European Journal of Operational Research 282 (2020) 741–752

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Scheduling surgery groups considering multiple downstream resources

A. J. Thomas Schneider[a,b,*], J. Theresia van Essen[c], Mijke Carlier[d], Erwin W. Hans[a]

[a] Center for Healthcare Operations Improvement & Research (CHOIR), University of Twente, P.O. Box 217, AE Enschede 7500, the Netherlands
[b] Department of Quality & Patient Safety, Leiden University Medical Center, Leiden, the Netherlands
[c] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands
[d] Rhythm, Zoetermeer, The Netherlands

ARTICLE INFO

Article history:
Received 14 September 2018
Accepted 16 September 2019
Available online 24 September 2019

Keywords:
OR in health services
Resource allocation
Master surgery scheduling
Ward and ICU occupancy
Surgery type clustering

ABSTRACT

Surgery groups are clustered surgery procedure types that share comparable characteristics (e.g. expected duration). Scheduling OR blocks leaves many options for operational surgery scheduling and this increases the variation in usage of both the OR and downstream beds. Therefore, we schedule surgery groups to reduce the options for operational scheduling, ultimately bridging the gap between tactical and operational scheduling. We propose a single step mixed integer linear programming (MILP) approach that approximates the bed and OR usage and a simulated annealing approach. Both approaches are compared on a real-life data set and results show that the MILP performs best in terms of solution quality and computation time. Furthermore, the results show that our model may improve the OR utilization from 71% to 85% and decrease the bed usage variation from 53 beds to 11 beds compared to historical data. To show the potential and robustness of our model, we discuss several variants of the model requiring minor modifications. The use of surgery groups makes it easier to implementation our model in practice and, for operational planners, it is instantly clear where to schedule different types of surgery.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The operating room (OR) is one of the most expensive resources (Guerriero & Guido, 2011) and a central hub in hospital patient flow. Therefore, the OR gets a lot of attention to improve productivity. By focusing on OR improvements, other resources get out of sight and are therefore easily forgotten. After surgery, patients are transferred to downstream departments such as the intensive care unit and inpatient wards (hereafter referred to as wards). Therefore, the performance of these downstream departments is directly influenced by the OR (Fügener, Hans, Kolisch, Kortbeek, & Vanberkel, 2014). Focusing solely on OR improvements results in large fluctuations in downstream resources, and therefore, requires overcapacity. To optimize all resources involved in the flow of surgical patients, a holistic approach is required. In other words, while improving the productivity of the OR (e.g. optimizing surgery planning), it is crucial to also consider the effect on downstream departments.

According to the organizational decision hierarchy described in Hans, van Houdenhoven, and Hulshof (2012) and Fügener et al. (2014), surgery planning consists of three stages: (1) the strategic case mix planning, (2) the tactical master surgery scheduling (MSS) and (3) the operational surgery planning. In the first stage of planning, OR capacity is roughly divided among surgical specialties via blocks (e.g. a day or half a day). Then, the assigned OR blocks are scheduled in a cyclic schedule, which means that the schedule is repeated (bi)weekly, and this results in the tactical MSS. Finally, on an operational level, patients are scheduled within the OR blocks of their surgical specialty.

In this paper, we discuss the tactical MSS problem while optimizing the effect on the downstream inpatient resources (e.g. bed usage in wards and the ICU). Although contra-intuitive, from hospital data, we observe that the fluctuations in bed occupancy are mostly caused by artificial (e.g. self induced) variation, and are therefore a result of planning. For this reason, we focus on elective surgery planning. We propose a single step model where bed usage variation is minimized and the OR utilization is maximized. Different from recent research where surgical specialties are assigned to OR blocks in the MSS, we schedule surgery groups within these OR blocks. Surgery groups are clusters of surgery types that share comparable characteristics (e.g. duration, specialty, and/or expertise of surgeons). As a result of the wide variety of surgery types, some surgery types are not performed (bi-)weekly. Therefore, these surgery types cannot be taken into account individually, which makes is necessary to cluster several surgery types within a

* Corresponding author at: Center for Healthcare Operations Improvement & Research (CHOIR), University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands.

E-mail address: a.j.schneider@utwente.nl (A.J. Thomas Schneider).

https://doi.org/10.1016/j.ejor.2019.09.029
0377-2217/© 2019 Elsevier B.V. All rights reserved.

surgery group. By scheduling surgery groups instead of OR blocks, we want to bridge the gap between the tactical and operational level. Scheduling OR blocks on a tactical level leaves many options for scheduling different surgery types with different expected durations on the operational level, which increases the probability of variation in OR utilization and bed usage. Therefore, we show that scheduling surgery groups reduces the probability of overtime and variation in bed usage.

In the remainder of this paper, we start with an overview of available literature on OR scheduling and position our paper (Section 2). In Section 3, we discuss three elements of our model: (1) the constraints, (2) the probability distributions of bed usage in the downstream departments for a given cyclic schedule of surgery groups, and (3) the objective function. Section 4 describes our global and local search approach, and Section 5 discusses the results of both approaches. We analyze several variants of our model in Section 6. Finally, we discuss the implications of our approach in Section 7.

2. Literature review and paper positioning

OR planning and scheduling literature is broadly available. For an overview on general OR scheduling literature, we refer to the systematic reviews of Cardoen, Demeulemeester, and Beliën (2010) and Guerriero and Guido (2011). Here, we solely consider OR planning and scheduling literature that take downstream resources into account.

Beliën and Demeulemeester (2007) use two approaches to model bed occupancy of a single ward while creating an MSS: (1) a mixed integer programming (MIP) based approach, linear as well as quadratic, and (2) simulated annealing (SA). In Beliën, Demeulemeester, and Cardoen (2009), this model is extended with multiple wards. Furthermore, Beliën and Demeulemeester (2007) assume the number of patients per OR block to be deterministically dependent on the type of surgery and fixed for each surgeon, while Beliën et al. (2009) assume a multinomial distribution function for this. Beliën et al. (2009) develop two hierarchical goal programming approaches that both consist of two goal programming models that are solved successively.

Santibanez, Begen, and Atkins (2007) developed a MIP using average values for the LoS (Length of Stay: the sojourn time at wards). Their model has two objectives: maximizing daily bed utilization and maximizing throughput and mix of patients. A mixed integer linear programming (MILP) model by Yahia, Eltawil, and Harraz (2016) levels the daily beds and nurse workload, while considering surgeons preferences.

Van Oostrum et al. (2008) plan elective surgical types that are frequently performed in a cyclic schedule. The solution approach consists of two steps: (1) an integer linear program (ILP) which ignores the required number of beds and that is solved by an implicit column generation approach and (2) a MILP with the objective to minimize the required number of beds. They incorporate three types of beds which can be prioritized. Adan, Bekkers, Dellaert, Vissers, and Yu (2009) also assign surgeries to a day in the cyclic schedule, as in Van Oostrum et al. (2008). However, they use a stochastic LoS that outperforms a deterministic LoS. The extension of Van Oostrum et al. (2008) in Adan, Bekkers, Dellaert, Jeunet, and Vissers (2011) also accounts for emergency patients. They use simulation to create an operational schedule based on the obtained tactical schedule with emergency patients.

Vanberkel et al. (2011a) and Vanberkel et al. (2011b) assign OR time to specialties, just as Beliën and Demeulemeester (2007), by computing the ward occupancy distributions, the patient admission/discharge distributions, and the distributions for the ongoing interventions/treatments required by recovering patients. In Vanberkel et al. (2011a), they swap OR blocks and surgical specialty assignments to find a good solution. This model is extended in other literature. van Essen, Bosch, Hans, van Houdenhoven, and Hurink (2014) use the analytical approach of Vanberkel et al. (2011a) to determine the number of required beds. Two solution methods are used: (1) ILP and (2) SA. To be able to use an ILP, the objective function is replaced by the maximum of the expected number of required beds. Fügener et al. (2014) extend the approach of Vanberkel et al. (2011a) by taking multiple wards and the ICU into account and consider several heuristic solution methods. In Fügener (2015), this is even further extended by including multiple ICUs and outpatient flows in downstream resources. Another extension by Fügener et al. (2016) includes outpatients and emergency surgeries during the weekends.

Min and Yih (2010) use simulation to investigate a stochastic surgery scheduling problem while considering ICU beds. Surgery durations and LoS on the ICU are assumed stochastic with known distributions. Chow, Puterman, Salehirad, Huang, and Atkins (2011) combine Monte Carlo simulation and a MIP to predict the impact of an MSS on bed occupancy. The simulation model predicts the daily demand of beds and the MIP (based on Beliën et al., 2009) optimizes the bed occupancy by scheduling surgery blocks and patient types within each block. Banditori, Cappnera, and Visintin (2013) propose a MIP model to find an MSS. Their objective is to maximize the number of surgeries planned while minimizing the violation of due dates. Next to the MIP model, they also simulate the MIP solution for robustness.

Bekker and Koeleman (2011) analyze the impact of variability in admissions and LoS on the required amount of bed capacity with an approximation method. Given an admission pattern, their quadratic programming model determines the mean bed occupancy of each day. The Markov Decision Process (MDP) model in Astaraky and Patrick (2015) provides scheduling policies for all surgeries, given an MSS, that minimize the time a patient spends on the waiting list, OR overtime and ward congestion. They use approximate dynamic programming to solve the MDP of a realistic problem.

We extend the previous work of Van Oostrum et al. (2008) and Fügener et al. (2014) by scheduling surgery groups within OR blocks and by developing an single step solution method instead of decomposition approaches. Scheduling surgery groups complicates modeling the overtime constraint and utilization of the OR, because there are multiple options for scheduling surgery groups within an OR block. To cluster surgical procedure types into surgery groups, we use techniques from data mining. Furthermore, we linearize the overtime constraint by a piecewise linear function and the objective function by using the expected variation in bed occupancy.

