Robust Surgery Scheduling: A Model-Based Overview

Maarten Otten, Jasper Bos, Aleida Braaksma, and Richard J. Boucherie

Abstract In this chapter we give a model-based overview of robust surgery scheduling literature. A robust schedule maintains to perform well in case of disturbances affecting the schedule. We distinguish three types of disturbances that affect the surgery schedule. First, internal disturbances, such as variations in surgical time. Second, external disturbances, such as non-elective surgeries. Third, disturbances due to artificial variability, such as unavailable operating rooms. For each of these disturbances, we provide an overview of models, described in literature, which reduce the effect of the disturbance on the schedule by making it robust. Furthermore, we identify relevant open problems.

1 Introduction

Performing surgeries is one of the key tasks of a hospital. A large share of the patients is treated by a surgeon during their hospital visit. Operating rooms (ORs) are among the most expensive resources of the hospital. Furthermore, hospitals are increasingly struggling to attract sufficient qualified OR personnel. Utilizing the ORs efficiently is therefore of key importance.

In an ideal world, surgeries would be scheduled such that the ORs are perfectly utilized, but due to, among others, stochasticity, this is impossible. For various reasons surgeries can be advanced, delayed, or canceled, such that the daily practice at the ORs rarely matches the schedule made beforehand.

Deviation from the original schedule can have many causes. In this chapter we divide these into three types. First, we consider internal causes due to natural variability of the processes at the OR, such as variability of OR setup times, of OR cleaning times, or of the surgery durations. These mostly affect the schedule

M. Otten (✉) · J. Bos · A. Braaksma · R. J. Boucherie
Center for Healthcare Operations Improvement and Research, University of Twente, Enschede, The Netherlands
e-mail: j.w.m.otten@utwente.nl

M. E. Zonderland et al. (eds.), *Handbook of Healthcare Logistics*, International Series in Operations Research & Management Science 302, https://doi.org/10.1007/978-3-030-60212-3_4

with elective surgeries. Second, we consider external causes, such as no-shows, illness of patients, or non-elective surgeries. Since only elective surgeries can be scheduled beforehand, non-elective patients that arrive during the day will be scheduled within the existing schedule, possibly causing later surgeries to be delayed or even canceled. Similarly, elective surgeries that are canceled due to unavailability of the patient will cause either subsequent surgeries to be advanced or a gap in the OR schedule. Third, we consider causes due to artificial variability, i.e., inefficient and counteracting processes at the ORs, such as unavailability of clean ORs, well equipped ORs, or OR personnel. The distinction between natural and artificial variability is convenient in our model-based approach of the literature. Natural variability typically cannot be eliminated, and the approach for this type of disturbance will therefore be anticipatory, whereas for artificial variability the approach will be to optimize the processes.

The literature on OR scheduling in general is abundant. Hulshof et al. [9] and Cardoen et al. [3] both provide extensive reviews of OR scheduling literature. Hans and Vanberkel [8] present a number of approaches to OR planning and scheduling on a strategic, tactical, and operational level. However, the amount of papers where variability is explicitly taken into account is limited. Van Riet and Demeulemeester [23] is a recent and up to date review of the literature where non-elective patients are taken into account. Ferrand et al. [5] also provide a review of literature where both elective and non-elective surgeries are considered.

In this chapter we give an overview of several mathematical models described in literature to obtain a robust schedule, i.e., a (near) optimal schedule that under some deviation will still perform reasonably well. Our contribution with this chapter is threefold. First, we provide a model-based overview of the literature that takes the various sources of disturbances described above into account. For relevant combinations of one or more sources of disturbance, we state an objective and describe the models that are used in literature. Second, we not only describe the models but also state them in a concise way. Third, we give an overview of open problems.

In Sect. 2 we discuss how internal disturbances caused by natural variation can be taken into account when scheduling the elective surgeries. In Sect. 3 we discuss the objectives and related models when external disturbances are taken into account. In Sect. 4 we discuss models described in literature that take artificial variation into account to optimize the processes at the OR. In Sect. 5 we discuss several models where combinations of the disturbances are considered. Finally in Sect. 6 we present an overview of the models we considered and draw conclusions based on our findings.

2 Internal Variability

Most patients that are treated at the surgical department are scheduled beforehand. In addition to these elective patients, there are non-elective patients that arrive during the day and need treatment within a short time span. In this section we focus on scheduling elective surgeries.

The duration of similar surgical procedures can fluctuate significantly due to various reasons. Some of these differences, such as variations between surgeons or between types of patients, can be accounted for. Other effects, like deviations from the original surgical procedure, cannot. In order to obtain a robust schedule, it is therefore important to take these variations into account when scheduling the elective surgeries. In this section we discuss several models for different objectives when the variability of surgery durations is taken into account.

2.1 Overtime

Scheduling elective surgeries based on their expected duration only will generally lead to a high OR utilization. However, the probability of overtime (surgeries performed outside scheduled hours) and the number of surgeries canceled at the end of the day will increase with increasing variability of the surgery duration. Taking into account this variability is a trade-off between maximizing the OR utilization and minimizing the probability of OR overtime. This section describes several approaches and models that aim to assign surgeries to ORs such that the expected utilization is maximized while at the same time minimizing the probability of overtime.

Scheduling surgeries can be divided into two related subproblems: the advance scheduling and the allocation scheduling problem. In the advance scheduling problem, surgeries are assigned to a date and an OR block. In the allocation problem, the sequence of the surgeries that are assigned to a certain date is determined.

