Operations Research for Health Care 22 (2019) 100196

Contents lists available at ScienceDirect

Operations Research for Health Care

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

Chance-constrained admission scheduling of elective surgical patients in a dynamic, uncertain setting

Wim Vancroonenburg [a,b,*,1], Patrick De Causmaecker [b], Greet Vanden Berghe [b]

[a] Research Foundation Flanders - FWO Vlaanderen, Belgium
[b] KU Leuven, Department of Computer Science, CODeS, Belgium

HIGHLIGHTS

- A chance-constrained model for admission scheduling of surgical patients is proposed.
- Sample average approximation is employed for approximating the stochastic model.
- A Late Acceptance Hill Climbing meta-heuristic is implemented for solving the models.
- The flexibility of the model is illustrated by implementing four admission policies.
- Computational results show that the model enables managing the risk on bed shortages.

ARTICLE INFO

Article history:
Received 20 July 2018
Accepted 11 July 2019
Available online 15 July 2019

Keywords:
Surgery admission scheduling
Chance-constrained stochastic optimization
Sample average approximation
Meta-heuristic

ABSTRACT

In the present contribution, a chance-constrained scheduling model is presented for determining admission dates of elective surgical patients. The admission scheduling model is defined considering a dynamic, stochastic decision-making environment. The primary aim of the model concerns the minimization of operating theatre costs and patient waiting times, while simultaneously avoiding bed shortages at a fixed certainty level through a chance-constrained formulation. This stochastic model is implemented by means of sample average approximation and is solved by a meta-heuristic algorithm. To illustrate the applicability of the model, the approach is used to implement four admission scheduling policies on this dynamic decision-making setting that are evaluated on different criteria in a computational study using simulation. The results show that the stochastic approach is able to account for the uncertainty in patients' length of stay and surgical procedure duration, enabling it to avoid bed shortages while still optimizing operating theatre costs and patient waiting times.

1. Introduction

Over the past decades, globally rising healthcare expenditure has forced governments to re-evaluate healthcare funding. To reduce public spending on healthcare, budgetary pressure on hospitals has increased tremendously. Spurred by this trend, hospital managers are seeking ways to improve the efficiency of their most costly resources. Faced with the uncertain and highly variable nature of the care process, they attempt managing patient flow to their best extent while rationalizing resource usage. One key instrument available to help manage this flow is scheduling elective patient admissions. Contrary to urgent or emergency patients, elective patients represent the only set of patients whose procedures may be planned and thus 'tuned' to the hospital's available capacity and resources. However, implementing such a system requires careful planning, an overview of all considered resources, forecasts concerning urgent/emergency admissions and other uncertain factors.

Admission scheduling for patients is generally performed by manual planners (e.g. an admissions office). An elective patient's admission request typically occurs after a physician or surgeon has determined that the patient may require admission for treatment in the near future. This process is very often myopic. For example, surgeons may schedule surgery for an elective patient by only considering their own availability and the availability of the operating theatre. Due attention is rarely given to post-operative patient recovery needs, namely their requirement for an available bed for an unknown (though estimable) period of time. This lack of foresight results in *bed shortages* at nursing wards, requiring patients to be admitted to different wards, admissions to be postponed or even, worst-case, to be cancelled.

* Corresponding author at: KU Leuven, Department of Computer Science, CODeS, Belgium.
E-mail address: wim.vancroonenburg@kuleuven.be (W. Vancroonenburg).
1 KU Leuven, Ghent Technology Campus, Gebroeders De Smetstraat 1, 9000 Gent, Belgium.

https://doi.org/10.1016/j.orhc.2019.100196

This paper presents a heuristic optimization approach for admission scheduling employing stochastic information and sampling to consider variations in patients' recovery times and surgery durations. A chance-constrained formulation models the risk of bed shortages and maintains a fixed certainty level. The scope of this study is restricted to the admission scheduling of elective surgical patients. Patients being admitted for surgery represent as much as 60% of hospital admissions [1]. The surgical patient flow is of particular interest as it involves the hospital's operating theatres. Operating theatres may be considered a key hospital resource, generating significant revenue, though, due to their resource intensive nature at considerable cost.

The scheduling model assumes a *dynamic* decision-making setting: as is the case in practice, a hospital is never empty to begin with. Hence, patients may already be present in nursing wards, some admissions may already be planned and some may be queued on a waiting list. The scheduling model accounts for all these elements, enabling the flexible application of the approach. To exemplify the flexibility and to validate the approach in various settings, four different admission scheduling strategies are developed and compared. These four strategies differ in both their assumed scheduling horizons, as well as the frequency of decision-making (daily vs weekly). Consequently, these four strategies have different degrees of flexibility for optimization and will therefore achieve varying levels of performance.

The paper is structured as follows. Section 2 discusses the relevant literature regarding admission scheduling and surgery scheduling, and positions the present study in this context. Next, in Section 3 the admission scheduling problem is defined and formulated as a Mixed-Integer Linear Programming (MILP) model: firstly, in a static, deterministic setting to illustrate the main elements; and secondly, in a dynamic, stochastic setting in full detail. In Section 4, the solution approach for solving the scheduling model is described. The stochastic MILP model is approximated using Sample Average Approximation (SAA) and a heuristic local-search algorithm is described for solving it. Section 5 describes the aforementioned admission strategies that are used to exemplify the approach. These are tested in Section 6 which discusses the setup and results of our experimental study. Finally, Section 7 concludes our study.

2. Literature review & contribution

2.1. Related work

This paper addresses admission scheduling decisions related to both surgical care services and inpatient care services in hospitals. According to the taxonomic classification of healthcare planning decisions by Hulshof et al. [2], the former services concern the provision of operative procedures to patients, whereas the latter are concerned with the provision of room, bed and board to inpatients (hospitalized patients requiring an overnight stay). Both inpatient and surgical care services have received considerable attention in the literature over the past decades. It is beyond the scope of this paper to thoroughly review such work; for this purpose there already exist several excellent, recent surveys in [2–6].

Planning decisions in hospitals are typically made in a hierarchical system. First, long-term strategic decisions are made concerning structural direction on how a hospital can cope with service demand for the years to come (based on forecasts). Secondly, and less abstractly, tactical decisions concern setting up guidelines, mid-term schemes (weeks, months) to organize work and facilitate operational planning. Finally, operational planning involves short-term decisions on how to execute the service process [2].

With respect to the admission process and inpatient care services, both the strategic and tactical decision-making have been the subject of numerous studies. Capacity dimensioning and allocating wards/beds and staff have been particularly well studied with a variety of techniques (simulation, mathematical programming, Markov processes, queueing theory). One approach to admission scheduling at the tactical decision level concerns the development of admission control schemes to set admission quota. Such quotas serve as rules by which admission planners can assign individual admissions. For example, Bekker and Koeleman [7] investigated approximations for determining the impact of daily variability in admissions on fluctuations in bed demand and blocking probabilities. Using these approximations, Bekker and Koeleman employed a quadratic programming model to determine an admission scheme that minimizes weighted deviations of the bed usage from a defined target load. Hulshof et al. [8] developed a mixed integer linear programming (MILP) model to construct tactical resource allocations and admission plans. The model determines a selection of patients to be served that are in a particular stage of their care process. Their main aims were to achieve equitable access for patients, to meet admission targets and to use resources efficiently.

Several studies have examined the link between admission scheduling and downstream nursing ward bed usage in the context of operating theatre planning. Mostly the connection between the master surgical schedule (MSS) and the impact on bed usage in surgical nursing wards has been considered. The MSS is a typically cyclical, weekly/bi-weekly/monthly timetable, that allocates operating rooms and time slots to individual surgeons/surgical disciplines based on their allocated capacity. The distribution of these OT slots over the MSS can be related to the bed usage in the surgical wards (through consideration of the length of stay (LOS) distributions and arrival patterns of the different surgical disciplines, see e.g. [9–12]). By manipulation of the MSS, it is possible to reduce fluctuating bed usage and possibly reduce the total number of required beds.

The admission scheduling problem considered by the present study is concerned with the operational decision level, dealing with the scheduling and execution of the admission process for individual patients. Vissers et al. [13] developed a platform for comparing different admission strategies, considering resources such as the operating theatre, nursing requirements, bed usage and intensive care unit (ICU) usage. Different admission strategies such as *maximum resource usage* that employ waiting lists, *booked admissions* (determining an admission date at request time) and *zero wait* (admitting a patient at request time) were compared with respect to several performance measures. Mazier et al. [14] discussed the problem of scheduling inpatient admissions in a hospital with a highly uncertain LOS and a significant number of emergency admissions. Their main concern was to assure enough beds are available for unforeseen emergency patients and future elective patients. Mazier et al. modelled this admission scheduling scenario as a stochastic programming problem, studying and proposing different estimation techniques for assessing the number of beds required by emergency patients. Schmidt et al. [15] presented an admission scheduling decision support system that determines both admission dates and bed assignments for elective patients. The system considers patient gender, priorities, and preferences, the availability of LOS estimates and dynamic adjustment of the LOS estimate if patients stay longer than expected. Both an exact and heuristic methods are presented for the decision support system and are compared through simulation. Gartner and Kolisch [16] presented an approach to schedule hospital wide patient flow at the operational level. Patient admissions are assumed to be classified by their diagnosis-related group and corresponding clinical pathway. The clinical pathway gives a blueprint of the sequence of activities, such as diagnostics or surgeries that will occur during a patient's stay, in addition to their corresponding resource requirements. Two mixed-integer programming models are presented that aim to maximize the total contribution margin of performing these activities, one in which admission dates are assumed fixed and another in which admission dates are decision variables. The models are tested and compared on real world data from a mid-size hospital in both a static and rolling horizon approach. Ceschia and Schaerf [17] studied a patient admission scheduling problem as an extension to a patient-to-room assignment problem, which was originally defined by Demeester et al. [18] and subsequently studied by [17,19–23]. Their extension expands the assignment problem to an admission scheduling problem that accounts for the availability of suitable hospital rooms in nursing wards for patients' admissions. Of particular interest is that their model assumes a dynamic context, making scheduling decisions on a daily basis. Furthermore, they account for some stochastic variation in patients' LOS, and minimize overcrowding risks. They published several instances for academic study, to which Lusby et al. [24] hold the best results. Ceschia and Schaerf [25] expand on their previous study to further consider the availability of operating theatre capacity for surgical patients.

For surgical patients, admission dates are generally derived from their assigned surgery dates — as mentioned earlier in the introduction. Considerable attention has been devoted to surgery scheduling at the operational level (see [3,5,6] for a thorough review). Assigning surgery dates to patients for an upcoming period is achieved via a process step denoted *advance scheduling* [26,27], though other authors have also coined terms such as "intervention assignment" [28] and "surgical case assignment" [29]. Many studies focus solely on the OT, assigning dates to individual surgical cases. This is often a weekly process whereby surgical cases are selected from a waiting list of those to be performed in the upcoming week and (pre-) assigned to individual ORs. The main objectives considered are minimizing overtime and underutilization of ORs [30,31], patient related costs and improving quality of service measures such as waiting time [30,32] and due date tardiness [28].

Some studies particularly relevant to the work presented in this paper, concern the application of stochastic optimization methods to advance scheduling. Hans et al. [33] presented a robust surgery loading study for the advance scheduling problem. Their aim was to minimize overtime and to maximize free capacity, whilst considering uncertainty in surgery durations and varying flexibility with respect to a base schedule. They minimize the required slack for avoiding overtime by exploiting the *portfolio effect* –decreasing risk by increasing diversity (by means of non-correlated portfolio components). Bruni et al. [34] presented a heuristic approach to a stochastic programming model for the advance scheduling problem considering different recourse methods. On a weekly basis, an advance schedule is constructed that maximizes an abstract priority-weighted profit of performing surgeries, decreased by the expected recourse cost. The different recourse methods considered either performing surgical cases in overtime, redistributing surgical cases between ORs or completely rescheduling surgical cases.

Very few of the aforementioned studies regarding surgery/advance scheduling consider downstream bed usage at nursing wards. One notable exception is the consideration of the intensive care unit (ICU), which is considered a bottleneck resource for certain surgical procedures. Min and Yih [35] presented a model to determine an optimal surgery schedule for elective patients, considering uncertainty on surgery durations and LOS in downstream care units (specifically the ICU). A stochastic optimization approach employing the sample average approximation method (SAA) is proposed and is proven superior at minimizing overtime to using solely estimated values. However, this comes at the expense of additional underutilization. More recently, Jebali and Diabat [36] also recognized the need for considering other scarce resources (ICU, nursing ward beds) during surgical scheduling, to balance utilization of hospital resources and to prevent system congestion. They account for post-surgery patient recovery in the ICU and nursing wards, and present a stochastic model that deals with uncertainty on surgery durations and length of stay in ICU and nursing wards. Again, the SAA method is applied for solving their two-stage stochastic programming formulation. Later on, Jebali and Diabat [37] presented a two-stage chance-constrained stochastic programming formulation for the surgical scheduling problem. In this study, Jebali and Diaba provide a parametrizable risk factor that upper-bounds the probability of cancellation due to a shortage of beds in the ICU. SAA is applied for solving this two-stage chance-constrained formulation.

