Noname manuscript No.
(will be inserted by the editor)

Constructing and Evaluating a Master Surgery Schedule using a Service-Level Approach

Received: date / Accepted: date

Abstract In this paper, we consider the problem to build a cyclic master surgery schedule for a case study of a small-to-medium-sized Belgian hospital. The problem considers the objectives of different stakeholders and aims to maximise the adjacency of operating room blocks assigned to an individual surgeon (group) and to optimise the bed capacity usage in the connected downstream units. To account for different sources of uncertainty surrounding the patient admissions and the required bed capacity in the downstream units, we formulate and solve a deterministic optimisation model relying on the service level concept to define higher-than-expected values for the patient demand. The yielded master surgery schedule establishes suitable bed capacity buffers in the downstream ward units, which improve the robustness to variations in the stochastic variables avoiding bed capacity overruns. The model is solved using a hierarchical two-stage procedure as a priority is established between the different objectives. In the first stage only the block adjacency objective is considered, whereas the second stage applies the $\mathcal{E}$-Constraint Method to find a set of solutions lying on or close to the Pareto front, making the trade-off between the usage of the bed capacity and the workload levelling in the hospital wards. The computational experiments provide insight in the (mutual) impact of the different considered objectives from a deterministic and stochastic point-of-view. We show that the scheduling process of surgeons can be improved by using an automated approach. The proposed method increased both the adjacency between OR blocks and the bed capacity usage significantly as all yielded solutions outperformed the master surgery schedule currently in use in the visited hospital. Findings have been derived using both a training and test real-life dataset to assess the resulting schedule robustness properly.

Keywords: Operating room department, Master Surgery Scheduling, Health Services, Multi-Objective Optimisation, Case study

1 Introduction

Healthcare facilities are continuously striving for maximising patient quality of care while controlling costs and minimising the waste of resources (e.g. hospital beds) (Toussaint and Berry, 2013; Lawal et al., 2014). Healthcare managers are trying to optimise their resource usage by making thorough decisions on three different levels, i.e. the strategic, tactical and operational planning level. The levels follow a top-down hierarchical approach with feedback loops to allow adjustments if necessary. The key differentiator between the different levels is the degree of 'in advance' versus 'reactive' decision-making (Hans et al., 2011). Decisions taken on the strategic planning level are made once a year, over a long time horizon, whereas decisions taken on the operational planning level can be seen as reactions to operational variability on the day of execution. More than 60% of hospital admissions are related to the operations within the operating room (OR) department, accounting for about 40% of the hospital costs (Pham and Klinkert, 2008; Villarreal and Keskinocak, 2016). Crucial in the organisation of the OR department is the decision-making on the tactical level with the purpose of steering the allocation of resources efficiently and effectively via the so-called master surgery schedule (MSS), which defines a cyclic timetable, managing the operating theatre by organising the OR time in blocks. These blocks are assigned to specific medical disciplines or surgeons (Beliën and Demeulemeester, 2007; Houdenhoven et al., 2008). In other words, the MSS administers the OR resources and incoming OR patients. Moreover, the master surgery structure and timetable determine to a large extent the usage of downstream resources, such as hospital beds and nurses. Only a few researchers have

Address(es) of author(s) should be given

linked the bed capacity usage in downstream units to the construction of the master surgery scheduling problem and optimised an integrated problem accordingly. However, one should understand the complexity of this approach due to the uncertainty embedded in a real-life hospital environment, involving admissions of (non-)OR patients, transfers of patients to ward units and length of stay, amongst other factors. This uncertainty may cause disruptions in the bed capacity usage at the downstream units and leads to divergences between expected and actual usage (Cardoen et al., 2010a; Schrijvers et al., 2012).

In this paper, we optimise the MSS to improve the assignment of surgeons to OR blocks while focusing on the effect of the MSS on the bed usage in the different involved hospital wards. The choice for this subject was induced for two main reasons. First, many master surgery scheduling approaches proposed in the literature consider only the impact on the operating theatre and its staff. Only a minority of research has been dedicated to the relationship between the deployment of a master surgery scheduling approach and the (stochastic) usage of resources in downstream units (Cardoen et al., 2010a; Fügener et al., 2014). The second driver is business-related as the problem definition is based on a real-life problem encountered in a Belgian hospital. In Belgium, the average bed occupancy level oscillates between 70 and 80%, which is, according to many hospital managers, too low (WHO, 2020). Via this study, the visited hospital wanted to increase their insight in the linkage between the MSS and the usage of bed resources in related downstream ward units, taking into account the hospital's patient-related data and the uncertainty surrounding the bed capacity usage. The goal is to develop an intelligent automated decision support system to determine the most suitable MSS based on historical hospital data.

To solve the real-life scheduling problem under study, we employ a stepwise approach based upon three elements, i.e. (i) the formulation of a multi-objective deterministic model relying on the service level concept, creating a schedule insensitive to operational variability such that bed capacity overruns in the downstream units are kept to a minimum, (ii) the development of a hierarchical two-stage optimisation approach that establishes a priority between the different encountered objectives, solving the problem in two stages, and uses in the second step an $\mathcal{E}$-Constraint Method, to find a set of solutions lying on or close to the Pareto front and (iii) an evaluation procedure to decide which MSS is preferred taking into account the schedule quality from a deterministic and stochastic point-of-view.
When constructing a MSS, we utilise a deterministic optimisation model that considers adequate proxy values for the stochastic variables based on a higher-than-expected patient demand. This higher patient demand is input to optimise the bed capacity occupancy in the downstream units and establishes adequate bed capacity buffers over the planning horizon. The augmented demand is calculated based on the provided service level in accordance to the strategic capacity decisions related to the allocated OR time. The objective function of the model is formulated as a preemptive goal function to consider the perspectives of multiple stakeholders, accounting for the adjacency of OR blocks assigned to a particular surgeon, the unused and overused bed capacity and the levelling of the bed capacity usage. The latter two objectives concern the resources in the downstream ward units. Although the optimisation of the adjacency of OR blocks is often neglected in the literature, the objective was raised in the discussions with the hospital's stakeholders as the main objective to organise the OR timetable. Consequently, this objective is considered as the only objective in the first stage of the preemptive solution approach. Furthermore, including this objective maximises the efficiency in the OR department (e.g. Penn et al., 2017). In order to consider the two objectives related to the bed capacity usage in the second stage of the solution procedure, we use the $\mathcal{E}$-Constraint Method of Chankong and Haimes (2008) to find a set of schedules lying on or close to the Pareto front and induce a focused decision to select the most suitable MSS in correspondence with the hospital manager. The latter is indispensable to improve the acceptance of the new MSS by the different stakeholders of the hospital and support the negotiation process (Guerriero and Guido, 2011). The employment of an $\mathcal{E}$-Constraint Method avoids that subjective weights should be quantified for the different objectives. To the best of our knowledge, this is the first study that yields a set of (approximated) Pareto-optimal solutions to determine the most desired MSS. The different stages of the solution approach are solved using mixed-integer programming.
In order to support the selection of the most suitable MSS, we apply two methods to evaluate the performance of the yielded MSSs. The first method is based upon the distance to the ideal point, which is often referred to as the Zeleny point (Zeleny, 1975). This method evaluates each MSS based on a single objective metric, combining the values of the two objectives considered in the second stage. The second method concerns the calculation of the global service level provided by the available bed capacity and measures the number of bed capacity overruns, considering different sources of uncertainty surrounding the patient admissions and the required bed capacity in the downstream units. The robustness of the retrieved MSSs is tested using a trace-driven simulation experiment with real-life data input. Computational experiments have been conducted to study the impact of the different objectives and to show the validity of the approach. Different data sets have been employed to set up a MSS and to validate the robustness of the yielded schedules.

The further sections are organised as follows. In Section 2, we present the related literature considering the master surgery scheduling problem and the uncertainty surrounding the problem under study. In Section 3, we present a profound problem description and a mathematical model formulation for the investigated master surgery scheduling problem. In Section 4, we discuss the proposed multi-objective solution procedure in detail. In Section 5, we discuss the computational experiments showing the validity of our approach and the computational findings exploring the impact of the different objectives, leading to the most appropriate MSS for the hospital under consideration. Section 6 provides concluding remarks.

2 Literature Review

In this section, we discuss the relevant literature related to the problem under study. In Section 2.1, we elaborate on the commonly studied features of the master surgery scheduling problem. Note that the usage of the MSS as a coordination tool has been widely discussed in the literature (see Cardoen et al., 2010a; Guerriero and Guido, 2011; Zhu et al., 2018; Magerlein and Martin, 1978; Van Riet and Demeulemeester, 2015). In Section 2.2, we characterise the different sources of uncertainty impacting the bed capacity usage in detail. Section 2.3 discusses stochastic solution methodologies accounting for uncertainty in the construction of a suitable MSS. Section 2.4 gives insight in the contribution of this study.

2.1 The Master Surgery Scheduling Problem

Accounting for the strategic long-term decisions, which stipulate the case-mix planning that distributes OR time among surgeons, the tactical OR decisions address the OR resource usage over a medium-term planning horizon of several weeks (Hall, 2012; Marques et al., 2019). In the literature, different approaches have been discerned to organise the OR resources, i.e. block scheduling, open scheduling and modified block scheduling (Guerriero and Guido, 2011). A survey by Cardoen et al. (2010b) has shown that 94% of the 52 Flemish hospitals uses a block scheduling approach to establish a MSS, and, thereby, govern resource usage over a medium-term. In case of block booking, the tactical decision level of the OR planning and scheduling process defines a so-called MSS, pre-allocating OR capacity to different surgeons or surgery groups (Zhu et al., 2018). In other words, each surgeon or group of surgeons is assigned blocks of time in ORs during which he/she can perform surgeries, reducing the scheduling complexity (Zhu et al., 2018; Rachuba and Werners, 2017). Beliën and Demeulemeester (2007) define the MSS as a cyclic timetable that defines the number and type of ORs available, the hours that rooms will be open and the surgical groups or surgeons who are given priority for the OR time. In the literature, OR blocks have been planned with different durations, e.g. two hours, half a day (4 hours) or an entire day (8 hours), defining the smallest time unit for which an OR can be allocated to a surgeon (group) (Cardoen et al., 2009; Choi and Wilhelm, 2014; Ma and Demeulemeester, 2013). Research suggests that an improved MSS can realise a better usage of the involved resources (Guerriero and Guido, 2011).

Due to its popularity, many researchers have tackled the problem of constructing an appropriate MSS. Different constraints have been considered related to the surgical demand, the preferences of surgical groups, avoiding parallel sessions, downstream bed and nursing resources, etc. The considerations related to the surgical demand refer to demand requested by the different surgical groups and is usually modelled via a lower/upper bound on the number of OR blocks that can be assigned to a surgical group and/or via a priority-based objective such that more capacity is allocated to higher prioritised surgery types (see Santibáñez et al., 2007; Kharraja et al., 2007; Samanlioglu et al., 2010; Agnetis et al., 2014; Fügener, 2015). The preferences of surgical groups are typically modelled via hard constraints and are either related to the timing of an OR block (see Agnetis et al., 2012; Spratt and Kozan, 2016) and/or to the equipment in a specific OR (see Spratt and Kozan, 2016; Penn et al., 2017). A limited number of studies consider the constraint avoiding the assignment of a parallel session for a particular surgical group (e.g. Agnetis et al., 2014; Fügener, 2015). The allocation of OR blocks may be further restricted by the availability of scarce operating room and downstream resources. Examples are the number of beds or nursing resources for post-operative processes in the OR department (see Tànfani and Testi, 2010; Abedini et al., 2017; M'Hallah and Visintin, 2019) and in downstream units (see Santibáñez et al., 2007; Beliën and Demeulemeester, 2008; Tànfani and Testi, 2010).
Since many stakeholders (i.e. surgeons, nurses and the hospital's management) are involved (Cardoen et al., 2010a; Zhu et al., 2018), multiple, often conflicting, objectives have been proposed related to the master surgery scheduling problem, which can be grouped in different categories. The first type of performance criteria is related to the hospital's operations. Typical objectives are optimising utilisation of ORs (e.g. Li et al., 2017), workload levelling in the OR department and downstream hospital wards (e.g. Beliën and Demeulemeester, 2007; Ma and Demeulemeester, 2013; van Oostrum et al., 2008), minimising the number of times surgeons of the same specialty perform surgeries in different rooms (Beliën and Demeulemeester,

2008) and the equity among surgeons by minimising the difference between the actual time assigned to surgeons and their target value (Blake et al., 2002). A second group is characterised by financial performance measures such as minimising the operational cost structure (e.g. Fügener et al., 2014). The last group of performance objectives is related to human factors such as the maximisation of surgeon preferences (e.g. Penn et al., 2017). In the literature, different methodologies consider multiple of these objectives together (e.g. Vissers et al., 2005; Guerriero and Guido, 2011; Bovim et al., 2020). According to the multi-objective optimisation literature, there are three categories of methods for solving multi-objective problems, i.e. a priori methods, interactive methods and a posteriori methods (Hwang and Masud, 1979). Using an a priori method, the decision-maker expresses his/her preferences before the solution process, e.g. by assigning (subjective) weights to the objective functions upfront. The second type of solution methods comprises an iterative approach, which solves the problem several times, each time converting to the most preferred solution. This type of solution method leads to a tunnel vision, meaning that the decision-maker will be focused on only one path for finding the most preferred solution. An a posteriori method creates several (Pareto-optimal) solutions and afterwards the decision-maker needs to select the most preferred solution among this set. The main advantages of the latter method compared to the other methods are threefold. First, an a posteriori method is not biased since we do not have to determine weights. Second, whereas a weighting method will only generate efficient extreme solutions, a posteriori methods like the $\mathcal{E}$-Constraint Method are also able to produce non-extreme solutions, which provides a more rich representation of the efficient set. Finally, we can control the number of generated efficient solutions by adjusting the number of grid points (Chankong and Haimes, 2008). Most of the approaches for constructing a MSS employ a set of linear weights to consider the different objectives into a weighted single objective function (e.g. Bovim et al., 2020). Only a limited number of studies propose a multi-objective solution procedure generating multiple solutions to improve real-life decision-making. In this respect, Marques et al. (2019) determine different weight settings to propose a set of schedules to the management of a real-life hospital. Khalfalli et al. (2019) demonstrate the application of the $\mathcal{E}$-Constraint Method to a surgery scheduling problem to minimise session completion time and staff overtime.

2.2 Uncertainty in Master Surgery Scheduling

Several authors have considered operational variability when tackling the master surgery scheduling problem. In the following, we discuss the relevant sources of uncertainty and discern appropriate distributions. Uncertainty surrounding the bed capacity usage in the ward units may involve elective and non-elective patient admissions, the transfer of OR patients to the different wards and the length of stay for the different patient types. Other types of uncertainty related to operational decision-making not impacting the bed capacity usage (e.g. the duration of surgeries (Bovim et al., 2020)) are not discussed.

Patient Admissions
Most approaches considering master surgery scheduling model the admission of elective patients using averages as an estimate based on historical data (Marques et al., 2019; van Oostrum et al., 2010, 2008). In other words, these models will use the expected values related to patient admissions to test the legitimacy of the models. Only some studies propose to model the uncertainty related to the number of elective patient admissions by a discrete probability distribution (Beliën and Demeulemeester, 2007; Vanberkel et al., 2011; Van Huele and Vanhoucke, 2014). Many researchers today agree on the importance of uncertainty caused by non-elective patients (Cardoen et al., 2010a). Litvak and Long (2000) argue that the prominent part (i.e. more than 80%) of the uncertainty is due to the arrival of non-elective patients leading to the rescheduling of elective patients (Figueira and Almada-Lobo, 2014). Although this source of uncertainty can cause major issues in a hospital's day-to-day operating environment (e.g. lack of human resources causing overtime, increased waiting time and stress), most of the published works do not consider this type of uncertainty in the construction of the MSS (Guerriero and Guido, 2011). Only Bovim et al. (2020) consider the uncertainty related to the arrivals of non-elective patients, impacting the utilisation of the ORs and bed capacity in the downstream units. To that purpose, they designed a simulation model mimicking the daily operational decisions in the OR department. The arrival of non-elective patients is modelled using a negative exponential distribution.