Hans and Vanberkel [8] use the term robust surgery loading for the variant of the advance scheduling problem they consider. Elective surgeries are scheduled in advance over a discrete planning horizon, $t = 1, \ldots, T$. On each day t, each specialty z has a number of ORs at its disposal, denoted by the set K_{zt}. An element k of the set K_{zt} is termed an OR-day. The set of surgeries V must be allocated to the OR-days. Each element $i \in V$ is a surgery of a certain specialty, and its duration has mean μ_i and variance σ_i^2. The set of surgeries of specialty z is denoted by V_z. To be determined are V_{ztj}, the set of surgeries assigned by specialty z to OR j, at day t. Together with the surgeries, there is slack capacity assigned to each OR-day to accommodate for the effect of randomness in the surgery times. The amount of slack capacity assigned to OR-day K_{zt} is:

$$\delta_{ztj} = \beta(p) \sqrt{\sum_{i \in V_{ztj}} \sigma_i^2}, \quad (1)$$

which is the standard deviation of the sum of the surgery durations scaled with $\beta(p) \geq 0$, a safety margin determined by the maximal tolerated probability of overtime p. For example, if we assume, by the central limit theorem, that the sum of the surgery durations is normally distributed, a probability $p = 0.69$ of having no overtime corresponds to $\beta(p) = 0.50$. Note that this approach assumes that the surgery durations are independent. Hans and Vanberkel [8] translate (1) into constraints on the number of surgeries that can be scheduled:

$$\sum_{i \in V_{ztj}} \mu_i + \delta_{ztj} \leq c_{tj} + O_{ztj} \quad \forall z, j \in K_{zt}, t, \tag{2}$$

where c_{tk} is the capacity needed per OR-day and O_{ztk} the maximum overtime for each specialty per OR-day. As objectives [8] use (in order of importance): minimizing the total overtime, maximizing the number of free OR-days, and maximizing the total free capacity. They use the base schedule, i.e., the schedule that the hospital scheduler makes, and propose several heuristics to improve the base schedule. They propose a list scheduling approach where surgeries are sorted in decreasing order of their expected duration. According to the order of the list, surgeries are added to the schedule. A surgery is added to the schedule when the constraints (2) are not violated; otherwise the next surgery on the list is added to the schedule. They furthermore propose local search methods to improve this base schedule. They find that in optimized schedules the portfolio effect (surgeries with similar variation characteristics are clustered on the same OR-day) plays an important role in reducing the amount of necessary slack capacity.

Denton et al. [4], Landa et al. [13], and Molina-Pariente et al. [16] model the advance scheduling problem as a type of stochastic programming problem. This stochastic programming problem, a two-stage simple recourse model, divides the decision variables in two groups, those that are to be determined here and now (assigning surgeries to OR-days) and those that can be adjusted, at a cost, later on (amount of overtime at an OR). We state and discuss the problem as formulated by [4].

Suppose there are n surgeries to be scheduled at m ORs, where m itself is part of the decision. Let $D_i(\omega)$ denote the duration of surgery i, a random variable where ω denotes a possible realization of the durations, and T the planned session length at each OR. g^f denotes the fixed costs of opening an OR and g^v the variable costs per time unit of keeping an OR open past time T. Let X_j be a binary decision variable whether OR j is opened, Y_{ij} whether surgery i is assigned to OR j and O_j the amount of overtime of OR j. Then the assignment of surgeries to ORs can be formulated as the following two-stage simple recourse problem:

$$Z^* = \min \left\{ \sum_{j=1}^{m} \left(g^f X_j + \mathbb{E}\left[g^v O_j(\omega) \right] \right) \right\}, \tag{3}$$

$$\text{s.t.} \quad Y_{ij} \leq X_j, \qquad \forall i, j, \tag{4}$$

$$\sum_{j=1}^{m} Y_{ij} = 1, \qquad \forall i, \qquad (5)$$

$$\sum_{i=1}^{n} D_i(\omega)Y_{ij} - O_j(\omega) \leq TX_j, \qquad \forall j, \omega, \qquad (6)$$

$$Y_{ij}, X_j \in \{0, 1\}, \qquad \forall i, j, \qquad (7)$$

$$O_j(\omega) \geq 0, \qquad \forall j, \omega. \qquad (8)$$

The decisions whether an OR is opened and to which OR a surgery is assigned, i.e., the binary decision variables X_j and Y_{ij}, are first-stage decisions. The variable for overtime at the OR, O_j, is a recourse variable and is determined once the realization of the surgery durations, ω, is known. This is a two-stage simple recourse model which can be easily solved [1]. In addition to a stochastic programming approach, [4] add a robust programming approach for which they use a polyhedral uncertainty set. The stochastic programming approach is very suitable, provided there is enough statistical information available for the possible outcome scenarios of the realized surgery durations $D_i(\omega)$. If there is only information about the possible outcomes of $D_i(\omega)$, the robust programming approach is more suitable.

2.2 Deviation from the Schedule

For the admission process to run smoothly, patients are asked to come to the hospital some time before their surgery is scheduled to start. In order to prevent that patients are not yet available for surgery or that their surgery is delayed for hours, it is important that the actual starting time does not deviate too much from the scheduled time. In this section we present the earliness/tardiness (E/T) model that aims to reduce the variability of the start times of surgeries.