2.2. Positioning of the paper and its contribution

In past research efforts, there has been a clear focus to optimize the usage of the operating theatre or nursing wards (and other resources) as independent units. However, their interdependence cannot be ignored to achieve a globally better performing hospital. This is supported by more recent studies that explicitly take into consideration extra resources, for example considering downstream ICU beds for surgery scheduling (e.g. [35–37]), or by tuning the MSS to also account for downstream nursing ward occupancy (e.g. [9,11]). Another observation is the growing body of research on methods that account for the stochastic nature of the problem (e.g. [36,37]), by assuming knowledge of surgery/recovery distributions to obtain more robust solutions or more accurate estimations of resource performance.

Our study contributes to both these research directions. Firstly, it explicitly shifts the focus away from operating theatre/surgical scheduling to surgical *admission* scheduling. Hence, the perspective is that of inpatient scheduling that requires taking into account beds for patients to stay in after surgery, for possibly a significant period of time (a few days to several weeks) depending on the patients' medical conditions. The scope of the planning problem thus extends over the mid-term, e.g. a month of planning admissions. This contrasts with for example the studies of Jebali and Diabat [36,37] who mostly discuss their model in the context of operating theatre scheduling with a planning horizon of one week. Given such a lengthy period, it is impossible to ignore the *dynamics* of the problem: patients enter and leave the hospital and the planning model must account for the partial occupancy of wards by patients (expected to be or currently) present. Similarly to Ceschia and Schaerf [25,38], this study applies a dynamic scheduling approach belonging to a class of predictive–reactive scheduling approaches [39] "in which the schedule is revised in response to real-time events". In our study, events (patient admission requests, admissions, discharges) are grouped on a daily basis. We demonstrate the flexibility and the usability of the presented model in such an environment by defining several different admission strategies that have various degrees of planning flexibility and planning frequency.

Secondly, this study employs a chance-constrained stochastic programming formulation that accounts for the stochastic variation in patients' surgery durations and LOS in nursing wards. Similarly to Jebali and Diabat [37], we formulate a chance constraint with parametrizable risk factor that upper-bounds the probability of cancellation due to a shortage of beds. In our view, the main benefit of this approach is to provide a more intuitive parameter to decision makers, rather than including hard-to-tune, instance-specific weighted penalties in the objective function of the optimization model. The objective function of our model only concerns costs related to operating theatre costs and patient access (waiting) time. Since we have experienced that decision makers struggle with weighting such different terms, a lexicographic set of weights is used: operating theatre costs are the primary minimization goal, secondary to patient access time. Note that, the chance-constrained stochastic model by Jebali and Diabat [37] is solved using SAA and a MILP solver. Due to the increased scope of the problem in the present study, this approach becomes computationally costly. Therefore, we show how such a chance-constrained model can be embedded in and solved by a metaheuristic algorithm.

3. Problem definition

In this section, we detail the problem under consideration in two sections. Firstly, in Section 3.1 we describe the admission scheduling problem as an MILP in a static, deterministic setting to focus on its main elements and purpose. Secondly, in Section 3.2 we further detail how the previous model should be adopted in a more realistic, dynamic uncertain setting.

3.1. The static, deterministic problem

This paper considers a setting in which a set of elective surgical patients P of a single[2] surgical discipline must be admitted to a hospital for treatment over a certain planning horizon, represented as a set of days $\{1, 2, \ldots, D\}$. Each patient $p \in P$ requires a certain surgical procedure to be performed and requires some days for recovery in a surgical ward before being discharged. The surgical discipline has a certain ward capacity, B, that is used by admitted patients (inpatients) for preoperative stays and recovery. In addition, the surgical discipline utilizes a set of operating rooms O for performing surgeries, in which the regular (i.e. not considering overtime) allocated surgery capacity (in minutes) for OR $o \in O$ on day d $(1 \leq d \leq D)$ is denoted by CAP_{od} (determined by an MSS). Overtime is also allowed, but total surgery time per OR and day is restricted by $CAP_{od}^{\max}$.

Each patient p is characterized by:

- the instant at which the request for admission was made r_p ($r_p \geq 0$, in fractional days);
- the earliest possible day for admission rd_p, which is at least the day after the admission request was made (i.e. $rd_p \geq \lceil r_p \rceil$);
- the pre-operative period before the surgery day the patient must be admitted $preop_p$ (≥ 0, in days);
- the duration of the surgical procedure d_p (in minutes). It is assumed that $d_p \leq CAP_{od}^{\max}$;
- the length of the recovery period los_p (in days).

The aim is to provide each patient an admission day a_p, ($rd_p \leq a_p \leq D$), considering that a bed must be available, and simultaneously optimizing two performance measures:

1. *Operating theatre usage*: irrespective of individual surgery/treatment costs and profits (i.e., some treatments may be more profitable than others), the aim is to utilize available operating theatre capacity as good as possible. Deviation from this available capacity should be minimized as both underutilization and overtime increase operating costs: underutilization decreases revenue and incurs opportunity costs, whereas overtime results in additional staffing costs, often at a higher overtime-rate.
2. *Patient access time*: a secondary aim is to admit patients in a timely fashion. Although elective patients are considered not urgent, long access times increase the risk of a deteriorating medical condition, which must be avoided. Moreover, longer access times negatively impact patient satisfaction.

The problem may be modelled by a MILP formulation as follows. Define the following decision variables:

$$x_{pd} = \begin{cases} 1 & \text{if patient } p\text{'s admission day is } d, \text{ i.e. } a_p = d, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$$y_{pdo} = \begin{cases} 1 & \text{if patient } p \text{ undergoes surgery on day } d \text{ in OR } o, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The model's objectives are to both minimize and balance operating theatre costs and patient access time. Let ot_{od}^{+}, ot_{od}^{-} denote real variables representing the overtime resp. undertime in OR o on day d. Finally, let AD_p denote the set of possible admission days for patient p, i.e. $AD_p := \{d | rd_p \leq d \leq D\}$. Thus, the objective is:

$$\text{Minimize } W^{\text{OT}} \cdot \sum_{1 \leq d \leq D} \sum_{o \in O} \alpha \cdot ot_{od}^{+} + ot_{od}^{-} + W^{\text{WAIT}} \cdot \sum_{p \in P} \sum_{o \in O} \sum_{d \in AD_p} (d - rd_p) \cdot x_{pd} \tag{3}$$

W_{OT}, W_{WAIT} denote weights reflecting the relative importance of different costs and α denotes the cost ratio of overtime to underutilization. As mentioned earlier, while any weights can be applied, in this study a lexicographic scheme is applied, i.e. $W_{\text{OT}} \gg W_{\text{WAIT}}$. Hence, running the operating theatre as cost-efficient as possible is the primary goal, while still aiming for a compact schedule by minimizing patient access time.

The model is subject to the following constraints. Firstly, to ensure each patient is only assigned one surgery day and OR, and to ensure the relation between y_{pdo} and x_{pd} variables:

$$\sum_{d \in AD_p} x_{pd} = 1 \qquad \forall p \in P \tag{4}$$

$$\sum_{o \in O} y_{p\delta o} = x_{pd} \qquad \forall p \in P, d \in AD_p, \delta = d + preop_p \tag{5}$$

Note that the admission day and surgery day are separated by the required number of pre-operative days. Secondly, the following constraints concern, for each operating room o and each day d of the planning horizon, the relation between the surgical load and the deviations (overtime/undertime) from the operating room's target capacity CAP_{od}. It also restricts the maximum overtime allowance.

$$\sum_{p \in P} d_p \cdot y_{pdo} \leq CAP_{od} + ot_{od}^{+} \qquad \forall o \in O, 1 \leq d \leq D \tag{6}$$

$$CAP_{od} - \sum_{p \in P} d_p \cdot y_{pdo} \leq ot_{od}^{-} \qquad \forall o \in O, 1 \leq d \leq D \tag{7}$$

$$\sum_{p \in P} d_p \cdot y_{pdo} \leq CAP_{od}^{\max} \qquad \forall o \in O, 1 \leq d \leq D \tag{8}$$

In a deterministic setting, admission scheduling should be such that all elective patients that are admitted for surgery can be admitted to the hospital ward. Therefore, the following expression relates admission scheduling to its corresponding bed usage and constrains it to be at most equal to the available bed capacity B.

[2] Note that the subsequently described problem may easily be extended to a multi-discipline setting, when provided with decision makers' preferences on how balance between different disciplines should be handled. Since we are interested in the model behaviour, rather than fine-tuning this discipline trade-off, we opt for the single discipline setting.

Let Ω_{pd} denote the set of possible admission days of patient p that lay within los_p days of day d, i.e. $\Omega_{pd} := \{\delta | \delta \in AD_p \wedge \delta \leq d \leq \delta + los_p\}$:

$$\sum_{p \in P} \sum_{\delta \in \Omega_{pd}} x_{p\delta} \leq B1 \leq d \leq D \tag{9}$$

Fourth, for each patient, the decision variables are constrained to the only days relevant for that patient — those after their earliest admission date rd_p:

$$x_{pd} = 0 \qquad \forall p \in P, 1 \leq d < rd_p \tag{10}$$

$$y_{pdo} = 0 \qquad \forall p \in P, o \in O, 1 \leq d < rd_p + preop_p \tag{11}$$

Finally, the domains of the decision variables x_{pd}, y_{pdo} and aid variables ot^+_{od}, ot^-_{od} are defined.

$$x_{pd} \in \{0, 1\} \qquad \forall p \in P, 1 \leq d \leq D \tag{12}$$

$$y_{pdo} \in \{0, 1\} \qquad \forall p \in P, o \in O, 1 \leq d \leq D \tag{13}$$

$$ot^+_{od}, ot^-_{od} \geq 0 \qquad \forall o \in O, 1 \leq d \leq D \tag{14}$$

3.2. The dynamic, stochastic problem

Clearly, the previous problem formulation and MILP model is not applicable to *dynamic* decision-making as it assumes knowledge of uncertain parameters. The release date rd_p, surgery duration d_p and length of stay los_p of all patients $p \in P$ only become known when patient admission requests arrive, when patients undergo surgery and, finally, when they are discharged. Knowledge of these uncertain parameters cannot be assumed *a priori*. A hospital (ward) is also rarely empty to begin with. In practice, several patients that arrived prior to the current planning period will remain in the hospital. Therefore, decision-making can only be done in a dynamic setting as new events occur. However, this does not imply that decision-making must take place immediately. Decisions concerning when to admit patients may be postponed (put on a waiting list) and gathered to make better decisions than when instantaneously deciding an admission date when an admission request arrives. Furthermore, the presumption that there is absolutely no information present concerning uncertain parameters must not be made. While a patient's LOS may remain unknown until they are discharged, there is information available regarding previous patients that have undergone similar procedures. A similar argument holds for surgery duration. Therefore, knowledge concerning the expected LOS and surgery duration of any patient may be assumed, in addition to an idea of their distribution.

The dynamic problem considers a day d' (with $1 \leq d' \leq D$) at which patient admissions must be scheduled within the planning horizon. At this point, a set of patients has already been assigned admission dates (and are possibly already admitted to the hospital ward), whilst a waiting list of patient requests for admission may have accumulated. Let $A_{d'} \subseteq P$ denote the patients that have already received an admission date, i.e. $A_{d'} = \{p \in P | r_p \leq d' \text{ and } a_p \neq \text{unassigned}\}$ and let $P_{d'} \subseteq P$ denote patients that are in the waiting list, i.e. $P_{d'} = \{p \in P | r_p \leq d' \text{ and } a_p = \text{unassigned}\}$.

For these patients $p \in P_{d'}$, we only assume the following is known:

- the earliest possible date for admission rd_p,
- the number of days before surgery the patient must be admitted $preop_p$,
- the distribution of random variable ξ_p, the unknown duration of p's surgical procedure,
- the distribution of random variable ζ_p, the unknown LOS of patient p.

In addition, at day d' information on departures of admitted patients becomes known if departures occur (i.e. the true LOS of any admitted patient is only known at time of departure).

Consider again model (3)–(14). In the dynamic, stochastic setting the problem's decision variables remain the same:

$$x_{pd} = \begin{cases} 1 & \text{if patient } p\text{'s admission day} \\ & \text{is day } d \text{, i.e. } a_p = d, \\ 0 & \text{otherwise.} \end{cases} \qquad \forall p \in P_{d'}, d' \leq d \leq D \tag{15}$$

$$y_{pdo} = \begin{cases} 1 & \text{if patient } p\text{'s surgery day} \\ & \text{is day } d \text{ in OR } o, \\ 0 & \text{otherwise.} \end{cases} \qquad \forall p \in P_{d'}, o \in O, d' \leq d \leq D \tag{16}$$

By adapting the model to minimize and balance *expected* operating theatre costs and patient access times, the uncertain nature of operating theatre usage is accounted for. The ultimate aim is to minimize the usage of the operating theatre from its target utilization, in the average case. Let $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)$ denote a vector of random variables representing the surgery durations of $N = |P|$ patients and let $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_N)$ denote a random vector representing the patients' LOS.

