of category i scheduled to subspecialty j in an elective surgery slot in OR k on day d in scenario s
u_{iwds}	Number of elective patients of patient category i resting in ward w on day d in scenario s
u_{iwds}^{SU}	Number of SU emergency patients from category i resting in ward w on day d in scenario s
x_{ikds}^C	Number of elective patients of patient category i scheduled to an elective slot within OR k on day d that are cancelled
β_{wds}	Number of additional beds staffed in ward w on day d in scenario s

Table B.10: Values obtained for the subspecialties and the patient categories in the optimization model for the big instance. T_i is the target number of elective patient category i to be scheduled in each cycle, S_i is the expected surgery duration of patient category i , E_{id} is the expected length of stay of patient category i that receive surgery on day d, R_i^{EL} is the gain for scheduling more patients from patient category i than the lower limit, C_i^C is the penalty for cancelling an elective patient of category i, N_{jd}^D is the maximum number of slots available to subspecialty j on day d, and N_j is the maximum number of slots available to subspecialty j in one cycle.

Subspecialty	T_i	S_i	E_{id}	R_i^{EL}	C_i^C	N_{jd}^D	N_j
Elective foot						4	5
Aggregated group	4	143	3	3	3		
Hand						4	7
Aggregated group	8	94	0	2	3		
Carpal tunnel syndrome	3	85	1	3	3		
Plastic						4	14
Aggregated group	15	95	2	3	3		
Plateepitelkarsinom	2	73	1	3	3		
BCC	5	142	1	3	3		
Malingt melanom	4	68	0	2	3		
Cancer mammae	4	97	1	3	3		
Arthroscopic						4	12
Aggregated group	6	123	2	3	3		
ACL	2	186	2	3	3		
Meniscus	3	173	0	2	3		
Back						4	6
Aggregated group	4	295	6	3	3		
Prostheses						4	16
Hip	7	177	4	3	6		
Knee	11	174	4	3	6		
Tumour						2	2
Aggregated group	2	76	1	3	3		
emergency patients							
SU inpatients		192	2				
SU outpatients		131	0				
U patients		165	0				

Table B.11: Values of the parameters related to the ORs in the optimization model for the big instance. M_{kd}^{OR} represents the number of surgery slots available at OR k on day d, and B_{kd} is the time available in each slot in OR k on day d.

OR	M_{kd}^{OR}	B_{kd}	Patient category
1	2	240	Elective foot, hand, plastics, arthroscopic
2	2	240	Elective foot, hand, plastics, arthroscopic
3	2	240	Elective foot, hand, plastics, arthroscopic
4	2	240	Elective foot, hand, plastics, arthroscopic
5	2	240	Back, tumour
6	2	240	Prosthesis
7	2	240	Prosthesis

Table B.12: The ward capacities obtained in the optimization model for the big instance. A_d represents the number of staffed beds available on day d, while A_w^{MAX} is the maximum number of beds available in ward w.

Ward	Name	A_d (week)	A_d (weekend)	A_w^{MAX}	Patient category hosted
1	Trauma			32	Elective foot, hand and SU inpatients
2	Reconstructive			16	Plastic, tumour
3	Elective			12	Arthroscopic, back
4	Fast-track			16	Prosthesis
5	Hotel-day			5	None, buffer capacity
		67	44		

Table B.13: Values obtained for the subspecialties and the patient categories in the optimization model for the test instances. T_i is the target number of elective patient category i to be scheduled in each cycle, S_i is the expected surgery duration of patient category i , E_{id} is the expected length of stay of patient category i that receive surgery on day d, R_i^{EL} is the gain for scheduling more patients from patient category i than the lower limit, C_i^C is the penalty for cancelling an elective patient of category i, N_{jd}^D is the maximum number of slots available to subspecialty j on day d, and N_j is the maximum number of slots available to subspecialty j in one cycle.