Suppose there are n surgeries to be scheduled in m ORs. Let n_j denote the number of surgeries in OR j, and let $\Pi = [\pi_1, \pi_2, \ldots, \pi_m]$ denote the schedule for the m ORs for which $\pi_j = (\pi_{j,1}, \ldots, \pi_{j,n})$, $ij1, 2, \ldots, m$, is the sequence in which the surgeries are performed in OR j. Further, let $s^{\pi_j} = (s_1^j \ldots s_n^j)$ denote the scheduled start times of the surgeries in OR i and $D_{i,j}$ the duration of surgery i in OR j, a random variable. The realized start time of surgery i in OR j, also a random variable, is denoted by

$$S_j^i = \sum_{k=1}^{i-1} D_{\pi_{jk},j}. \qquad (9)$$

The earliness of surgery i at OR j is defined as $E^i_j(s^i_j) = (s^i_j - S^i_j)^+$ and its tardiness as $T^i_j(s^i_j) = (S^i_j - s^i_j)^+$, where $(x)^+ = \max\{x, 0\}$. The unit earliness and tardiness costs parameters of surgery i in OR j are given by ϵ^i_j and τ^i_j, respectively. The E/T costs of surgery i in OR j are denoted by

$$A^i_j(s^i_j) = \mathbb{E}\big[\epsilon^i_j E^i_j(s^i_j) + \tau^i_j T^i_j(s^i_j)\big], \tag{10}$$

for $i = 1, \ldots, n$, $j = 1, \ldots m$. The total expected E/T costs for a schedule Π with start times s^Π is denoted by $ET(\Pi, s^\Pi)$. The objective is to select a schedule Π and corresponding start times s^Π in such a way that the total expected E/T costs are minimized. If the surgeries are preassigned to the ORs, it suffices to optimize the schedule for each OR separately. The single OR E/T problem is defined as

$$\text{(SET)} \qquad \min_{\pi,s} \quad ET(\pi, s) = \min_{\pi,s} \sum_{i=1}^{n} \mathbb{E}\big[\epsilon_i E_i(s_i) + \tau_i T_i(s_i)\big]. \tag{11}$$

When assigning the surgeries to an OR is also part of the decision, the joint schedule has to be optimized. The multiple OR E/T problem is defined as

$$\text{(MET)} \quad \min_{\Pi,s^\Pi} \quad ET(\Pi, s^\Pi) = \min_{\Pi,s^\Pi} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbb{E}\big[\epsilon^i_j E^i_j(s^i_j) + \tau^i_j T^i_j(s^i_j)\big]. \tag{12}$$

The problem is termed symmetric if the cost parameters are equal for each surgery, i.e., $\epsilon_i = \epsilon$ and $\tau_i = \tau$, for $i = 1, \ldots, n$. Weiss [25] describes an extension of this model in which the earliness costs are infinite, i.e., advancing a surgery is impossible. This variant of the problem is significantly harder because starting times of subsequent surgeries are not linked together in a straightforward way anymore due to the fact that idle time may occur between surgeries. Gupta [7] shows that for this variant of the E/T problem with one OR and two surgeries, it is optimal to schedule the surgeries in increasing order of the variances of the duration, the so-called Smallest Variance First (SVF) rule. For more than two surgeries and for multiple ORs, this remains an open problem.

Whenever idle time is not allowed, the optimal scheduled start time of a surgery is fully determined by the probability distributions of the preceding surgeries. In this case, given a schedule Π, the optimal scheduled start time of surgery i in OR j with actual start time S^i_j is

$$s^*_i = G^{-1}\left(\frac{\tau_i}{\epsilon_i + \tau_i}\right), \tag{13}$$

where G is the distribution function of the realized start time, S^i_j. We now give a concise overview of the results described in literature and the open problems for the various versions of the E/T problem.

For the case when the surgeries are preassigned to the ORs and when the earliness and tardiness costs are equal for each surgery, i.e., the symmetric single OR E/T problem, [6] obtain an optimal schedule under the assumption that the mean and variance of the surgery duration S are finite. In this case it is optimal to schedule the surgeries according to the SVF rule.

Otten and Boucherie [17] consider an optimal schedule for the multiple OR version of the E/T problem under some technical assumptions on the distribution of the surgery duration, which they show to hold for at least the normal distribution. For the multiple OR E/T problem, they extend the SVF rule, i.e., they show that it is optimal to not only apply the SVF rule to the local sequence of surgeries at each OR but also to the global sequence of all surgeries at all ORs.

In conclusion, we see that the SVF rule is optimal for several variants of the E/T problem when idle time is not allowed. For the nonsymmetric and multiple OR variants, it remains an open problem if this result holds for arbitrary distributions of the surgery duration. When idle time is allowed, very few results are known, but [7] and [6] show that the SVF rule still performs reasonably well. Furthermore, [17] show that an optimal schedule for the multiple OR variant is not unique. This leaves room for additional optimization of the schedule.

3 External Variability

An important external cause of disturbance to the OR schedule is the arrival of non-elective patients during the day. In this section we discuss various relevant objectives when anticipating this type of disturbance, both from the perspective of OR utilization and from the perspective of the non-elective patients.

3.1 Overtime

Ideally, the utilization of the ORs is high. However, as non-elective patients arrive during the day, a high OR utilization will inevitability lead to a high probability of overtime or cancellation, which is undesirable. When the objective is to maximize OR utilization but avoid overtime, the trade-off will be between using OR capacity for elective patients and reserving slack OR capacity for possible non-elective patients.