3. Problem formulation

In this section, we formulate our problem of creating a schedule that specifies which surgery groups should be scheduled in an OR block. First, we explain our clustering approach for defining the surgery groups in Section 3.1. From Section 3.2 to 3.4, we explain the mathematical model.

3.1. Clustering surgery types into surgery groups

We cluster surgery types into surgery groups using data mining techniques. Data mining improves the understanding of the relations between predictor and response variables, underlying structures and/or distributions of the input data. Therefore, data mining potentially improves the results of the considered model. Data mining techniques can be split into two main categories: supervised learning and unsupervised learning. Supervised learning makes use of labeled training and predicts a response variable

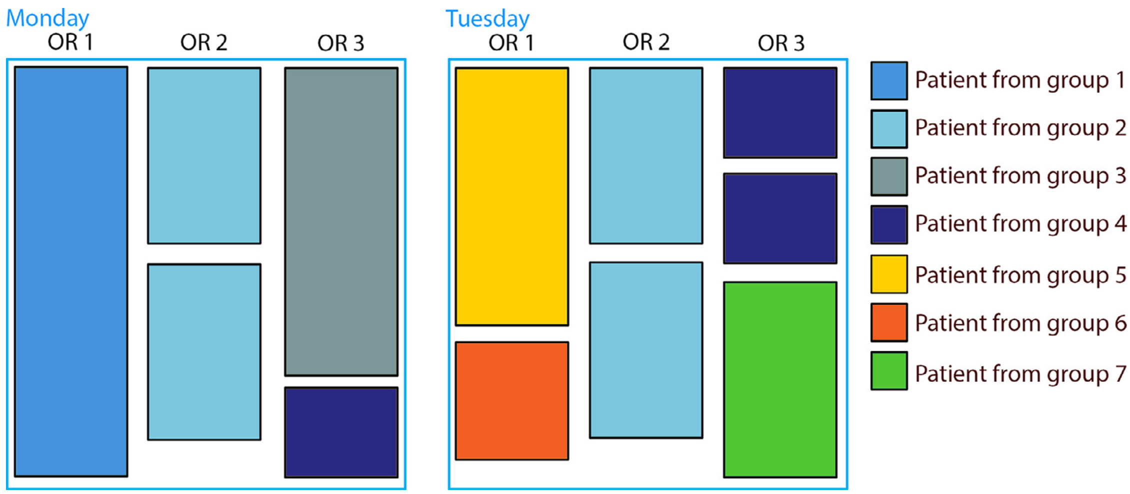

Fig. 1. OR schedule example with surgery groups.

with predictor variables (van der Aalst, 2011). Unsupervised learning only uses unlabeled (e.g. predictor) variables and analyzes the underlying structure or distribution of the data (e.g. clustering or association). For our model, we want to use the predictor variables *specialty* and *surgery* type to predict the response variables *surgery duration* (for OR utilization) and *LoS* (for bed usage) as is done in supervised learning. The response variable could then be split into certain classes such as short LoS and short surgery duration. However, these labels are dependent on the classification we would like to make, and are therefore not available. The other category, unsupervised learning, assumes unlabeled data and does not split the variables into response and predictor variables.

Clustering algorithms examine the data to find groups of similar instances. We would like instances with the same specialty and surgery type to be in one cluster, so they account for surgeon specialization. However, most clustering algorithms (unsupervised learning) assume independent instances. Moreover, in most clustering algorithms we cannot specify what type of clusters we want. This means that one cluster could contain instances where the dispersion of LoS is small and the dispersion of surgery duration is large, and vice versa in another cluster. Therefore, we combine supervised and unsupervised learning techniques in our approach: first, we divide the surgery types of a specialty into short and long stay clusters based on the *median* LoS of the surgery type. This means that the cut-off point between short and long stay clusters depends on the specialty. The cut-off point is determined by maximizing the precision, based on all instances, of both clusters. Precision is an evaluation measure of the confusion matrix and is defined as the fraction of correct positive predictions among the total number of positive predictions (van der Aalst, 2011). In our study, this equals the number of instances in a cluster that were indeed lower (for short stay) or higher (for long stay) than the cut-off point among all instances of a surgery type. Next, we further divide each short and long stay cluster into three sub clusters based on the surgery duration. The clustering for the surgery duration is similar as for the LoS, although now we take the *mean* of each surgery type. The cut-off points are again determined by maximizing the precision of each cluster. This means that our clustering approach results in six groups per surgical specialty.

To ensure that the sizes of the resulting surgery groups do not become too small, we set the cut-off point such that the number of instances assigned to a group is at least 20% of the number of instances that can be divided. In addition, we use a two-sample *t*-test with a 5% significance level to determine whether two groups are significantly different. When the two groups fail this test, i.e., when they are not significantly different, we decrease the number of groups.

3.2. Conceptual model

In our approach, we assign surgery groups to OR blocks instead of surgical specialties. See Fig. 1 for a graphical example of scheduling surgery groups. Assigning a surgery group to an OR block allows for a single surgery type of that group to be scheduled during the next planning stage. Multiple surgery groups can be assigned multiple times to the same OR on the same day as long as the surgery groups belong to the same specialty of the allocated OR block. The order in which individual patients of the surgery groups are scheduled on the operational level is undefined. For example, our MSS specifies that surgery group X and Y are scheduled on the same day and OR, but does not specify if surgery group X must be scheduled before or after surgery group Y during that day. Hence, a variable amount of surgery groups can be scheduled in an OR block of the MSS. The objective of our model is to find an optimal schedule of surgery groups that maximizes OR utilization while minimizing the variance of bed usage at the wards.

3.3. Constraints

Multiple constraints are taken into account for our model, e.g., the need for specific ORs, the need for specific equipment, the total available OR time during opening hours and the number of scheduled surgery groups. Let $\mathcal{O}$ be the set of given ORs and $\mathcal{K}$ the set of days in the MSS. Then, an OR block (o, k) is defined as a combination of day $k \in \mathcal{K}$ of the MSS and OR $o \in \mathcal{O}$. The set of given surgery groups is denoted by set $\mathcal{J}$.

The integer decision variable z_{okj} specifies the number of surgeries from surgery group $j \in \mathcal{J}$ that are scheduled in OR block (o, k). To ensure equitable access for each surgery group, we set a lower bound β_j on the number of scheduled surgeries per surgery group $j \in \mathcal{J}$ and assume that waiting lists are inexhaustible. The following constraints ensure that all groups $j \in \mathcal{J}$ are scheduled a minimum of β_j times.

$$\sum_{o\in\mathcal{O},k\in\mathcal{K}} z_{okj} \geq \beta_j, \qquad \forall j \in \mathcal{J}. \tag{1}$$

Let $\mathcal{S}$ be the set of specialties and $\mathcal{J}_s \subseteq \mathcal{J}$ the set of surgery groups belonging to specialty $s \in \mathcal{S}$. We introduce binary parameters ϵ_{oks} that are one when specialty $s \in \mathcal{S}$ can be allocated to OR block (o, k) in the MSS, and zero otherwise. Furthermore, we introduce binary decision variables u_{oks} which are one when a surgery group of specialty $s \in \mathcal{S}$ is scheduled in OR block (o, k) and zero otherwise. Now we can ensure that only surgery groups of the specialty that is allocated to OR block (o, k) can be scheduled:

$$u_{oks} \leq \epsilon_{oks}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}, s \in \mathcal{S}. \tag{2}$$

The relation between z_{okj} and u_{oks} is given by constraints (3), where M_s is the maximum number of surgeries of a specialty $s \in \mathcal{S}$ that fit in one OR block:

$$\sum_{j \in \mathcal{J}_s} z_{okj} \leq M_s \cdot u_{oks}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}, s \in \mathcal{S}. \tag{3}$$

The total surgery duration of the surgery groups we assign to an OR block is limited by the opening hours of the OR. The surgery duration ζ_j of surgery group $j \in \mathcal{J}$ is a stochastic variable with mean μ_j and variance σ_j^2. Let g_{ok} denote the stochastic variable representing the total duration of the surgery groups that are scheduled in OR block (o, k). The available OR time on day $k \in \mathcal{K}$ in OR $o \in \mathcal{O}$ is denoted by τ_{ok}. We introduce constraints (4) to ensure that the probability of overtime is below α with $0 \leq \alpha \leq 1$. Overtime occurs when the total sum of the duration of the scheduled groups exceeds the available time of that OR block:

$$P(g_{ok} \leq \tau_{ok}) \geq 1 - \alpha, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}. \tag{4}$$

To ensure that only one specialty $s \in \mathcal{S}$ can be assigned to each OR block, we introduce the following constraints and binary parameters χ_{ok} which are one when OR $o \in \mathcal{O}$ is open on day $k \in \mathcal{K}$ and zero otherwise:

$$\sum_{s \in \mathcal{S}} u_{oks} \leq \chi_{ok}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}. \tag{5}$$

Some surgery groups require specific equipment that is not available in every OR, and therefore, have to be scheduled in specific ORs, while other surgery groups can be scheduled in every OR. To model this, we define a set of OR types $\mathcal{R}$ and we denote the subset of surgery groups that can be performed in OR type $r \in \mathcal{R}$ by $J_r \subseteq J$. Binary parameters υ_{okr} are one when OR $o \in \mathcal{O}$ on day $k \in \mathcal{K}$ is of type $r \in \mathcal{R}$ and zero otherwise. This leads to the following constraint:

$$\sum_{j \in \mathcal{J}_r} z_{okj} \leq N_r \upsilon_{okr}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}, r \in \mathcal{R}. \tag{6}$$

where N_r is the maximum number of surgeries belonging to OR type r in one OR block.