Transfer of OR Patients to the Hospital Wards
Fügener et al. (2014) consider the transfer of OR patients to the different hospital wards as a source of uncertainty. They observed that the vast majority of patients (98%) follow three different care paths, of which the path *'OR→ward→discharge'* was the most frequent one (92%). Traditionally, OR patients are transferred to hospital wards based on surgery type. However, over the years, other classifications are introduced based on the patient's length of stay or level of care (Walley et al., 2006; Bekker et al., 2017). Schrijvers et al. (2012) argue that factors such as the patient's age or the hospital's ward capacity are also

considered when specifying a patient's care path and allocating patients to wards. Therefore, obtaining the underlying distribution based on historical data is of critical importance in achieving a reliable linkage between the MSS and the bed occupancy in downstream ward units (van Oostrum et al., 2010).

Patient Length of Stay
The length of stay (LOS) can be defined as the average number of days that patients spend in the hospital and is often used as a performance indicator for costing and hospital management (OECD, 2017). Most researchers specify the average LOS as an estimate to calculate the bed capacity usage in the hospital downstream units (e.g. van Oostrum et al., 2008). Adan and Vissers (2002), Ma and Demeulemeester (2013), Li et al. (2017) and M'Hallah and Visintin (2019) consider the uncertainty related to LOS following surgery and define a multinomial probability distribution, which can be based on historical data.

2.3 Solution methodologies considering uncertainty for Master Surgery Scheduling

A large number of stochastic variables entangles the application of a methodology relying on stochastic programming. However, according to van Oostrum et al. (2010) and Penn et al. (2017), incorporating operational uncertainty in the decision-making is one of the success factors when implementing a MSS. Servaux and Sörensen (2004) suggest that constructing a robust schedule that is able to absorb some of the uncertainty is in many real-life optimisation problems as important as finding the best possible solution and cannot be neglected in the real-life decision-making process. Different methods have been proposed relying on simulation and optimisation to solve stochastic optimisation problems (Figueira and Almada-Lobo, 2014). These methods rely mainly on either (i) the evaluation of various solutions using a comprehensive simulation model; (ii) the interaction between the evaluation and generation of solutions in an iterative manner; and (iii) the inclusion of the simulation results or stochastic variables in the optimisation model. Hence, several authors apply simulation relying on theoretical or empirical distributions to evaluate the performance of (optimised) MSSs in a more operational setting and assess different scheduling policies. For example, (Zhang et al., 2009) use simulation to mimic patient demand, patient arrivals, surgery times and no-shows, considering a longer time horizon to assess the quality of the MSS yielded using deterministic MIP. Adan et al. (2011) propose a goal programming approach to minimise deviations between expected and target resource utilisations as well as the overuse of resources. Optimisation and simulation of elective and emergency patient arrivals are applied in an iterative manner to determine the best-performing slack and flexibility planning rules. Banditori et al. (2013) use simulation in view of uncertainty related to LOS and surgery duration to assess the robustness of a surgery schedule yielded using a MIP model. The latter incorporates predefined resource slacks to reduce the number of cancellations and overtime. Cappanera et al. (2014) use simulation and optimisation to compare alternative objective functions related to number of surgeries scheduled, workload balancing among OR resources and robustness involving overruns of OR and bed utilisation. Lamiri et al. (2008), Min and Yih (2010) and Bovim et al. (2020), amongst others, propose sample average approximation methods, relying on simulation to generate scenarios as input to formulate a deterministic optimisation model for a two-stage stochastic problem. Kumar et al. (2018) developed a multi-stage stochastic optimisation approach to find the MSS that leads to the smallest number of cancellations over multiple sequential periods. The proposed method relies on deterministic optimisation based on different sequences of randomly generated LOS scenarios. Fügener et al. (2014) use stochastic programming and propose an exact branch-and-bound method and several heuristics for the stochastic optimisation problem to minimise downstream costs.
Apart from simulation or stochastic optimisation approaches, a limited number of authors developed a robust optimisation approach to mitigate the impact of uncertainty. In this perspective, Denton et al. (2010) relies on the definition of uncertainty sets to model the patient demand using a simple heuristic founded on the newsvendor problem. Mannino et al. (2012) propose an optimisation approach relying on the concept of light robustness to mitigate the impact of demand uncertainty. Their approach does not consider the maximum (or worst-case) demand but allows for unsatisfied demand that is penalised. Cappanera et al. (2014) minimise the sum of positive deviations, so-called overruns, of daily utilisation of ORs and ward beds to create robust schedules. These former robust approaches consider only the patient demand uncertainty and no downstream resources. Only Neyshabouri and Berg (2017) propose a two-stage robust optimisation solution methodology to account for the inherent uncertainty in surgery duration and length-of-stay in the downstream units, following the definition of user-definable arbitrary uncertainty sets.

2.4 Contribution to the literature

The research in this paper distinguishes itself in several aspects from the literature. First, the primary optimisation objective, considered in this paper, involves the adjacency of OR blocks in the same OR for a particular surgeon and has not been addressed before in the related literature. We study its implications on the usage of downstream bed resources. Second, we propose a multi-objective optimisation method to provide a suitable solution to the hospital management. Instead of assigning weights to each of the optimisation criteria, we propose a hierarchical two-stage optimisation approach where in the first stage only the primary objective to schedule the OR physicians is considered. The second stage considers two objectives related to the downstream resources and relies on the $\mathcal{E}$-Constraint Method to find a set of MSSs lying on or close to the Pareto front. Third, we modelled a deterministic equivalent of a stochastic master surgery scheduling problem. To increase the schedule robustness, we define adequate higher-than-expected proxy values for the stochastic patient demand, which are taken into account to optimise the MSS and realise adequate bed capacity buffers in the downstream units over the planning horizon. These proxy values are calculated in line with the strategic capacity decisions involving the allocated OR time, guided by the postulated service level. The evaluation of the yielded MSSs relies on a trace-driven simulation experiment based on a real-life validation dataset taking three different sources of uncertainty into account, i.e. patient admissions of both OR and non-OR patients, transfers to hospital wards, and patients' LOS

3 Problem Description and Formulation

In this section, we focus on the problem statement and formulation of the studied master surgery scheduling problem. The research has been inspired by the real-life practices of a private non-for-profit hospital (Belgium). In Section 3.1, we describe the context encountered at the visited hospital, expressing the need for optimising the MSS. Section 3.2 discusses the constraints and objectives of the problem under study. In Section 3.3, we formulate a MIP model, aligned with the specific problem definition as stipulated by the hospital. Note that the proposed MIP model has a probabilistic nature as different input parameters surrounding the inflow of patients and bed usage represent random variables.

3.1 Problem Context and Motivation

The visited hospital uses a block scheduling approach for establishing the MSS. The hospital disposes of 8 ORs where surgeries can be performed during time blocks comprising the entire morning or afternoon, i.e. an OR day in a room consists out of two blocks of four hours. After surgery, inpatients are transferred and treated in one of the 19 hospital wards (e.g. 'Long Stay', 'Paediatrics', 'Maternity', 'Geriatrics'). The hospital's staff uses three indicators to allocate patients to specific hospital wards, namely (i) surgery type, (ii) estimated LOS and (iii) age. The total bed capacity summed over the wards is equal to 240 justified beds, which are beds financially supported by the Belgian Government based upon patient-related activities (Deschodt et al., 2015). The hospital aims to investigate how to improve the utilisation of the bed capacity and to smoothen the workload in their inpatient wards over the time horizon. In order to obtain insight in the as-is situation and encountered issues in the visited hospital, a dashboard has been constructed based on the hospital data. The average bed occupancy level in 2018 was equal to 83.75%, which was 0.33% higher than the level in 2017, but was not considered as satisfactory by the hospital's stakeholders. Note that in these calculations, outpatients are not accounted since they do not stay in the hospital for more than a day. In addition, we show the daily evolution of the aggregate bed occupancy level for the year 2018 in Figure 1. The x-axis shows the different days of the week whereas the primary y-axis indicates the bed occupancy level as a percentage of the available capacity. Further insight in the fluctuations of the bed occupancy is provided by the secondary y-axis, showing the number of incoming and outgoing transfers on a yearly basis. Incoming transfers are patients who are going to a certain hospital ward. Outgoing transfers are either patients going to another ward or patients who are leaving the hospital permanently.

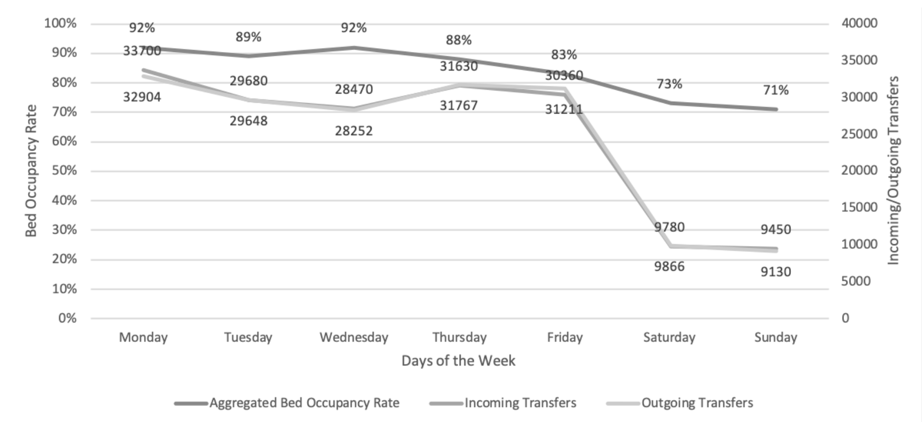

Fig. 1: Aggregated bed occupancy rate: weekly overview (2018)

Figure 1 leads to a number of observations. First, the bed occupancy level is high from Monday until Wednesday but shows a decreasing trend starting from Thursday. Although this is rather logic since the ORs are closed during the weekend, the occupancy level in the weekend has been assessed as far too low compared to the other days of the week due to (i) a peak workload that is too high on some week days, leading to a substantial number of bed capacity overruns; (ii) the misalignment between the patient workload and the scheduled nursing capacity in the wards that is organised in a cyclic manner, more or less levelling the number of nurses over the days; and (iii) an unbalanced ward occupancy entangling the organisation of the wards, i.e. the staff scheduling and the operational execution. Second, most transfers take place on Monday, Thursday and Friday. Based on a thorough discussion with the hospital's planning manager, we concluded that the incoming and outgoing flows of patients were the root cause of the oscillation in the bed occupancy level, which is steered prominently by the MSS. Third, the hospital suffered from a number of bed capacity overruns on particular days due to the workload variability, causing severe stress situations for personnel and endangering the patient quality of care, despite the slack bed capacity at other points in time during the week. The hospital requested to resolve these issues by constructing a new MSS taking hospital-specific data into account, linking the MSS to the hospital's bed occupancy rate in the downstream ward units. In order to decide on the composition of the MSS, different objectives have been postulated in the discussions with different stakeholders. The hospital's planning manager proposed two objectives, i.e. to maximise the usage of the current bed capacity and to level the bed occupancy rate during the week. If we could level the hospital's bed occupancy rate, the staff's workload would be better spread. The physicians, however, desired to operate in the same OR on the same day as much as possible.

3.2 Problem Description

In the visited hospital, the majority of the patients (97%) follows the care path *'OR→ward→discharge'*, which is conform to the observations of Fügener et al. (2014). To improve the bed capacity usage in the downstream ward units, we focus on this care path, which is predominant in the visited hospital. Hence, we make abstraction of other possible care paths, involving patients transferred first to the Intensive Care Unit, from one ward to another or immediately discharged after recovery. Note that all inpatients who enter a ward at some point in time and are not transferred directly from the OR department to the ward, are labelled as non-OR patients in our analysis. These patients are taken into account as an external factor to accurately calculate the bed capacity usage in the downstream units. Outpatients, who are immediately discharged after recovery, are treated in the surgery day clinic and are not considered in this study.
Correspondingly, in the problem under study we aim to optimise the organisation of the OR resources (i.e. OR time and surgeons) and the usage of downstream resources in the ward units (i.e. beds). We neglect the recovery beds in the OR and the Intensive Care Unit. The problem encompasses the construction of a MSS taking into account the requirements and the objectives of the different stakeholders, i.e. the surgeons related to different disciplines, the downstream ward units and the involved hospital. The MSS is composed by the hospital's planning manager and embodies a cyclic schedule that defines the allocation of OR time blocks in different ORs to a specific surgeon or group of surgeons. While constructing the MSS, the hospital's planning manager considers the availability of scarce OR resources (i.e. OR availability, surgeon availability and nurse availability) and downstream resources (i.e. hospital beds). Figure 2 depicts the MSS currently in use in the visited hospital.

	Monday		Tuesday		Wednesday		Thursday		Friday	
	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM
OR 1	Urology - General	Urology - General	Gynaecology - Physician 1	Urology - General	Urology - General	CLOSED	Urology - General	Urology - General	Gynaecology - General	Gynaecology - General
OR 2	CLOSED	Surgery - General	Gynaecology - Physician 2	Surgery - Physician 1	Surgery - Physician 2	Surgery - Physician 2	Surgery - Physician 1	Surgery - Physician 1	CLOSED	Plastic Surgery - Physician 1
OR 3	Orthopaedics - Physician 1	Orthopaedics - Physician 1	Orthopaedics - Physician 2	Orthopaedics - Physician 3	Orthopaedics - General	Orthopaedics - General	Plastic Surgery - General	Plastic Surgery - General	Orthopaedics - Physician 6	Orthopaedics - Physician 6
OR 4	Neurology - Physician 1	Neurology - Physician 1	Vascular surgery - Physician 1	Vascular surgery - Physician 1	Orthopaedics - Physician 4	Orthopaedics - Physician 4	Vascular surgery - Physician 1	Orthopaedics - General	Orthopaedics - Physician 5	Orthopaedics - Physician 5
OR 5	Stomatology - Physician 1	Stomatology - Physician 2	Stomatology - General	Stomatology - General	ORL	ORL	ORL	ORL	Stomatology - General	Stomatology - General
OR 6	Urology - Physician 2	Endocrinology	Anasthesea	Endocrinology	Endocrinology	Endocrinology	Pulmonology	Endocrinology	Anasthesea	Endocrinology
OR 7	Gynaecology - Physician 2	Endocrinology	Gynaecology - Physician 3	Gynaecology - Physician 4	Gynaecology - Physician 2	Endocrinology	Anasthesea	Endocrinology	Gynaecology - Physician 4	Urology - General
OR 8	Neurology - General	CLOSED	Neurology - General	CLOSED	CLOSED	CLOSED	Neurology - General	CLOSED	Surgery - General	Neurology - General

Fig. 2: MSS currently in use in the visited hospital

The planning horizon of the MSS, which is repeated over time, consists of a limited set of days I (index i). The OR department disposes of a set of operating rooms K (index k). A set of surgeons S (index s), which may also refer to a group of surgeons, is allocated OR time in these rooms to perform surgical cases. These surgical cases involve a set of surgery types T (index t). The binary parameter b_{st} indicates if surgeon s performs surgeries of type t. The organisation of the OR department follows a block booking framework. This implies that an OR day comprises of a set of OR time blocks J (index j), which embody a large span of OR time that is exclusively reserved for an individual surgeon or a group of surgeons related to a specific medical discipline. Surgeons may schedule a series of surgical cases in their assigned block, which can be finished within the associated regular block time. In the visited hospital, blocks of 4 hours are allocated to a specific (group of) surgeon(s). The set O defines the available OR blocks indicated by the tuple (day i, period j, room k) that are closed in the current MSS as specified by the visited hospital. After surgery, the patients are transferred to a set of downstream ward units W (index w), allowing the recovery and further care for patients in the hospital. In the following, we discuss the relevant constraints (Section 3.2.1) and objectives (Section 3.2.2) considered to construct the cyclic MSS for the visited hospital. We make use of the following notation to describe the problem under study:

Sets and indices

Y	The set of master surgery schedules (index y)
W	The set of hospital wards (index w)
S	The set of surgeons (index s)
T	The set of surgery types (index t)
I	The set of days relevant in the cyclic MSS horizon (index i)
J	The set of time blocks in an OR day (index j)
K	The set of ORs (index k)
O	The set of closed OR blocks in the current MSS, indicated by the tuple (day i, period j, room k)

Parameters

A_{sij}	1, if surgeon s is available on day i during time block j; 0, otherwise
b_{st}	1, if surgeon s performs surgeries of type t; 0, otherwise
N_t	Number of required OR nurses to support surgeries of type t
N_{ij}	Number of OR nurses available on day i during time block j
R_s	Number of required time blocks for surgeon s
B_{wi}	Number of available beds in hospital ward w on day i
LOS^{max}	The maximum LOS measured by the hospital

Stochastic input variables

$\tilde{I}_t$	The number of OR patients per surgery type t treated in a single block
$\tilde{P}_t$	The transfer of an OR patient of surgery type t to a particular hospital ward
$\tilde{I}_{wt}$	The number of patients of surgery type t treated in a single OR block transferred to hospital ward w, determined by the function $\tilde{f}_1(\tilde{I}_t, \tilde{P}_t)$
$\tilde{I}'_{wj}$	The number of non-OR patients entering hospital ward w on day i
$L\tilde{O}S_{wt}$	LOS of a patient of surgery type t in hospital ward w

Decision variables

x_{sijk}	1, if surgeon s is assigned to day i, time block j, room k; 0, otherwise
$\tilde{I}_{wit}$	The number of OR patients of surgery type t entering hospital ward w on day i, determined by the function $\tilde{f}_2(y \in Y, \tilde{I}_{wt})$
$\tilde{D}_{wi}$	The number of occupied beds in hospital ward w on day i, determined by the function $\tilde{f}_3(L\tilde{O}S_{wt}, \tilde{I}_{wi^*t}, \tilde{I}'_{wi^*})$ with $i^* \leq i$
T^+_{sijk}	1, if morning block (i, j, k) assigned to surgeon s is not followed by an assignment to the consecutive afternoon block (i, $j+1$, k); 0, otherwise
T^-_{sijk}	1, if afternoon block (i, $j+1$, k) assigned to surgeon s is not preceded by an assignment to the morning block (i, j, k); 0, otherwise
G^+_{wi}	Amount of overused beds reflecting the additional beds required to meet demand in ward w, day i
G^-_{wi}	Amount of unused beds in ward w, day i
L^+_{wi}	Positive slack difference in number of required beds between subsequent days i and $i+1$ in ward w
L^-_{wi}	Negative slack difference in number of required beds between subsequent days i and $i+1$ in ward w

Objective Function Values

T'	Objective function value related to the maximisation of the adjacency of OR blocks
G'	Objective function value related to the maximisation of the bed usage in ward units
L'	Objective function value related to the bed usage levelling

3.2.1 MSS Restrictions

In this section, we discuss relevant types of constraints involving (i) the internal organisation of the OR department and the assignment of surgeons to blocks and (ii) the bed usage in the downstream ward units. These constraints describe the rules and relationship between the different types of inputs to define a feasible MSS for the visited hospital.