3.1.1 Non-elective Surgery Policy

The main decision is how non-elective surgeries are accommodated. We distinguish three possible policies [2, 23]: (1) a dedicated policy, where there are one or more dedicated ORs for non-elective patients; (2) a flexible policy, where slack capacity is fragmented and added to each of the regular ORs, and non-electives are then fitted into the elective schedule; and (3) a hybrid policy, where policies (1) and (2) are combined.

Literature shows that it highly depends on the hospital's characteristics which policy is optimal. For instance, [26] find a decrease in overtime using the flexible policy, while [23] report an increase in overtime using the same policy. Van Veen-Berkx et al. [24] find that the flexible policy is optimal by simulation while after implementation the opposite was the case. Borgman [2] carries out a simulation study in which he tests the different policies for many different hospital characteristics. His main conclusion is that the dedicated policy works best for small hospitals with at most eight ORs and that the flexible policy is best for larger hospitals.

If the dedicated policy is adopted, the remaining question is how many ORs should be dedicated to non-elective patients. There is no research that addresses this question because this could easily be determined based on the expected capacity needed for non-elective surgeries.

In the remainder of this section, we discuss some of the problems that hospitals face anticipating arriving non-electives during the day, when the flexible policy is adopted.

3.1.2 Required Capacity

Van Houdenhoven et al. [22] develop a model to determine the amount of required slack capacity for emergency patients. They assume that the elective surgeries are planned in blocks per specialty. Suppose that the expected number of emergency surgeries of specialty z is n_z^{el} at an arbitrary OR-day, each having a mean duration μ_z^{el} and standard deviation σ_z^{el}. The expected duration of the elective surgeries in this block is therefore $n_z^{el}\mu_z^{el}$ with a standard deviation of $\sqrt{n_z^{el}(\sigma_z^{el})^2}$. The total required capacity to perform all the expected non-elective surgeries of specialty z is

$$c_z(p_z) = n_z^{el}\mu_z^{el} + \beta(p_z)\sqrt{n_z^{el}(\sigma_z^{el})^2}, \tag{14}$$

where p_z is the allowed probability of overtime and $\beta(p_z)$ is a function that gives a factor that yields a probability of overtime of p_z, similar to the model by [8], described in Sect. 2.1. The function $\beta(p_z)$ depends on the distribution of the sum of surgery durations. Van Houdenhoven et al. [22] assume that this sum is normally distributed, which implies that $\beta(p_z) = \Phi^{-1}(p_z)$, the standard-normal quantile function. For each block of surgeries of specialty z, an additional amount of capacity $c_z(p_z)$ is needed in order to have a probability of overtime of p_z. For arbitrary distributions of the surgeries, this approach still works, albeit that the function $\beta(p_z)$ will have to be approximated.

Zonderland et al. [27] develop a queuing model to determine, on a strategic level, how much slack capacity needs to be reserved for semi-urgent patients, patients that do not need surgery immediately but within a few days. On a tactical level they develop a decision support tool, based on Markov decision theory, for scheduling these patients within the coming days.

3.1.3 Scheduling Elective Surgeries Anticipating Emergencies

Lamiri et al. [11] and Min and Yih [15] use a stochastic programming approach to model the scheduling of elective surgeries anticipating non-elective surgeries, under the flexible policy. We state and discuss the model as formulated by [11]. Suppose there are n^{el} elective surgeries to be planned over a horizon of T days. To be determined is to which day each of the elective surgeries is assigned. Let c_t^{el} be the OR capacity in units of time available for elective surgeries on day $t = 1, \ldots, T$ and g_t^o the cost of overtime per time unit. During the day emergency patients arrive and are to be treated at the same day. C_t^{em} is a random variable having probability density function $f_{C_t^{em}}(x)$ and denotes the capacity needed for emergency patients on day $t = 1, \ldots, T$. The elective surgeries have a duration d_i, a release date r_i, and surgery costs g_{it} associated with them, which are assumed to be deterministic. Let Y_{it} denote the binary decision variable whether elective surgery i is assigned to day t. The goal is to assign the elective patients such that the total surgery costs and the overtime costs are minimized. This is formulated in the following optimization problem:

$$\min \quad J(Y) = \quad \min \sum_{i=1}^{n^{el}} \sum_{t=r_i}^{T+1} (g_{it} Y_{it}) + \sum_{t=1}^{T} g_t^o \mathbb{E}\left[\left(C_t^{em} + \sum_i d_i Y_{it} - c_t^{el}\right)^+\right], \tag{15}$$

$$\text{s.t.} \qquad \sum_{t=r_i}^{T+1} Y_{it} = 1, \quad \forall i, \tag{16}$$

$$Y_{it} \in \{0, 1\}, \quad \forall i, t. \tag{17}$$

Lamiri et al. [11] show that the solution can be analytically obtained by numerical integration:

$$J(Y) = \sum_{i=1}^{n^{el}} \sum_{t=r_i}^{T+1} (g_{it} Y_{it}) + \sum_{t=1}^{T} g_t^o \int_{q_t}^{\infty} (z - q_t) f_{C_t^{em}}(z) dz, \tag{18}$$

where $q_t = C_t^{el} - \sum_i d_i Y_{it}$ is the remaining regular capacity at day t. This approach is only tractable for small instances. For larger instances they propose a Monte Carlo approach:

$$J(Y) \approx J_L(Y) = \sum_{i=1}^{n^{el}} \sum_{t=r_i}^{T+1} (g_{it} Y_{it}) + \sum_{t=1}^{T} g_t^o \frac{1}{L} \sum_{l=1}^{L} \left[\left(C_t^{em,l} + \sum_i d_i Y_{it} - c_t^{el} \right)^+ \right], \quad (19)$$

where $C_t^{em,1}, \ldots, C_t^{em,L}$ are L independent samples generated for the random variable C_t^{em}.