3.4. Bed usage distributions

Next, we want to determine the bed usage distributions of the wards in three steps: (1) we calculate the bed usage distribution for the wards per surgery group, (2) we calculate the bed usage distribution for overlapping cycles, and (3) we calculate the bed usage distribution for an entire OR block. This final step needs to be repeated for every cyclic schedule. The first two steps can be done beforehand. Section 3.4.4 describes the resulting objective function.

As mentioned before, we further extent the work of Vanberkel et al. (2011b) and Fügener et al. (2014) by assuming that patients from the same surgery group can be admitted at different wards (e.g. the ICU or different wards belonging to the same surgical specialty). Therefore, we take into account all wards where patients of a certain surgical specialty can be admitted.

We assume that patients can take two paths after surgery: (1) directly to a ward or (2) first to the ICU followed by a transfer to a ward (see Fig. 2). Finally, patients are discharged and leave the system. Let set $\mathcal{I}$ denote all ICUs and let set $\mathcal{W}$ denote all wards. For all $j \in \mathcal{J}$, we define subsets $\mathcal{J}_i \subseteq \mathcal{J}$ and $\mathcal{J}_w \subseteq \mathcal{J}$ for the surgery groups

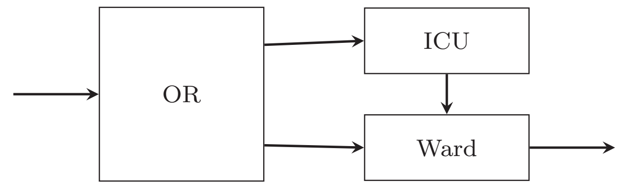

Fig. 2. Main hospital flows for surgical patients.

that are transferred to the ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$, respectively. The LoS (in days) in the ICU $i \in \mathcal{I}$ or ward $w \in \mathcal{W}$ of each surgery group is modeled by discrete empirical distributions based on historical data. The empirical distribution of the LoS is determined per surgery group, regardless of the ward they are transferred to. The following input parameters are required for every surgery group $j \in \mathcal{J}$:

- a_{ij} represents the probability that a patient of surgery group $j \in \mathcal{J}$ is transferred to ICU $i \in \mathcal{I}$ after surgery.
- b_{wj} represents the probability that a patient from surgery group $j \in \mathcal{J}$ is transferred to ward $w \in \mathcal{W}$ after surgery or ICU.
- c^I_{jn} represents the probability that a patient from surgery group $j \in \mathcal{J}$ stays exactly n days in the ICU after surgery.
- c^{WS}_{jn} represents the probability that a patient from surgery group $j \in \mathcal{J}$ stays exactly n days in the ward after surgery.
- c^{WI}_{jn} represents the probability that a patient from surgery group $j \in \mathcal{J}$ stays exactly n days in the ward after a stay in the ICU.

The probability that a patient from surgery group $j \in \mathcal{J}$ is transferred to the ICU is given by $\sum_{i \in \mathcal{I}} a_{ij}$ and a transfer to the ward is given by $1 - \sum_{i \in \mathcal{I}} a_{ij}$. The probabilities c^I_{jn}, c^{WS}_{jn} and c^{WI}_{jn} are not given separately for every ward or ICU, because for every surgery group $j \in \mathcal{J}$, the probability of a patient staying exactly n days is independent of the ward or ICU. We also assume a bed is occupied a whole day if a patient is discharged on that day.

3.4.1. Single surgery group

The first step of our approach is similar to the approach presented in Fügener et al. (2014). As Fügener et al. (2014), we start by calculating conditional probabilities d^I_{jn+1} that a patient from surgery group $j \in \mathcal{J}$ is transferred from the ICU to a ward on day $n+1$ (which is n days after surgery). In a similar way, the conditional probabilities d^{WS}_{jn+1} that a patient from surgery group $j \in \mathcal{J}$, who is in the ward on day n, is discharged on day n can be determined. Conditional probabilities d^{WI}_{jn+1} represent the probability that a patient from surgery group $j \in \mathcal{J}$, who is in the ward on day n after being transferred from the ICU, is discharged on day n, where we assume that the patient is transferred from the ICU on day 1.

As Fügener et al. (2014), we can now calculate probabilities e^I_{jn} that a patient from surgery group $j \in \mathcal{J}$, who had surgery on day 1, is still occupying a bed on day n. For $n = 1$ and the ICU, this is simply the probability that the patient is transferred to the ICU after surgery. We assume a patient stays at least one day in the ICU, otherwise, a patient is transferred directly to the ward. Therefore, for $n = 2$, we have the same probability as for $n = 1$. For $n \in \{3, \ldots, N^I_j + 1\}$, where N^I_j is the maximum number of days that a patient from surgery group $j \in \mathcal{J}$ stays in the ICU after surgery, this is the probability that the patient was not transferred to the ward the day before, i.e., day $n - 1$, multiplied by the probability that the patient was still in the ICU the day before.

Similarly, probabilities e^{WS}_{jn} and e^{WI}_{jnm} are determined, i.e., the probabilities that a patient from surgery group $j \in \mathcal{J}$ who had surgery on day 1, is still occupying a bed in the ward on day n and the probability that after an ICU stay of m days, a patient from surgery group $j \in \mathcal{J}$ is still in the ward on day n, respectively. Probabilities e^{WS}_{jn} and e^{WI}_{jmn} are combined to calculate the probability e^{W}_{jn} that a patient of surgery group $j \in \mathcal{J}$ is in the ward on day n.

Different from Fügener et al. (2014), we consider multiple ICUs. Therefore, we also need to calculate the probability that a patient from surgery group $j \in \mathcal{J}$ is in ICU $i \in \mathcal{I}$, given that this patient is in the ICU. $\sum_{i\in \mathcal{I}} a_{ij}$ is the probability that a patient of surgery group $j \in \mathcal{J}$ is in the ICU. So, for all $j \in \mathcal{J}$ and $i \in \mathcal{I}$, we have conditional probability $\hat{a}_{ij} = \frac{a_{ij}}{\sum_{i\in \mathcal{I}} a_{ij}}$ that a patient of surgery group $j \in \mathcal{J}$ is in ICU $i \in \mathcal{I}$, given that this patient is in the ICU. For the wards, this probability is given by b_{wj}. We do not need to normalize this probability, since every patient in our model is transferred to the ward. Patients who do not stay at the ward are represented using a ward LoS of zero days.

The probability distributions of the number of patients from surgery group $j \in \mathcal{J}$ in ICU $i \in \mathcal{I}$ or ward $w \in \mathcal{W}$ on day n are denoted by f^{I}_{ijn} and f^{W}_{wjn}. The discrete stochastic variables that are associated with these probability distributions are given by $\mathfrak{f}^{I}_{ijn}$ and $\mathfrak{f}^{W}_{wjn}$, respectively. Since different from Fügener et al. (2014), we schedule surgery groups instead of OR blocks, the number of patients in an ICU or ward can only equal zero or one. So, the probability that there is one patient in the ward or ICU is calculated by multiplying the probability that a patient from surgery group $j \in \mathcal{J}$ goes to ICU $i \in \mathcal{I}$ (ward $w \in \mathcal{W}$), given this patient is in the ICU (ward), with the probability that this patient is in the ICU (ward) on day n. The probability that there are zero patients is equal to one minus the probability that there is one patient.

$$P(\mathfrak{f}^{I}_{ijn} = 0) = 1 - \hat{a}_{ij} e^{I}_{jn}, \qquad i \in \mathcal{I}, j \in \mathcal{J}, n \in \{1, \ldots, N^{I}_{j}\}; \tag{7}$$

$$P(\mathfrak{f}^{I}_{ijn} = 1) = \hat{a}_{ij} e^{I}_{jn}, \qquad i \in \mathcal{I}, j \in \mathcal{J}, n \in \{1, \ldots, N^{I}_{j}\}; \tag{8}$$

$$P(\mathfrak{f}^{W}_{wjn} = 0) = 1 - b_{wj} e^{W}_{jn}, \qquad w \in \mathcal{W}, j \in \mathcal{J}, n \in \{1, \ldots, N^{W}_{j}\}; \tag{9}$$

$$P(\mathfrak{f}^{W}_{wjn} = 1) = b_{wj} e^{W}_{jn}, \qquad w \in \mathcal{W}, j \in \mathcal{J}, n \in \{1, \ldots, N^{W}_{j}\}. \tag{10}$$

3.4.2. Cyclical surgery group

Now that we have all probabilities for single surgery groups, we can calculate the bed usage distribution for overlapping cycles using the approach of Vanberkel et al. (2011b), since the maximum LoS of a patient can exceed the cycle length. The distribution of the number of patients in overlapping cycles is denoted by F^{I}_{ijl} and F^{W}_{wjl} for surgery group $j \in \mathcal{J}$ in ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$, respectively, on the lth day of a cycle, when the surgery group is scheduled on day one of the cycle. The number of overlapping cycles depends on the maximum LoS in the ICU and wards, N^{I}_{j} and N^{W}_{j}, respectively, and on the cycle length L, which is the number of elements in $\mathcal{L}$. Depending on the day l in the cycle, we have $\lfloor (N^{I}_{j} - l)/L \rfloor + 1$ overlapping cycles for the ICU and $\lfloor (N^{W}_{j} - l)/L \rfloor + 1$ overlapping cycles for the ward.