Constraints related to the internal organisation of the OR department

The MSS should ensure different assignment restrictions. First, an OR time block j in room k on day i can only be allocated to at most one surgeon and a surgeon i can be assigned to at most one OR during a particular block j on day i, avoiding the assignment of surgeons to parallel blocks. The surgeon availability, indicated by A_{sij}, may prevent the assignment of an OR block to surgeon s when the timing of block j on day i does not fit the agenda of the surgeon (e.g. due to consultations planned). Some pre-specified OR time blocks (i, j, k) are closed and cannot be allocated to surgeons, as stipulated by the current MSS in the hospital. For example, the constraint assures that on Monday OR 2 is closed during the first time block of the visited hospital. Further, the allocated number of blocks to a particular surgeon (group) s should be aligned with the allotted capacity R_s determined in the strategic case-mix planning. In addition to the limited amount of OR time and surgeon availability, the allocation of OR blocks is guided by the nurse schedule. The latter defines the amount of available OR nurses N_{ij}, scheduled to assist the planned surgeries during a particular block j on day i. This available number of nurses N_{ij} during a block restricts the block allocation of the MSS as the nurses should be able to cope with the workload associated with the surgery types planned in a specific block, indicated by the required number of nurses N_t.

Constraints related to the bed usage in the downstream ward units

The MSS impacts the number of patients admitted to the downstream ward units and, correspondingly, the number of beds occupied over the time horizon in these wards. The number of patients in a particular ward w on day i is restricted by the number of available beds B_{wi} in the ward. In the following, we specify the relationship between the bed usage and the MSS, which defines the moment (day) patients of a particular surgery type are treated. To that purpose, we first explain the characterisation of the number of incoming patients for a particular ward and day. Second, we discuss the bed usage taking the patients' LOS into account.

Calculation of number of incoming patients for a particular ward and day - The number of patients $\tilde{I}_{wt}$, originating from a single OR block reserved for patients of surgery type t and transferred to ward w, is a random variable determined by a function with as arguments the number of patients $\tilde{I}_t$ treated in the OR block and the transfer of patients to the ward characterised by the variable $\tilde{P}_t$, i.e. $\tilde{I}_{wt} = \tilde{f}_1(\tilde{I}_t, \tilde{P}_t)$. The number of patients that are treated in a single OR block, i.e. $\tilde{I}_t$, is a random variable for which the underlying distribution is case-specific and the associated parameter values are dependent on the surgery type t. The transfer of an OR patient to one of the hospital wards after one surgery block is determined by a random variable $\tilde{P}_t$, which is dependent upon the surgery type t treated in the considered block. $\tilde{P}_t$ is characterised by a discrete distribution with $P(\tilde{P}_t = w)$, indicating the probability a patient of surgery type t is transferred to ward w. This distribution reflects the outcome of the heuristic rules utilised by staff to allocate patients to wards. The sum of probabilities associated with the transfer of a patient of a specific surgery type over all wards in the hospital is equal to 1, i.e. $\sum_{w \in W} P(\tilde{P}_t = w) = 1$ $(\forall t \in T)$. Based upon these random variables and the surgeon assignments in the MSS, the OR patients $\tilde{I}_{wit}$ of surgery type t entering ward w on a particular day i is determined by the function $\tilde{I}_{wit} = \tilde{f}_2(y \in Y, \tilde{I}_{wt})$. Note that patients enter a ward on the day of surgery in the OR department. Apart from the patients treated in the OR department, other patients may enter the wards on a particular day, i.e. the so-called non-OR patients, which also impact the bed usage in the wards. The number of non-OR patients $\tilde{I}'_{wi}$ is a random variable that indicates the number of patients entering ward w on day i. As a result, the total number of incoming patients in ward w and day i comprehends the summation of the calculated incoming OR patients and the non-OR patients, i.e. $\tilde{I}_{wit} + \tilde{I}'_{wi}$. The involved random input variables $\tilde{I}_t$, $\tilde{P}_t$ and $\tilde{I}'_{wi}$ can be characterised based on the real-life observations and input data from the visited hospital.

Calculation of bed usage for a particular ward and day - In order to estimate the patient's demand for care in a particular ward w and day i, we need to take the patients' LOS into account, which is defined as the average number of days patients spend in a hospital. In this paper, the LOS is considered as one of the main sources of uncertainty, which is primarily dependent on the department a patient is transferred to and the conducted type of surgery. Following the research of Gallivan and Utley (2005) and Beliën and Demeulemeester (2007), the LOS for a patient is modelled as a stochastic variable $L\tilde{O}S_{wt}$, characterised by a multinomial probability distribution with $P(L\tilde{O}S_{wt} = n)$ indicating the probability a patient undergoing a surgery of type t entering ward w and is still resident n days later. This random variable can be characterised based on real-life input data observed in the visited hospital. For example, the probability that a patient in the department 'Short stay' is resident 1 day (2 days) later after entering this department is 0.94 (0.36). Consequently, the total number of occupied beds $\tilde{D}_{wi}$ in ward w on day i is a random variable determined by a function, i.e. $\tilde{D}_{wi} = \tilde{f}_3(L\tilde{O}S_{wt}, \tilde{I}_{wi^*t}, \tilde{I}'_{wi^*})$ $(\forall i^* \leq i \in I)$, which is dependent upon the patients' LOS, the incoming number of OR patients of surgery type t and the non-OR patients entering ward w on day i^*, prior to day i, who are still resident $i - i^*$ days later, at day i, in hospital ward w.

This calculated number of occupied beds per ward and day allows to evaluate the performance of a particular

MSS based on the overall bed capacity utilisation rate C, which is a relevant key performance indicator (KPI) for the hospital. To calculate this indicator, we need to determine first the bed occupancy rate C_{wi} of a hospital ward w and day i by dividing the realised demand per hospital ward on a particular day by the number of justified hospital beds. The bed occupancy rate for a particular day (C_i) is determined by averaging these capacity rates over the wards. The overall bed capacity rate C is obtained by averaging the daily utilisation rates over the days of the week.

3.2.2 MSS Objectives

Optimising the MSS often involves multiple stakeholders, which postulate different criteria to evaluate a MSS. The visited hospital considers three criteria to construct an appropriate MSS, which are explained in detail in the following paragraphs. Note that the criteria are presented in order of importance. In other words, in correspondence with the hospital's stakeholders, the criterion of maximising the adjacency of OR blocks is identified as the most important. The other two objectives are more or less even important and no priority has been distinguished.

Objective 1: Maximising the adjacency of OR blocks
The first optimisation criterion aims to maximise the adjacency of OR blocks for the surgeons. In other words, we minimise the number of changes related to the surgeon assignments in the MSS, i.e. when the surgeon assignment of the morning block differs from the surgeon assignment of the afternoon block. In the visited hospital, blocks with a duration of 4 hours are allocated to different surgeons or surgeon groups. When maximising the adjacency of blocks on the same day, surgeons are working the entire daily horizon, composed out of two consecutive blocks. The allocation of an OR during an entire day (i.e. a so-called OR day) to a surgeon (group) has been observed in different papers as the common practice when constructing a MSS (e.g. Cardoen et al., 2009; Ma and Demeulemeester, 2013; Fügener et al., 2014; Marques et al., 2019; Schneider et al., 2020), modelling the allocation of consecutive blocks as a hard constraint (Mannino et al., 2012). The allocation of an entire OR day to a surgeon group or maximising the adjacency of consecutive blocks can be primarily motivated on organisational grounds in order to minimise the divergence in the OR schedule or to maximise the efficiency, taking considerations into account such as e.g. the requirements for specialised equipment and (multi-employable) personnel (Van Houdenhoven et al., 2008), reducing surgeon waiting time and idle time (Choi and Wilhelm, 2014), avoiding (long) changeover times between different surgery types (Penn et al., 2017) and increased flexibility to schedule surgical cases with a longer duration (Roland and Riane, 2011). Similar findings have been presented by e.g. Van Houdenhoven et al. (2008) and Hans et al. (2008) involving the operational scheduling of surgical cases. They state that the OR efficiency can be further increased when lengthening the duration of OR blocks or combining subsequent OR blocks, as similar case types can be treated together on the condition that the patient demand is large enough. An analogous argumentation has been put forward by the surgeons of the visited hospital. The hospital's surgeons prefer to operate for a whole day compared to a single morning or afternoon block as their individual schedule is better aligned, facilitating the planning of non-operating activities such as patient consultations. In addition, maximising adjacency of OR blocks reduces the sharing of ORs, avoiding possible delays for the surgeon assigned to the afternoon session when the surgeries in the morning block take more time than expected.
In this way, setting a proper duration for an OR block embodies a trade-off for a hospital between the OR efficiency and the flexibility to schedule surgical cases from multiple specialties in the same OR on the same day in line with their patient demand. The visited hospital stipulated the policy of having 4 hours as the minimum length for an OR block, which implies that not all surgeon(s) (groups) need to be allocated to work an entire OR day, embodying a higher granularity and thus flexibility in the allocation of OR blocks. However, by defining the objective to maximise the adjacency of blocks, the hospital considers the OR efficiency as well when constructing a MSS. This allows the scheduling of surgeons to consecutive morning and afternoon OR blocks, which lengthens the duration of OR time allotted to a surgeon to an entire OR day.

Objective 2: Maximising the usage of the bed capacity
The second optimisation criterion optimises the usage of the bed capacity, focusing on both the unused bed capacity and the additional number of beds required to meet demand. In other words, we minimise the difference between the occupied bed capacity and the available (justified) bed capacity.

Objective 3: Levelling the usage of the bed capacity
The third optimisation criterion aims to level the bed usage over the days of the planning horizon and to avoid severe jumps in the bed occupancy on subsequent days of the week. An unbalanced ward occupancy makes staff scheduling and ward operations difficult (Hall, 2012). The latter is confirmed by the fact that the nurse schedule in the visited hospital is organised in a cyclic manner and the planned number of nurses

is more or less levelled over the days. A more levelled workload ensures a better alignment with the nurse schedule. Note that this objective is complementary to the second objective as a (too) high bed usage on a particular day may increase the possibility for a capacity shortage in a particular week due to operational uncertainty (Houdenhoven et al., 2008).

3.3 Mathematical Formulation

In this section, we provide a mathematical MIP formulation of the master surgery scheduling problem under study. This optimisation problem is inherently stochastic due to the presence of uncertainty related to the number of patients treated in a single OR block ($\tilde{I}_t$), the transfer of patients of a particular surgery type to the wards ($\tilde{P}_t$), the amount of non-OR patients admitted to a particular ward ($\tilde{I}'_{wi}$) and the patients' LOS ($L\tilde{O}S_{wt}$). The formulation relies on the assignment variables x_{sijk}, which indicate the allocation of surgeon s to day i, block j and OR k, and different auxiliary variables to calculate the objective function values and the resulting workload.

Mathematical problem formulation

$$\text{Minimise} \quad T' = \sum_{s\in S}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}(T^+_{sijk} + T^-_{sijk}) \tag{1}$$

$$\text{Minimise} \quad G' = \sum_{i\in I}\sum_{w\in W}(G^+_{wi} + G^-_{wi}) \tag{2}$$

$$\text{Minimise} \quad L' = \sum_{i\in I}\sum_{w\in W}(L^+_{wi} + L^-_{wi}) \tag{3}$$

subject to

$$\sum_{s\in S} x_{sijk} \leq 1 \qquad \forall i \in I, \forall j \in J, \forall k \in K \tag{4}$$

$$\sum_{k\in K} x_{sijk} \leq 1 \qquad \forall s \in S, \forall i \in I, \forall j \in J \tag{5}$$

$$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} x_{sijk} = R_s \qquad \forall s \in S \tag{6}$$

$$\sum_{s\in S} x_{sijk} = 0 \qquad \forall (i,j,k) \in O \tag{7}$$

$$\sum_{s\in S}\sum_{k\in K}(b_{st} \cdot x_{sijk} \cdot N_t) \leq N_{ij} \qquad \forall i \in I, \forall j \in J \tag{8}$$

$$\sum_{k\in K} x_{sijk} \leq A_{sij} \qquad \forall s \in S, i \in I, j \in J \tag{9}$$

$$x_{sijk} - x_{si(j+1)k} \leq T^+_{sijk} \qquad \forall s \in S, i \in I, j \in J, k \in K \tag{10}$$

$$-x_{sijk} + x_{si(j+1)k} \leq T^-_{sijk} \qquad \forall s \in S, i \in I, j \in J, k \in K \tag{11}$$

$$\tilde{D}_{wi} - B_{wi} \leq G^+_{wi} \qquad \forall w \in W, \forall i \in I \tag{12}$$

$$B_{wi} - \tilde{D}_{wi} \leq G^-_{wi} \qquad \forall w \in W, \forall i \in I \tag{13}$$

$$\tilde{D}_{wi} - \tilde{D}_{w,i+1} \leq L^+_{wi} \qquad \forall w \in W, \forall i \in I \tag{14}$$

$$\tilde{D}_{w,i+1} - \tilde{D}_{wi} \leq L^-_{wi} \qquad \forall w \in W, \forall i \in I \tag{15}$$

$$T^+_{sijk}, T^-_{sijk} \text{ binary} \qquad \forall s \in S, i \in I, j \in J, k \in K$$

$$G^+_{w_i}, G^-_{wi} \geq 0 \qquad \forall w \in W, \forall i \in I$$

$$L^+_{wi}, L^-_{wi} \geq 0 \qquad \forall w \in W, \forall i \in I$$

$$x_{sijk} \text{ binary} \qquad \forall s \in S, i \in I, j \in J, k \in K \tag{16}$$

Equations (1)-(3), together with the goal constraints (10)-(15), represent the mathematical formulation of the different objective function criteria. The *first optimisation criterion* (eq. (1)) aims to maximise the adjacency of OR blocks for the surgeons. In other words, we minimise the number of changes related to the surgeon assignment in the MSS between consecutive assignments on a particular day, which are modelled via the goal constraints (10)-(11) and the binary auxiliary variables T^+_{sijk} and T^-_{sijk}. Constraints (10) and (11) determine the value of T^+_{sijk} and T^-_{sijk}, respectively. If surgeon s operates in the same OR k during the whole day (i.e. the subsequent morning and afternoon blocks are assigned to the same surgeon in an OR), the values of the variables T^+_{sijk} and T^-_{sijk} will be equal to 0. However, if this is not the case, meaning that surgeon s does not operate in the same OR during subsequent time frames, either T^+_{sijk} or T^-_{sijk} will be equal to 1. Consequently, the first objective function (eq. (1)) tries to minimise the sum of the variables T^+_{sijk} and

T^{-}_{sijk} and, therefore, strives to schedule the same surgeon s in the same OR k during two subsequent time blocks j and $(j+1)$ on the same day i. The *second optimisation criterion* (eq. (2)) optimises the use of the bed capacity, focusing on both the unused and overused beds. In other words, we relax the bed capacity constraint, stipulating that the workload measured by the number of required beds should be lower than the number of available beds on any day and in any ward ($\tilde{D}_{wi} \leq B_{wi}, \forall w \in W, \forall i \in I$), and minimise the difference between the required (stochastic) bed usage and the available (justified) bed capacity summed over the days of the week and wards in the hospital. To that purpose, we include the goal constraints (12) and (13), giving insight in the positive absolute value of the difference between the number of occupied beds and the number of justified beds. The *third optimisation criterion* (eq. (3)) optimises the levelling of the bed usage over the days of the planning horizon. To model this objective, we include constraints (14) and (15), which calculate the (non-negative) workload difference in the wards on subsequent days. As a result, the objective minimises the differences in patient demand for care between two subsequent days summed over the wards and the days.