3.2 *Waiting Time of Emergency Patients*

Emergency patients arriving at the OR need to be treated within a short time span. However, ongoing surgeries cannot be interrupted, so when the hospital does not have dedicated emergency ORs or these ORs are occupied, an emergency patient has to wait until the first of the ongoing surgeries is completed. It is therefore a reasonable idea to schedule the elective surgeries while taking into account the waiting time of emergency patients that will arrive during the day. In this section we discuss two models for the case when emergency patients are to be accommodated into the elective surgery schedule and the objective concerns the waiting time of these patients.

3.2.1 Maximum Waiting Time

Emergency patients of the most urgent category require surgery as soon as possible. In order to obtain a schedule that satisfies this requirement, the elective surgeries have to be sequenced in such a way that the maximum time until an emergency patient can be inserted into the schedule is minimized. Van Der Lans et al. [20] and Van Essen et al. [21] described this problem as a scheduling problem. In this section we state and discuss this model as described by [21].

Suppose there are n elective surgeries to be scheduled in m ORs. The surgeries are preassigned to an OR, Y_i denotes the OR surgery i is assigned to and n_j denotes the number of surgeries in OR j. Assume that at time $t = 0$ all elective patients are available for surgery and that preemption of surgeries and idle time for the ORs are not allowed. To be decided is the order in which surgeries are performed at each OR. Let $\Pi = [\pi_1, \ldots, \pi_m]$ denote the schedule, where $\pi_j = [\pi_{j,1}, \ldots, \pi_{j,n_j}]$ is the order of surgeries in OR j. The duration of surgery i is denoted by d_i and its completion time by

$$F_i^{\Pi} = \sum_{k=1}^{\pi_{Y_i,i}} d_{\pi_{Y_i,k}}. \quad (20)$$

A break-in moment (BIM) is any point in time when either a surgery starts or finishes. The set of BIMs is defined as

$$\mathscr{B}_{\Pi} = \left\{t \in \mathbb{R}_{+} | t = F_i^{\Pi}, i = 1, \ldots, n\right\}. \tag{21}$$

A break-in interval (BII) is the time between two consecutive BIMs. Since preemption and idling are not allowed, these are precisely the moments at which emergency patients can be fitted into the schedule, hence the name. See Fig. 1 for a visualization. The set of BIIs is defined as

$$\mathscr{I}_{\Pi} = \{(a, b) \subset \mathbb{R}_{+} | a, b \in \mathscr{B}_{\Pi}, a \leq b, \nexists c \in \mathscr{B}_{\Pi} \text{s.t.} a \leq c \leq b\}. \tag{22}$$

The objective is to minimize the maximum waiting time of an emergency patient. This is equivalent with minimizing the maximum BII over all possible schedules. The BIM problem is defined as

$$\text{(BIM)} \qquad \min_{\Pi} \left(\max_{I \in \mathscr{I}_{\Pi}} (|I|) \right).$$

Note that BIIs are the intervals between BIMs; therefore, the set of BIIs forms a partition of the time interval that the ORs are in operation. The total time the ORs are in operation is not influenced by the sequencing of the surgeries because the surgeries are preassigned to the ORs and idle time is not allowed. This implies that if a schedule has a lot of relatively short BIIs, it will most likely have a long maximum BII. Therefore, intuitively we see that the goal of the BIM problem is to spread the BIMs as evenly as possible over the interval the ORs are in operation. Ideally, the BIMs are spread equidistantly over the interval, and all BIIs would have length

$$\lambda = \frac{m \cdot \max_i \left(F_i^{\Pi}\right)}{n}. \tag{23}$$

The value λ is a lower bound for the optimal BIM value. Note that $\max_i \left(F_i^{\Pi}\right)$ is the time that surgeries are performed at the OR(s), given a schedule Π.

Van Essen et al. [21] show that the BIM problem is strongly NP-hard and therefore intractable for realistic size instances. In order to obtain reasonably good schedules for this problem fast enough for practical purposes, they describe several heuristics. First, they propose to sort the surgeries in increasing order of duration and add the first surgery to the schedule such that the new resulting BII is close enough to λ, where close enough depends on the scheduler's preferences. Second, they propose to add the surgery for which the resulting BII is closest to λ to the schedule first. Third, they propose the previous heuristic with the addition to update the value of λ every time a surgery is added to the schedule. Based on simulation they conclude that the third heuristic performs best and results in a reduction of

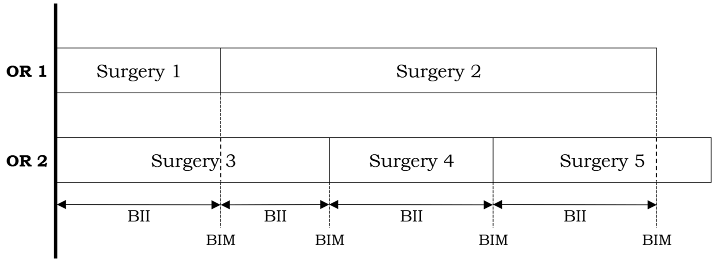

Fig. 1 Break-in moments and break-in intervals

about 10% of the maximum BII length compared to the base schedule provided by the hospital.