Let $ot^+_{od}(\boldsymbol{\xi}), ot^-_{od}(\boldsymbol{\xi})$ correspond to positive, real-valued random variables representing the overtime resp. undertime in operating room o at day d. The objective thus becomes:

$$\text{Minimize } W_{\text{OT}} \cdot \mathbf{E}_{\Xi}\left[\sum_{d' \leq d \leq D} \sum_{o \in O} \alpha \cdot ot^+_{od}(\boldsymbol{\xi}) + ot^-_{od}(\boldsymbol{\xi})\right] + W_{\text{WAIT}} \cdot \sum_{p \in P_{d'}} \sum_{rd_p \leq d \leq D} (d - rd_p) \cdot x_{pd} \tag{17}$$

where $\mathbf{E}_{\Xi}[\ldots]$ denotes the expected value of $(\ldots)$ over all possible realizations of $\boldsymbol{\xi}$. Note that in the dynamic setting only patients in the waiting list $P_{d'}$ are scheduled, while patients that have been scheduled previously (and who may already be present in the hospital) are taken into account regarding their bed and surgery usage.

Assume that one may consider all possible realizations of random vector $\boldsymbol{\xi}$, represented by $\boldsymbol{\xi}^k \in \Xi$ with $\Xi \subseteq \mathbb{R}^N$ denoting the sample space of the probability distribution of $\boldsymbol{\xi}$. The model is subject to the same range of constraints discussed earlier, but now taking into account the realization of random variables $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, as well as the presence of previously scheduled patients. Firstly, the following constraints enforce that each patient is admitted exactly once and operated upon once:

$$\sum_{rd_p \leq d \leq D} x_{pd} = 1 \qquad \forall p \in P_{d'} \tag{18}$$

$$\sum_{o \in O} y_{p\delta o} = x_{pd} \qquad \forall p \in P_{d'}, rd_p \leq d \leq D, \delta = d + preop_p \tag{19}$$

Secondly, the following constraints concern, for each operating room o and each day d of the scheduling horizon, the relation between the surgical load to the deviations (overtime/underutilization) from the operating room's target capacity CAP_{od}. In the dynamic setting, the expressions consider both patients that are still to be scheduled and those in $A_{d'}$ already admitted (if $a_p \leq d'$) or whose future admission date is already fixed (if $a_p > d'$).

$$\sum_{p \in P_{d'}} \xi^k_p \cdot y_{pdo} + \sum_{\substack{p \in A_{d'}: \\ d = a_p + preop_p}} \xi^k_p \cdot y_{pdo} - CAP_{od} \leq ot^+_{od}(\boldsymbol{\xi}^k) \qquad \forall o \in O, d' \leq d \leq D,$$

$$\xi^k \in \Xi \quad (20)$$

$$CAP_{od} - \sum_{\substack{p \in A_{d'}: \\ d = a_p + preop_p}} \xi_p^k \cdot y_{pdo} - \sum_{p \in P_{d'}} \xi_p^k \cdot y_{pdo} \leq ot_{od}^-(\xi^k) \qquad \forall o \in O, d' \leq d \leq D, \; \xi^k \in \Xi \quad (21)$$

where ξ_p^k is the realization of patient p's surgery duration under scenario k, and $ot_{od}^+(\xi^k)$, $ot_{od}^-(\xi^k)$ the realization of the resp. overtime, underutilization random variables on day d, OR o under scenario k.

To limit hospital ward usage for each day of the planning horizon, a chance-constraint [40] is introduced to maintain usage to be below or equal to the number of available beds with probability η.

$$\mathbf{Pr}_Z \left\{ \sum_{\substack{p \in A_{d'} \\ a_p \leq d < a_p + \zeta_p}} 1 + \sum_{p \in P_{d'}} \sum_{\substack{\delta : rd_p \leq \delta \\ d \in [\delta, \delta + \zeta_p - 1]}} x_{p\delta} \leq B \right\} \geq \eta \qquad d' \leq d \leq D \quad (22)$$

where $Z \subset \mathbb{R}^N$ denotes the sample space of the probability distribution of ζ, and $\mathbf{Pr}_Z\{\ldots\}$ denotes the probability of event $(\ldots)$ occurring.

Again, the decision variables associated with each patient are constrained to the relevant days of the scheduling horizon — those after each patient's earliest admission date rd_p:

$$x_{pd} = 0 \qquad \forall p \in P_{d'}, d' \leq d < rd_p \quad (23)$$

$$y_{pdo} = 0 \qquad \forall p \in P_{d'}, o \in O, d' \leq d < rd_p + preop_p \quad (24)$$

Finally, the domains of decision variables x_{pd}, y_{pdo} and aid variables $ot_{od}^+(\xi^k)$, $ot_{od}^-(\xi^k)$ are defined:

$$x_{pd} \in \{0, 1\} \qquad \forall p \in P_{d'}, d' \leq d \leq D \quad (25)$$

$$y_{pdo} \in \{0, 1\} \qquad \forall p \in P_{d'}, o \in O, d' \leq d \leq D \quad (26)$$

$$ot_{od}^+(\xi^k), ot_{od}^-(\xi^k) \geq 0 \qquad \forall o \in O, d' \leq d \leq D, \xi^k \in \Xi \quad (27)$$

4. Solution approach

In this section, we detail how the stochastic MILP model presented in the previous section is solved. Firstly, Section 4.1 describes how the stochastic model (17)–(27) is converted into a deterministic model using SAA. Secondly, Section 4.2 describes how the SAA model, which is difficult to solve using MILP solvers, can be solved using a local search approach.

4.1. Sample average approximation

Formulation (17)–(27) is difficult to solve in general as it depends on random vectors ξ and ζ and all their possible realizations (potentially infinite).

Therefore, SAA [41] can be used to approximate the stochastic model, transforming it into a deterministic one. Informally, SAA approximates a stochastic programming problem such as model (17)–(27), by considering a sample realizations of ξ and ζ of size K to represent the entire set of possible outcomes. Each realization (ξ^k, ζ^k) is taken with a probability $p_k = \frac{1}{K}$.

Let $\xi^1, \xi^2, \ldots, \xi^K$ be K i.i.d. samples of the random vector ξ and, likewise, $\zeta^1, \zeta^2, \ldots, \zeta^K$ be K i.i.d. samples of random vector ζ. The following decision aid variables are introduced to serve as indicators for exceeding bed capacity:

$$z_{dk} = \begin{cases} 1 & \text{if bed usage in sample } k \text{ exceeds the available capacity on day } d, \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

The SAA of model (17)–(27) is then defined by:

$$\text{Min } W_{\text{OT}} \cdot \sum_{k=1}^{K} \frac{1}{K} \sum_{d' \leq d \leq D} \sum_{o \in O} \alpha \cdot ot_{od}^+(\xi^k) + ot_{od}^-(\xi^k) + W_{\text{WAIT}} \cdot \sum_{p \in P_{d'}} \sum_{rd_p \leq d \leq D} (d - rd_p) \cdot x_{pd} \quad (29)$$

Subject to:

$$\sum_{rd_p \leq d \leq D} x_{pd} = 1 \qquad \forall p \in P_{d'} \quad (30)$$

$$\sum_{o \in O} y_{p\delta o} = x_{pd} \qquad \forall p \in P_{d'}, rd_p \leq d \leq D, \delta = d + preop_p \quad (31)$$

$$\sum_{p \in P_{d'}} \xi_p^k \cdot y_{pdo} + \sum_{\substack{p \in A_{d'} \\ d = a_p + preop_p}} \xi_p^k \cdot y_{pdo} - CAP_{od} \leq ot_{od}^+(\xi^k) \qquad \forall o \in O, d' \leq d \leq D, \; k = 1, \ldots, K \quad (32)$$

$$CAP_{od} - \sum_{p \in P_{d'}} \xi_p^k \cdot y_{pdo} - \sum_{\substack{p \in A_{d'} \\ d = a_p + preop_p}} \xi_p^k \cdot y_{pdo} \leq ot_{od}^-(\xi^k) \qquad \forall o \in O, d' \leq d \leq D, \; k = 1, \ldots, K \quad (33)$$

$$\left| p \in A_{d'} : a_p \leq d < a_p + \zeta_p^k \right| + \sum_{p \in P_{d'}} \sum_{\substack{\delta : rd_p \leq \delta \\ d \in [\delta, \delta + \zeta_p^k - 1]}} x_{p\delta} \leq B + M \cdot z_{dk} \qquad \forall d' \leq d \leq D, k = 1, \ldots, K \quad (34)$$

$$\sum_{k=1}^{K} z_{dk} \leq \lfloor (1 - \eta) \cdot K \rfloor \qquad \forall d' \leq d \leq D \quad (35)$$

$$x_{pd} = 0 \qquad \forall p \in P_{d'}, d' \leq d \leq rd_p \quad (36)$$

$$y_{pdo} = 0 \qquad \forall p \in P_{d'}, o \in O, d' \leq d \leq rd_p + preop_p \quad (37)$$

Expressions (29)–(33) represent straightforward conversions of their counterparts in model (17)–(27) that consider a limited set of realizations ξ^k.

To model chance-constrained bed usage, additional aid variables z_{dk} are introduced in constraints (34) to indicate whether B has been exceeded in sample k on day d. This is modelled by means of a Big M formulation (with M a sufficiently large constant, for example $M := |P|$). Expression (35) subsequently constrains, for any day d of the scheduling horizon, the portion of samples in which bed usage exceeds capacity ($z_{dk} = 1$) to be less than or equal to $\lfloor (1 - \eta) \cdot K \rfloor$ (the proportion of samples that may exceed capacity at certainty level η).

Also note that for patients $p \in A_{d'}$ for which $a_p < d'$ (i.e. patients that are currently admitted to a hospital bed), ζ_p^k should be sampled from ζ_p conditioned on the fact that $\zeta_p > d' - a_p$. Therefore, a conditional sampling approach must be employed for patients already admitted.

4.2. Local search

Preliminary testing with the SAA model revealed that its computational execution times were too slow, even for small-sized instances. This can be related to two parameters, namely the planning horizon D and the sample size K. D is presented as a known parameter, however, in fact it should be chosen sufficiently large to allow scheduling of all patients over the planning horizon However, if D is large, the MILP SAA model suffers from the dimensionality of the problem. Reducing D, the model may become infeasible (due to insufficient bed capacity over the period) or may lead to high OT costs (overtime in the OT to fit all admissions within the period). So D should be set sufficiently large. On the other hand, the sample size K further increases the dimensionality of the problem, but it also negatively impacts the linear relaxation of the MILP model. Many solutions with only slight increases in the linear relaxation lower bound can be found, due to the averaging of the overtime/underutilization of all samples. Therefore, a local search-based meta-heuristic which does not suffer from the impact of these parameters was implemented to solve the model.

A Late Acceptance Hill Climbing (LAHC) procedure was developed to solve SAA model (29)–(37). Late Acceptance Hill Climbing [42] is a simple but effective, list-based threshold accepting meta-heuristic. General pseudocode is presented in Algorithm 1. The algorithm requires only one parameter (list length L), an initial solution s_0, an objective function $f(s)$, a set of local search operators/neighbourhoods $\mathcal{N}$ and a stopping criterion. The main principle behind the algorithm is that a candidate solution s' (obtained by applying a local search operator to the current solution s) is compared to the solution that was 'current' L iterations ago. The gap between $f(s')$ and $laList[i\,\textbf{mod}\,L]$ leaves room for diversification and escaping from local optima.

Algorithm 1 Pseudocode for the LAHC meta-heuristic

```
Require: L, s_0, f : s ↦ ℝ, 𝒩, termination criterion
  s* ← s_0, s ← s_0        ▷ s*, s maintain best found/current solution
  laList ← (f(s*), f(s*), ..., f(s*))   [L entries]
  i ← 0
  while termination criterion not met do
      N ← SelectNeighbourhood(𝒩)
      s' ← N(s)                ▷ Sample a neighbouring solution of s
      if f(s') ≤ f(s) or f(s') ≤ laList[i mod L] then
          s ← s'
          if f(s') < f(s*) then
              s* ← s'
          end if
      end if
      laList[i mod L] ← f(s)
      i ← i + 1
  end while
  return s*
```

4.2.1. Solution representation and objective function

The local search approach employs a straightforward solution representation based on decision vectors $ad = (ad_1, ad_2, \ldots, ad_N)$ and $ot = (ot_1, ot_2, \ldots, ot_N)$ that maintain the admission date and the operating room assignment for each patient $p \in P_{d'}$. In addition, the solution representation maintains a fixed set of samples of size K for the LOS vector $\boldsymbol{\zeta}^k$ and surgery duration vector $\boldsymbol{\xi}^k$.

For any given assignment of ad and ot and each sample $k = 1, \ldots, K$, the bed usage, operating theatre cost and total waiting time may be easily evaluated. However, the bed usage may exceed the available bed capacity at a level higher than $(1-\eta)\%$.