Constraints (4)-(9) embody the restrictions related to the internal organisation of the OR department and the assignment of surgeons to OR blocks, following the resource requirements and guidelines of the hospital. Constraint (4) stipulates that an OR k can only be occupied by at most one surgeon during a particular time block j on day i. Constraint (5) ensures that a particular surgeon s can be assigned to at most one OR on day i during a certain time block j. Constraint (6) indicates that each surgeon has a sufficient number of OR slots in accordance to the strategic case-mix planning. Constraint (7) imposes the closure of a certain OR, day and time block, following the current MSS in the hospital. Constraint (8) imposes that the number of requested nurses during a particular block is limited by the number of available OR nurses N_{ij} on day i during time block j as stipulated by the nurse schedule. Constraint (9) restricts the MSS by imposing that a surgeon cannot operate when he/she is not available. Constraints (16) denote the domain constraints for the different types of variables.

4 Solution Methodology

We propose a multi-stage solution methodology to find an appropriate MSS for the multi-objective optimisation problem under study. The problem is characterised by a large number of stochastic input variables related to the patient arrivals and LOS. The considered uncertainty directly impacts the performance of a MSS involving the bed capacity usage in the downstream units. The methodology builds upon a deterministic model that relies on the service level concept related to the number of patients treated in a single OR block. Applying this concept helps to establish suitable bed capacity buffers in the downstream ward units such that the MSS becomes rather insensitive to variations in the stochastic input variables (Section 4.1). In the visited hospital, a hierarchical ordering is present between the different objectives, such that we rely on a preemptive optimisation approach to solve the problem under study (Section 4.2). In a first stage, we seek for the optimal value of solely the first objective to maximise the adjacency of block assignments allocated to a particular surgeon. Note that the nature of this objective gives rise to a large number of alternative optimal solutions. In a second stage, the two objectives related to the usage of bed capacity in the downstream ward units, i.e. the optimisation and levelling of the bed capacity usage, are considered to build a MSS without degrading the objective value for the objective of the first stage. However, as this second stage contains two different objectives, the difficulty is to set appropriate weights and retrieve a balanced solution. For that purpose, we aim to construct a set of solutions lying on or close to the Pareto front using a multi-objective decomposition method, allowing the decision-maker to make a focused trade-off. A multi-objective optimisation approach offering an (approximated) Pareto front has not been explored before in the literature for the master surgery scheduling problem but accomodates the schedule acceptance by all parties. However, in order to support the hospital management in the selection of a schedule, we utilised two performance metrics for evaluating the different schedules and selecting the most suitable MSS for the hospital (Section 4.3), i.e. (i) the distance to the ideal point, a concept from the multi-objective optimisation literature, to measure the quality from a deterministic point-of-view; and (ii) the attained service level provided by the available bed capacity in the downstream ward units, to measure the quality from a stochastic point-of-view. Figure 3 provides the different steps of the solution methodology, represented via a flowchart. The software tools used to effectuate the proposed solution method are discussed in Section 4.4.

4.1 Formulation of a Deterministic Model using a Service Level Approach

To cope with the stochastic variables in the model (1)-(16), we rely on the service level concept to formulate a deterministic optimisation model, which will be used to define a suitable MSS that is rather insensitive to changes in the stochastic input variables. To that purpose, we do not rely on the expected or mean patient demand to calculate the expected bed capacity usage. Instead, we employ a higher-than-expected

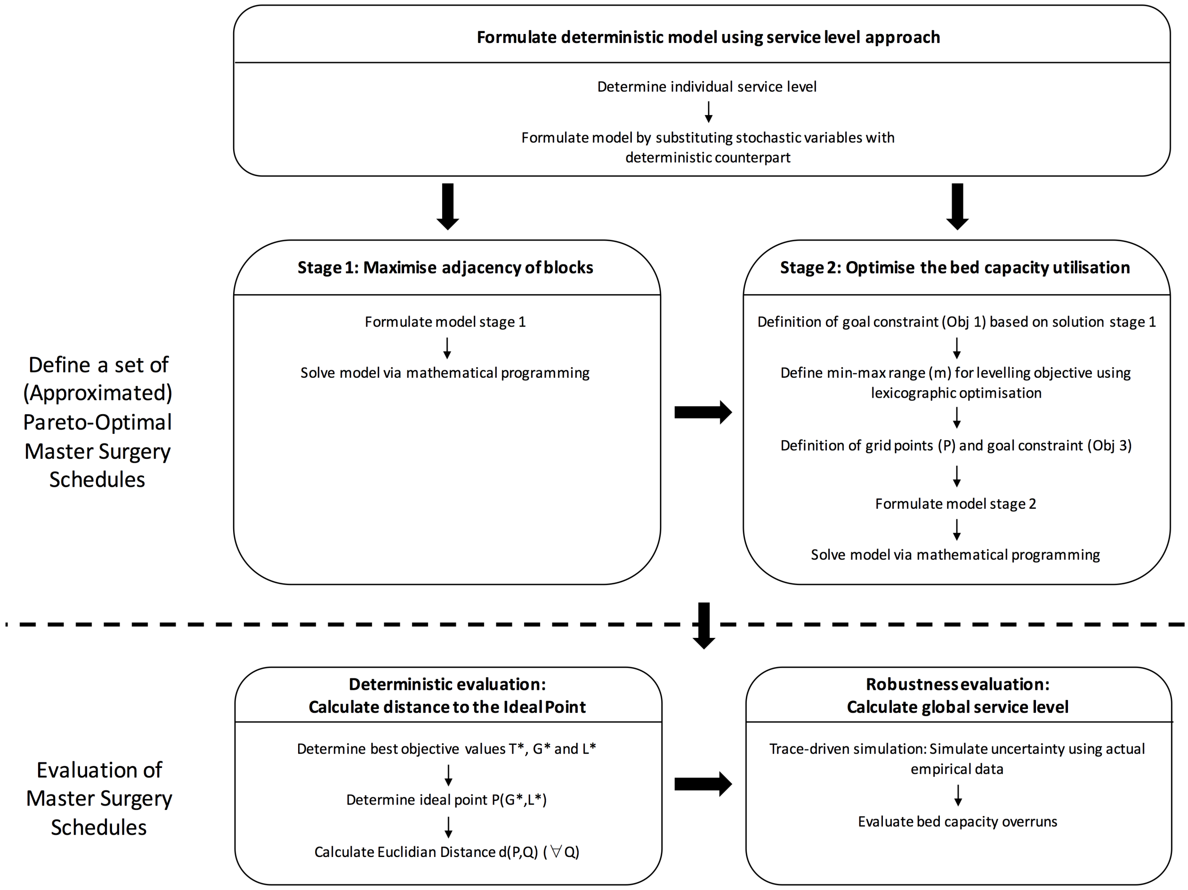

Fig. 3: Overview of the solution methodology

deterministic patient demand measured in terms of the number of patients, which relates directly to the bed capacity usage in the downstream ward units. In this way, an adequate bed buffer capacity is installed over the planning horizon, increasing the robustness of the MSS. The augmented patient demand is calculated in line with the strategic capacity decisions of the involved hospital, guided by the postulated service level. Based upon this deterministic patient demand, we formulate a deterministic optimisation model for the problem under study by calculating the associated bed capacity usage, which is a function of the allocations of surgeons to OR blocks. In the following, we first explain the underlying service level approach to calculate a higher-than-expected number of patients. Second, we discuss the calculations of the associated deterministic inputs for the different stochastic variables and formulate the problem as a deterministic optimisation model, which can be solved using mixed-integer programming.

Service level approach

In healthcare, the term 'service level' is often associated with the so-called service level agreement, which can be described as an agreement between the provider of a service and its customers which quantifies the minimum quality of service (Hiles, 1994). A service level is an important parameter in strategic capacity decisions, indicating the minimum level of scarce resources the hospital should guarantee to deliver satisfactory patient care (Berbe et al., 2008). A service level defines the probability of satisfying the service demand arising during a certain period given the limited available capacity (Nahmias and Olsen, 2015). This capacity is defined in accordance to the mean demand for service plus $F^{-1}(SL)$ times the standard deviation of the demand for service. $F^{-1}(SL)$ refers to the inverse of the cumulative standard normal distribution ($N(0,1)$) associated with the postulated service level (SL), following the Central Limit Theorem. The service level in the OR department refers to the provision of a particular level of care by a surgeon (group) who disposes of the necessary OR block time (capacity). This concept is reflected in terms of the number of patients that can be treated in a single OR block, which is relevant to measure the bed capacity usage in the downstream ward units. In order to construct a MSS, we consider the augmented patient demand $\hat{I}_t$ following this particular service level and calculate its repercussions on the usage of the bed capacity to allocate the OR blocks to surgeon(s) (groups). The augmented patient demand $\hat{I}_t$ is calculated via equation (17), adding a certain buffer to the mean number of patients $\bar{I}_t$ that can be handled during a certain OR time block. The size of this buffer depends on the provided service level. In Section 5.1.2, we identify the relevant service level per surgery type based upon the actual input data for the visited hospital. Similar calculations are conducted via equation (18) to calculate the augmented number of non-OR patients $\hat{I}'_{wi}$ that enter ward w on day i and are not directly transferred from the OR department. Via these calculations, we determine a higher-than-expected number of patients requiring bed capacity in the downstream ward units, which is input to the optimisation model to construct a robust MSS, creating an adequate bed buffer capacity.

$$\hat{I}_t = \bar{I}_t + F^{-1}(SL) \cdot \sigma_t \qquad \forall t \in T \tag{17}$$

$$\hat{I}'_{wi} = \bar{I}'_{wi} + F^{-1}(SL) \cdot \sigma'_{wi} \qquad \forall w \in W \tag{18}$$

with

$\hat{I}_t$	Number of patients of surgery type t that can be treated in one single OR block following a particular service level SL
$\bar{I}_t$	Average number of patients of surgery type t that can be treated in one single OR block
σ_t	The standard deviation related to the number of patients during a single OR block
$\hat{I}'_{wi}$	Augmented number of non-OR patients entering ward w on day i following a particular service level SL
$\bar{I}'_{wi}$	Average number of non-OR patients entering ward w on day i
σ'_{wi}	The standard deviation related to the number of non-OR patients entering ward w on day i
$F^{-1}(SL)$	Standard z-score calculated following the inverse of the cumulative standard normal distribution ($N(0,1)$) taking the postulated service level (SL) into account

Deterministic model

To formulate a deterministic model, deterministic inputs are required for the stochastic variables $\tilde{I}_t$, $\tilde{P}_t$, $\tilde{I}_{wt}$, $\tilde{I}'_{wi}$ and $L\tilde{O}S_{wt}$ to accurately calculate the variables $\tilde{I}_{wit}$ and $\tilde{D}_{wi}$, which are dependent upon the composition of the MSS. As indicated, the calculations to find a suitable MSS are based upon the augmented patient demand $\hat{I}_t$ and $\hat{I}'_{wi}$, replacing the respective stochastic variables $\tilde{I}_t$ and $\tilde{I}'_{wi}$ by a deterministic counterpart. As a proxy for the random variable $\tilde{I}_{wt}$, we utilise the deterministic number of OR patients $\hat{I}_{wt}$ of a particular surgery type t transferred to a ward w. This measure is calculated as a function of the augmented number of patients $\hat{I}_t$ treated in the OR block and the expected proportion of OR patients $P(\tilde{P}_t = w)$ of surgery type t transferred to ward w, i.e. $\hat{I}_{wt} = \hat{I}_t \cdot P(\tilde{P}_t = w)$. Based upon the number of OR patients $\hat{I}_{wt}$ and the surgeon assignments in the MSS, i.e. the decision variables x_{sijk}, we calculate the estimated inflow of OR patients $\hat{I}_{wit}$ of surgery type t in ward w and on day i via equation (19). The total number of incoming patients in ward w on day i comprehends the summation of the calculated incoming OR patients and the non-OR patients in the ward, i.e. $\hat{I}_{wit} + \hat{I}'_{wi}$. In order to calculate the total number of occupied beds $\hat{D}_{wi}$ on day i in ward w, we sum the number of patients of surgery type t entering ward w on day i^* prior to day i, i.e. $\hat{I}_{wi^*t} + \hat{I}'_{wi^*}$, multiplied by the expected proportion of patients $P(L\tilde{O}S_{wt} = i - i^*)$ that are still resident in hospital ward w $i - i^*$ days later at day i. The summation is done over a period of time with length LOS^{max}, indicating the maximum LOS. The relevant calculations are indicated in equation (20). The decision variables $\hat{D}_{wi}$ are used in the deterministic optimisation model as a proxy for the stochastic demand for beds $\tilde{D}_{wi}$.

$$\sum_{t \in T} \sum_{s \in S} \sum_{j \in J} \sum_{k \in K} b_{st} \cdot \hat{I}_{wt} \cdot x_{sijk} = \hat{I}_{wit} \qquad \forall w \in W, \forall i \in I \tag{19}$$

$$\sum_{i^* = i - LOS^{max}+1}^{i} (\hat{I}_{wi^*t} + \hat{I}'_{wi^*}) \cdot P(L\tilde{O}S_{wt} = i - i^*) = \hat{D}_{wi} \qquad \forall w \in W, \forall i \in I \tag{20}$$

with

$P(\tilde{P}_t = w)$	Expected proportion of OR patients of surgery type t originating from a single block transferred to ward w
$P(L\tilde{O}S_{wt} = n)$	Expected proportion of patients of surgery type t that are still resident n days later in hospital ward w
$\hat{I}_{wt}$	Deterministic proxy for $\tilde{I}_{wt}$, equal to $\hat{I}_t \cdot P(\tilde{P}_t = w)$
$\hat{I}_{wit}$	Deterministic proxy for $\tilde{I}_{wit}$ (eq. (19))
$\hat{D}_{wi}$	Deterministic proxy for $\tilde{D}_{wi}$ (eq. (20))

The deterministic model considers thus the different objectives (eqs. (1)-(3)) subject to the restrictions related to the internal organisation of the OR department (4)-(9), the goal constraints (10)-(15), the constraints related to the calculation of the deterministic proxy for the bed usage in the downstream ward units (17)-(20) and the domain constraints (16). In all these constraints, deterministic proxies are used as presented in this section, substituting their stochastic counterparts.

4.2 Define a Set of (Approximated) Pareto-Optimal MSSs

In this section, we describe the hierarchical two-stage methodology used to derive a set of relevant MSSs. In the first stage, we determine the optimal solution value for the objective related to the adjacency of OR blocks assigned to individual surgeons, which is discussed in Section 4.2.1. In Section 4.2.2, we consider the two objectives related to the bed capacity usage in the downstream units using the multi-objective optimisation $\mathcal{E}$-Constraint Method of Chankong and Haimes (2008), creating a set of (approximated) Pareto-optimal MSSs upon which the stakeholders can decide in a focused manner which schedule is preferred.

4.2.1 Stage 1: Maximise Adjacency of Blocks

In order to construct a set of MSSs, we determine in the first step of the solution methodology the optimal solution value for the objective related to the maximisation of the adjacency of OR blocks, preferably scheduling surgeons in the same OR during subsequent time blocks on the same day. We decided to consider this hierarchy in the objectives based on the following considerations, i.e. (i) the large priority given to this objective by the different stakeholders; (ii) the simplicity and comprehension of this approach for practitioners, improving their decision-making ability; and (iii) the large number of alternative solution points as a result of the objective function structure. To that purpose, we consider in this first stage objective function (1), goal constraints (10)-(11) and constraints related to the construction of a MSS (4)-(9), (17)-(20) and (16). We solve the resulting deterministic optimisation problem to optimality using a commercial IP software package.