The BIM model provides an elegant framework to reduce the maximum waiting time of emergency patients. However, it also has its shortcomings. First, the BIM model aims to minimize the maximum waiting time of an emergency patient, but it does not take into account that the schedule will change when an emergency patient is accommodated into the schedule. In a sense the BIM model only optimizes for the first arriving emergency patient. Second, the BIM model only takes the maximum waiting time into account. In Sect. 5 we discuss a way to address the first shortcoming, and in Sect. 3.2.2 we discuss a slightly modified problem that addresses the second shortcoming.

3.2.2 Average Waiting Time

The BIM problem described in the previous section provides an upper bound on the waiting time of arriving emergency patients, but it does not provide insight in the expected waiting time. The two schedules in Fig. 2 both have a maximum BII of length 2 and are therefore, according to the BIM problem, equivalent. However, the expected waiting time of a patient arriving at an arbitrary time during the day is 0.8 in the left schedule but only 0.6 in the right schedule. This example illustrates that only taking the largest BII into account may result in unnecessary waiting time for emergency patients. As we noted in Sect. 3.2.1, for the BIM problem it is optimal if all BIIs have length λ (see equation (23)). This also holds if we consider the average waiting time. The intuition for this is the following: suppose there is a large BII, then, because the sum of the lengths of the BIIs is constant, at least one of the other BIIs will be small. The probability that an emergency patient, who arrives at an arbitrary moment during the day, arrives during the large BII is higher than the probability of arriving during the small BII because this probability is proportional to the length of the interval. So we have that the expected waiting time for an emergency patient arriving at an arbitrary moment during the day increases if we deviate from a schedule with equidistant BIMs. The problem to find the schedule with the minimum expected average waiting time for emergency patients is therefore

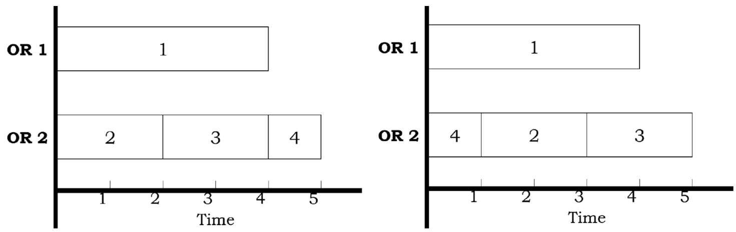

Fig. 2 Two schedules, both with a maximum BII of 2. The expected waiting time is 0.8 for the left schedule and 0.6 for the right schedule

equivalent to the problem of finding a schedule with a minimal sum of squared BII lengths. The minimal squared BII (MSB) problem is defined as

$$\text{(MSB)} \qquad \min_{\Pi} \left\{ \sum_{(a,b)\in \mathscr{I}_{\Pi}} (b-a)^2 \right\}.$$

Note that, although the BIM and MSB problems are quite similar and both try to approach the equidistant distribution of the BIMs, they are not equivalent. The reason for this is that the BIM problem focuses only on the maximum BII, whereas the MSB problem takes all BIIs into account. Unlike the BIM problem, the MSB problem cannot be formulated as an integer linear program (ILP), since the objective function is quadratic and not linear. This problem is not further described in literature. Therefore, further analysis of this problem and finding useful heuristics remains an open problem.

4 Artificial Variability

A surgery can only be conducted as planned if all necessary resources are available at the right time. For example, if the surgical team or the anesthesiologist is unavailable, the surgery will be delayed or canceled. Also when the OR is not clean yet, the surgery will be delayed until the cleaning staff has finished. In this section we discuss several models that consider the variability in the OR schedule due to unavailability of necessary resources.

4.1 Blocking Time

Usually, several resources necessary for surgery are shared between ORs. For example, an anesthesiologist that provides care at several ORs or an OR-cleaning team. If during the day the demand for such a resource is higher than the available capacity, e.g., if there are two cleaning teams but three ORs finish at the same time, some ORs will be blocked for some time, causing surgeries to be delayed. With this in mind, it is reasonable to schedule the surgeries in such a way that not too many surgeries either start or end at the same time. This problem is first described by [19]. In their paper they describe the Scheduling with Safety Distances (SSD) problem in which surgeries are scheduled such that the minimal time between start times is at least v. Below we state and discuss this problem.

We use the notation and definitions stated in Sect. 3.2.1. Suppose there are n surgeries to be scheduled at m ORs. Before a surgery can start, a certain procedure has to be performed, e.g., anesthesiology. For this procedure there is only one available resource. The aim is to schedule the surgeries such that there is no blocking of the ORs due to unavailable resources.

The Scheduling with Safety Distances (SSD) problem is to find an optimal schedule Π^* such that

$$\text{(SSD)} \qquad \Pi^* = \arg \min_{I \in \mathscr{I}_\Pi} (|I|) \geq v$$

for a certain threshold value v, where $\mathscr{I}_\Pi$ is the set of all BIIs in schedule Π (see equation (22)). Spieksma et al. [19] prove that the SSD problem is NP-complete. They identify some special cases for which the problem is solvable in polynomial time, but these are very stylized cases that are not useful in practice.

The variant of the SSD problem described by [19] assumes that there is only one resource available that is necessary for each surgery (e.g., if there is only one cleaning team at the ORs). In practice, there are often multiple, albeit limited, resources. We can adjust the model for this situation by instead of requiring that all start times are at least v apart, requiring that at most n start times are less than v apart, when there are n available resources.