3.4.3. Cyclic schedule

We now have all the elements for the final step: calculating the bed usage distributions for a cyclic surgery group schedule. The calculations in this step differ from Vanberkel et al. (2011b) and Fügener et al. (2014), since we schedule surgery groups instead of OR blocks. A cyclic schedule is given by the integer decision variables z_{okj}, which represent the total number of surgeries from surgery group $j \in \mathcal{J}$ that are scheduled in OR $o \in \mathcal{O}$ on day $k \in \mathcal{K}$. Let $\mathbf{1}_{z_{okj}}$ be an indicator function that is equal to one if z_{okj} is greater than zero and equal to zero if z_{okj} is zero. The bed usage distribution in ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$ on day l of a cyclic schedule when scheduling surgery group $j \in \mathcal{J}$ once in OR block (o, k) is given by G^{I}_{iokjl} and G^{W}_{wokjl}, respectively.

Next, we shift both distributions F^{I}_{ijl} and F^{W}_{wjl} to the day on which the surgery group is scheduled. Here, l is the day for which we are determining the bed usage distribution and k is the day on which the surgery group is scheduled in the cyclic schedule. If $l \geq k$, we shift F^{I}_{ijl} and F^{W}_{wjl} by $k-1$ days. If $l < k$, the bed usage distribution on day l results only from surgery groups scheduled on day k of previous cycles. Thus, we shift by $k-1-L$ days. We multiply these distributions by $\mathbf{1}_{z_{okj}}$, which is only non-zero if the surgery group $j \in \mathcal{J}$ is assigned to OR block (o, k).

$$G^{I}_{iokjl} = \begin{cases} F^{I}_{ijl-k+1} \mathbf{1}_{z_{okj}}, & l \geq k \\ F^{I}_{ijl-k+1+L} \mathbf{1}_{z_{okj}}, & \text{otherwise.} \end{cases} \tag{11}$$

$$G^{W}_{wokjl} = \begin{cases} F^{W}_{wjl-k+1} \mathbf{1}_{z_{okj}}, & l \geq k \\ F^{W}_{wjl-k+1+L} \mathbf{1}_{z_{okj}}, & \text{otherwise.} \end{cases} \tag{12}$$

Next, we obtain the bed usage distributions for an OR block. We use the indicator function $\mathbf{1}_{z_{okj}}$ to indicate that a surgery group $j \in \mathcal{J}$ is assigned at least once to OR block (o, k). However, a surgery group might be assigned multiple times to one OR block. To obtain the distribution of patients from an entire OR block, we need the convolution of all distributions G^{I}_{iokjl} and G^{W}_{wokjl} of the surgery groups scheduled in that OR block. If a surgery group is assigned n times to one OR block, we need to convolute the distribution n times with itself, before convolving it with the distributions of other surgery groups assigned to that OR block. Therefore, we use the convolution power, which is defined as the n-fold iteration of the convolution with itself. For h, a function $\mathbb{Z} \to \mathbb{R}$ and $n \in \mathbb{N}_{>0}$, we have:

$$h^{*n} = \underbrace{h * h * \ldots * h * h}_{n}, \qquad h^{*0} = \delta_0, \tag{13}$$

where δ_0 is Dirac's delta function. Dirac's delta function focuses the mass of a function around zero. When we convolve a distribution zero times, the probability of being zero is equal to one.

The bed usage distribution in ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$ on day l of the cyclic schedule per surgery group $j \in \mathcal{J}$ in OR block (o, k) is given by $\hat{G}^{I}_{iokjl}$ and $\hat{G}^{W}_{wokjl}$:

$$\hat{G}^{I}_{iokjl} = {G^{I}_{iokjl}}^{*z_{okj}}, i \in \mathcal{I}, o \in \mathcal{O}, k \in \mathcal{K}, j \in \mathcal{J}_i, l \in \mathcal{L}. \tag{14}$$

$$\hat{G}^{W}_{wokjl} = {G^{W}_{wokjl}}^{*z_{okj}}, w \in \mathcal{W}, o \in \mathcal{O}, k \in \mathcal{K}, j \in \mathcal{J}_w, l \in \mathcal{L}. \tag{15}$$

Now we can define distributions H^{I}_{iokl} and H^{W}_{wokl}, which represent the bed usage distributions on day l at ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$, resulting from all surgery groups $j_1, j_2, \ldots, j_{\max} \in \mathcal{J}_i$ and $j_1, j_2, \ldots, j_{\max} \in \mathcal{J}_w$, respectively.

$$H^{I}_{iokl} = \hat{G}^{I}_{iokj_1l} * \hat{G}^{I}_{iokj_2l} * \ldots * \hat{G}^{I}_{iokj_{\max}l}, \quad i \in \mathcal{I}, o \in \mathcal{O}, k \in \mathcal{K}, l \in \mathcal{L}, \tag{16}$$

$$H^{W}_{wokl} = \hat{G}^{W}_{wokj_1l} * \hat{G}^{W}_{wokj_2l} * \ldots * \hat{G}^{W}_{wokj_{\max}l}, \quad w \in \mathcal{W}, o \in \mathcal{O}, k \in \mathcal{K}, l \in \mathcal{L}. \tag{17}$$

Following the approach of Vanberkel et al. (2011b) and Fügener et al. (2014), we convolve the distributions of all the OR blocks in the cyclic schedule to obtain the bed usage distributions resulting from the complete cyclic schedule. $\hat{H}^{I}_{il}$ denotes the distribution of patients in ICU $i \in \mathcal{I}$ on day l of the cyclic schedule and $\hat{H}^{W}_{wl}$ denotes the distribution of recovering patients in ward $w \in \mathcal{W}$ on day l of the cyclic schedule. The last OR and the last day in the cyclic schedule on which surgeries take place are denoted by $\max\{\mathcal{O}\}$ and $\max\{\mathcal{K}\}$, respectively.

$$\hat{H}^{I}_{il} = H^{I}_{i11l} * H^{I}_{i12l} * \cdots * H^{I}_{i1\max\{\mathcal{K}\}l} * H^{I}_{i21l} * H^{I}_{i22l} * \cdots * H^{I}_{i\max\{\mathcal{O}\}\max\{\mathcal{K}\}l}, \quad i \in \mathcal{I}, l \in \mathcal{L}, \qquad (18)$$

$$\hat{H}^{W}_{wl} = H^{W}_{w11l} * H^{W}_{w12l} * \cdots * H^{W}_{w1\max\{\mathcal{K}\}l} * H^{W}_{w21l} * H^{W}_{w22l} * \cdots * H^{W}_{w\max\{\mathcal{O}\}\max\{\mathcal{K}\}l}, \quad w \in \mathcal{W}, l \in \mathcal{L}. \qquad (19)$$

We define the probability of having n patients in ICU $i \in \mathcal{I}$ or ward $w \in \mathcal{W}$ on day l by $\hat{H}^{I}_{il}[n]$ and $\hat{H}^{W}_{wl}[n]$.

For a given cyclic schedule ψ, we want to determine the variation in bed occupancy. This means that we calculate for each day l and with probability p that there are at most n patients, thus n required beds, by summing over the probabilities that there are at most n patients in the ICU or ward. The required number of beds $\gamma_{il}(\psi)$ on day l in ICU $i \in \mathcal{I}$ for a given solution $\psi \in \Psi$ is then given by:

$$\gamma_{il}(\psi) = \min\left\{ n \,\middle|\, \sum_{m=0}^{n} \hat{H}^{I}_{il}[m] \geq p \right\}. \qquad (20)$$

The required number of beds $\gamma_{wl}(\psi)$ on day l in ward $w \in \mathcal{W}$ for a given solution ψ is given similarly by:

$$\gamma_{wl}(\psi) = \min\left\{ n \,\middle|\, \sum_{m=0}^{n} \hat{H}^{W}_{wl}[m] \geq p \right\}. \qquad (21)$$

Peaks in bed occupancy occur during weekdays since new patients arrive to undergo scheduled surgeries. These peaks may cause surgery cancellations, because not enough beds are available. Therefore, we are interested in minimizing the variation in bed occupancy during weekdays. As no surgeries are scheduled during the weekends, the bed occupancy is lower. The variation in bed occupancy, denoted by $\gamma_i(\psi)$ and $\gamma_w(\psi)$, in ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$ is given by the difference between the maximum and minimum number of required beds during the week and are given by:

$$\gamma_i(\psi) = \max_{l \in \mathcal{K}} \gamma_{il}(\psi) - \min_{l \in \mathcal{K}} \gamma_{il}(\psi), \qquad (22)$$

$$\gamma_w(\psi) = \max_{l \in \mathcal{K}} \gamma_{wl}(\psi) - \min_{l \in \mathcal{K}} \gamma_{wl}(\psi), \qquad (23)$$

where $\mathcal{K}$ is the set of all workdays as defined in Section 3.3.

3.4.4. Objective function

Our model has two main goals: (1) to maximize the OR utilization and (2) to minimize the variation in bed occupancy. Because the available OR time is determined at the strategical level, it is constant in our model. Hence, maximizing the OR utilization is equal to maximizing the time allocated for scheduled surgery groups. The utilized OR time is the sum of the mean surgery durations μ_j of the scheduled surgery groups. Furthermore, we want to minimize the variation in bed occupancy, γ_i and γ_w. Finally, we include weights θ_i and θ_w, so we can manage the balance between the variation in bed occupancy and the OR utilization. The objective function is now given by:

$$\max \sum_{o \in O} \sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} \mu_j \cdot z_{okj} - \sum_{i \in I} \theta_i \cdot \gamma_i(\psi) - \sum_{w \in W} \theta_w \cdot \gamma_w(\psi), \qquad (24)$$

where the objective function value for a given schedule ψ is denoted by $OB(\psi)$.