4.2.2 Stage 2: Optimising the Objectives related to the Bed Capacity Usage

The second step in the solution procedure is more complicated since we are now considering two objectives simultaneously. We tackle this multi-objective problem using an a posteriori method, aiming to construct a set of Pareto-optimal solutions that are typically depicted using a Pareto front, which is a convex (concave) function in a minimisation (maximisation) problem. This set of solutions cannot be improved in one objective function without deteriorating the performance in the other objective. More precisely, we rely on the $\mathcal{E}$-Constraint Method proposed by Chankong and Haimes (2008), which optimises one of the objective function components using the other objective as constraint. To construct a set of solutions, we first derive a min-max range of values for every objective function. The best or max value of this range is easily computed as the optimum of the single objective optimisation problem. The worst or min value is obtained via the use of lexicographic optimisation, which has been implemented via a commercial MIP solver thriving on the deterministic model (eqs. (1)-(20)) proposed in Section 4.1, omitting objective (1) and goal constraints (10)-(11) related to the first stage. Based upon the min-max range for each objective, the decision-maker determines a set of relevant grid points P for each objective, equally distributed over the associated range. Note that for computational reasons, a limit may be imposed on the solution time to solve for a Pareto-optimal solution, such that the proposed methodology finds solutions lying on or close to the exact Pareto front.

In other words, the $\mathcal{E}$-Constraint Method constructs a set of solutions by considering the objective function (2), the associated goal constraints (12)-(13) and the other constraints (4)-(9), (16) and (17)-(20). Note that, in order to not deteriorate the optimal objective value related to the adjacency of OR blocks, constraint (21) is also added to the model, characterising the hierarchical component of this multi-objective solution methodology. T' indicates the objective value related to the adjacency of OR blocks obtained via the first stage model. Different solution points are obtained by relaxing the optimal objective function value for the levelling objective via the inclusion of constraint (22) and solving the relevant model to construct a MSS for different values of $p \in P$. The number of items in the set P is defined by the decision-maker. This decision will have impact on the size of the retrieved set of schedules since, in most cases, a new value of p will lead to a different solution. The parameter m depicts the distance between minimum and maximum objective function values related to the levelling objective, obtained via lexicographic optimisation. L'_{min} represents the best possible objective function value for the levelling of the workload. This method gives us the opportunity to define a concise set of (approximated) Pareto solutions, of which the solutions are derived stepwise, from fully focusing on objective (2) ($p = 1$) towards fully focusing on objective (3) ($p = 0$).

$$\sum_{s\in S}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}(T^+_{sijk} + T^-_{sijk}) = T' \qquad \forall w \in W \tag{21}$$

$$\sum_{i\in I}(L^+_{wi} + L^-_{wi}) \leq L'_{min} + m \cdot p \qquad \forall w \in W \tag{22}$$

4.3 Evaluation of MSSs

In this section, we provide objective measures to support the selection of the MSS for the involved hospital. We rely on two measures to assess the bed capacity usage of the downstream units in connection to the master surgery scheduling problem. The first measure values the bed capacity usage and its levelling over the days in a deterministic manner, calculating the distance to the ideal solution point, which is explained in Section 4.3.1. The second measure assesses the number of bed capacity shortages of the retrieved set of MSSs in a stochastic manner. The latter, which is discussed in Section 4.3.2, takes into account the uncertainty arising in the OR department and downstream units related to the demand for care.

4.3.1 The Distance to the Ideal Point

One method to objectively decide which MSS is the most preferred, is proposed by Zeleny (1975). According to this study, the process of selecting the 'most attractive' alternative out of a feasible set of solutions is based on the concept of the ideal alternative. This ideal alternative, which is in most cases an infeasible solution, provides the best possible solution quality with respect to the individual objectives, indicated by T^*, G^* and L^*. To that purpose, we can objectively measure which feasible solution lies the closest to the ideal point by calculating the Euclidian distance between the ideal point and the objective values related to a generated schedule lying on the (approximated) Pareto front, indicated by T', G' and L'. Equation (23) calculates the Euclidian distance between the ideal point $P(G^*, L^*)$ and a point $Q(G', L')$ lying on the (approximated) Pareto front. Note that we do not include T^* and T' in the expression of the Euclidian distance since constraint (21) assures that both values are equal.

$$d(P,Q) = \sqrt{(L' - L^*)^2 + (G' - G^*)^2} \tag{23}$$

However, the selection of a solution based on the procedure of the ideal point has one major drawback. Indeed, one may argue that if the scales of both objective functions are different, the distance to the ideal point will not reflect a balanced solution. For example, if the scale of objective value G is much smaller than the scale of objective value L, the solution point with the closest value to the optimal value of L will be selected as the 'most attractive' solution. To solve this issue, researchers have proposed several options (e.g. Ishizaka and Nemery, 2013). However, in consultation with the hospital we decided to use Zeleny's initial approach. In doing so, we assumed that the scales of the second objective function and the third objective function are comparable.

4.3.2 Robustness Evaluation

A second performance measure is related to the robustness of a MSS in order to come forward to operational uncertainty. As discussed in Section 3, the problem under study is complicated by many stochastic variables impacting the bed occupancy in the downstream ward units. This has been taken into account in the solution methodology by constructing MSSs following a deterministic optimisation model relying on a service level approach. The evaluation of the embedded robustness is based on a trace-driven simulation giving insight in the number of bed capacity shortages as would have been encountered in real-life. In contrast to the service level accounting for a higher-than-expected inflow of patients in the wards, employed to construct suitable MSSs, a global service level is used to evaluate the outcome on the bed occupancy after a particular MSS has been implemented. In order to define the global service level, we rely on a concept in inventory management and utilise the cycle service level to measure product availability (Chopra, 2019). Cycle service level is defined as the fraction of replenishment cycles that end with all customer demand being met (Chopra, 2019). In other words, it indicates the fraction of replenishment cycles without stock-outs. If we apply this definition to the healthcare sector, we could define the cycle (or global) service level as the fraction of periods (i.e. days or weeks) without capacity overruns. A capacity overrun occurs when the amount of scarce resources (hospital beds) is lower than the required number caused by patient demand. The bed capacity restriction is satisfied whenever the demand for bed capacity is lower than the number of available beds, which is calculated using an indicator function (eq. (24)). The global service level (GSL) represents the fraction no capacity overrun is observed over the number of days ($|N|$) considered in the horizon of the test dataset (eq. (25)), aggregated over the different wards.

$$\mathbb{I}_{(0,+\infty)}\left(g_i(\tilde{D}_{wi})\right) = \begin{cases} 0, & \text{if } \sum_{w\in W} \tilde{D}_{wi} - \sum_{w \in W} B_{wi} > 0 \\ 1, & \text{if } \sum_{w\in W} \tilde{D}_{wi} - \sum_{w \in W} B_{wi} \leq 0 \end{cases} \forall i \in I \tag{24}$$

$$GSL = \frac{1}{|N|} \sum_{i \in N} \mathbb{I}\left(g_i(\tilde{D}_{wi})\right) \tag{25}$$

Important to note is that, based on the terms discussed above, we can define a direct relationship between the individual service level of the different surgeons and the global service level of the hospital. In conclusion, the individual service level creates a buffer upon the estimated number of patient arrivals, i.e. the capacity provided to individual surgeons. Consequently, the MSS and the corresponding bed capacity in the downstream ward units will take a more-than-average number of incoming patients into account, and, will be, therefore, robust against shocks in patient demand and arrival for care. Note that the global service level is established not only via the available bed capacity but also via the composition of the MSS, impacting the possible number of capacity overruns.

Trace-driven simulation
To evaluate the global service level, we will make use of real-life data that is used directly in the simulation, which is referred to as trace-driven simulation. This approach increases the validation of the proposed methodology and the acceptance of the eventually retained schedule by the stakeholders (Servaux and Sörensen, 2004). This technique is already used by other researchers to test whether their approach can cope with real-life events (e.g. Lin et al., 2013). To that purpose, we retrieved data from the visited hospital related to the different sources of uncertainty surrounding the parameters $\tilde{I}_t$, $\tilde{I}'_{wi}$, $\tilde{P}_t$ and $\tilde{LOS}_{wt}$. In the experimentation to find the most suitable MSS, a distinction is made between the training dataset, i.e. the data related to the years 2017 and 2018, and the test dataset, i.e. the data related to the year 2019. The training dataset is implemented to build the model, calculating the mean input values for the relevant parameters (i.e. $\bar{I}_t$, $\bar{I}'_{wi}$, $P(\tilde{P}_t = w)$ and $P(\tilde{LOS}_{wt} = n)$), which are input to the deterministic model proposed in Section 4.1. The test dataset is used to validate the performance of the constructed MSSs. Indeed, instead of using estimates, we used the actual empirical hospital data for the first 30 weeks of 2019. In other words, $\tilde{I}_t$, amongst the other stochastic variables, shows the actual inflow of incoming patients for each surgery type. By using two different data sets, we lower any possible bias in the robustness evaluation of the selected schedule. This approach is often used in data mining and machine learning applications to verify a model's accuracy (e.g. Liu and Cocea, 2017). Note that an alternative approach to define the global service level is based upon Monte Carlo simulation relying on empirical or theoretical distributions. However, since we obtained a large test dataset, we rely solely on a trace-driven simulation and use this data to estimate the global service level for the involved hospital.

4.4 Implementation

To effectuate the proposed solution procedure, we used Microsoft Excel with the OpenSolver extension, which is coded in VBA. The embedded solver engine is the COIN Branch and Cut solver. The Excel-based system can be easily integrated in the organisation's OR systems, enabling the construction of a MSS, which can be repeated easily by practitioners on a periodic basis. The tool consists, on the one hand, out of different data input sheets (see Section 5.1.1 for further details). On the other hand, we included a data processing sheet to prepare the data as input for the solver model, a solver sheet that implements the deterministic optimisation model and a simulation sheet to conduct the trace-driven simulation based upon the MSS retrieved via the solver sheet. Note that the steps executed on the data processing sheet, solver sheet and simulation sheet have been documented in Section 4. The reasons to implement the solution procedure in Microsoft Excel are twofold, i.e.

(i) As Microsoft Excel is a widespread software system, the results can be displayed in an understandable manner for practitioners and researchers. The suggested MSS and a graphical representation of the associated bed occupancy level during the week can be easily shown;

(ii) the ease of use of Microsoft Excel, which stimulates the comprehension of the process. Most optimisation software can be seen as a 'black box' approach to solve a problem. By using Microsoft Excel, the user can see how every constraint is linked to the other. In other words, using Microsoft Excel can be seen as a 'white box' approach and practitioners can also easily learn to use the software such that they are able to easily upload new data and make changes to the model when necessary.

5 Computational Experiments

In this section, we effectuate the solution and evaluation procedure proposed in Section 4. In Section 5.1, we describe the input data retrieved from the hospital. Section 5.2 shows the validity of our approach using real-life data, generating a set of MSSs lying on or close to the Pareto front and assessing the quality of these schedules in different ways. In addition, we give insight in the mutual impact of the postulated objectives on the performance of the devised MSSs from a deterministic and stochastic perspective. Although the block adjacency objective has been selected as the primary objective, we performed in Section 5.3 a sensitivity analysis and robustness evaluation for different levels of non-adjacent surgery blocks to verify how this impacts our analysis. Lastly, in Section 5.4, we discuss the managerial insights from our analysis.

5.1 Test Design

In this section, we identify the necessary data to generate a set of MSSs and select the most attractive MSS. In Section 5.1.1, we elaborate on the hospital-specific input data. Thereafter, in Section 5.1.2, we discuss in detail the different sources of uncertainty modelled via specific probability distributions. Note that all parameters and probability distributions are based upon actual data and practical guidelines we received from the visited hospital. These inputs for the optimisation model are described following the organisation of the data sheets in Microsoft Excel.

5.1.1 Hospital-Specific Data

The visited hospital is a small-to-medium sized hospital in Belgium, which has treated 10102 inpatients, 14447 outpatients and 16944 emergency patients during the year 2018. Patients can enjoy a large range of highly specialised quality care. The hospital employs 631 people, of which 92 are surgeons, which are organised into 12 groups (e.g. 'Endocrinology', 'Urology', 'Neurosurgery'). The hospital disposes of 8 operating rooms. Afterwards, these patients are further cared for in one of the 19 hospital wards (e.g. 'Long Stay', 'Paediatrics', 'Maternity', 'Geriatrics'). The average hospitalisation LOS in the hospital equals 6.62 days. According to Penn et al. (2017), there are five essential categories of input factors that should be obtained from the hospital under discussion to establish relevant MSSs, i.e. (i) the amount of available theatre time, (ii) the case-mix, (iii) the number of available hospital beds, (iv) the availability of surgeons, and (v) the availability of other resources. Each one of these categories are discussed in detail below, assigning real-life values to the required model parameters based upon the hospital-specific data of the visited hospital.

Available Theatre Time
Based upon the hospital guidelines, different rules are stipulated the set of potential MSSs should adhere to. First, the MSS should only plan surgeries from Monday until Friday. Consequently, the ORs are closed during the weekend. Second, we derived the number of ORs that are available. In the hospital under consideration, there are 8 ORs available during each time block. Note that on certain weekdays, some ORs are closed. Third, on each day, the MSS is divided into two blocks, namely a morning and afternoon block. Disciplines cannot be assigned to certain blocks as the MSS foresees some unused blocks to cope with non-elective OR admissions.

Case-Mix Planning
Since we focus on tactical decision-making, we assume that the decisions taken on a strategic planning level are fixed. Consequently, we consider the case-mix planning, i.e. the distribution of the available OR time among surgeons, as a given. The case-mix planning of the visited hospital can be found in the Online Appendix A. This table is based upon the number of operating blocks each surgeon occupies in the historical MSS.

Number of Available Hospital Beds
A third category we should consider is the number of available hospital beds. Important to note is that we make a distinction between the number of justified beds and the number of actual beds, as hospitals need to buffer for several types of uncertainty. The number of actual beds will be higher than the number of justified beds. Hospitals want to keep their number of actual beds as low as possible since they do not receive any funding for this type of beds on top of the justified beds. The number of justified hospital beds amounts to 240. When running the optimisation model to construct a MSS, we only consider the number of justified beds. This is because reducing the hospital's actual beds to the level of justified beds is the ultimate goal of every hospital manager. The number of actual beds is important to effectuate the robustness evaluation of the set of MSSs considered. The hospital under discussion administers 292 actual beds during the week. During the weekend, this number is slightly lower, amounting to 267, since the hospital ward 'Short Stay', which has 25 beds, is closed. An overview of the number of justified and actual hospital beds over the different wards can be found in the Online Appendix B and the Online Appendix C, respectively.

Availability of Surgeons
A fourth important input factor is the availability of surgeons, which is directly linked to his/her consultation planning. This means that a surgeon cannot operate while he/she has consultations planned, which is imposed by constraint (9) in the model formulation. The consultation planning of the surgeons has been retrieved and is assumed to be fixed.

Availability of Other Resources
The last parameter relevant for the redesign of the MSS is the availability of other resources. Although there are different types of other resources, we focus on the availability of OR nurses such that the MSS is aligned with the cyclic nurse schedule. Based upon the hospital guidelines, we assume that we need a fixed number of OR nurses to carry out a surgery of a particular surgery type. For example, surgeons belonging to the category 'General Surgery' need on average 2.5 OR nurses. To allocate ORs to a particular surgery type on a specific day and block, a sufficient number of OR nurses is needed for assistance. A summary of the OR nurse roster can be found in the Online Appendix D.

5.1.2 Characterisation of Uncertainty

In the following, we estimate the different model parameters necessary to construct a MSS, surrounding the different sources of uncertainty, i.e. (i) the admissions of elective and non-elective OR and non-OR patients, (ii) the transfer of these patients to the different hospital wards and (iii) the LOS of patients in a particular hospital ward, and characterise the associated probability distributions. Note that these estimations are based on input data related to the years 2017 and 2018. In order to evaluate the robustness of a MSS, we make use of the hospital's real-life data obtained during the first 30 weeks of the year 2019 and apply a trace-driven simulation approach. Hence, for this evaluation, we do not estimate the number of patient admissions based on a probability distribution and the individual service level but rely on the observed actual input data.