If the threshold value v is not known or it is variable, we can adjust the problem such that it aims to schedule the surgery start times not at least v apart but as far apart as possible. This modified SSD problem has a lot of resemblance with the BIM problem. We use the name Scheduling with Maximum Safety Distances (SMSD) for this problem, and it is defined as

$$\text{(SMSD)} \qquad \max_{\Pi} \left(\min_{I \in \mathscr{I}_\Pi} (|I|) \right).$$

Where the BIM problem minimizes the maximum BII and the SMSD problem maximizes the minimum BII. By the same reasoning as for the BIM and MSB problem, we see that all three problems in fact aim to spread the BIMs as evenly as possible over the day. However, the BIM, MSB, and SMSD problems are not equivalent. In Sect. 5 we will discuss the relation between these problems.

5 Multiple Sources of Disturbance

In the previous sections we discussed models that consider one type of disturbance to the OR schedule in isolation. However, in practice most, if not all, of the disturbances described will be of influence simultaneously. Furthermore, most of the objectives we considered are influenced by more than one source of disturbance. In this section we therefore discuss models that take multiple sources of disturbances into account.

5.1 Overtime

In Sects. 2.1 and 3.1 we discussed models that minimize overtime anticipating emergency arrivals and variable surgery duration, respectively. All these models share the same objective, i.e., reducing the probability of overtime. Moreover, they use similar techniques, i.e., estimating the mean capacity needed and, based on the variance, additional slack capacity to reduce the probability of overtime. It is therefore reasonable to incorporate them in one model. Van Houdenhoven et al. [22] extend their model to determine the additional capacity necessary for emergency patients; see Sect. 3.1 to incorporate the variability of the elective surgeries. For this, suppose that there are n_z^{el} elective surgeries of specialty z, each having mean duration μ_z^{el} with standard deviation σ_z^{el}. Then the formula for the slack capacity needed (14) becomes

$$\delta_z(p_z) = n_z^{em}\mu_z^{em} + \beta(p_z)\sqrt{n_z^{el}(\sigma_z^{el})^2 + n_z^{em}(\sigma_z^{em})^2}, \quad (24)$$

where n_z^{em} is the number of expected emergency surgeries of specialty z with expected duration μ_z^{em} and variance $(\sigma_z^{em})^2$. So for each specialty z there are $n_z^{el}\mu_s$ time units of planned capacity, and additionally an amount of $\delta_z(p_z)$ of slack capacity is added to the schedule. Like in Sect. 3.1, $\beta(p_z)$ is a factor depending on the probability distributions of the surgeries and the maximal tolerated probability of overtime p_z.

The model proposed by [11] (see equations (15)–(17)) can be extended to incorporate the variability in the durations of the surgeries by modeling the duration d_i as a random variable like the required capacity for emergency surgeries C_t^{em}. This, however, reduces the tractability of the model. Lamiri et al. [10] propose a stochastic programming approach in which both the durations of the elective surgeries and the demand for emergency surgeries are modeled as a random variable. Lamiri et al. [12] and Razmi et al. [18] describe a column generation approach to solve this model efficiently.

5.2 Waiting Time Emergency Patients

In Sects. 2.1 and 3.1 we discussed several ways how a scheduler can determine the amount of slack capacity to anticipate emergency patients or variation in the surgery duration. Usually, the slack capacity is scheduled at the end of the day to act as a buffer to prevent overtime. However, if we take other objectives into account, it may be better to schedule the slack capacity at other moments. In Sect. 3.2.2 we discussed the BIM model. By dividing the slack capacity around these BIMs, the time until emergency patients can go into surgery can be reduced. Furthermore, by introducing slack intervals in the schedule, emergency patients arriving during these intervals will have zero waiting time.

5.3 Deviation from the Schedule

Analysis of the models on minimizing deviation from the schedule, discussed in Sect. 2.2, shows that the Smallest Variance First (SVF) rule is in many cases optimal, and otherwise it still yields very good results. The intuition for this is that since surgeries are scheduled consecutively, the variability of earlier surgeries will accumulate for the later surgeries. Therefore, by scheduling highly variable surgeries near the end of the day, the effect on the other surgeries will be limited. Again, as discussed above, the accumulating effect of the variability can be further reduced by adding slack capacity at the end of each surgery, based on the variation of the surgery duration.

The BIM problem discussed in Sect. 3.2.2 and the SMSD problem discussed in Sect. 4.1 both aim to spread the start and end times of the surgeries evenly over the day. However, since the BIM problem only considers the maximum BII, it accepts schedules where some of the BIIs are small as long as the maximum BII remains the same. Similarly the SMSD problem can produce schedules that have large BIIs. We can avoid this by extending the problem slightly. For this we define the following two BII vectors:

$$b_{\Pi}^{+}=[b_1^{+}, b_2^{+}, \ldots, b_n^{+}], \quad \text{where } b_1^{+} \geq \cdots \geq b_n^{+} \text{and } b_k^{+}=|I_k|,\, I_k \in \mathscr{I}_{\Pi},\, k=1\ldots n, \tag{25}$$

$$b_{\Pi}^{-}=[b_1^{-}, b_2^{-}, \ldots, b_n^{-}], \quad \text{where } b_1^{-} \leq \cdots \leq b_n^{-} \text{and } b_k^{-}=|I_k|,\, I_k \in \mathscr{I}_{\Pi},\, k=1\ldots n, \tag{26}$$