4. Solution methods

The calculations in Sections 3.4.1 and 3.4.2 can be performed beforehand. However, the calculations in Section 3.4.3 still involve the convolution of several probability distributions, and the minimum and maximum operators in Eqs. (22) and (23) are non-linear operators. Moreover, the constraints in (4) are nonlinear which makes the model nonlinear. Therefore, we use two different approaches to solve our problem: (1) approximate the objective function and the nonlinear constraint and use these approximations in a MILP and (2) use simulated annealing (SA) as local search approach based on the given constraints and objective function. MILP and SA are widely used for solving the MSS problem and are also compared on the trade-off between the objective function value and computational performances by Cardoen et al. (2010).

4.1. Global approach

Our global approach uses an approximation of the objective function and a linearized version of nonlinear constraints (4) in order to formulate a MILP which we can solve with a commercial solver. In Section 4.1.1, we linearize the overtime constraints (4). Because there is no direct relation between a given OR-schedule and the number of required beds, we also linearize the objective function in Section 4.1.2.

4.1.1. Linearization of the surgery duration constraint

In the problem formulation introduced in Section 3.3, we have nonlinear constraints that make the surgery schedule more robust against overtime. We linearize the overtime constraint using the same approach as Bosch (2011). The overtime constraint is given by:

$$P(g_{ok} \leq \tau_{ok}) \geq 1 - \alpha, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K} \qquad (25)$$

where g_{ok} is the stochastic variable representing the total session time of all surgery types scheduled in OR block (o, k) and τ_{ok} is the total available time to schedule surgeries in OR block (o, k).

May, Strum, and Vargas (2000) have shown that the 3-parameter lognormal distribution is the best fit for surgery duration distributions. However, since there is no known exact result for the distribution of the sum of 3-parameter lognormal distributed stochastic variables, we approximate the distribution of the sum of the surgery durations with a normal distribution as is done by Hans, Wullink, van Houdenhoven, and Kazemier (2008) and Van Oostrum et al. (2008). Therefore, the total duration of OR block (o, k) is normally distributed with mean μ_{ok} and variance σ^2_{ok}. Thus, $g_{ok}(x) \sim \mathcal{N}(\mu_{ok}, \sigma_{ok})$. Then, the overtime constraints can be written as:

$$P(g_{ok} \leq \tau_{ok}) = \Phi\left(\frac{\tau_{ok} - \mu_{ok}}{\sigma_{ok}}\right) \geq 1 - \alpha, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}. \qquad (26)$$

Rewriting Eq. (26) gives:

$$\mu_{ok} + \Phi^{-1}(1-\alpha)\sigma_{ok} \leq \tau_{ok}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}. \qquad (27)$$

The mean and variance of the total surgery duration g_{ok} of OR block (o, k) can be written as:

$$\mu_{ok} = \sum_{j \in \mathcal{J}} z_{okj}\mu_j \quad \text{and} \quad \sigma^2_{ok} = \sum_{j \in \mathcal{J}} z_{okj}\sigma^2_j. \qquad (28)$$

Substituting the latter two expressions into the overtime constraints (27) gives:

$$\sum_{j \in \mathcal{J}} z_{okj}\mu_j + \Phi^{-1}(1-\alpha)\sqrt{\sum_{j \in \mathcal{J}} z_{okj}\sigma^2_j} \leq \tau_{ok}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}. \qquad (29)$$

To linearize this constraint, we approximate the square root function $f(x) = \sqrt{x}$ by a piecewise linear function. The square root function needs to be approximated on the interval $[x_{\min}, x_{\max}]$. We do not want to underestimate the function $f(x)$, so the approximation function must be greater than or equal to $f(x)$ for all $x \in [x_{\min}, x_{\max}]$. The intervals of the piecewise linear functions are determined by breakpoints $n \in N$, where $N = \{0, 1, \ldots, m\}$. Here, x_n is the value on the x-axis of breakpoint $n \in N$. We define x_0 as the first

x-value and x_m as the last x-value for which we approximate the square root function. The other x-values, x_n for $n = \{1, \dots, m-1\}$, are intersection points of the linear approximations. Let y_n be the function value of the linear approximation function at breakpoint n, so $y_n = \sqrt{x_n}$. See for more details the appendix in the supplementary materials.

Once the breakpoints are known, we can use the λ-formulation by Bisschop (2016) to model piecewise linear functions together. The function value of any point between two breakpoints is the weighted sum of the function values of these two breakpoints. Let λ_{okn} denote n nonnegative weights for each OR block (o, k) such that their sum equals one. Then, the piecewise linear approximation of the overtime constraint can be written as:

$$\sum_{j\in\mathcal{J}} z_{okj}\mu_j + \Phi^{-1}(1-\alpha)\sum_{n\in N}\lambda_{okn}y_n \le \tau_{ok}, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}, \tag{30}$$

$$\sum_{n\in N}\lambda_{okn}x_n = \sum_{j\in\mathcal{J}} z_{okj}\sigma_j^2, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}, \tag{31}$$

$$\sum_{n\in N}\lambda_{okn} = 1, \qquad \forall o \in \mathcal{O}, k \in \mathcal{K}. \tag{32}$$

Considering overtime constraints (29), we show that when scheduling surgery groups instead of surgical specialties, we can at least assign the same number of surgeries to one OR block. Assuming that there exists a surgery group $j \in \mathcal{J}_s$ with $\mu_j \le \mu_s$ and $\sigma_j^2 \le \sigma_s^2$, where μ_s and σ_s^2 represent the mean and variance of the surgery duration for surgical specialty $s \in \mathcal{S}$, we have that

$$z\mu_j + \Phi^{-1}(1-\alpha)\sqrt{z\sigma_j^2} \le z\mu_s + \Phi^{-1}(1-\alpha)\sqrt{z\sigma_s^2} \le \tau \tag{33}$$

where z denotes the number of assigned surgeries to a given OR block. This means that the cyclic schedule obtained when scheduling surgery groups instead of surgical specialties allows us to schedule at least the same number of surgeries and possibly more.

4.1.2. Linearization of the objective function

Our approach for linearizing the objective function is an extension of the approach of Beliën and Demeulemeester (2007). Instead of using γ_i and γ_w, we use the expected number of beds at ward $w \in \mathcal{W}$ and ICU $i \in \mathcal{I}$ on day l of the cycle. For a solution ψ, this is given by $\bar{\gamma}_{wl}(\psi)$ and $\bar{\gamma}_{il}(\psi)$, respectively. We use the expected value of the distribution functions $\hat{H}^I_{il}$ and $\hat{H}^W_{wl}$, which are defined as the probability distributions of the bed usage in the ICU and ward, respectively. The expected value of $\hat{H}^I_{il}$ is given by:

$$\begin{aligned}\bar{\gamma}_{il} &= \mathbb{E}\left(\hat{\mathcal{H}}^I_{il}\right)\\ &= \sum_{o\in O}\sum_{\substack{k\in\mathcal{K}\\ l\ge k}}\sum_{j\in\mathcal{J}_i}\sum_{n=0}^{\lfloor D^I_{jkl}/L\rfloor} \hat{a}_{ij}e^I_{j(l-k+1+nL)}\cdot z_{okj}\\ &\quad+\sum_{o\in O}\sum_{\substack{k\in\mathcal{K}\\ l< k}}\sum_{j\in\mathcal{J}_i}\sum_{n=1}^{\lfloor (D^I_{jkl}-L)/L\rfloor+1} \hat{a}_{ij}e^I_{j(l-k+1+nL)}\cdot z_{okj}\end{aligned} \tag{34}$$

with $\lfloor D^L_{jkl}/L\rfloor = \lfloor (N^I_j-(l-k+1))/L\rfloor$ for the number of overlapping cycles on day $l \in \mathcal{L}$ when a surgery group is scheduled on day k and $l \ge k$ and $\lfloor (D^I_{jkl}-L)/L\rfloor + 1 = \lfloor (N^I_j-(l-k+1+L))/L\rfloor + 1$ the number of overlapping cycles on day $l \in \mathcal{L}$ when $l < k$. The expected number of required beds on day l is given by the sum over all surgery groups of the probability that a patient from surgery group $j \in \mathcal{J}$ is in an ICU on day l, accounting for all cycles, multiplied by the number of times this surgery group is scheduled in all OR blocks (o, k). Similarly, we obtain:

$$\begin{aligned}\bar{\gamma}_{wl} &= \mathbb{E}\left(\hat{H}^W_{wl}\right)\\ &= \sum_{o\in O}\sum_{\substack{k\in\mathcal{K}\\ l\ge k}}\sum_{j\in\mathcal{J}_w}\sum_{n=0}^{\lfloor D^W_{jkl}/L\rfloor} b_{wj}e^W_{j(l-k+1+nL)}\cdot z_{okj}\\ &\quad+\sum_{o\in O}\sum_{\substack{k\in\mathcal{K}\\ l< k}}\sum_{j\in\mathcal{J}_w}\sum_{n=1}^{\lfloor (D^W_{jkl}-L)/L\rfloor+1} b_{wj}e^W_{j(l-k+1+nL)}\cdot z_{okj}.\end{aligned} \tag{35}$$