To enforce the chance-constrained bed usage constraint, the ηth percentile of bed usage is compared against the available bed capacity B for each day d (see Fig. 1 for an example with $\eta = 95\%$). The bed shortages are then additionally penalized in the objective function. The objective function for the local search algorithm thus becomes:

$$\begin{aligned}\text{Min } f(s) :=& W_{BED} \sum_{d' \le d \le D} \min\left(\sum_{k=1}^{K} z_{dk} - \lfloor (1-\eta)\cdot K \rfloor, 0\right) \\ &+ W_{\text{OT}} \cdot \sum_{k=1}^{K} \frac{1}{K} \sum_{d' \le d \le D} \sum_{o \in O} \alpha \cdot ot^{+}_{od}(\boldsymbol{\xi}^k) + ot^{-}_{od}(\boldsymbol{\xi}^k) \\ &+ W_{\text{WAIT}} \cdot \sum_{p \in P_{d'}} \sum_{rd_p \le d \le D} (d - rd_p) \cdot x_{pd}\end{aligned} \tag{38}$$

and dropping Expression (35). W_{BED} should be given a significantly higher weight than W_{OT} and W_{WAIT} since it denotes a hard constraint violation. If W_{BED} is not at the highest level, the chance-constrained nature of the model would be nullified.

4.2.2. Initial solution generation

Algorithm 1 requires an initial solution which determines the starting point of the search. To enable diversification at the start of the LAHC algorithm, a randomly generated initial solution is provided. By doing so, it is easy to find improvements upon this solution. Combined with a suitably large L value, at the start of the search on average the gap between $f(s')$ and $laList[i\,\textbf{mod}\,L]$ will be large, thus allowing the LAHC algorithm to easily explore many solutions. By contrast, when a strong local optimum would be provided as initial solution, $laList$ would be filled with this local optimum's objective function value, making it difficult to escape from.

Algorithm 2 provides pseudocode on how this initial solution s_0 is randomly generated. The algorithm starts by setting up an initial solution s_0 with all admitted patients' corresponding admission dates and operating room assignments. Then, for each patient in $P_{d'}$, a random admission date from the remaining planning horizon is sampled, along with a random operating room. It checks if adding this patient on this admission date and in this operating room will not violate the capacity limits, considering only the expected surgery durations and length of stays. If the capacity limits are violated, it will iterate over subsequent admission dates and operating rooms, until a suitable date and room have been found. Note that this assumes D is sufficiently long. In practice, we extend D if it is not the case.

4.2.3. Local search operators

The local search employs three operators to perturb a solution s to a solution s'.

- Change Admission (**CA**): this local search operator samples a random patient $p \in P_{d'}$ and assigns a random new admission date $ad_p \leftarrow ad'_p$. ad'_p is sampled from an exponential distribution (with rate parameter λ) that is constructed such that a certain percentage, denoted by γ_{CA}, of all samples fall within the interval $[rd_p, lastSurgery)$, with $lastSurgery = \max_{p \in P_{d'}} \left\{ad_p + preop_p\right\}$ (see Fig. 2 for an example with $\gamma_{CA} = 95\%$). If $ad'_p > D$, meaning it falls out of the scope of the current scheduling horizon, then we extend D to

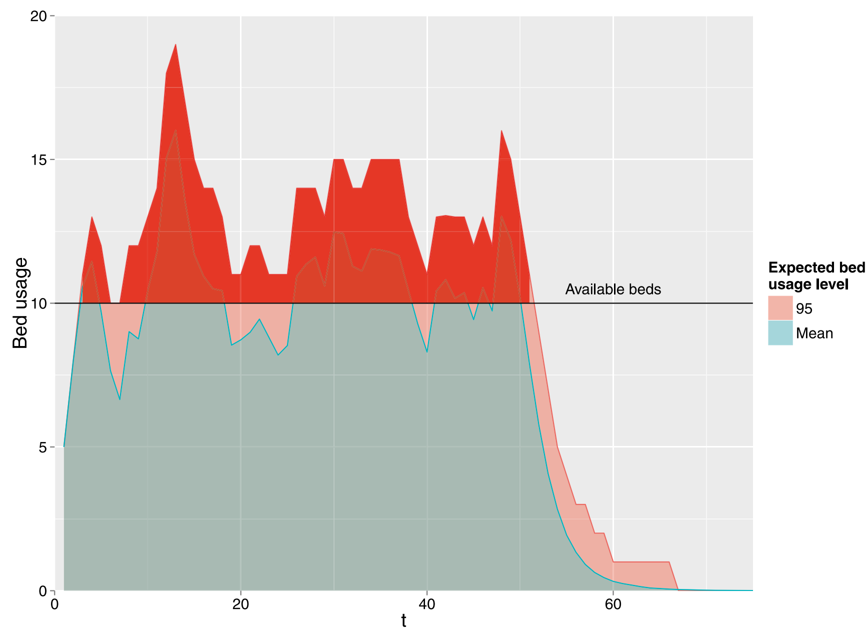


Fig. 1. Example showing bed usage for 100 patients arriving at a rate of 2 per time unit (Poisson distributed arrivals), with log-normally distributed length of stay with mean = 5 and stdev. = 3. The local search objective function minimizes the ηth percentile bed shortage (dark red colour) with a high cost. In this example $\eta = 95\%$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Algorithm 2 Initial solution generation

Require: d',$A_{d'}$,$P_{d'}$
  Initialize s_0 with admitted patients from $A_{d'}$
  for $p \in P_{d'}$ **do**
    $ad_p \leftarrow$ UniformSample($\max(d'+1, rd_p), D$)
    $ot_p \leftarrow 1$
    $ok \leftarrow$ true
    repeat ▷ Increase ad_p, ot_p until capacity limits are respected
      $ok \leftarrow$ true
      if bed occupancy on any day $d \in [ad_p, ad_p + E[\zeta_p] - 1] > B - 1$ **then**
        $ok \leftarrow$ false
        $ad_p \leftarrow ad_p + 1$ ▷ Try next day
      end if
      if surgery occupancy on day $d := ad_p + preop_p$ in OR $o := ot_p > Cap_{od} - E[\xi_p]$ **then**
        $ok \leftarrow$ false
        $ot_p \leftarrow ot_p + 1$ ▷ Try next OR
        if $ot_p > |O|$ **then**
          $ad_p \leftarrow ad_p + 1$ ▷ Try next day, reset OR to 1
          $ot_p \leftarrow 1$
        end if
      end if
    until ok
  end for
  return $s_0 = (ad\text{-vector}, ot\text{-vector})$

accommodate for this, i.e. $D := ad'_p + \zeta_p^{\max}$, thus enabling the sampling method to sample admission dates beyond the current last-planned surgery date, though at an exponentially decreasing probability. This enables the local search to postpone admissions far enough, thereby respecting the bed usage constraint.
The λ parameter is determined by:

$$1 - e^{\lambda \cdot (lastSurgery - rd_p)} = \gamma_{CA}$$
$$1 - \gamma_{CA} = e^{\lambda \cdot (lastSurgery - rd_p)}$$

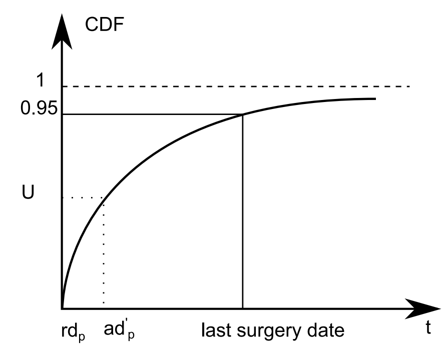
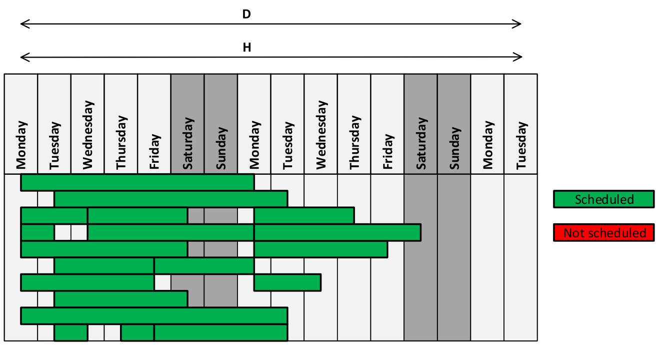


Fig. 2. Sampling ad'_p according to an exponential distribution, with a 95% probability of sampling before the last surgery date (i.e. $\gamma_{CA} = 95\%$). U denotes a uniform random number between 0 and 1.

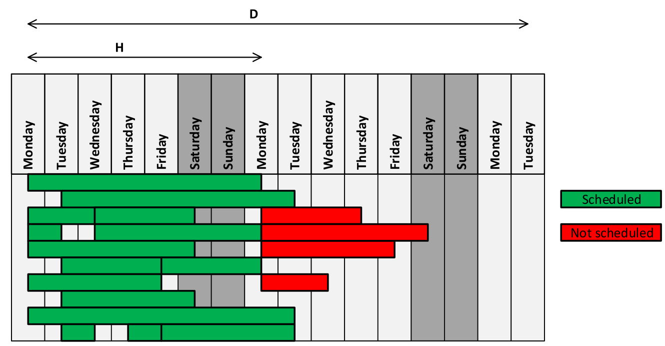
$$\ln(1 - \gamma_{CA}) = \lambda \cdot (lastSurgery - rd_p)$$
$$\lambda = \frac{\ln(1 - \gamma_{CA})}{(lastSurgery - rd_p)} \tag{39}$$

Additionally, ad'_p must not be a weekend-day, as in practice, generally, no elective admissions are planned for weekends. If that is the case, the sample is rejected and a new one is taken.

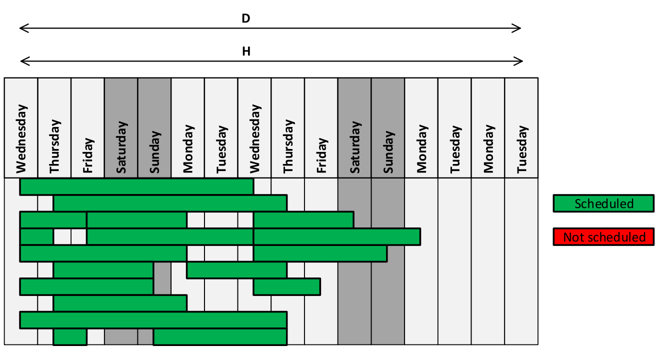
- Change OR (**CO**): this local search operator samples a random patient $p \in P_{d'}$ and assigns a random new operating room $o' \in O$ (if $|O| > 1$). This operator is not used if $|O| = 1$.
- Swap admissions (**SA**): this local search operator randomly selects two patients $p_1, p_2 \in P_{d'}$ and swaps their admission dates (if feasible with respect to rd_{p_1}, rd_{p_2}).
- Move OT Block (**MOB**): this local search operator samples a non-empty OR and day (with planned surgeries) and moves all related admissions to an earlier, empty OR and day in the scheduling horizon. This local search operator was added to compact schedules, thus minimizing waiting time and speeding up algorithm convergence.



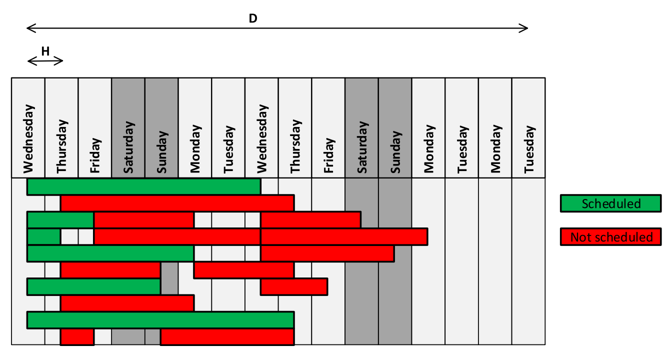
(a) Week scheduling fixing all admissions.



(b) Week scheduling fixing only current week.



(c) Daily scheduling fixing all admissions.



(d) Daily scheduling fixing only current day.

Fig. 3. Illustration of four different admission strategies, visualizing patient admissions with expected LOS. D denotes an upper bound on the scheduling horizon required to schedule all patients. H denotes the interval in which admission dates are effectively fixed. Green coloured patient admissions have an admission date $ad_p \in [d', H)$ and are fixed after the optimization algorithm has finished. Red coloured patient admissions having $ad_p \notin [d', H)$ are not fixed.. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Admission strategies

To illustrate the flexibility of the SAA model (29)–(37) and local search approach, four different admission scheduling strategies have been developed while varying two different parameters: the *frequency of decision-making* and the *scheduling window*.

- *Frequency of decision-making*: a factor that determines how long patients accumulate on a waiting list until a new admission schedule is constructed for the upcoming scheduling period. This work considers two frequency levels: daily and weekly.
  - In the daily setting, every day the admission schedule is recomputed. For example, imagine that on a Tuesday three requests for admission were added to the waiting list. On Tuesday eve, a new admission schedule for Wednesday (and onwards) would be constructed that considers these three admission requests and any other request still in the waiting list.
  - In the weekly setting, the admission schedule is constructed only once a week (in practice, usually a Friday) considering any admission requests made during the past week and all other admission requests still remaining on the waiting list.
- *Scheduling window H*: this factor determines how far into the future admission dates are considered fixed (they cannot be altered in future recomputations). Note that SAA formulation (29)–(37) plans all patients in the waiting list over the planning horizon D (which is an upper bound to plan all patients). H by contrast determines that portion of planned admissions that is effectively fixed. If only a portion of admission requests are fixed, more flexibility is given to upcoming scheduling periods to achieve higher efficiency. However, this also results in some admission requests remaining in the waiting list for several iterations, ultimately increasing their waiting time. This work considers two possible configurations:
  - If $H = D$, then after computing an admission schedule, all scheduled admissions in it are effectively fixed (a_p is set for all patients) to their scheduled admission date and removed from the waiting list (i.e. those patients p transition from P'_d to A'_d).
  - If $H = 1$, then only admissions planned on the next day of the upcoming scheduling period are fixed. This leaves all other planned admissions beyond $H = 1$ on the waiting list. Obviously, for week scheduling (see above), this makes no sense. Only one day would be planned, leaving all others open. Therefore, for daily scheduling $H = 1$ is used, whereas for weekly scheduling $H = 7$ is employed. In this case, daily admission scheduling with $H = 1$ only fixes admission dates for the upcoming day, whereas weekly admission scheduling with $H = 7$ only fixes admission dates for the upcoming week.