Admissions of OR and non-OR patients
To characterise the hospital admissions of elective and non-elective OR patients, we need to estimate the mean number of patients $\bar{I}_t$ treated in a single OR block and identify the individual service level SL associated with the capacity of a single OR block. These estimations are used to calculate the augmented number of patients $\hat{I}_t$, which is input to the optimisation model to construct a suitable MSS, creating an adequate bed buffer capacity to avoid capacity overruns. The empirical number of admissions are normally distributed, which is validated in both a graphical and statistical manner. Figure 4 shows the z-scores for the incoming OR patients per week. The black line defines the theoretical normal distribution. The graph reveals a similar shape between both curves, which has been verified via the Kolmogorov-Smirnov test (p = 0.770) using statistical software. Based upon this test result, we can state that the number of patient admissions follows a normal distribution. Following this distribution, the individual service level (SL) of the OR time capacity has been identified at a level of 55%. This means that, for each group of OR patients linked to a specific surgery type, the hospital foresees a buffer of scarce OR time to cope with an additional number of patients. This buffer equals the z-score $F^{-1}(0.55)$ of the standard normal distribution times the standard deviation of patient demand for that particular surgery type. This value for the individual service level is estimated based upon its resemblance with the actual hospital figures.

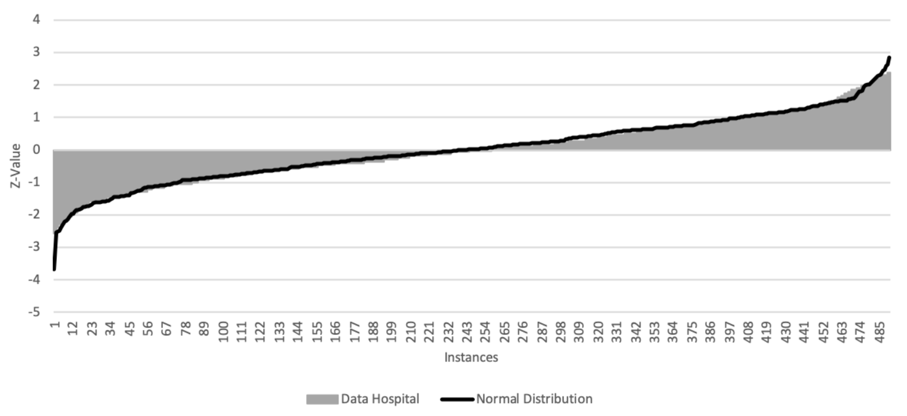

Fig. 4: Comparison between a theoretical normal distribution and the empirical z-values related to the incoming number of OR patients

Table 1 gives insight in the calculation of the augmented number of patients per OR block, following equation (17). The historical data delivered by the hospital related to the hospital admissions was accumulated per week and per surgery type. Consequently, calculations are conducted on the aggregated level per week and afterwards reduced to the augmented number of patients per OR block. The table displays the mean and standard deviation related to the incoming OR patients per surgery type and per week. The last two columns calculate the augmented number of incoming OR patients per week and per block. To derive the latter, we divide the weekly augmented number of patient admissions for a specific surgery type by the total number of OR blocks planned per week for a particular type of surgery. The total number of time blocks can be derived from the strategic case-mix decision (cf. Online Appendix A). For example, the surgery type 'Urology' disposes of six time blocks. This suggests that the total number of urology-related admissions is divided by 6 to calculate the number of incoming OR patients per block, assuming that for every OR block of the same surgery type, the number of treated patients is the same.

Table 1: Inflow of OR-related patients (2017-2018)

	Historical data		**Incoming # patients (SL = 55%)**	
Specialism	Mean/Week	St.Dev./Week	Inflow/Week	Inflow/Block ($\hat{I}_t$)
Urology	19.12	17.75	21.35	3.56
Gynaecology	7.31	6.50	8.12	1.02
Neurology	25.35	22.23	28.14	3.13
General Surgery	17.17	15.66	19.14	2.73
Plastic Surgery	1.88	2.28	2.17	0.72
Orthopaedics	27.19	24.46	30.27	2.33
Vascular Surgery	4.35	4.79	4.95	1.65
Anaesthesia	20.10	19.85	22.59	7.53
Endocrinology	27.23	24.49	30.31	3.37
Stomatology	4.27	4.85	4.88	0.81
ORL	6.56	6.80	7.41	1.85
Pulmonology	3.50	3.87	3.99	3.99

For the non-OR patients, the procedure is different since the care pathway for these patients is different. Therefore, instead of estimating the number of OR admissions per surgery type, we estimate the number of admissions per hospital ward. To do so, we made a distinction between week and weekend days because of the different pattern in the number of admissions. After dividing the non-OR patients into week and weekend groups, we can estimate the incoming flow of patients per hospital ward w. We again make use of a normal distribution with an individual service level of 55% to estimate the augmented number of non-OR patients per hospital ward w and day i (i.e. $\hat{I}'_{wi}$). An individual service level of 55%, which corresponds to a more-than-average number of non-OR patients per day, implies that an additional buffer is accounted for each hospital ward to cope with the oscillating behaviour of non-OR patient admissions. The values for $\hat{I}'_{wi}$ during week and weekend days can be found in Table 2 and Table 3, respectively. Note that we assumed that the number of admissions of non-OR patients for a particular hospital ward is equal for every day of the week or weekend.

Table 2: Inflow of non-OR-related patients during weekdays (2017-2018)

Hospital Ward	**Historical data**		**Incoming # patients**
	Mean/Day	St.Dev./Day	Inflow/Day ($\hat{I}'_{wi}$)
Long Stay	2.01	1.14	2.16
Short Stay	3.14	2.24	3.42
Intern A	3.99	2.07	4.25
Paediatrics	4.83	2.44	5.14
Intern B	5.62	2.50	5.93
Maternity	4.04	2.12	4.31
Geriatrics A	2.22	1.53	2.41
Geriatrics B	2.00	1.34	2.17
PAAZ	5.66	3.53	6.11
MIDC	1.98	1.02	2.11

Transfer of Patients to the Hospital Wards

A second input factor that is subject to uncertainty is the transfer of OR patients to hospital wards. In patients' care pathways, there is no one-on-one relationship between the type of surgery and the hospital ward they are transferred to after surgery in the OR. To that purpose, we use the historical data of the period 2017-2018 to determine the probability an OR patient is transferred to a certain hospital ward ($P(\tilde{P}_t = w)$). The discrete distribution probabilities for these patient transfers can be found per surgery type in the Online Appendix E.

Length of Stay

Due to a patient's clinical requirements, a patient's LOS in a certain hospital ward can vary significantly. Similar to previous studies in literature, we define, based on historical data, for each hospital ward and surgery type a multinomial distribution with probability $P(L\tilde{O}S_{wt} = n)$ associated with the patient's LOS.

Table 3: Inflow of non-OR-related patients during weekend days (2017-2018)

Hospital Ward	Historical data		Incoming # patients
	Mean/Day	St.Dev./Day	Inflow/Day ($\hat{I}'_{wi}$)
Long Stay	2.02	1.08	2.15
Short Stay	0.00	0.00	0.00
Intern A	2.61	1.51	2.80
Paediatrics	2.65	1.49	2.84
Intern B	3.82	2.14	4.09
Maternity	2.47	1.41	2.64
Geriatrics A	1.42	0.70	1.51
Geriatrics B	1.47	0.76	1.56
PAAZ	10.96	3.81	11.43
MIDC	1.75	0.86	1.86

The probability $P(\tilde{LOS}_{wt} = n)$ denotes the probability that a patient, undergoing a surgery of type t and entering hospital ward w on a particular day, is still resident n days later, where n is a strictly positive integer value. To use this method, we need to make two important assumptions. First, we expect that there is sufficient capacity available. Second, due to the cyclical nature of the MSS, we can also assume that the weekly or biweekly probabilities are cyclical. In other words, we can state that the probabilities remain the same during the whole planning process (Gallivan and Utley, 2005). An overview of the probability distribution functions associated with the LOS can be found per surgery type and hospital ward in the Online Appendix F.

5.2 Computational Results

In this section, we carry out the proposed procedure relying on the estimated input parameters to define the most suitable MSS for the visited hospital. In Section 5.2.1, we construct a set of MSSs by solving the problem according to the hierarchical two-stage methodology presented in Section 4.2. In correspondence with the hospital, we have decided to establish five different MSSs, including the extreme scenarios from a scenario fully focused on optimising the bed capacity usage to a scenario fully focused on levelling the utilised bed capacity. Although one may argue that establishing different MSSs enhances the negotiation process (Guerriero and Guido, 2011), the hospital's planning manager still has to decide which MSS is the most attractive one. The results to support this selection process are discussed in Section 5.2.2.
All tests were carried out on a MacBook Air with Intel Core i5 - 1.60GHz CPU and 4 GB RAM. The implemented models are characterised by a vast amount of constraints (14496) and decision variables (7830), which are determined by the dimensions of the problem. For computational reasons, we applied a time limit of 1800 seconds as stop criterion for the procedure to compose a MSS. If no schedule has been retrieved during this time span, we run the solver until the first (integer) schedule has been found. The latter was only required once and resulted in a very long computation time that has been caused as a result of the goal constraint imposed, which increases the empirical complexity significantly. This longer CPU time is not regarded by the hospital practitioners as an issue because of the longer-term horizon of the decision taken, which is typically a number of months. In addition, the practitioners agree that the advantages of the spreadsheet approach outweigh the advantages of a computationally more advanced approach. For the majority of the instances, this stop criterion was reached when applying the $\mathcal{E}$-Constraint Method in the second stage of the procedure. In this way, the retrieved set of schedules approximates the exact Pareto front. Note that any hospital manager can run the algorithm, if desired, without the time limit, such that Pareto-optimal solutions are yielded.

5.2.1 Construction of an approximated Pareto Front of MSSs

The first step in the solution procedure is to obtain a MSS that optimises the adjacency of OR blocks, clustering surgeries of an individual surgeon to the highest extent. Note that not all surgeries can be clustered since some surgeons have an odd number of scheduled time blocks or that a time block of a surgeon is preceded or followed by a closed block. The overall MSS contains 80 time blocks. However, we do know, based on the organisation of the current schedule, that during the week 8 blocks are closed. Therefore, we will assume that only 72 time blocks are distributed between the different surgeons when constructing the MSS. The MSS that is currently used, shown in Figure 2, counts 38 non-consecutive time blocks. This means that in 38 cases the allocated surgeon differs between the morning and afternoon time block. In other words, only 34 of all time blocks are clustered, which is equal to 47.2%. Instead of solving this problem manually, we used a mathematical programming approach following the model presented in Section 3.3. The resulting schedule, shown in the Online Appendix G, counts only 16 non-adjacent time blocks. Consequently, in 56 cases, the time blocks of a surgeon (group) were clustered, which is equal to 77.8%. Hence, the proposed formulation enables to increase the adjacency of time blocks by more than 30% compared to the current MSS.

The second step of the proposed solution procedure applies the $\mathcal{E}$-Constraint Method in order to consider the objectives related to the bed capacity usage and levelling. Note that, from now on, we will focus solely on the bed occupancy in the downstream units since the adjacency of the surgeon blocks is controlled in the second stage via constraint (21). First, we need to determine the value of p, which indicates the number of retrieved solutions. In our case, based on a discussion with the hospital's planning manager, we agreed that five different schedules should be sufficient. The distance m between minimum and maximum objective function values related to the levelling objective is calculated via lexicographic optimisation. In a first (extreme) scenario, we optimise solely the bed capacity usage by minimising the difference between occupied beds and justified beds, yielding an optimised objective function value for the capacity usage equal to 2567.83. This value is 17% lower than the value for the bed capacity usage (3113.16) obtained in the first stage of the hierarchical procedure, where we only focused on the adjacency of surgeon blocks. In this scenario, the value for the levelling objective amounts to 503.79, which is the maximum value for this objective using lexicographic optimisation. The second (extreme) scenario focuses only on the levelling of the bed capacity over the cyclic horizon and yields an objective value of 393.67 for this levelling objective, whereas the value for the bed capacity objective is obviously worse and equals 3506.15. Since the two objective functions conflict with each other, this observation seems logical. Based upon the observed maximum and minimum value for the levelling objective, we determine the distance m as $503.79 - 393.67 = 110.11$. To obtain five schedules lying on or close to the Pareto front, we impose constraint (22) and vary the parameter p from 1 to 0 in steps of 0.25. Following this procedure, we move from a MSS fully focusing on the bed capacity usage to a MSS focusing only on the levelling objective. The quality of the resulting MSSs are depicted in Figure 5. These MSSs were presented to the hospital manager. Note that, we can already conclude that the current MSS is not Pareto efficient since the solution can be improved in one objective without deteriorating the performance of the other objective.

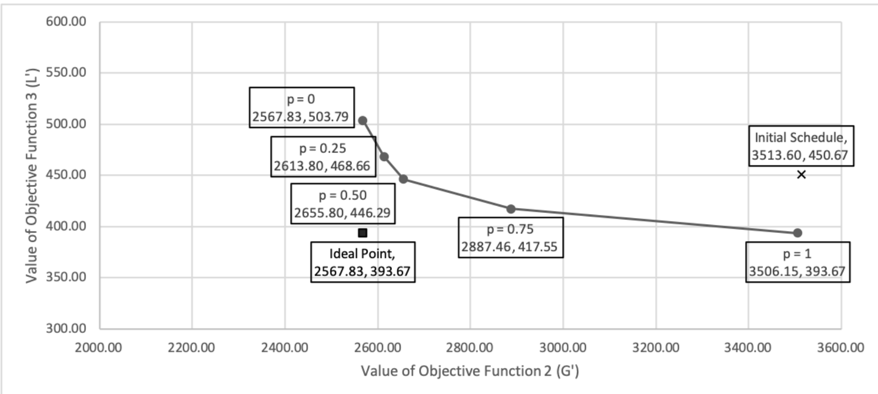

Fig. 5: Approximated Pareto front resulting from the proposed two-stage hierarchical optimisation method

The performance of these five MSSs is compared by looking at how the bed occupancy rate changes during the week (i.e. C_i) based on the input data related to the years 2017-2018. Visualising the bed occupancy rate of the obtained MSSs gives a first indication of which MSS is able to cope with the issues presented in Section 3.1. Figure 6 illustrates the bed occupancy rates of the two extreme scenarios. We can observe that both MSSs do balance the capacity usage in a more attractive way than the MSS that is currently used (cf. Figure 1). It should be clear that the new schedules balance the workload better by pushing workload units towards the end of the week. Since the hospital uses nearly the same amount of resources during week and weekend days, utilisation is better spread compared to the current scenario. Note that not all newly constructed MSSs are illustrated since their estimated occupancy graphs show exactly the same trends as the ones presented.

5.2.2 Selection of a MSS for Implementation

In this section, we discuss the results for two metrics to facilitate the decision-maker's choice of selecting a MSS. First, we discuss the distance to the ideal point in order to assess a MSS in an objective and unambiguous manner using the model input data of the years 2017 and 2018. Second, the selected MSS should be robust to the uncertainty surrounding the bed capacity utilisation in the downstream units,

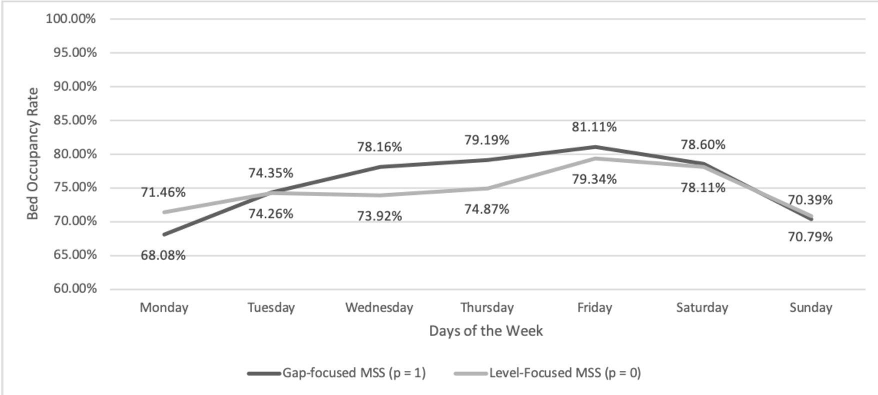

Fig. 6: Estimated bed occupancy rates of the two extreme cases based on model input data (years 2017-2018)

featuring a high global service level. In order to evaluate the robustness of the retrieved schedules, we use the test dataset related to the year 2019.