Since $\sum_n \hat{a}_{ij}e^I_{j(l-k+1+nL)}$ and $\sum_n b_{wj}e^W_{j(l-k+1+nL)}$ are constant, the new objective function is linear in the decision variables z_{okj}. Again, we want to obtain the maximum and minimum of both $\bar{\gamma}_{il}(\psi)$ and $\bar{\gamma}_{wl}(\psi)$ to determine the variation in bed occupancy during the week. The maximum and minimum operator are not linear. Therefore, we add the following constraints:

$$\bar{\gamma}_i^{\max} \ge \bar{\gamma}_{il}, \qquad \forall i \in \mathcal{I}, l \in \mathcal{L}, \tag{36}$$

$$\bar{\gamma}_w^{\max} \ge \bar{\gamma}_{wl}, \qquad \forall w \in \mathcal{W}, l \in \mathcal{L}, \tag{37}$$

$$\bar{\gamma}_i^{\min} \ge -\bar{\gamma}_{il}, \qquad \forall i \in \mathcal{I}, l \in \mathcal{L}, \tag{38}$$

$$\bar{\gamma}_w^{\min} \ge -\bar{\gamma}_{wl}, \qquad \forall w \in \mathcal{W}, l \in \mathcal{L}. \tag{39}$$

Additionally, let

$$\hat{\gamma}_i = \bar{\gamma}_i^{\max} + \bar{\gamma}_i^{\min}, \qquad \forall i \in \mathcal{I}, \tag{40}$$

$$\hat{\gamma}_w = \bar{\gamma}_w^{\max} + \bar{\gamma}_w^{\min}, \qquad \forall w \in \mathcal{W}. \tag{41}$$

The resulting MILP model is now given by:

$$\max \sum_{o\in O}\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}}\mu_j\cdot z_{okj} - \sum_{i\in I}\theta_i\hat{\gamma}_i - \sum_{w\in\mathcal{W}}\theta_w\hat{\gamma}_w \tag{42}$$

$$\text{s.t. } (1)-(3), (5)-(6), (30)-(41)$$

We refer to this problem as the linear OR schedule problem, which is NP-hard as proven by van Essen et al. (2014). Note that a solution obtained by solving the linear OR schedule problem will still be evaluated by using the original objective function (24).

4.2. Local search approach

Similarly to Beliën et al. (2009), Hans et al. (2008), Beliën and Demeulemeester (2007), and van Essen et al. (2014), we use SA as local search approach, and therefore, we need to specify a cooling scheme, using the following parameters: the initial temperature, the final temperature, the reduction (e.g. cooling) factor and the length of the Markov chain. The effectiveness of this method depends on the configuration of these parameters and the local search strategy. First, we explain how we define neighbor solutions, and then, we describe how we determine the cooling scheme.

To obtain feasible neighbor solutions, we set a generator function that uses the current solution as input and produces a new solution. We consider four strategies to generate a neighbor solution:

- *Removing a surgery group*
  We find a neighbor solution by removing one surgery group from a certain OR-day. To find a feasible new solution, it is important to only remove a surgery group if it is scheduled more often than the required minimum amount.

- *Adding a surgery group*
  Similarly, adding one surgery group to a certain OR-day leads also to a neighbor solution. To find a feasible new solution, it is important to only add surgery groups from the specialty assigned to the selected OR-day and to check if adding this surgery group does not violate the overtime constraint.
- *Swap two OR blocks*
  Similar to Beliën et al. (2009), Beliën and Demeulemeester (2007), and van Essen et al. (2014), we define neighbor solutions by swapping all surgery groups between two OR blocks. This can only be done if they have the same available time for surgeries and the same specialty can operate in the ORs. We do not swap two OR blocks that take place on the same day, because this leads to a symmetric solution.
- *Swap two groups*
  Similar to Hans et al. (2008), we define neighbor solutions by swapping two surgery groups that have been scheduled in the current solution. They can only be swapped if either the OR or the day on which they are scheduled is different. Furthermore, the new solution is only feasible when surgery groups from the same specialty are swapped and the overtime constraint is not violated.

Per iteration, one strategy is selected with equal probability $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ to find the next neighbor solution.

We follow a similar approach as van Essen et al. (2014) to select appropriate values for the initial temperature T_{in} and final temperature T_f. In our preliminary tests, we used $\theta = \theta_w = \theta_i$, so each ward was given the same weight. The maximum possible decrease of the objective function is given by $\max_{j\in\mathcal{J}} \mu_j + \theta$, which depends on the parameters θ and the surgery groups $\mathcal{J}$. At the start of the procedure, we want to accept this maximum decrease with probability 0.5. Thus, the initial temperature is given by:

$$T_{in} = \frac{-\left(\max_{j\in\mathcal{J}} \mu_j + \theta\right)}{\ln(0.5)}. \tag{43}$$

We determine the final temperature using the same approach. Near the end of the procedure, we want to accept negative changes in the objective function with a low probability. This way, the procedure converges to a local minimum. Our minimum negative change is given by removing the surgery group with the shortest surgery duration, while not influencing the variation in bed occupancy. We set the probability of accepting this change to 0.001 and this gives:

$$T_f = \frac{-\min_{j\in\mathcal{J}} \mu_j}{\ln(0.001)}. \tag{44}$$

Next to the initial and final temperature values, we also need to set the reduction factor, the number of iterations per temperature and the maximum number of accepted solutions per temperature. We used sensitivity analysis to determine the best combination of parameters considering both computational time and solution quality.

5. Computational results

In this section, we present the results of our two approaches. To compare the performance of the global approach and the SA approach, we use a real-life data set. This data includes a master surgery schedule where each OR block is assigned to a specialty. The cycle length is 14 days with 13 ORs where 9 surgical specialties operate. We have 11 wards and one ICU. Data was gathered from interviews with surgeons involved with planning, OR management and the hospital data warehouse. As a result of missing time stamps, 75% of the data set is used. For each surgery group obtained from the data, the mean and variance in surgery duration and LoS are determined. Furthermore, we determine the probability of patients from a surgery group going to ICU $i \in \mathcal{I}$ and ward $w \in \mathcal{W}$. With the model described in Section 3.4, we determine the bed usage distribution resulting from scheduling the surgery groups. The changeover time between surgeries is set to 15 minutes.

In the global approach, we calculate the objective function value differently from the SA approach. The objective function of the MILP is an approximation of the original objective function and only depends on the expected number of beds, while the SA approach considers the original objective function. In order to make a fair comparison, we also determine the original objective function value for the solution given by the MILP. We also combine both approaches by starting with the global approach and then try to improve that solution with SA. Finally, we compare the performance of the best solution of both approaches with the performance of the real-life data set in Section 5.6. For analysis, we also consider computation time as a performance indicator.

We start this section with the results of our clustering approach and parameter settings for both the global and local approach. In Section 5.3, we compare the results of both approaches and try to further improve the value of the objective function by combining both approaches. In Section 5.5, we compare the result of our approach with the commonly used block scheduling approach. Finally, we validate our model using historical data in Section 5.6.

Solving the MILP model is done by using version 4.2.3 of AIMMS. For our MILP model, we use CPLEX version 12.6.3. The SA procedure is implemented in MATLAB R2016b. All computational experiments are performed on a PC with an Intel Core i7 6700K 4.20 gigahertz with 16 gigabytes RAM.

5.1. Clustering

For each specialty, we use the clustering approach as described in Section 3.1. First, we determine the threshold between the short stay group and the long stay group per surgical specialty. The procedures with a median LoS of less than the threshold are denoted as short stay, while the procedures with a median LoS higher than the threshold are in the long stay group. Next, each LoS group is divided into three surgery groups based on the surgery duration of the surgery types. Two thresholds are determined and procedures are put into a short, medium or long surgery duration group depending on the mean surgery duration. However, some LoS groups did not contain enough procedures to be split into three significantly different surgery duration groups. In these cases, only two surgery duration groups are defined. This approach leads to a total of 62 different surgery groups. Four medium surgery duration groups have a precision of less than 0.6 and all belong to different specialties. For these groups, the interval between the two thresholds defining the three surgery duration groups is small (less than 30 minutes). Therefore, the mean surgery duration of certain procedures may fall into the interval between the two thresholds, but many realized instances are outside these bounds, which leads to a low precision. However, the three surgery duration groups have significantly different means, and therefore, our method does define three groups instead of two. Defining less thresholds would increase cluster variance, and therefore, we decided not to adjust our clustering approach for groups with low precision.

5.2. Parameter settings

In this section, we discuss the input parameters for both the global and SA approach.

5.2.1. Global approach

In our MILP model, we only have to define the input parameters based on managerial decisions. These consist of parameter α

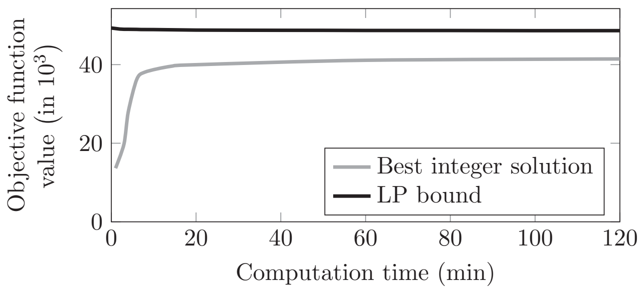

Fig. 3. The value of the LP bound and best integer solution with increasing computation time.

Table 1
Parameter setting for SA approach.

Symbol	Value	Description
T_{in}	1000	Initial temperature
T_f	1	Final temperature
ρ	0.97	Reduction factor
ω	450	Number of iterations for one temperature
ω_{new}	150	Maximum number of new solutions accepted for one temperature

that denotes the overtime probability and parameters θ_w and θ_i to balance the OR utilization and the variation in bed usage at the wards. Preliminary results indicated that setting $\theta_w = \theta_i = 500$ provided the best trade-off between the variation in required number of beds and OR utilization. This means that we would remove scheduled surgery groups with a total OR time of 500 minutes if this would reduce the variation in required number of beds by one. The overtime probability α is set to 0.3.

In Fig. 3, the LP bound and current best solution are shown when increasing the computation time. We see that the solution improves with longer computation times, however, the speed of improvement decreases rapidly after 20 minutes. Since we are creating a tactical schedule, which in theory should only be calculated a couple of times per year, we decided to set the computation time to 90 minutes.