Fig. 3 exemplifies these four different strategies and Table 1 provides a summary of their intention. In the weekly scheduling strategies (Figs. 3a and 3b), the scheduling period's first day is always a Monday (assuming elective admissions are planned only

Table 1
Summary and explanation of the different admission strategies tested.

	More planning flexibility →			
	Strategy 1	Strategy 2	Strategy 3	Strategy 4
Frequency	Daily	Weekly	Weekly	Daily
Scheduling horizon	All queued patients	All queued patients	Next week	Next day
Explanation	Patients get AD at consultation day	Patients get AD at Friday of consultation week	Patients get AD Friday before admission week	Patients get AD the day before admission day
	← More patient friendly			

Table 2
Characteristics of log-normally fitted UZL surgical data of 2013.

Discipline	rel. freq.	Surg. Dur. (min.)		LOS (days)		# unique procedure
	(%)	Mean	Stdev.	Mean	Stdev.	types[a]
Abdominal surg. (ABD)	19.8	137.04	90.24	7.11	10.04	105
Cardiac surg. (CAH)	8.3	292.70	128.54	11.97	14.83	39
Gynaecology (GYN)	3.6	189.42	109.97	3.10	1.84	27
Neurosurg. (NCH)	7.6	223.47	136.66	10.13	24.55	49
Otorhinolaryngology (NKO)	5.4	173.98	106.04	4.99	10.64	35
Oncology (ONC)	5.4	106.13	86.93	7.64	15.71	30
Thoracic surg. (THO)	7.6	238.48	165.93	9.46	11.74	60
Abdominal transplant surg. (TRA)	2.2	264.31	181.74	12.97	12.18	9
Traumatology (TRH)	13.4	141.30	79.48	7.02	12.34	84
Urology (URO)	10.4	109.88	83.48	4.58	7.80	66
Vascular surg. (VAT)	8.2	125.66	102.54	4.93	10.43	39
Plastic, reconstructive and cosmetic surg. (RHK)	4.5	195.10	171.94	7.67	14.36	27
Oral and maxillofacial surg. (MKA)	2.0	236.95	197.21	4.24	13.23	13
Orthopaedic surg. (ORT)	1.8	218.35	179.00	11.62	17.02	11

[a]Unique proc. types with >10 surgeries performed in 2013.

on weekdays, which is common practice). If $H = D$ (Fig. 3a), then all planned admissions are effectively fixed. If $H = 7$ (Fig. 3b), then only admissions planned in the upcoming week are fixed. In the daily scheduling strategies (Figs. 3c and 3d), the scheduling period's first day may be any weekday. Again, if $H = D$ (Fig. 3c), then all planned admissions are effectively fixed. If $H = 1$ (Fig. 3d), then only admissions for the upcoming day are fixed. Algorithm 3 presents pseudocode describing the strategies.

This final admission strategy (daily, $H = 1$) is clearly the least patient-friendly, as patients can only be notified one day before admission. The daily admission strategy with $H = D$ is most patient-friendly, as patients are given their admission date on the day of their request (although it must be noted this may potentially be very short notice if some capacity is still available in the upcoming day). Weekly admission strategies are more moderate in this sense than their daily counterparts, as either a patient is assigned an admission date at the end of the week in which the admission request was made (if $H = D$), or the admission date is given the week prior to their admission.

Algorithm 3 Pseudocode: application of the different admission strategies in an dynamic setting.

Require: P, $freq \in \{daily, weekly\}$, $H \in \{D, 1/7\}$
$d' \leftarrow 0$
$A_{d'} \leftarrow \emptyset$
$P_{d'} \leftarrow \emptyset$
while $P \neq \emptyset$ **do**
$P_{d'} \leftarrow P_{d'-1} \cup \{p \in P : r_p = d'\}$
$P \leftarrow P \backslash \{p \in P : r_p = d'\}$
if $freq = daily$ and (d' **mod** $7 \neq 5 \wedge d'$ **mod** $7 \neq 6$) **then** ▷ Solve every workday (not weekends)
Solve SAA-model (29)–(37) using LAHC heuristic starting from next day
(Monday if d' **mod** $7 = 4$).
else if $freq = weekly$ and d' **mod** $7 = 4$ **then** ▷ Solve every Friday
Solve SAA-model (29)–(37) using LAHC heuristic starting from Monday.
end if
$A_{d'+1} \leftarrow A_{d'} \cup \{p \in P'_d | a_p < H\}$
$P_{d'+1} \leftarrow P_{d'} \backslash \{p \in P'_d | a_p < H\}$
$d' \leftarrow d' + 1$
end while

6. Computational study

In this section, we validate our problem formulation, the solution approach and the exemplary admission strategies by means of a computational study. Our main aim with the latter is to illustrate how the different admission strategies cope with an increasingly restrictive bed capacity, indicated further by a certain bed blocking probability. Section 6.1 describes the setup of our study, by detailing the dataset used (Section 6.1.1), the specifics of the admission strategies that were tested (Section 6.1.2), and how the results will be evaluated (Section 6.1.3). Section 6.2 then goes on to list and discuss the results of the study.

6.1. Experimental setup

6.1.1. Experimental data

The university hospital of Leuven, UZ Leuven (UZL), has provided a dataset of hospital admissions and surgery plans. The hospital comprises 1995 beds spread over four campuses and represents one of the largest hospitals in Belgium. The dataset

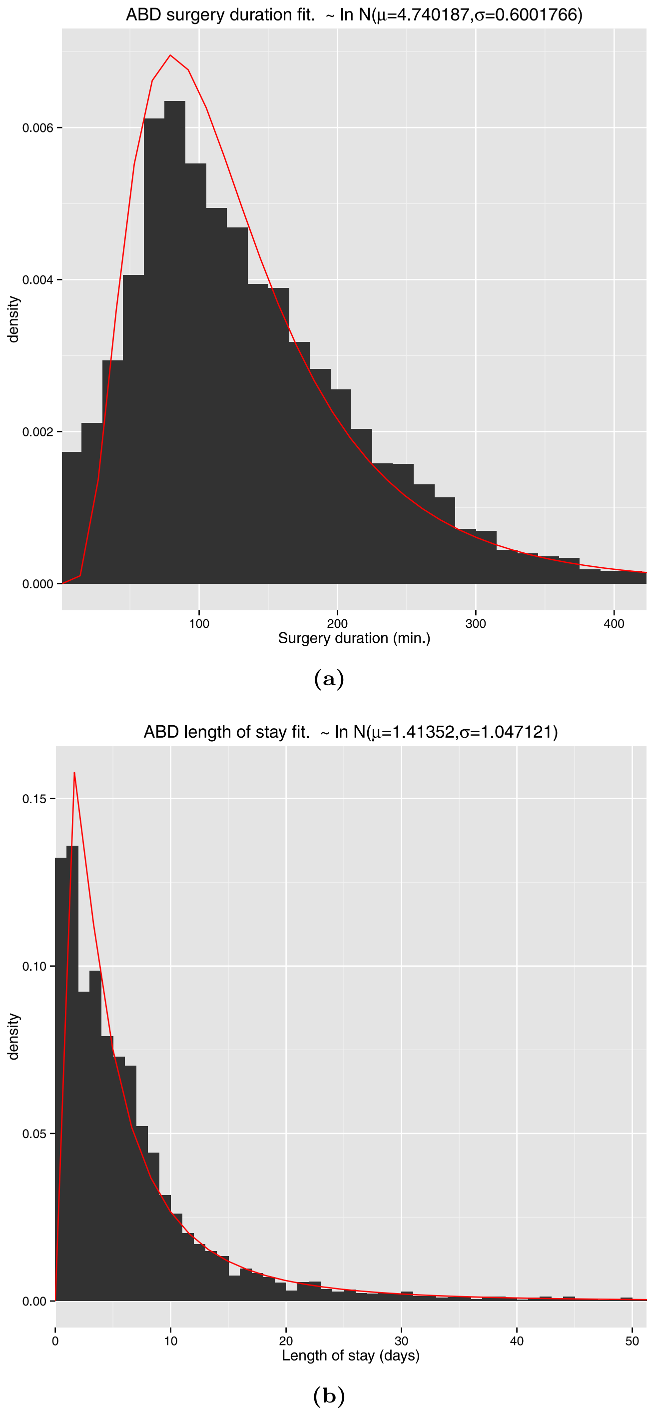


Fig. 4. Log-normal fitting of surgery duration (a) and length of stay (b) over all abdominal surgery (ABD) procedures.

spans the year 2013, during which 62 871 patient were admitted. From this set, 33 663 admissions associated with a surgical pathway were considered. The following relevant data on surgical admissions was extracted: disciplines, procedures per discipline, procedure durations and LOS per procedure. For each surgical procedure within each discipline, a log-normal distribution was fitted to both the procedure duration and LOS. The log-normal distribution has been proven to match well with both procedure duration [43] as LOS [44,45] (although for the latter, alternative positive skewed distributions have also been proposed: Exponential [46], Weibull [44], Phase-type [47]). Figs. 4 and 5 show for example, log-normal fittings of both the surgery duration and LOS for abdominal surgery (ABD) and thoracic surgery (THO). Some general information concerning the characteristics of this data and the corresponding fitted distributions is presented in Table 2.

A dataset of test instances was generated using the following method. Assume a given surgical discipline *dis*, a fixed number of operating rooms $\#OT$ and a fixed number of patients $\#P$. Let $j \in J_{dis}$ denote the procedures of surgical discipline *dis* and $f_{dis}(j)$ the relative frequency/occurrence of procedure j. Let ξ_j, ζ_j denote duration and LOS distributions of procedure j, respectively. The dataset was subsequently generated with the following properties:

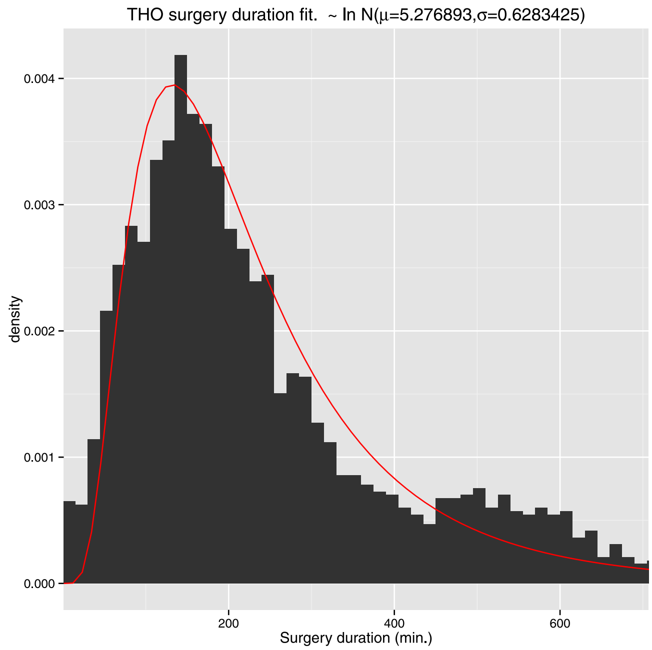
1. $\#P$ patient admission requests are generated according to a Poisson arrival process. The mean arrival rate is determined by $\lambda_{dis} := \frac{\#OT * CAP_{OT}}{E[\xi_{dis}]}$, with $E[\xi_{dis}]$ denoting the expected value of the surgical durations for discipline *dis* and $CAP_{OT} = \max_{o \in O, 1 \leq d \leq D}\{CAP_{od}\}$.
   $CAP_{od} := 480$ min for all $o \in O$, $1 \leq d \leq D$, except for weekend days, where $CAP_{od} := 0$. Thus, operating capacity is assumed to be constant during weekdays and operating rooms are closed for elective surgeries during the weekend. Furthermore, no patient admission requests are generated during weekends, and $CAP^{\max}_{od} := 720$ min.
2. For each patient admission request, a procedure j is randomly selected from J_{dis} according to their relative frequencies $f_{dis}(j)$. The corresponding surgical duration and LOS distributions are associated with the patient request ($\xi_p := \xi_j$ and $\zeta_p := \zeta_j$).
3. Next to defining the arrival process and the dimensioning of the OT, the dimensioning of the downstream nursing ward must also be specified. Since our main aim is to be able to shift the admission scheduling bottleneck away from the OT capacity towards nursing wards' bed capacity, we employ a configurable dimensioning scheme. de Bruin et al. [48] show that the Erlang loss queueing model with general service time distribution (also denoted M/G/c/c model in Kendall's notation [49]) is a good fit for dimensioning hospital wards to meet a certain blocking (cancellation) probability/risk. The model therefore proves useful to determine bed capacity for the test instances, under different blocking probabilities.
   Let μ_{dis} denote the mean length of stay for surgical discipline *dis*. The Erlang loss formula

$$P_b = \frac{(\lambda_{dis}\mu_{dis})^b / b!}{\sum_{k=0}^{b} (\lambda_{dis}\mu_{dis})^k / k!} \tag{40}$$