Distance to the ideal point

Since visual inspection is subject to interpretation, the graphs, which are shown in Figure 6, do not give a decisive answer to the question which MSS is most preferred. Therefore, this decision is objectified by calculating the distance to the ideal point, for which the coordinates correspond to the two (lowest) objective function values obtained via lexicographic optimisation. Hence, the coordinates of the ideal point are (2567.83,393.67), assuming that the capacity usage objective is represented on the x-axis and the levelling objective on the y-axis. Note that this ideal point is shown in Figure 5. This ideal solution will be infeasible since it lies beyond the efficiency frontier. Next, we calculate the Euclidian distance between the ideal point and the solutions on the approximated Pareto front, for which the results are displayed in Table 4.

Table 4: Distance to the ideal point for the set of retrieved schedules

Scenario	**Euclidian distance**
Gap Focused ($p = 1$)	110.11
$p = 0.75$	87.69
$p = 0.50$	102.50
$p = 0.25$	320.52
Level Focused ($p = 0$)	938.32
Current	947.48

Table 4 reveals that any schedule on the approximated Pareto front outperforms the current MSS. The distance between the current MSS and the ideal point is equal to 947.48. In addition, the schedule corresponding to $p = 0.75$ lies closest to the ideal point. Therefore, based on the assumption that the decision-maker is rational, we can predict that he/she will identify the corresponding MSS as the most attractive one. However, note that also other scenarios (e.g. $p = 0.50$ or $p = 1$) seem to be valuable alternatives.

Robustness evaluation using a service level approach

In this section, we validate the robustness of the identified set of schedules using a trace-driven simulation experiment. This simulation is based on the input of the real-life test data related to the year 2019. In this way, we verify the performance of the schedules obtained using the model input parameters estimated based on the training data related to the years 2017 and 2018. This means that, instead of using an aggregate demand for patient care with an individual service level of 55%, we will now use actual demand for patient care on a weekly basis. This demand will then be divided over the number of surgeons as defined by the case-mix. In other words, the parameters $\tilde{I}_t, \tilde{I}_{wt}, \tilde{I}'_{wi}, \tilde{P}_t$ and $L\tilde{O}S_{wt}$ will be interchanged with the actual values of the year 2019 and we evaluate the MSSs obtained in the previous section to test how the bed capacity utilisation reacts to new input.

To evaluate the performance of the newly established MSSs, we first assess the behaviour of the current MSS, which can be used as a benchmark. The trace-driven simulation using the input of the first 30 weeks from the year 2019 leads to a total of 17 cases of capacity overruns. A capacity overrun is here defined as the situation when the number of occupied beds exceeds the number of actual beds. The global service level (GSL) provided by the current MSS is assessed using equations (24) and (25) and is equal to 91.9%. Other relevant benchmarks, based on the trace-driven simulation experiment, are the obtained values for the bed capacity usage and levelling objectives, which are for the current schedule equal to 4382.72 and 577.84 respectively. These results for the current MSS reveal that the performance on the test data compared to the performance on the training data is inferior as both objective function values are significantly worse, which is obvious as the training data relies on estimated values, averaging out the variability in the real-life data, in contrast to the test data. Moreover, the simulation experiment shows that, compared to the results on the training data, the majority of the workload is shifted to the end of the week, which implies that the current MSS is heavily impacted by uncertainty, making the organisation of the OR department and the scheduling of resources very difficult.

These benchmarks give us an indication on the performance of the current MSS allowing the comparison with other schedules. In the following paragraphs, we make the comparison with three of the MSSs identified earlier, namely (i) the MSS focusing on the bed capacity usage objective ($p = 1$), (ii) the MSS focusing on the levelling objective ($p = 0$) and (iii) the 'most attractive' MSS that is closest to the ideal point ($p = 0.75$). Note that we also checked the performance of the other two schedules. However, since these schedules do not lead to substantial other findings, we opted not to discuss these scenarios in depth. The schedules associated with the below-discussed results can be found in the Online Appendix G. The observations are as follows:

(i) *The MSS focusing on the bed capacity usage objective ($p = 1$)*
This MSS is designed to minimise the difference between the number of occupied beds and the number of available beds. Figure 7 shows the average bed capacity utilisation rate for every day (i.e. C_i) of the cyclic horizon resulting from the trace-driven simulation experiment and makes the comparison with the current MSS. The resulting schedule ($p = 1$) leads to a better utilisation of the hospital's resources. Indeed, in contrast to the current schedule, the utilisation rate keeps increasing during the week to reach its peak on Friday. During the weekend, the utilisation rate decreases again due to the lower number of patient admissions. The objective value related to the bed usage (2959.07) is significantly lower compared to the value of the current MSS, whereas the levelling objective value is worse, amounting to 685.98. The latter is caused by quite significant changes in the workload from day to day (e.g. the bed occupancy level rises from 80.59% on Monday to 92.49% on Tuesday). In addition, during the thirty-week period, we counted 33 capacity overruns, which is significantly worse compared to the current MSS, which accounts for only 17 capacity overruns. Consequently, the global service level related to the gap-focused MSS is equal to 84.29%.

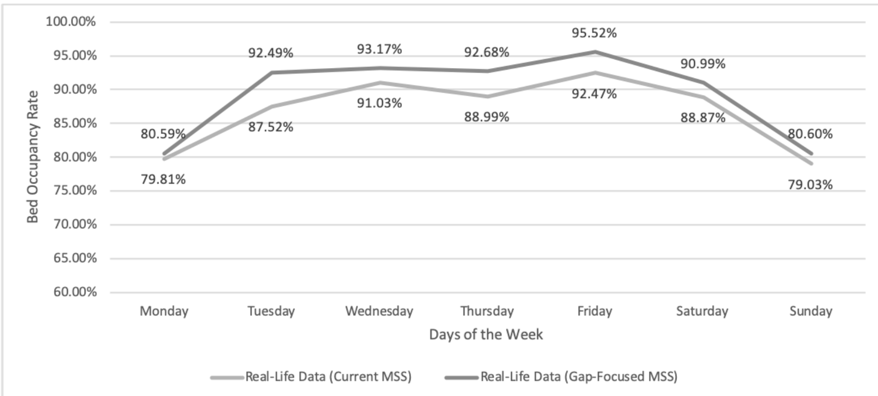

Fig. 7: Evolution of the bed utilisation rate of the gap-focused MSS ($p = 1$) using test data (year 2019)

(ii) *The MSS focusing on the levelling objective ($p = 0$)*
The resource levelling objective aims to ensure that the bed utilisation rate does not exhibit too large differences from one day to another. Figure 8 shows the average daily bed capacity utilisation rate (i.e. C_i) for the MSS ($p = 0$) and the current MSS. As a result of the resource levelling objective, the bed utilisation rate does not exhibit large jumps from one day to another on weekdays. The largest increase is 6.53% from Thursday to Friday, whereas, in the previous case of a schedule focusing on capacity usage ($p = 1$), the largest increase was 11.9%. In addition, the simulation experiment reveals that the bed capacity levelling reduces the number of capacity overruns to only 11. This is a relative reduction in capacity overruns of 35.3% compared to the current schedule, outperforming all other retrieved schedules on this criterion. Consequently, the global service level increases up to 94.76%.

(iii) *The 'most attractive' MSS that is closest to the ideal point ($p = 0.75$)*
From a deterministic point-of-view, the MSS keeping the balance between the capacity usage objective and the levelling objective with $p = 0.75$ is preferred as it lies closest to the ideal point. Figure 9 displays

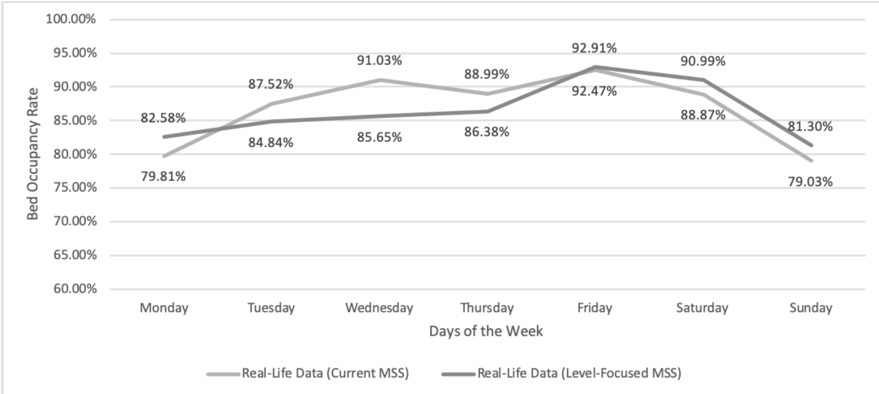

Fig. 8: Evolution of the bed utilisation rate of the level-focused MSS ($p = 0$) using test data (year 2019)

the capacity utilisation over the days for both the most attractive MSS and the current schedule. The former MSS ($p = 0.75$) demonstrates a worse levelling of the workload (623.08 vs 577.84) and leads to 21 capacity overruns or a GSL equal to 90.0%. Note that this is slightly higher than the number of capacity overruns obtained by the current MSS.

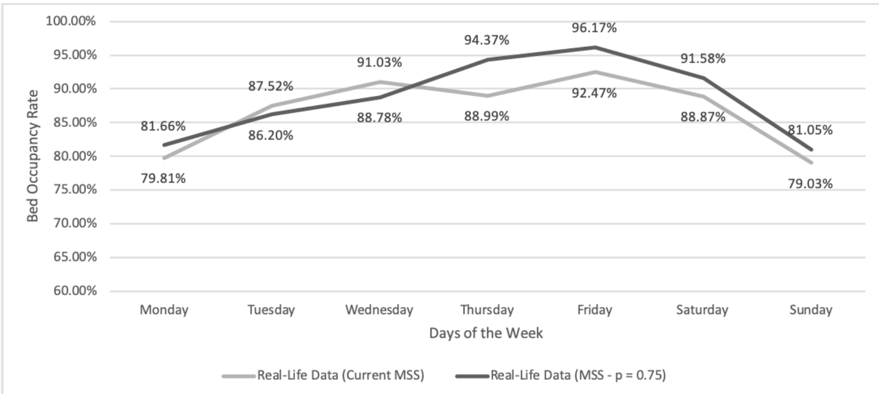

Fig. 9: Evolution of the bed utilisation rate of the most attractive MSS ($p = 0.75$) using test data (year 2019)

To sum up, we argue that solely relying on an objective measure to select a preferred MSS will lead to inferior decisions. As can be derived from the above-described results, the level-focused MSS is preferred if we only take GSL into account. Nevertheless, the distance to the ideal point can still be used to point the decision-maker in the right direction. For example, although the GSL of the 'most attractive' MSS ($p = 0.75$) is lower compared to the GSL of the level-focused MSS, the average bed occupancy rate of the former (88.55%) is higher than that of the latter (86.38%). In other words, although the GSL is slightly lower, resources (e.g. hospital beds) are deployed more efficiently. Therefore, it is important to use both methods in conjunction with each other in search for the most desired MSS.

5.3 Sensitivity Analysis of the Adjacency Criterion

In this section, we verify the impact of the case-specific objective to maximise the adjacency of OR blocks in the first hierarchical stage on the second and third objective criteria related to the capacity usage in the downstream units. The choice to define the adjacency of OR blocks as a preemptive objective was induced by the hospital's stakeholders. However, as discussed in Section 3.2.2, this choice embodies a trade-off for the hospital between OR efficiency and flexibility. Indeed, by defining the postulated objective to maximise the adjacency of blocks, the hospital wants to take into account efficiency considerations when constructing a MSS to allow the scheduling of surgeons to consecutive OR blocks. However, due to the lack of flexibility in the proposed schedules, the impact on the second and third criterion may be severe since matching supply and demand appropriately will be more complex.

In order to assess the impact of the adjacency objective on the other objective function criteria following the postulated hierarchical setting, we vary the required degree of adjacency of OR blocks according to different levels. More precisely, we determine the range for this adjacency criterion by minimising and maximising the adjacency of consecutive blocks, leading to respectively the minimum (adj^{min}) and the maximum (adj^{max}) adjacency. Based on this range and a stipulated value ADJ, we define different objective values for the adjacency criterion as follows:

$$T' = adj^{min} + ADJ \cdot (adj^{max} - adj^{min}) \tag{26}$$

For each value ADJ, we derive an approximated Pareto front of MSSs based on the bed capacity usage and levelling objective, yielded via the second stage of the solution methodology as proposed in Section 4. Note that in this approach equation (21) is substituted by equation (26) taking a specific value for the parameter ADJ into account. Following this approach, we first need to obtain the range for the adjacency criterion. As indicated in Section 5.2.1, the maximum number of adjacent surgery blocks is equal to 56 (adj^{max}). To determine the minimum number of adjacent surgery blocks, we apply the same mathematical formulation as defined in Section 3.3. However, instead of minimising equation (1), we maximise the number of non-adjacent surgery blocks. The maximum value we obtained was equal to 70, which means that the number of adjacent blocks is equal to $72 - 70 = 2$ (adj^{min}). Based on these values for adj^{max} and adj^{min}, the range for the adjacency criterion is equal to 54. Second, we need to define different (arbitrary) values for the variable ADJ. We decided to construct 5 different approximated Pareto fronts following five equidistant values for ADJ. The values for ADJ with their corresponding level of (non-)adjacent surgery blocks are summarised in Table 5. Figure 10 shows the approximated Pareto fronts for each value of ADJ. Again, we display the second objective function criterion on the x-axis, while the third objective function criterion is shown on the y-axis.

Table 5: Number of adjacency blocks for different levels of ADJ

Scenario	**Number of Non-Adjacent Surgery Blocks (T')**	**Number of Adjacent Surgery Blocks**
$ADJ = 1$	70	2
$ADJ = 0.75$	57	15
$ADJ = 0.50$	43	29
$ADJ = 0.25$	30	42
$ADJ = 0$	16	56

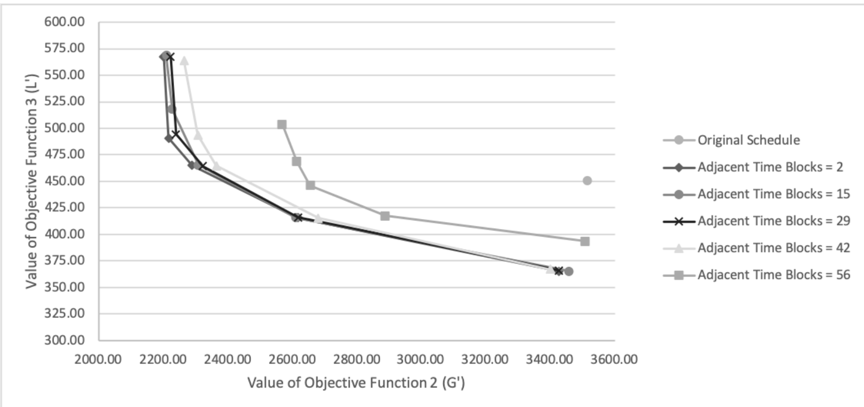

Fig. 10: Approximated Pareto fronts for different levels of ADJ

Figure 10 reveals that the yielded approximated Pareto fronts with a higher degree of adjacency, i.e. a lower value for ADJ, are dominated by fronts with a lower degree of adjacency. Hence, the proposed methodology, which first maximises the number of adjacent surgery blocks ($ADJ = 0$), leads to inferior values for the second and third objective in comparison to other, lower levels of adjacency. In general, a lower adjacency leads to a better performance in terms of bed occupancy usage and levelling. This appears to be logical since a higher extent of flexibility can be incorporated into the MSS, resulting in a better match between supply and demand in the downstream units. However, do note that this comes at a cost, i.e. a loss of OR efficiency (cf. Section 3.2.2). Therefore, Figure 10 is a good visualisation of the trade-off between OR efficiency and flexibility. Note that including a limited degree of flexibility ($ADJ = 0.25$), improves the bed capacity usage and levelling significantly. Additional benefits are decreasing and only marginal when further increasing this value of ADJ.

Apart from a deterministic analysis for different levels of ADJ, we also performed a stochastic evaluation of the retrieved MSSs to evaluate whether a higher level of flexibility, i.e. a lower postulated adjacency, would have a beneficial impact on the hospital's global service level. However, after performing the trace-driven simulation discussed earlier in this paper, we could not see any significant differences between the GSLs for different levels of adjacency. The only pattern we could identify is that the GSL of the gap-focused MSS is significantly worse than the GSL of the level-focused MSS, which confirms our robustness analysis performed based on a maximum of 56 adjacent surgery blocks (cf. Section 5.2.2).
Based on the fact that there is no significant difference between the GSLs for varying levels of adjacency, we can motivate the choice of maximising the number of adjacent blocks before optimising the other two criteria. However, if the hospital's planning manager requires a better alignment between supply and demand in the downstream units, flexibility can be induced into the model by decreasing the number of adjacent OR blocks using an appropriate value for ADJ.