5.2.2. SA approach

As initial solution for SA, we use the incumbent solution obtained after solving our MILP for 60 seconds. Furthermore, we have to set the following parameters: the initial temperature, reduction factor, final temperature and the maximum number of iterations within one temperature. As in Section 5.2.1, we use $\alpha = 0.3$ and $\theta_w = \theta_i = 500$. Table 1 gives an overview of the parameter settings for the SA approach. Using our data, $T_{in} \approx 1000$ and $T_f \approx 5$. However, preliminary results showed that we should set the stopping temperature to $T_f < 1$ to make sure SA converges to a local optimum. Furthermore, the preliminary results showed that we should set the number of iterations for one temperature, given by ω, to 450 and the maximum number of new solutions accepted for one temperature, denoted by ω_{new}, to 150 to obtain acceptable solutions.

5.3. Comparing the global and local approach

We compare the best solutions of both approaches to determine which approach performs best, using five key performance indicators (KPIs): (1) objective value, (2) OR utilization, (3) total number of used beds, (4) total difference in used beds during the cycle, and (5) computation time. The objective function values for both the MILP and SA are calculated using the 90-percentile and 85-percentile of the probability distribution of the number of required beds.

Table 2
Results for the best solution of MILP and SA procedure with 90-percentile.

KPI	MILP	SA
Objective value	41,778	38,699
OR utilization	0.839	0.855
Number of beds	152	159
Difference in beds	12	20
Computation time (hour)	1.5	7

Table 3
Results for the best solution of MILP and SA with 85-percentile.

KPI	MILP	SA
Objective value	41,278	36,518
OR utilization	0.839	0.843
Number of beds	146	149
Difference in beds	13	23
Computation time (hour)	1.5	6

Given the parametrization used in our SA procedure, SA is slower than the MILP approach. The best obtained solution is shown in Table 2 and required seven hours to compute. This can be explained by the large amount of convolutions needed to calculate the objective function value. Recall that we set the computation time of the MILP to 90 minutes.

The MILP also performs best compared to the SA approach for the 85-percentile, see Table 3.

5.4. Improving MILP solution with SA

We also test whether the MILP solution can be improved by the SA procedure. With the initial temperature at $T_{in} = 1000$, we did not obtain better solutions. Therefore, we analyzed different initial temperatures. Results improve slightly for $T_{in} = 10$: the OR utilization improves with 0.66 percentage point to 84.57% and the variation in required number of beds decreases by 1 bed to 11 beds. We need an additional 30 minutes of computation time to obtain this solution.

5.5. Scheduling surgical specialties instead of surgery groups

In our introduction, we state that scheduling surgery groups instead of surgical specialties reduces the OR overtime probability and variation in bed usage. In Section 4.1.1, we have already shown that under the assumption that there exists a surgery group $j \in \mathcal{J}_s$ with $\mu_j \leq \mu_s$ and $\sigma_j^2 \leq \sigma_s^2$, we can schedule at least the same number of surgeries in one OR block when compared to scheduling surgical specialties. This assumption holds for our data.

In addition, our results show that we can even schedule more surgeries when scheduling surgery groups instead of surgical specialties. If we evaluate the solution obtained by scheduling surgery groups on data on surgical specialty level, we see that the OR utilization increases from 84.57% to 100.71%, which means that the obtained solution is not feasible when aggregating the data on surgical specialty level. In addition, we see that the bed variation increases from 12 to 36 beds and the maximum number of required beds increases from 152 to 192. This means that by scheduling the same number and type of surgeries, we need to reserve more OR and bed capacity when aggregating the data on surgical specialty.

Next to this, if we schedule surgical specialties, we see that we cannot meet the restriction on the minimum number of surgeries that should be scheduled per surgical specialty. By relaxing this constraint, i.e., by setting $\beta_s := 0.75\beta_s$, we do obtain a feasible solution with an OR utilization of 61.53%, bed variation of 23 and maximum number of required beds equal to 122. This means that this solution performs worse in terms of maximizing OR utilization

Table 4
Comparison between the historical mean bed variation and the bed variation given by the model.

Ward	Bed variation historical	Bed variation model
Day treatment	9	0
Weekday ward	13	2
Long stay 1	5	0
Long stay 2	4	2
Long stay 3	4	1
Long stay 4	5	2
Long stay 5	2	1
Long stay 6	1	0
Long stay 7	2	0
Long stay 8	4	4
ICU	4	1

and minimizing bed variation and that it is not feasible according to our original constraints.

5.6. Historical versus model performance

The average OR utilization registered in the data set was 71%. The weekly variation in bed occupancy over all wards was 53 beds. When we compare this with our best solution, the variation in bed occupancy can be improved by 42 beds, while the OR utilization can be improved to 84.57%. Given that the available OR capacity has not changed, these results show that more surgeries can be performed while the variation in the number of required beds decreases. In Table 4, we see the historical mean bed variation and the bed variation resulting from our best solution for each ward and the ICU. We also see that for each ward the variation in bed occupancy decreases, or in case of long stay ward 8, stay the same.

6. Problem and model variants

To show the robustness and potential of our model, we analyze several variants of the model using minor modifications. In the first variant of our model, discussed in Section 6.1, we try to avoid occupied beds during the weekends at the weekday ward, as this ward closes during the weekends. In Section 6.2, we describe and test a variant of our model that minimizes the total number of required beds instead of the variation in number of required beds. In Section 6.3, we analyze a relaxation of the MILP. Finally, we apply our MILP model to data from another hospital in Section 6.4.

6.1. Closure during the weekends

The weekday ward (WDW) is intended to be only used during weekdays. When there are still patients admitted at this ward when the weekend starts, these patients have to be transferred to other wards. This setting has not yet been included in our current model. However, we could schedule the surgery groups in such a way that no patients are admitted at the weekday ward during the weekend. We do so by adding a penalty Q for each patient admitted at the weekday ward during the weekend. We introduce variable r which denotes the number of patients admitted at the weekday ward during the weekend. The weekday ward is abbreviated by WDW and the Saturdays and Sundays in the planning horizon are given by set $\mathcal{L}_r \subset \mathcal{L}$. The resulting MILP model is given by

$$\max \sum_{o\in\mathcal{O}}\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}} \mu_j \cdot z_{o,k,j} - \sum_{i\in\mathcal{I}} \theta_i \hat{\gamma}_i - \sum_{w\in\mathcal{W}} \theta_w \hat{\gamma}_w - Q \cdot r \quad (45)$$

$$\text{s.t. } (1)-(3),\ (5)-(6),\ (30)-(41)$$
$$r \geq \bar{\gamma}_{WDW,l}, \qquad \forall l \in \mathcal{L}_r \quad (46)$$

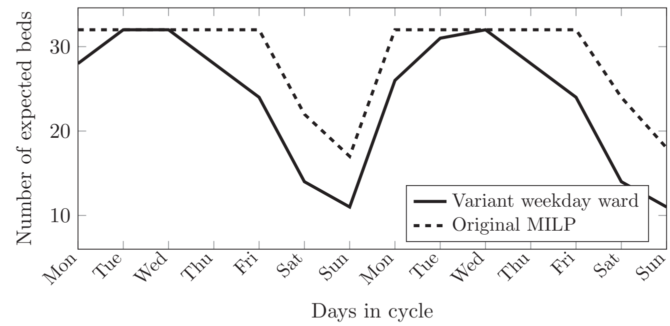

Fig. 4. The expected number of beds at the weekday ward.

Table 5
Results for two variants of the model: (1) minimizing variation in bed utilization and (2) minimizing the number of beds.

KPI	Minimize variation	Minimize required beds
OR utilization (%)	83.9	85.2
Number of beds	152	147
Difference in beds	12	26

In Fig. 4, the results of the model with $Q = 10,000$ and a computation time of 90 minutes is compared to the best solution obtained by the initial MILP model. We see that the expected number of beds during the weekend is reduced, but does not reach zero. It also affects the OR utilization, which decreases by 7.5 percentage point and the difference in beds, which increases by 12 beds. For higher values of Q, the results for the weekday ward do not improve. These results can be explained by the fact that each surgery group has to be scheduled a minimum number of times. In the solution provided when $Q = 10,000$, each surgery group for which patients are admitted to the weekday ward are scheduled the minimum number of times. However, the used surgery groups are not the best predictor for the ward the patients need to be admitted, because every surgery group has some probability that a patient will be admitted at the weekday ward. Therefore, always some patients will be admitted at the weekday ward during the weekend given the used surgery groups.

6.2. Minimize the number of beds

In our model, we minimize the variation in the number of required beds. However, personnel to keep the beds open is expensive. Therefore, instead of minimizing the variation in bed usage, we can also minimize the number of required beds. Even though there might be more variation in bed usage, the number of required beds may decrease.

To minimize the number of required beds, we modify our linear model described in Section 4.1.2. In the modified model, we use the maximum values of $\bar{\gamma}_{i,l}$ and $\bar{\gamma}_{w,l}$ instead of using the difference between the maximum and minimum values of $\bar{\gamma}_{i,l}$ and $\bar{\gamma}_{w,l}$. The resulting MILP is:

$$\max \sum_{o\in\mathcal{O}}\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}} \mu_j z_{o,k,j} - \sum_{i\in\mathcal{I}} \theta_i \hat{\gamma}_i - \sum_{w\in\mathcal{W}} \theta_w \hat{\gamma}_w \quad (47)$$

$$\text{s.t. } (1)-(3),\ (5)-(6),\ (30)-(35)$$
$$\hat{\gamma}_i \geq \bar{\gamma}_{i,l}, \qquad \forall i \in \mathcal{I}, l \in \mathcal{L}$$
$$\hat{\gamma}_w \geq \bar{\gamma}_{w,l}, \qquad \forall w \in \mathcal{W}, l \in \mathcal{L}.$$

The results for this variant of the model are shown in Table 5. We cannot compare objective function values, since different objective functions are used for both methods. The variation in number of required beds is a lot higher, as was to be expected. However, the OR utilization is also higher while less beds are needed in total.