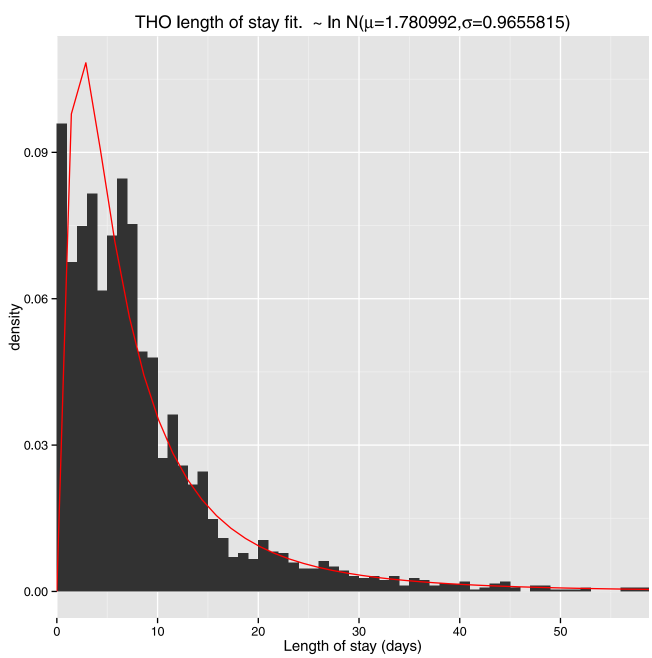
   determines the probability P_b of blocking (i.e. a bed shortage) in a hospital ward with b beds.
   Given a fixed blocking probability P_b^*, the required number of beds to meet that probability is found by enumeration on b until $P_b \leq P_b^*$.

Instances were generated in accordance with the following considerations:

- *Instance size*: determined by $\#P$ and $\#OT$. Clearly $\#P$ influences instance size as the overall scheduling horizon will increase, while $\#OT$ increases scheduling complexity as more operating rooms may be used.
- *Surgical discipline*: different characteristics with respect to mean and variation of procedure durations and LOS.



(a)



(b)

Fig. 5. Log-normal fitting of surgery duration (a) and length of stay (b) over all thoracic surgery (THO) procedures.

- *Blocking factor*: the bed shortage probability is a crucial factor determining the 'bottleneck' associated with admission scheduling. Low blocking probabilities shift the bottleneck towards the operating theatre capacity, whereas those higher shift the bottleneck towards ward capacity.

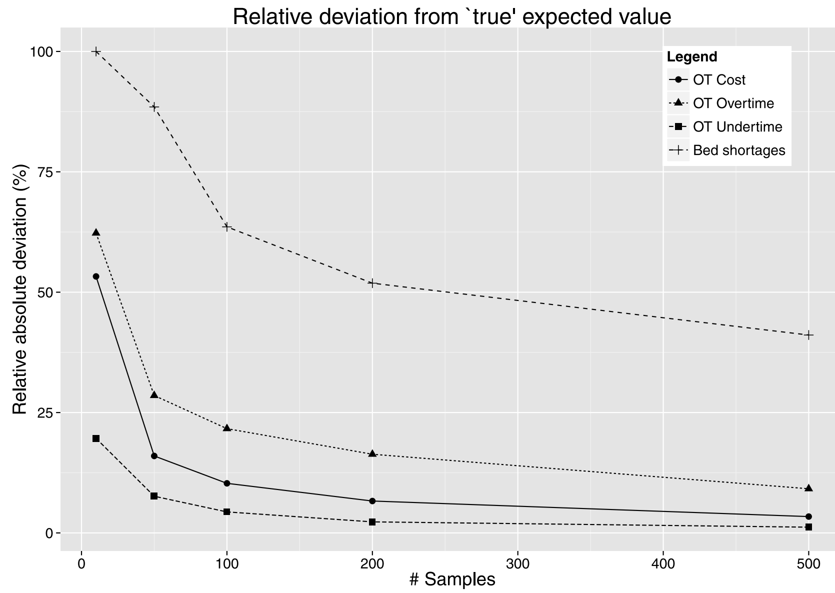
Table 3 provides an overview of the considered values for these factors. A full factorial set of instances was generated, with 5 different instances (generated using a different seed) per combination resulting in a total of 585 instances in the dataset. Note that cardiac surgery (CAH) was omitted given its high mean duration (292.7 min) which results in only one case being per OR per day, leaving no room for optimization.

Table 3
Instance generation values for the different factors under consideration. Refer to Table 2 for discipline abbreviations.

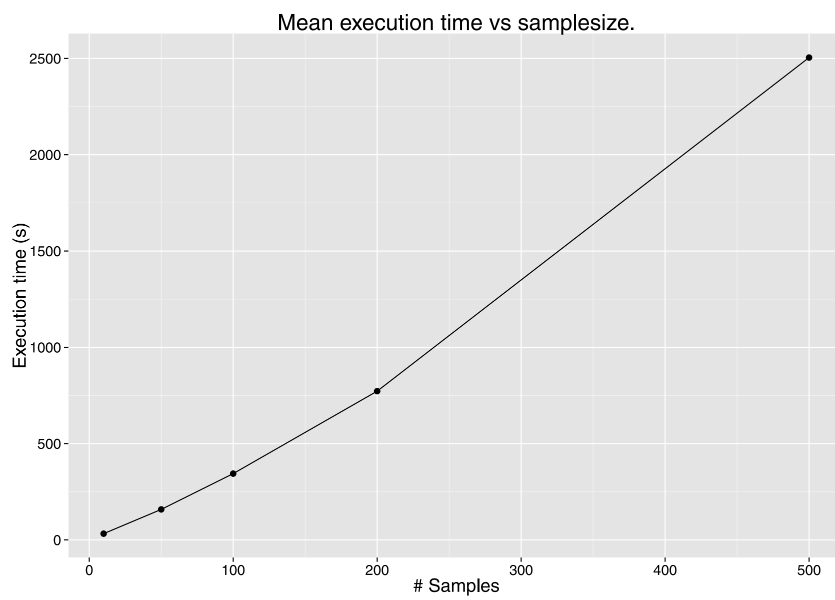
Factor	Considered values
Instance size ($\#OT$, $\#p$)	(1, 100); (2, 200); (4, 400)
Surgical disciplines	ABD,GYN,NCH,NKO,ONC,THO,ORT, TRA,TRH,URO,VAT,RHK,MKA
Blocking factor P_b^*	1%, 5%, 10%

6.1.2. Admission strategies

The four different admission strategies discussed in Section 5 were applied to all instances using the pseudocode presented in



(a)



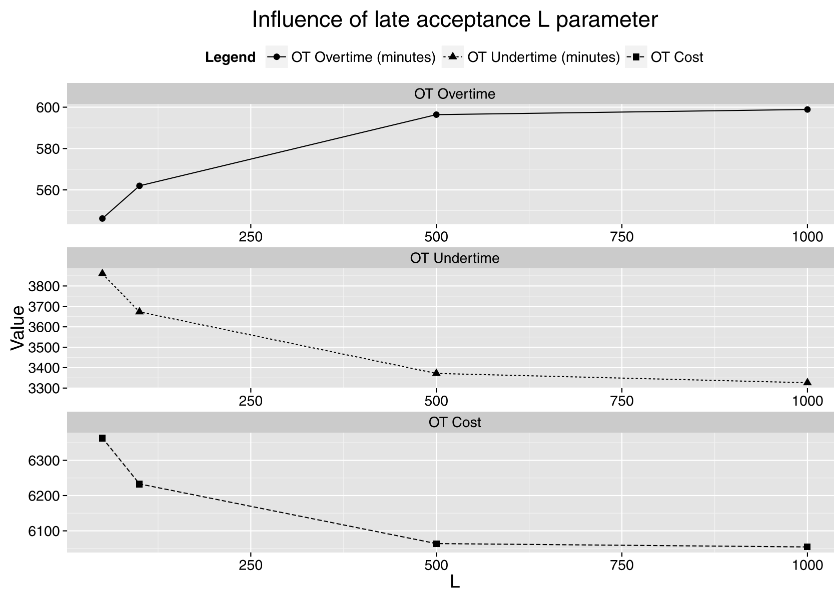
(b)

Fig. 6. (a) Accuracy of the average OT cost, OT overtime, OT undertime and bed shortages as a function of the number of samples, relative to their 'true' expected value as determined by 10 000 samples. The values represent average results obtained over a subset of the test instances (all ABD instances, 45 in total). (b) Execution time as a function of the number of samples.

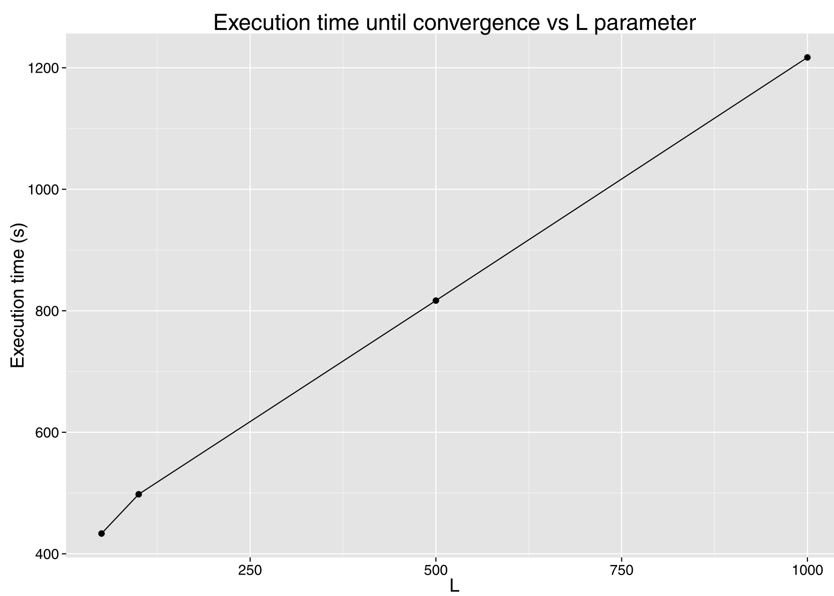
Algorithm 3. The following parameters were determined during preliminary testing to configure the local search algorithm and the SAA model:

- sample size K: a sample value of 100 was determined to give a good balance between statistical accuracy of the SAA model and the local search algorithm speed (iterations per second). Fig. 6 presents the relative deviation of the OT cost, OT overtime, OT undertime and bed shortages from their true expected value, as a function of the sample size K. Higher values (above 100) slightly increase statistical accuracy, but slow down the local search algorithm (as much more constraints/averages must be evaluated). For this experimental setup, the aim was an average execution time of at most 10 min.

  In addition, a sample size of '1' was also used, using the distributions expected value for both surgical duration and LOS. The benefit of the stochastic model over a deterministic model (with expected values) can thus be determined via comparison to the $K = 1$ approach.
- LA list length L: a list length of 500 was determined to give a good balance between early convergence and local search duration. Fig. 7 shows the influence of the L parameter on the OT cost after optimization. The L parameter clearly influences post-optimization solution quality: higher values of L enable more diversification and ultimately result in better solutions.
- Timeout criterion: the local search algorithm assumes convergence after 25000 non-improving solutions and consequently terminates.



(a)



(b)

Fig. 7. (a) Influence of the L parameter on the expected OT overtime, OT undertime and OT cost after optimization. The values represent average results obtained over a subset of the test instances (all ABD instances, 45 in total). (b) Execution time until convergence as function of L.

- This experimental setup assumes hospital management wants to meet bed availability at a certainty level $\eta = 95\%$.

In addition to the four admission strategies, tested with both $K = 1$ and $K = 100$, four 'real-time' baseline admission strategies that are frequently applied in practice, were also tested for comparison.

- First Fit OT (FFOT): when a new request for admission arrives, it is assigned the first OR slot with sufficient capacity to fit the mean surgical duration.
- Best Fit OT (BFOT): when a new request for admission arrives, it is assigned the best OR slot that has sufficient capacity and minimal slack to fit the mean surgical duration.
- FFOT-Bed and BFOT-Bed: the counterparts of the previous two strategies that additionally check whether or not there is sufficient bed capacity considering admission requests' mean LOS.

In particular, FF OT represents a useful means for comparison to current practice, where surgery admission dates are often assigned to the first available slot under the assumption that a bed is available.

6.1.3. Simulation and evaluation

The four admission strategies (each with $K = 1$ and $K = 100$) and the four 'real-time' admission strategies are tested in a straightforward simulation. It is beyond the scope of the present study to perform a completely accurate simulation of an admission process. In this study, the simulation steps through the days of the planning horizon until all admission request arrivals, determined in an instance generated in Section 6.1.1, have been

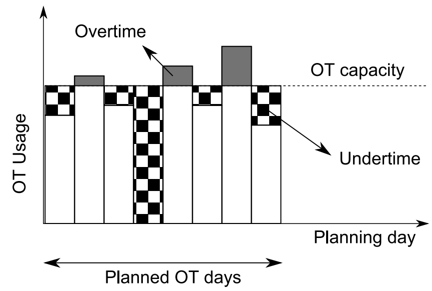


Fig. 8. Operating theatre performance measures for a single OR. For multiple ORs, total overtime and undertime is measured over all ORs.