Subspecialty	T_i^{LOW}	T_i^{HIGH}	S_i	E_{id}	R_i^{EL}	C_i^C	N_{jd}^D	N_j
Subspecialty 1							6	41
Patient category 1	8	12	143	3	3	10		
Subspecialty 2							6	41
Patient category 2	32	38	94	0	2	10		
Patient category 3	18	25	85	1	3	10		
Subspecialty 3							6	41
Patient category 4	21	25	95	2	3	10		
Patient category 5	12	15	73	0	3	10		
emergency patients								
SU inpatients			192	2				
SU outpatients			131	0				
U patients			150	0				

Table B.14: Values of the parameters related to the ORs in the optimization model for the test instances. M_{kd}^{OR} represents the number of surgery slots available at OR k on day d, and B_{kd} is the time available in each slot in OR k on day d.

OR	M_{kd}^{OR}	B_{kd}	Patient category
1	2	240	All
2	2	240	All
3	2	240	All
4	2	240	All
5	2	240	All

Table B.15: The ward capacities obtained in the optimization model for the test instances. The four configurations represent four different bed capacities, A_d represents the number of staffed beds available on day d, while A_w^{MAX} is the maximum number of beds available in ward w.

Configuration	Ward	A_d (week)	A_d (weekend)	A_w^{MAX}	Patient category hosted
W1	1			32	1, 3 and SU inpatients
W1	2			25	4
		45	40		
W2	1			32	1, 3 and SU inpatients
W2	2			25	4
		55	50		
W3	1			45	1, 3 and SU inpatients
W3	2			40	4
		85	50		
W4	1			60	1, 3 and SU inpatients
W4	2			50	4
		110	55		

Table B.16: Maximum number of ORs that may be covered by anesthesiologists each day in the optimization model for both the large and the test instances.

Instance size	M_1^A	M_2^A	M_3^A	M_4^A	M_5^A	M_6^A	M_7^A
Large (Section 5.4)	7	7	7	7	4	0	0
Small (Sections 5.2 and 5.3)	5	5	5	5	4	0	0

Table B.17: The values obtained for the other parameters in the optimization model. $C^W_{ww'}$ is the penalty of putting a patient meant for ward w in ward w', C^{SU} is the penalty of scheduling an SU patient to an elective slot, P^U is the gain for scheduling a U patient to a flexible slot, C^β is the penalty of having more patients resting in the wards than the total amount of staffed beds available, V_i is the share of patients belonging to the elective patient category i that needs to be scheduled for surgery, and M^{CYCLE} is the total amount of slots available through the cycle.

Instance size	$C^W_{ww'}$	C^{SU}	P^U	C^β	V_i	M^{CYCLE}
Large (Section 5.4)	0.1	2	0.5	1000	0.5	70
Small (Sections 5.2 and 5.3)	1	2	0.5	1000	0.5	70

Table C.18: The wards present in the simulation model

Ward	Patient category	Capacity
1	Elective foot, hand and emergencies	∞
2	Plastic, tumour, emergencies	∞
3	Arthroscopic, back, emergencies	∞
4	Prosthesis, emergencies	∞
5	No electives, emergencies	∞

Table C.19: The ORs present in the simulation model

OR-type	Number of ORs	Opening hours
Elective	7	08.00-16.00 (Monday to Friday)
Emergency	3	Varies dependent on day and OR

Table C.20: Stochastic processes

Process	Probability distribution
Emergency arrivals	Poisson
Surgery duration	Empirical
Length of stay	Empirical

Table D.21: The MSS generated in the case study. Green letters indicate flexible slots.

OR	Day of week				
	Monday	Tuesday	Wednesday	Thursday	Friday
1	Hand	Hand	El. foot/ Plastic	Plastic/ Arthroscopic	Plastic/ Plastic
2	Arthroscopic	Arthroscopic	El. foot	Plastic	Arthroscopic
3	Plastic	Plastic/ Arthroscopic	Arthroscopic	Hand	Arthroscopic
4	Plastic	Plastic/-	Plastic	Hand/-	El. foot
5	tumour/tumour	Back	Back	Back	-
6	Prosthesis	Prosthesis	Prosthesis/-	Prosthesis	-
7	Prosthesis	Prosthesis	Prosthesis/-	Prosthesis	-