Table 6
Results of relaxation variant disregarding the blocks in the MSS.

KPI	90 minutes	360 minutes
Objective value	40,561	41,847
OR utilization (%)	87.0	87.5
Number of beds	163	162
Difference in beds	18	16
Optimality gap (%)	45.8	39.3

Table 7
Results from other hospital data set.

KPI	10 minutes	90 minutes
OR utilization (%)	86.7	89.6
Number of beds	40	39
Difference in beds	3	3
Optimality gap (%)	15.4	12.2

6.3. Scheduling without blocks

Our model uses the OR blocks of the MSS as input. This means that surgery groups can only be scheduled within the OR blocks that are allocated to this surgical specialty. In this variant of the model, we relax our model by excluding the MSS, meaning that every surgery group can be scheduled at any day within the cycle. In Table 6, we see that the complexity of the problem increases when not using an MSS. After 90 minutes of computation time, the optimality gap is still 46%. The solution has a lower objective function value than the solution of the basic model, which has an objective function value of 41,778, an OR utilization of 83.9%, and a variation in number of required beds of 12. After six hours, the found solution has a higher objective function value than the solution of the basic model. The OR utilization has improved, but the variation in the number of required beds has increased.

Scheduling without OR blocks shows advantages, but ignores many other factors that affect the MSS, e.g. schedules of surgeons, staff and equipment availability, and therefore, implementation is challenging.

6.4. Applying the model to different instances

To analyze whether our approach also works for other real-life instances, we obtained a data set from another hospital. This data set contains 43 surgery groups for which the minimum, mean and standard deviation and LoS probability distributions per surgery group are given. Furthermore, one ward is taken into account. We use the 95-percentile to calculate the required number of beds. The results can be found in Table 7 for different computation times. Bosch (2011) uses the same data set and overtime probability. Their decomposition approach consists of: (1) maximizing the OR utilization and (2) minimizing the required number of beds. The best solution found in Bosch (2011) yields an OR utilization of 91% that needs 45 beds in total. Our solution has 1.4 percentage point lower OR utilization, however, the required number of beds decreases by 6 beds. In Bosch (2011), the OR blocks are formed beforehand, so there is no flexibility in assigning surgery groups to OR blocks when minimizing the required number of beds.

7. Discussion

In this paper, we show the positive impact of the holistic perspective on surgery scheduling. We introduce two single step approaches for scheduling surgery groups while taking into account the overtime constraint and maximizing the OR utilization and minimizing variation of bed usage. Scheduling surgery groups instead of OR blocks leaves fewer options on an operational level to schedule surgeries, and therefore, the probability of overtime and variation in bed usage as a result of surgery scheduling on an operational level decreases. We also added weights to the objective function, θ_i and θ_w, to balance the managerial trade-off between variation of bed usage and OR utilization.

We compare two approaches for finding a good feasible solution for large real-life instances. Both on computational results and on computation time, the MILP outperforms the SA approach. We also combine both approaches where we first optimize the MSS with the MILP, and then try to further improve the objective function value with SA. This combination leads to slightly better results. The MILP shows good results for large real-life instances without long computation times and is therefore suitable for practical applications. Comparing the results of the model with the historical performance derived from the data set, the variation in beds is improved from 53 beds to 11 beds and the OR utilization can be improved from 71% to 85%.

With the use of the MILP, we also analyze some model variants that can give more managerial insights. The first variant focuses on closing the weekday ward during the weekend, as in practice, weekday wards are only opened during weekdays. Weekday wards often struggle with patients that are still admitted during weekends. Without changing the surgery planning, the ward management has two options to solve this problem: (1) extend the opening hours of the weekday ward or (2) transfer these patients to other wards on Friday. So, in the first variant of our model, we extended the model by including a penalty in the objective function for patients being admitted on a weekday ward during the weekend. The computational results show that it is difficult to close such wards during weekends. The next variant of the model minimizes the usage of beds instead of the variation of bed usage which can be achieved by modifying the objective function. Results show that with this approach, the number of required beds can be further reduced and OR utilization increased. However, this also results in an increase in the variation of bed usage.

Furthermore, we relaxed our model such that every surgery group can be scheduled in any OR block in the cycle. The results show that OR utilization can be improved at the cost of an increase in bed variation and required number of beds. To analyze the robustness of our model, we compare our model with another solution approach and data set. Results show that our model has 1.4 percentage point lower OR utilization, however, the required number of beds decreases by 6 beds.

An important step in our approach is the clustering of surgery types into surgery groups. Our clustering approach has a major effect on the group variation, in terms of surgery duration and length of stay, and possible destination wards. With the surgery groups used for our model, we were not able to close the weekday ward during the weekend, because too many surgery groups may use the weekday ward after surgery. Therefore, we conclude that the groups at hand are still too aggregated for this model variant. Further research on applying data mining on such instances could increase the predictive value of clusters (in our case surgery groups), and therefore, improve the robustness of planning.

When clusters are only based on specialty, we obtain a model which schedules OR blocks similar as previous work (Fügener et al., 2014; van Essen et al., 2014; Vanberkel et al., 2011b). As shown here, our clustering approach results in more precise predictions for surgery duration and LoS, and therefore, results in a higher OR utilization and lower variation in bed usage. However, smaller clusters (e.g. clustered surgery types versus clusters based on specialty) require more data to attain similar precision levels. Therefore, our approach assumes no limitations in data availability. Since most hospitals nowadays have advanced electronic health record systems, this would be a fair assumption to make. Next to possible data limitations, our model assumes that the clusters also account for the biweekly number of realizations of surgery types (e.g. at least once every two weeks) such that each block can be filled with surgeries of that type on the operational level. Furthermore, the dispersion of durations within surgery groups should be limited. When this is not the case, it could result in under- or overutilization of resources, since on the operational level, surgery types with significant shorter or longer individual durations could be scheduled than was accounted for when scheduling surgery groups on the tactical level.

The model can also be extended to optimize the schedule of surgeons. To achieve this, the model should not only take the OR and its downstream resources into account, but also its upstream resources such as the outpatient clinic given that surgeons also work there. Another potential direction for further research is optimizing break-in moments for OR cleaning.

Overall, this research provides a way to bridge the gap between tactical and operational planning of surgeries. It reduces the variation in bed usage and improves the robustness of the schedules. The use of surgery groups makes it possible to easily implement our model into practice, and for operational planners, it is instantly clear where to schedule what type of surgery. With only minor model modifications, we show that a broad range of variants on OR scheduling can be analyzed to obtain valuable managerial insights.

Acknowledgments

The authors would like to thank the anonymous referees for their useful comments, which helped to improve the paper. We are also grateful for the valuable comments of our CHOIR colleague Maartje E. Zonderland.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2019.09.029.

References

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.

Adan, I., Bekkers, J., Dellaert, N., Vissers, J., & Yu, X. (2009). Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning. *Health Care Management Science, 12*(2), 129–141.

Astaraky, D., & Patrick, J. (2015). A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. *European Journal of Operational Research, 245*(1), 309–319.

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., & Koeleman, P. M. (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Management Science, 14*(3), 49–237.

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., & Cardoen, B. (2009). A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling, 12*(2), 147–161.

Bisschop, J. (2016). Aimms optimizating modelling.

Bosch, J. (2011). Better utilisation of the or with less beds. Master's thesis.Industrial Engineering & Management, University of Twente.

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

Chow, V., Puterman, M., Salehirad, N., Huang, W., & Atkins, D. (2011). Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management, 20*(3), 418–430.

Fügener, A. (2015). An integrated strategic and tactical master surgery scheduling approach with stochastic resource demand. *Journal of Business Logistics, 36*(4), 374–387.

Fügener, A., Edenharter, G., Kiefer, P., Mayr, U., Schiele, J., Steiner, F., … Blobner, M. (2016). Improving intensive care unit and ward utilization by adapting master surgery schedules. *A A Case Rep, 6*(6), 172–180.

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.

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

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

Hans, E. W., van Houdenhoven, M., & Hulshof, P. J. H. (2012). A framework for healthcare planning and control. In R. Hall (Ed.), *Handbook of healthcare system scheduling* (pp. 303–320). Boston, MA: Springer US.

May, J. H., Strum, D. P., & Vargas, L. G. (2000). Fitting the lognormal distribution to surgical procedure times. *Decision Sciences, 31*(1), 129–148.

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

Santibanez, 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.

van der Aalst, W. M. P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin, Heidelberg, Germany: Springer.

van Essen, J., Bosch, J., Hans, E., van Houdenhoven, M., & Hurink, J. (2014). Reducing the number of required beds by rearranging the or-schedule. *Or Spectrum, 36*(3), 585–605.

Van Oostrum, J., van Houdenhoven, M., Hurink, J., Hans, E., Wullink, G., & Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum, 30*(2), 355–374.

Vanberkel, P., Boucherie, R., Hans, E., Hurink, J., van Lent, W., & van Harten, W. (2011a). Accounting for inpatient wards when developing master surgical schedules. *Anesthesia and Analgesia, 112*(6), 1472–1479.

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

Yahia, Z., Eltawil, A., & Harraz, N. (2016). The operating room case-mix problem under uncertainty and nurses capacity constraints. *Health Care Management Science, 19*(4), 383–394.