Table 4
Objective function weight values.

Weight	Value
W_{BED}	10^5
W_{OT}	10
W_{WAIT}	10^{-6}
α	2

processed. Admission postponement/cancellation or shifting admissions to different wards/departments is not modelled in the simulation; rather, the resulting bed shortages (i.e. there is over-utilization of the beds) are allowed and measured as part of the simulation results. Three performance criteria are thus used for evaluating the performance of the different scheduling strategies: # bed shortages, OT cost and the mean admission waiting time.

Given different admissions strategies may plan surgeries over a different timespan, a relative measure is required to gain insight into performance between strategies. Therefore, when comparing OT cost, the total overtime/undertime is measured between the first surgery day (day of first planned surgery) and the last surgery day, from which the total OT cost can be calculated. This cost is then divided by the planned number of surgery days, resulting in an average daily OT cost:

$$\text{Daily OT Cost} = \frac{\alpha \cdot \text{Total overtime} + \text{Total undertime}}{\text{\#Planned surgery days}} \quad (41)$$

Fig. 8 illustrates these performance measures. The overtime to undertime ratio was set to $\alpha = 2$, under the assumption that overtime working hours are paid at a 100% additional rate.

To obtain an accurate estimate of the expected bed shortages, the OT overtime, undertime and cost, 10 000 evaluations were conducted on the final solutions obtained using samples drawn from ξ_p and ζ_p for each patient.

Concerning the relative weights between objectives, as mentioned earlier, this study applies a lexicographic weighting scheme to the OT cost and patient access time. Also mentioned earlier, in the heuristic approach, bed shortages may still occur at a certainty level η and should therefore be penalized in the objective function. Hence, the weighting scheme is such that $W_{BED} \gg W_{OT} \gg W_{WAIT}$. The specific values can be found in Table 4. Nevertheless, in principle, any weighting scheme for W_{OT} and W_{WAIT} could be applied, as long as both are still significantly smaller than W_{BED}. As mentioned earlier, if W_{BED} is not at the highest level, the chance-constrained nature of the model would be nullified. Increasing W_{WAIT} to the level of W_{OT} would increase OT costs, due to increases in overtime (due to admitting patients earlier). To achieve the correct balance between the two terms, an additional study would be needed on how patient access time and operating theatre usage costs could be translated in effective (monetary) costs. This is however out-of-scope for this study. When W_{WAIT} is at a much higher level than W_{OT}, overtime costs would be secondary to patient access time, filling the operating theatre schedule up to the maximum overtime allowance, if bed capacity is not limiting. This, again, mostly nullifies the stochastic nature of the model, since the expected OT costs are of lower importance.

6.1.4. Implementation and computational setup

The admission scheduling approach and all supporting code was implemented in Java 1.8 and employs Apache Commons Math 3.3 for stochastic sampling routines. All tests were performed on a workstation computer equipped with two eight-core Intel Xeon 2650 v2 2.6 GHz processors and 128 GB of main memory (RAM), running a Linux-based operating system. Only one processing thread is used per test and therefore this system was used to perform up to 16 tests in parallel (limiting available memory to 8 GB for each test).

6.2. Results and discussion

Tables 5–7 report on the average performance results of the admission strategies at a blocking level P_b of 1%, 5% and 10%, respectively. Reported are:

- the mean and 95th-percentile of the relative daily bed shortage and the average bed occupancy,
- the mean daily OT overtime, undertime and cost, as well as the mean operating room utilization and number of planned OT days (refer to Fig. 8),
- the mean waiting time per patient,
- the execution time per instance.

At a blocking level of 1% (Table 5), the number of bed shortages is low for all admission strategies, as expected. Differences can be seen in OT cost between real-time, daily and weekly scheduling strategies.

- Real-time strategies: FFOT and BFOT generally perform similar, having near-equal OT costs and utilization. However, since BFOT minimizes slack in each OR, it plans closer to the available capacity. This ultimately results in a higher OT overtime and lower OT undertime. Since the decrease in OT undertime does not compensate for the increase in OT overtime, this results in a (slightly) higher average daily cost. In addition, BFOT is more likely to postpone an admission slightly to minimize slack, therefore mean access time for patients is increased. A similar observation can be made for their variants considering bed availability. However, OT utilization is slightly lower due to postponing a fraction of surgeries because of bed shortages.
- Daily optimization: a distinction is possible between the stochastic approaches ($K = 100$) and average value approaches ($K = 1$). Average value approaches plan much closer to the available capacity and reach higher utilization. However, as they have no notion of variance, OT overtime cost is underestimated. This results in an increased daily OT overtime. The stochastic approaches, in contrast, do have information concerning variance and in the case where there is also sufficient flexibility ($H = 1$), a significant decrease in daily OT cost is identified. Finally, the least flexible approaches with $H = D$ perform worst. However, the stochastic approach $K = 100$, $H = D$ remains able to match the daily OT cost of real-time strategies, at a higher utilization rate. Thus, mean patient access time can be reduced while maintaining a similar cost level.
- Weekly optimization: the same distinction is possible between stochastic approaches ($K = 100$) and the average value approaches ($K = 1$) for weekly planning approaches.

Table 5
Average performance results for the different admission strategies, for instances with blocking probability $P_b^* = 1\%$. Notation: OTO = OT overtime, OTU = OT undertime.

Strategy		Rel. bed short. ($\frac{\%}{\text{day}}$)		Bed	OT performance (daily)				OT	Mean wait.	Exec. time
		Mean	95th-p	util. (%)	OTO	OTU	Cost	util. (%)	days	(days/pat.)	(ms)
Real-time	FFOT	0.0	0.1	56.2	31.94	78.27	142.16	90.3	98.57	9.88	27.60
	FFOTBed	0.0	0.0	56.1	31.74	79.25	142.74	90.1	98.86	9.94	13.98
	BFOT	0.0	0.1	56.3	32.48	77.85	142.80	90.5	98.34	10.01	28.42
	BFOTBed	0.0	0.0	56.1	32.26	79.04	143.55	90.3	98.70	10.07	27.86
Daily	$K = 1.\ H = 1$	0.0	0.1	60.1	43.72	56.01	143.45	97.4	90.15	7.04	31 182.08
	$K = 1.\ H = D$	0.0	0.1	59.6	43.67	57.70	145.05	97.1	90.42	8.00	110 279.90
	$K = 100.\ H = 1$	0.0	0.0	57.4	33.05	72.76	138.86	91.7	95.77	8.37	2 163 438.00
	$K = 100.\ H = D$	0.0	0.0	58.0	37.42	68.41	143.25	93.5	93.64	8.95	720 638.70
Weekly	$K = 1.\ H = 7$	0.0	0.1	60.9	44.23	54.19	142.65	97.9	89.93	12.19	9 625.79
	$K = 1.\ H = D$	0.0	0.1	60.6	44.44	54.34	143.22	97.9	89.81	13.07	6 779.82
	$K = 100.\ H = 7$	0.0	0.0	58.1	33.11	71.22	137.44	92.1	95.46	13.65	758 406.90
	$K = 100.\ H = D$	0.0	0.0	58.2	34.90	68.57	138.38	93.0	94.31	14.43	465 510.00

Table 6
Average performance results for the different admission strategies, for instances with blocking probability $P_b^* = 5\%$. Notation: OTO = OT overtime, OTU = OT undertime.

Strategy		Rel. bed short. ($\frac{\%}{\text{day}}$)		Bed	OT performance (daily)				OT	Mean wait.	Exec. time
		Mean	95th-p	util. (%)	OTO	OTU	Cost	util. (%)	days	(days/pat.)	(ms)
Real-time	FFOT	0.3	0.8	65.2	32.05	78.81	142.90	90.3	99.09	9.94	25.43
	FFOTBed	0.1	0.4	63.9	30.73	86.77	148.23	88.3	101.36	10.44	14.10
	BFOT	0.3	0.9	65.3	32.60	78.14	143.35	90.5	98.79	10.09	27.67
	BFOTBed	0.1	0.4	63.9	31.24	86.53	149.00	88.5	101.18	10.59	25.10
Daily	$K = 1.\ H = 1$	0.2	0.8	68.5	43.01	58.83	144.84	96.7	91.32	7.41	35 663.21
	$K = 1.\ H = D$	0.2	0.6	67.1	40.88	70.79	152.54	93.8	124.19	9.99	219 687.30
	$K = 100.\ H = 1$	0.0	0.2	64.9	31.83	77.00	140.66	90.6	97.51	8.79	2 743 409.00
	$K = 100.\ H = D$	0.1	0.3	65.6	35.18	78.40	148.75	91.0	97.24	9.54	845 337.20
Weekly	$K = 1.\ H = 7$	0.2	0.8	69.6	44.05	55.38	143.47	97.6	90.61	12.38	11 267.56
	$K = 1.\ H = D$	0.2	0.8	69.2	42.99	60.36	146.34	96.4	91.87	13.49	7 785.38
	$K = 100.\ H = 7$	0.0	0.2	65.8	32.29	74.33	138.90	91.2	96.91	13.87	927 730.40
	$K = 100.\ H = D$	0.0	0.2	66.1	33.02	75.49	141.53	91.2	97.06	14.89	573 612.40

Table 7
Average performance results for the different admission strategies, for instances with blocking probability $P_b^* = 10\%$. Notation: OTO = OT overtime, OTU = OT undertime.

Strategy		Rel. bed short. ($\frac{\%}{\text{day}}$)		Bed	OT performance (daily)				OT	Mean wait.	Exec. time
		Mean	95th-p	util. (%)	OTO	OTU	Cost	util. (%)	days	(days/pat.)	(ms)
Real-time	FFOT	1.2	2.7	72.7	31.94	78.04	141.91	90.4	98.58	9.99	26.38
	FFOTBed	0.3	0.9	69.1	28.79	98.73	156.31	85.4	104.32	11.29	14.83
	BFOT	1.2	2.7	72.8	32.45	77.43	142.34	90.6	98.33	10.11	28.48
	BFOTBed	0.3	0.9	69.2	29.26	98.86	157.39	85.5	104.37	11.40	25.16
Daily	$K = 1.\ H = 1$	0.6	1.9	74.1	40.86	63.80	145.51	95.2	92.61	7.93	123 060.10
	$K = 1.\ H = D$	0.4	1.3	72.0	37.23	87.00	161.45	89.6	131.24	10.66	114 062.90
	$K = 100.\ H = 1$	0.1	0.4	69.2	28.54	86.71	143.78	87.9	100.60	9.65	3 529 782.00
	$K = 100.\ H = D$	0.1	0.5	69.9	30.73	96.88	158.34	86.2	102.70	10.72	939 127.70
Weekly	$K = 1.\ H = 7$	0.6	1.9	75.3	41.80	60.46	144.06	96.1	91.84	12.93	13 068.21
	$K = 1.\ H = D$	0.5	1.6	74.5	39.17	74.05	152.40	92.7	95.80	14.34	8 059.26
	$K = 100.\ H = 7$	0.1	0.4	70.3	28.67	83.46	140.81	88.6	99.72	14.62	1 034 508.00
	$K = 100.\ H = D$	0.1	0.5	70.7	28.78	90.37	147.94	87.2	101.69	15.88	642 710.30

Average value approaches underestimate OT overtime cost and plan at a higher occupancy. The stochastic approaches on the other hand correctly estimate the expected OT cost. In addition, OT cost is lower for the weekly optimization approaches than for the daily optimization approaches due to increased scheduling flexibility. However, this is at the expense of mean patient access time which is higher due to postponing decision-making. Again, the most flexible, stochastic approach ($K = 100$, $H = 7$) performs best with respect to daily OT costs.

Week scheduling strategies, as expected, incur the highest mean access time per patient due to accumulating more patients on the waiting list before decision-making occurs. Interestingly, admission strategies that fix all admission dates ($H = D$) have a slightly higher mean access time than strategies that do not fix all planned patients ($H = 1$, $H = 7$). This is explained by the fact that in general all optimization-based strategies have a tendency to plan short surgical procedures with shorter LOS at earlier times and longer surgical procedures with longer LOS at later times. This results in decreased average access time and increases throughput at the scheduling horizon's beginning. This effect is less pronounced when all admission dates are fixed after scheduling them, as longer procedures are not indefinitely postponed.

As the blocking probability increases and the bottleneck shifts towards bed capacity (Tables 6 and 7), it is observed that FFOT and BFOT, which do not consider bed usage at all, incur non-negligible bed shortages. These bed shortages denote the average daily relative bed shortage. Bed occupancy generally increases during the weekdays due to new admissions and falls during the weekend. Therefore, peak bed shortages during the week may be considerable.

Non-stochastic approaches considering the expected LOS (FFOTBed, BFOTBed and admission strategies with $K = 1$) also underestimate expected bed usage. This effect is worse for the average value daily and weekly admission strategies ($K = 1$) that plan closer to available bed and OT capacity as they do not consider variance. This results in a higher utilization and admission rate, ultimately resulting in higher bed load and thus a higher probability for bed shortages.