5.4 Managerial Insights and Discussion

Table 6 gives an overview of the potential benefits regarding the proposed solution methodology. The results indicate that the newly calculated MSSs using an automated solution methodology perform better than the current MSS from a deterministic and a stochastic point-of-view. This results from the improvements related to the adjacency of surgeon blocks and the better spread of incoming OR patients. Via the clustering of OR blocks of a single surgeon, the hospital's stakeholders aim to increase OR efficiency and minimise the divergence such that the overall hospital performance will increase. In addition, the usage of historical data in combination with the two objectives related to the bed capacity utilisation in the downstream units also proved to be a significant asset to compute new MSSs. By using reliable estimates on the number of incoming patients and their LOS, the workload in the downstream units can be better planned, spreading the resource utilisation more efficiently and effectively. Moreover, if the daily utilisation levels are better balanced, peaks will occur less frequently, creating some buffer or spare bed capacity on each day to deal with unforeseen events. This justifies why the MSS focusing on the levelling objective has the highest level of robustness, i.e. the lowest number of bed capacity overruns. Moreover, a better performance in the downstream units with respect to the usage of the bed capacity and the workload levelling can be realised when the required degree of adjacency between OR blocks is lowered, making the trade-off between OR efficiency and flexibility. Note that most of the improvements can already be realised when introducing a limited amount of flexibility. However, the objective related to the degree of adjacency embedded in the MSS does not impact the schedule robustness related the usage of downstream resources. Based on all performance metrics, the MSS that lies closest to the ideal point is recommended. This schedule achieves the best deterministic performance and shows a large degree of robustness to operational variability related to the patient admissions. A large degree of robustness is correlated with a lower level of variation in the weekly bed occupancy levels of the different hospital wards, facilitating the coordination of scarce hospital resources. However, the hospital should be careful when implementing this proposed MSS. As discussed by van Oostrum et al. (2010), implementing a MSS should not happen from one day to the other. The implementation is a step-wise procedure for which continuous monitoring of the proposed benchmark performance measures, i.e. the bed capacity utilisation, bed capacity levelling and global service level, is necessary.
In addition, the approach can be utilised to gain insights on the bed capacity buffer realised in the hospital by the actual number of beds versus the justified number of beds. The number of justified beds seems insufficient to provide satisfactory care for the hospital's patients. If we only use the number of justified beds to evaluate the robustness of the set of MSSs, capacity overruns happen frequently. For example, if we use the number of justified hospital beds to assess the global service level for the schedule with $p = 0$, the global service level is equal to only 23.8%. Hence, the additional 52 hospital beds are of vital importance for the well-functioning of the visited hospital. Increasing or decreasing this capacity buffer will improve or deteriorate this service level.

Table 6: Overview managerial insights and consequences of the proposed methodology

Managerial Implications	**Consequences**
Automated Process	- simplifies the hospital's planning manager task of establishing a suitable MSS. - more efficient approach to attain certain objectives, e.g. increasing adjacent surgery blocks.
Usage of Historical Data	- historical insights can be used to spread the workload in the hospital wards, and thereby, increasing the bed occupancy rate. - peak workload can be forecasted, and the MSS can be adjusted appropriately. - including a levelling objective improves the schedule robustness related to the bed capacity usage in the downstream units.
Objective Decision-Making	- improves transparency and reduces conflicts between different stakeholders. - increases the sense of fairness between the involved stakeholders.
Tool for Richer Decision-Making	- scenario analysis, e.g. gaining insights on the bed capacity buffer. - aligning the different schedules used in the hospital context, e.g. MSS and nurse roster. - visualising the trade-off between OR efficiency and flexibility.

6 Conclusion

This paper presents a multi-objective MIP model and solution methodology to construct and evaluate MSSs under uncertainty. The optimisation model is based on the requirements of a private non-for-profit hospital and considers three optimisation criteria, i.e. (1) to optimise the adjacency of OR blocks for a particular surgeon in the same OR as much as possible, (2) to maximise bed capacity usage and (3) to level capacity-related workload in the downstream units. The problem is inherently a stochastic optimisation problem due to three different sources of uncertainty, i.e. patient admissions, the patient transfer between hospital wards, and the patient LOS, impacting the bed capacity usage in the downstream units. In order to improve the robustness of the yielded solutions regarding the (stochastic) bed usage, we embed adequate bed capacity buffers in the downstream units over the planning horizon. This is realised by formulating a deterministic model accounting for higher-than-expected proxy values to model the stochastic patient demand. These proxy values are determined based on the accounted service level, reflected in the higher-level strategic capacity decisions of the visited hospital. To support the negotiation process between the multiple stakeholders, a set of solutions are generated lying on or close the Pareto front by applying a multi-objective hierarchical solution procedure relying on the $\mathcal{E}$-Constraint Method. We sophisticated the process of selecting the most suitable MSS by calculating the distance to the ideal point and performing robustness evaluations related to the number of bed capacity overruns. The latter has been conducted by a trace-driven simulation experiment using a test dataset from real-life. The approach has been implemented using the widely accessible spreadsheet tool Microsoft Excel, which enhances the facile adoption of the proposed methodology by hospital practitioners.

An interesting topic for future research is to study the relationship between the individual service level associated with a single OR block and the global service level realised in the downstream units. In this way, the strategic service level decisions and the tactical decisions should be integrated to improve the organisation of the OR department in connection to the downstream units.

References

Abedini A, Li W, Ye H (2017) An Optimization Model for Operating Room Scheduling to Reduce Blocking Across the Perioperative Process. Procedia Manufacturing 10(859):60–70, DOI 10.1016/j.promfg.2017.07.022

Adan I, Vissers J (2002) Patient mix optimisation in hospital admission planning a case study. International Journal of Production and Operations Management 22(4):445–461

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

Agnetis A, Coppi A, Corsini M, Dellino G, Meloni C, Pranzo M (2012) Long term evaluation of operating theater planning policies. Operations Research for Health Care 1(4):95–104, DOI 10.1016/j.orhc.2012.10.001

Agnetis A, Coppi A, Corsini M, Dellino G, Meloni C, Pranzo M (2014) A decomposition approach for the combined master surgical schedule and surgical case assignment problems. Health Care Management Science 17(1):49–59

Banditori C, Cappanera P, Visintin F (2013) A combined optimization-simulation approach to the master surgical scheduling problem. IMA Journal of Management Mathematics 24(2):155–187

Bekker R, Koole G, Roubos D (2017) Flexible bed allocations for hospital wards. Health Care Management Science 20(4):453–466, DOI 10.1007/s10729-016-9364-4

Beliën J, Demeulemeester E (2007) Building cyclic master surgery schedules with leveled resulting bed occupancy. European Journal of Operational Research 176(2):1185–1204

Beliën J, Demeulemeester E (2008) A branch-and-price approach for integrating nurse and surgery scheduling. European Journal of Operational Research 189(3):652–668

Berbe RG, Gemmel P, Droesbeke B, Casteleyn H, Vandaele D (2008) Evaluation of hospital service level agreements. International Journal of Health Care 22(5):483–497

Blake JT, Dexter F, Donald J (2002) Operating Room Managers Use of Integer Programming for Assigning Block Time to Surgical Groups: A Case Study. Anesthesia & Analgesia 94(1):143–148, DOI 10.1213/00000539-200201000-00027

Bovim T, Christiansen M, Gullhav A, Range T, Hellemo L (2020) Stochastic master surgery scheduling. European Journal of Operational Research 285(2):695–711

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

Cardoen B, Demeulemeester E, Beliën J (2009) Sequencing surgical cases in a day-care environment: An exact branch-and-price approach. Computers and Operations Research 36(9):2660–2669

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

Cardoen B, Demeulemeester E, der Hoeven JV (2010b) On the use of planning models in the operating theatre: results of a survey in Flanders. The International journal of health planning and management pp 400–414

Chankong V, Haimes Y (2008) Multiobjective Decision Making: Theory and Methodology. Courier Dover Publications

Choi S, Wilhelm W (2014) On capacity allocation for operating rooms. Computers And Operations Research 44:174–184

Chopra S (2019) Supply Chain Management - Strategy, Planning, and Operation. Pearson, seventh Edition

Denton B, Miller A, Balasubramanian H, Huschka T (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. Operations Research 58(4):802–816

Deschodt M, Boland B, van Grootven B (2015) Report 245 from the Belgian Healthcare Knowledge Center (KCE): Comprehensive geriatric care in hospitals: the role of inpatient geriatric consultation teams. KCE Report pp 1–134

Figueira G, Almada-Lobo B (2014) Hybrid simulation optimization methods: A taxonomy and discussion. Simulation Modelling Practice and Theory 46:118–134

Fügener A (2015) An Integrated Strategic and Tactical Master Surgery Scheduling Approach With Stochastic Resource Demand. Journal of Business Logistics 36(4):374–387, DOI 10.1111/jbl.12105

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

Gallivan S, Utley M (2005) Modelling Admissions Booking in Elective In-patients into a Treatment Centre. IMA Journal of Management Mathematics 16(305):305–315

Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: a survey. Health Care Management Science 14(1):89–114

Hall R (2012) Handbook of Healthcare System Scheduling. Springer

Hans E, Wullink G, van Houdenhoven M, Kazemier G (2008) Robust surgery loading. European Journal of Operational Research 185(3):1038–1050

Hans EW, Houdenhoven MV, Hulsof PJ (2011) A Framework for Health Care Planning and Control. Springer

Hiles AN (1994) Service Level Agreements. The TQM Magazine 6(2):14–16

Houdenhoven MV, van Oostrum J, Wullink G, Hans E, Hurink J, Bakker J, Kazemier G (2008) Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. Journal of Critical Care 23(2):222–226

Hwang C, Masud A (1979) Multiple Objective Decision Making - Methods and Applications. Springer-Verlag Berlin Heidelberg

Ishizaka A, Nemery P (2013) Multi-criteria Decision Analysis: Methods and Software. Wiley

Khalfalli M, Abdelaziz F, Kamoun H (2019) Multi-objective surgery scheduling integrating surgeon constraints. Management decision 57(2):445–460

Kharraja S, Albert P, Chaabane S (2007) Block scheduling: Toward a master surgical schedule. Proceedings - ICSSSM'06: 2006 International Conference on Service Systems and Service Management 1(July 2016):429–435

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

Lamiri M, Xie X, Zhang S (2008) Column generation approach to operating theater planning with elective and emergency patients. IIE Transactions 40(9):838–852

Lawal AK, Rotter T, Kinsman L, Sari N, Harrison L, Jeffery C, Kutz M, Khan MF, Flynn R (2014) Lean management in health care: definition, concepts, methodology and effects reported. Systematic Reviews 3(103)

Li X, Rafaliya N, Baki M, Chaouch B (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

Lin G, Yi P, Si L, Zhu T, Jiang X, Li G, Begovic MM (2013) Robustness analysis on electric vehicle energy distribution networks. IEEE Power and Energy Society General Meeting pp 1–5

Litvak E, Long M (2000) Cost and Quality Under Managed Care: Irreconcilable Differences? The Amarican Journal of Managed Care 6(5):305–312

Liu H, Cocea M (2017) Semi-random partitioning of data into training and test sets in granular computing context. Granular Computing 2:357–386

Ma G, Demeulemeester E (2013) A multilevel integrative approach to hospital case mix and capacity planning. Computers and Operations Research 40(9):2198–2207, DOI 10.1016/j.cor.2012.01.013

Magerlein JM, Martin JB (1978) Surgical demand scheduling: a review. Health services research 13(4):418–433

Mannino C, Nilssen E, Nordlander T (2012) A pattern based, robust approach to cyclic master surgery scheduling. Journal of Scheduling 15(5):553–563, DOI 10.1007/s10951-012-0275-z

Marques I, Captivo E, Barros N (2019) Optimizing the master surgery schedule in a private hospital. Operations Research for Health Care 20(1):11–24

M'Hallah R, Visintin F (2019) A stochastic model for scheduling elective surgeries in a cyclic Master Surgical Schedule. Computers and Industrial Engineering 129(October 2017):156–168, DOI 10.1016/j.cie.2019.01.030

Min D, Yih Y (2010) Scheduling elective surgery under uncertainty and downstream capacity constraints. European Journal of Operational Research 206(3):642–652

Nahmias S, Olsen TL (2015) Production and Operations Analysis. Waveland Press, Inc., seventh Edition

Neyshabouri S, Berg B (2017) Two-stage robust optimization approach to elective surgery and downstream capacity planning. European Journal of Operational Research 260:21–40

OECD (2017) Length of Hospital Stay (Indicator). Accessed: 2020-03-12

van Oostrum J, Bredenhoff E, Hans E (2010) Suitability and managerial implications of a Master Surgical Scheduling approach. European Journal of Operational Research 178(1):91–104

van Oostrum JM, Houdenhoven MV, Hurink JL, Hans E, Wullink G, Kazemier G (2008) A master surgical scheduling approach for cyclic scheduling in operating room departments. Operations Research Spectrum 30:355–374

Penn ML, Potts CN, Harper PR (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, DOI 10.1016/j.ejor.2017.03.065

Pham DN, Klinkert A (2008) Surgical case scheduling as a generalized job shop scheduling problem. European Journal of Operational Research 185(3):1011–1025

Rachuba S, Werners B (2017) A Fuzzy Multi-Criteria Approach for Robust Operating Room Schedules. Annals of Operations Research 251:325–350

Roland B, Riane F (2011) Integrating surgeons' preferences in the operating theatre planning. European J of Industrial Engineering 5(2):232

Samanlioglu F, Ayag Z, Batili B, Evcimen E, Yilmaz G, Atalay O (2010) Determining master schedule of surgical operations by integer programming: A case study. IIE Annual Conference and Expo 2010 p 2010

Santibáñez P, Begen M, Atkins D (2007) Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. Health care management science 10(3):269–282, DOI 10.1007/s10729-007-9019-6

Schneider T, van Essen J, Carlier M, Hans E (2020) Scheduling surgery groups considering multiple downstream resources. European Journal of Operational Research 282(2):741–752

Schrijvers G, van Hoorn A, Huiskes N (2012) The care pathway: concepts and theories: an introduction. International Journal of Integrated Care 12

Servaux M, Sörensen K (2004) Robustness Analysis: Optimisation. EWG-MCDA Newsletter 3(10)

Spratt B, Kozan E (2016) Waiting list management through master surgical schedules: A case study. Operations Research for Health Care 10:49–64, DOI 10.1016/j.orhc.2016.07.002

Tànfani E, Testi A (2010) A pre-assignment heuristic algorithm for the Master Surgical Schedule Problem (MSSP). Annals of Operations Research 178(1):105–119

Toussaint JS, Berry LL (2013) The Promise of Lean in Health Care. Mayo Clinic Proceedings 88(1):74–82

Van Houdenhoven M, van Oostrum JM, Wullink G, Hans E, Hurink JL, Bakker J, Kazemier G (2008) Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. Journal of critical care 23(2):222–6, DOI 10.1016/j.jcrc.2007.07.002

Van Huele C, Vanhoucke M (2014) Analysis of the Integration of the Physician Rostering Problem and the Surgery Scheduling Problem. Journal of Medical Systems pp 38–43

Van Riet C, Demeulemeester E (2015) Trade-offs in operating room planning for electives and emergencies: A review. Operations Research for Health Care 7:52–69

Vanberkel P, Boucherie R, Hans E, Hurink J, van Lent W, van Harten W (2011) An exact approach for relating recovering surgical patient workload to the master surgical schedule. Journal of the Operational Research Society 62(10):1851–1860

Villarreal MC, Keskinocak P (2016) Staff planning for operating rooms with different surgical services lines. Health Care Management Science 19(2):144–169

Vissers J, Adan I, Bekkers J (2005) Patient mix optimization in tactical cardiothoracic surgery planning : a case study. IMA Journal of Management Mathematics 16(3):281–304, DOI 10.1093/imaman/dpi023

Walley P, Silvester K, Steyn R (2006) Managing variation in demand: lessons from the UK National Health Service. Journal for Healthcare Management 5(51):309–320

WHO (2020) Bed occupancy rate (in percentage), acute care hospitals only. Accessed: 2020-03-09

Zeleny M (1975) The Theory of the Displaced Ideal. Lecture Notes in Economics and Mathematical Systems pp 153–207

Zhang B, Murali P, Dessouky M, Belson D (2009) A mixed integer programming approach for allocating operating room capacity. Journal of the Operational Research Society 60:663–673

Zhu S, Fan W, Yang S, Pei J, Pardalos P (2018) Operating room planning and surgical case scheduling: a review of literature. Journal of Combinatorial Optimization pp 757–805