The stochastic approaches ($K = 100$) have a better estimate of the expected bed usage and its variation and are able to reduce bed shortage risks. Evidently, they are unable to completely avoid bed shortages at the 95th percentile, as the sampling approximation is limited to 100 samples and thus may still underestimate the 95th percentile. Nevertheless, bed shortages at the 95 percentile are maintained at an acceptable level. The most flexible stochastic approaches ($H = 1$ and $H = 7$) obtain both the lowest risk on bed shortages and have low OT cost. However, this is at the expense of patient friendliness (notification of admission date either the day before, or during the previous week) and increased mean patient access time (for week scheduling).

Interestingly, a stochastic week scheduling approach where admission dates are fixed ($H = D$) and notification of admission date occurs at the end of the week in which the patient admission request was made, has reduced risk of bed shortages and a daily OT cost lower than real-time scheduling strategies also considering bed availability.

7. Conclusion

This paper presented an admission scheduling approach for the scheduling of elective surgical patients, for which a stochastic, chance-constrained optimization model is introduced. It aims at minimizing the expected operating theatre cost (as a function of both over- and underutilization) and patient access time, while simultaneously limiting the risk of bed shortages to a fixed certainty level. The latter is an intuitive parameter to decision makers, whereas other approaches often require unintuitive, problem and context specific weights to be specified, in order to incorporate them in the objective function of a model. A sample average approximation of the model is proposed and solved by a meta-heuristic approach based on Late Acceptance Hill Climbing. Finally, different admission scheduling strategies are constructed based on this stochastic approach and are compared in a computational study.

The results indicate that a stochastic approach is appropriate in an uncertain scheduling setting, given its ability to consider the uncertainty concerning patients' length of stay and surgery duration, and the resulting variance in bed usages and operating theatre occupancy/costs. When the surgical ward capacity is not a bottleneck, stochastic approaches prove able to reduce expected operating theatre costs. This also depends on the admission scheduling strategy that is employed: the most flexible approach enables a reduction of expected costs, however this will be at the expense of both patient friendliness (late notification of admission date) and/or increased access time. By contrast, when the surgical ward capacity is a bottleneck, stochastic approaches are able to maintain a low risk of exceeding bed capacity. However, this is again at the expense of patient friendliness and patient access time.

In conclusion, there is *no such thing as a free lunch* in a capacity constrained uncertain environment such as a hospital. Hedging against bed shortages comes at the expense of less patient-friendly scheduling techniques (waiting lists, late notification) or at the expense of increased OT costs. Hence, the final decision lies with the decision- and policy-makers of hospitals, weighing the pros- and cons of each policy and its impact on hospital operations and profitability. However, the presented stochastic decision model for admission scheduling is an important contribution in that regard. It allows to devise *new* admission scheduling policies that are able to deal with uncertain data and provide targeted optimization of hospital KPI's (in contrast to rule-of-thumb policies), while not overlooking the impact of the surgery admission schedule on downstream wards.

The restricted scope of this study still leaves some open ends that will be investigated in future research efforts. Notably, the consideration of unforeseen, emergency patient admissions has not been addressed in this paper. Clearly, these must also be taken into account in order to leave sufficient remaining bed and OT capacity. This should not pose a problem for the current approach: additional random variables can be modelled to represent ward and operating usage by emergency patients and sampled in the SAA approach using appropriate methods. The latter implies the implementation of either a forecasting method to simulate emergency patient arrivals, or a sampling method for estimating daily ward and operating theatre usage directly from historical data. Finally, the study has also only explored four basic admission policies using the presented model. However, the possibilities are practically endless. One may imagine incorporating quota, using different disciplines' wards, or only fixing the admission week instead of the day (leaving room to optimize each week's schedule at a later time, when more data is known). These policies may even be specific to certain disciplines that have different characteristics in patient admissions. Such a broad spectrum of possibilities was out of scope for this study.

Acknowledgements

We would like to thank the university hospital of Leuven, UZ Leuven for the many fruitful discussions on the specific subject of this study, as well as for the invaluable provision of data. This work was supported by the Belgian Science Policy Office (BELSPO) in the Interuniversity Attraction Pole COMEX (http://comex.ulb.ac.be) and a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders, Belgium (IWT-Vlaanderen). Wim Vancroonenburg is now a postdoctoral research fellow at Research Foundation Flanders (FWO-Vlaanderen).

References

[1] J.M. van Oostrum, E. Bredenhoff, E.W. Hans, Suitability and managerial implications of a master surgical scheduling approach, Ann. Oper. Res. 178 (1) (2010) 91–104.
[2] P.J.H. Hulshof, N. Kortbeek, R.J. Boucherie, E.W. Hans, P.J.M. Bakker, Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS, Health Syst. 1 (2) (2012) 129–175.
[3] B. Cardoen, E. Demeulemeester, J. Beliën, Operating room planning and scheduling: A literature review, European J. Oper. Res. 201 (3) (2010) 921–932.
[4] A. Rais, A. Viana, Operations research in healthcare: a survey, Int. Trans. Oper. Res. 18 (1) (2011) 1–31.
[5] F. Guerriero, R. Guido, Operational research in the management of the operating theatre: a survey., Health Care Manag. Sci. 14 (1) (2011) 89–114.
[6] M. Samudra, C. Van Riet, E. Demeulemeester, B. Cardoen, N. Vansteenkiste, F.E. Rademakers, Scheduling operating rooms: achievements, challenges and pitfalls, J. Sched. 19 (5) (2016) 493–525.
[7] R. Bekker, P.M. Koeleman, Scheduling admissions and reducing variability in bed demand., Health Care Manag. Sci. 14 (3) (2011) 237–249.
[8] P.J.H. Hulshof, R.J. Boucherie, E.W. Hans, J.L. Hurink, Tactical resource allocation and elective patient admission planning in care processes., Health Care Manag. Sci. 16 (2) (2013) 152–166.

[9] J. Beliën, E. Demeulemeester, Building cyclic master surgery schedules with leveled resulting bed occupancy, European J. Oper. Res. 176 (2) (2007) 1185–1204.
[10] P.T. Vanberkel, R.J. Boucherie, E.W. Hans, J.L. Hurink, W.A.M. van Lent, W.H. van Harten, An exact approach for relating recovering surgical patient workload to the master surgical schedule, J. Oper. Res. Soc. 62 (10) (2010) 1851–1860.
[11] J.T. van Essen, J.M. Bosch, E.W. Hans, M. van Houdenhoven, J.L. Hurink, Reducing the number of required beds by rearranging the OR-schedule, OR Spectrum 36 (3) (2013) 585–605.
[12] A. Fügener, E.W. Hans, R. Kolisch, N. Kortbeek, P.T. Vanberkel, Master surgery scheduling with consideration of multiple downstream units, European J. Oper. Res. 239 (1) (2014) 227–236.
[13] J.M. Vissers, I.J. Adan, N.P. Dellaert, Developing a platform for comparison of hospital admission systems: An illustration, European J. Oper. Res. 180 (3) (2007) 1290–1301.
[14] A. Mazier, X. Xie, M. Sarazin, Scheduling inpatient admission under high demand of emergency patients, in: Proceedings of IEEE Conference on Automation Science and Engineering (CASE), IEEE, 2010, pp. 792–797.
[15] R. Schmidt, S. Geisler, C. Spreckelsen, Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources, BMC Med. Inf. Decis. Mak. 13 (3) (2013).
[16] D. Gartner, R. Kolisch, Scheduling the hospital-wide flow of elective patients, European J. Oper. Res. 233 (3) (2014) 689–699.
[17] S. Ceschia, A. Schaerf, Local search and lower bounds for the patient admission scheduling problem, Comput. Oper. Res. 38 (10) (2011) 1452–1463.
[18] P. Demeester, W. Souffriau, P. De Causmaecker, G. Vanden Berghe, A hybrid tabu search algorithm for automatically assigning patients to beds, Artif. Intell. Med. 48 (1) (2010) 61–70.
[19] B. Bilgin, P. Demeester, M. Misir, W. Vancroonenburg, G. Vanden Berghe, One hyper-heuristic approach to two timetabling problems in health care, J. Heuristics 18 (3) (2012) 401–434.
[20] T.M. Range, R.M. Lusby, J. Larsen, A column generation approach for solving the patient admission scheduling problem, European J. Oper. Res. 235 (1) (2014) 252–264.
[21] W. Vancroonenburg, P. De Causmaecker, G. Vanden Berghe, A study of decision support models for online patient-to-room assignment planning, Ann. Oper. Res. 239 (1) (2016) 253–271.
[22] A.M. Turhan, B. Bilgen, Mixed integer programming based heuristics for the patient admission scheduling problem, Comput. Oper. Res. 80 (2017) 38–49.
[23] R. Guido, M.C. Groccia, D. Conforti, An efficient matheuristic for offline patient-to-bed assignment problems, European J. Oper. Res. 268 (2) (2018) 486–503.
[24] R.M. Lusby, M. Schwierz, T.M. Range, J. Larsen, An adaptive large neighborhood search procedure applied to the dynamic patient admission scheduling problem, Artif. Intell. Med. 74 (2016) 21–31.
[25] S. Ceschia, A. Schaerf, Dynamic patient admission scheduling with operating room constraints, flexible horizons, and patient delays, J. Sched. 19 (4) (2016) 377–389.
[26] J.M. Magerlein, J.B. Martin, Surgical demand scheduling: a review, Health Serv. Res. 13 (4) (1978) 418–433.
[27] I. Ozkaraham, Allocation of surgeries to operating rooms by goal programing, J. Med. Syst. 24 (6) (2000) 339–378.
[28] A. Riise, E.K. Burke, Local search for the surgery admission planning problem, J. Heuristics 17 (4) (2011) 389–414.
[29] A. Agnetis, A. Coppi, M. Corsini, G. Dellino, C. Meloni, M. Pranzo, Long term evaluation of operating theater planning policies, Oper. Res. Health Care 1 (4) (2012) 95–104.
[30] A. Jebali, A.B. Hadj Alouane, P. Ladet, Operating rooms scheduling, Int. J. Prod. Econ. 99 (1–2) (2006) 52–62.
[31] H. Fei, C. Chu, N. Meskens, Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria, Ann. Oper. Res. 166 (1) (2008) 91–108.
[32] A. Guinet, S. Chaabane, Operating theatre planning, Int. J. Prod. Econ. 85 (1) (2003) 69–81.
[33] E. Hans, G. Wullink, M. van Houdenhoven, G. Kazemier, Robust surgery loading, European J. Oper. Res. 185 (3) (2008) 1038–1050.
[34] M.E. Bruni, P. Beraldi, D. Conforti, A stochastic programming approach for operating theatre scheduling under uncertainty, IMA J. Manag. Math. 26 (1) (2015) 99–119.
[35] D. Min, Y. Yih, Scheduling elective surgery under uncertainty and downstream capacity constraints, European J. Oper. Res. 206 (3) (2010) 642–652.
[36] A. Jebali, A. Diabat, A stochastic model for operating room planning under capacity constraints, Int. J. Prod. Res. 53 (24) (2015) 7252–7270.
[37] A. Jebali, A. Diabat, A chance-constrained operating room planning with elective and emergency cases under downstream capacity constraints, Comput. Ind. Eng. 114 (2017) 329–344.
[38] S. Ceschia, A. Schaerf, Modeling and solving the dynamic patient admission scheduling problem under uncertainty, Artif. Intell. Med. 56 (3) (2012) 199–205.
[39] D. Ouelhadj, S. Petrovic, A survey of dynamic scheduling in manufacturing systems, J. Sched. 12 (4) (2008) 417.
[40] A. Charnes, W. Cooper, G. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, Manage. Sci. 4 (3) (1958) 235–263.
[41] A.J. Kleywegt, A. Shapiro, T. Homem-de Mello, The sample average approximation method for stochastic discrete optimization, SIAM J. Optim. 12 (2) (2002) 479–502.
[42] E.K. Burke, Y. Bykov, The late acceptance Hill-Climbing heuristic, European J. Oper. Res. 258 (1) (2017) 70–78.
[43] D.P. Strum, J.H. May, L.G. Vargas, Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models., Anesthesiology 92 (4) (2000) 1160–1167.
[44] A. Marazzi, F. Paccaud, C. Ruffieux, C. Beguin, Fitting the distributions of length of stay by parametric models, Med. Care 36 (6) (1998) 915–927.
[45] P.R. Harper, A.K. Shahani, Modelling for the planning and management of bed capacities in hospitals, J. Oper. Res. Soc. 53 (1) (2002) 11–18.
[46] J.D. Griffiths, V. Knight, I. Komenda, Bed management in a Critical Care Unit, IMA J. Manag. Math. 24 (2) (2012) 137–153.
[47] M. Fackrell, Modelling healthcare systems with phase-type distributions, Health Care Manag. Sci. 12 (1) (2008) 11–26.
[48] A.M. de Bruin, R. Bekker, L. van Zanten, G.M. Koole, Dimensioning hospital wards using the Erlang loss model, Ann. Oper. Res. 178 (1) (2010) 23–43.
[49] D.G. Kendall, Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain, Ann. Math. Stat. 24 (3) (1953) 338–354.