where $\mathscr{I}_{\Pi}$ is the set of all BIIs (see equation (22)). The vector b_{Π}^{+} is a vector containing the lengths of the BIIs corresponding to schedule Π in descending order, and b_{Π}^{-} is the same vector only in the reverse order. With this notation we can define the BIM problem as finding a schedule Π for which the corresponding vector b_{Π}^{+} has a minimum first element. The extended BIM and SMSD problems are considering not only the first element of the b vector but all entries. A schedule Π^* is optimal with respect to the lexicographical break-in moments (LBIM) problem if its corresponding BII vector $b^+_{\Pi^*}$ is lexicographically minimal, i.e.,

$$b^+_{\Pi^*} \leq_{\text{lex}} b^+_{\Pi} \quad \forall \Pi. \tag{27}$$

A schedule Π^* is optimal with respect to the lexicographical Scheduling with Maximum Safety Distances (LSMSD) problem if its corresponding BII vector $b^-_{\Pi^*}$ is lexicographically maximal, i.e.,

$$b^-_{\Pi^*} \geq_{\text{lex}} b^-_{\Pi} \quad \forall \Pi. \tag{28}$$

Although LBIM and LSMSD appear to be very similar, both aim to the schedule the BIMs as evenly spread over the day as possible, they are not equivalent. Suppose we have a set of surgeries with durations 2, 2, 2, 11, 11, 11, 11, 12, 20, and 20. Let Π_1 denote the LBIM optimal schedule and Π_2 the LSMSD optimal schedule for this set of surgeries; see Fig. 3. The corresponding BII vectors are

$$\begin{aligned} b^+_{\Pi_1} &= [10, 10, 10, 10, 2, 1, 1, 1, 1], \\ b^-_{\Pi_1} &= [1, 1, 1, 1, 2, 10, 10, 10, 10], \\ b^+_{\Pi_2} &= [11, 11, 10, 7, 5, 4, 2, 2, 2, 2] \text{ and} \\ b^-_{\Pi_2} &= [2, 2, 2, 2, 4, 5, 7, 10, 11, 11]. \end{aligned}$$

We have that $b^+_{\Pi_1} \leq_{\text{lex}} b^+_{\Pi_2}$ but also $b^-_{\Pi_2} \leq_{\text{lex}} b^-_{\Pi_1}$. This shows that the LBIM problem and the LSMSD problem are not equivalent. As mentioned before there are clear similarities between the two problems; however, it remains an open problem what the implications of these similarities are for the optimal schedules for both problems.

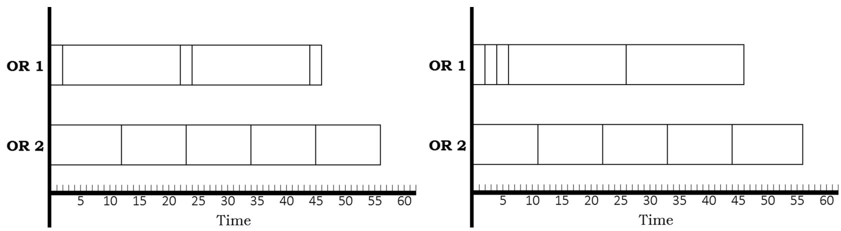

Fig. 3 The LBIM optimal schedule, Π_1 (left), and the LSMSD optimal schedule, Π_2 (right)

Table 1 Overview of the models of this chapter (between parenthesis are the subsections where the models are discussed) for each combination of objective and type of disturbance

	Internal variability	External variability	Artificial variability
Overtime	Slack capacity ILP and SP (Sect. 2.1)	Non-elective policy (Sect. 3.1.1) Non-elective capacity (Sect. 3.1.2)	SSD (Sect. 4.1)
Waiting time	E/T (Sect. 2.2)	BIM (Sect. 3.2.1), MSB (Sect. 3.2.2)	SSD (Sect. 4.1)
Utilization	Slack capacity ILP and SP (Sect. 2.1) E/T (Sect. 2.2)	Non-elective policy (Sect. 3.1.1) Non-elective capacity (Sect. 3.1.2) Assigning surgeries (Sect. 3.1.3)	SSD (Sect. 4.1)

6 Conclusion

In this chapter we presented a model-based overview of literature on robust surgery scheduling. We showed that robustness is a key aspect to improve OR performance. A non-robust surgery schedule performs well when all processes at the OR are going as expected but can significantly reduce the OR performance when there are small disturbances in the schedule during the day. There are several types of disturbances which we divide into three groups. First, those that affect the scheduling of elective surgeries. Second, those that affect the effectuation of the schedule, e.g., non-elective surgeries. Third, those that affect the availability of resources needed. The best suited approach to anticipate one of the described disturbances is highly dependent on the objective the scheduler has. Table 1 gives an overview of models described in literature for combinations of a disturbance and an objective. In the large majority of papers on robust surgery scheduling, one specific source of disturbance is analyzed in isolation. In Sect. 5 we described some models that incorporate multiple sources of disturbances. Additional research is needed in order to develop models anticipating multiple sources of disturbances. Furthermore, models described in literature are mostly applied to a specific case which complicates the comparison of different approaches. It would therefore be beneficial to apply the various approaches to a standardized case, like [14], in order to develop a benchmark for adequate comparison of the models proposed in literature.

References

1. Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer-Verlag, New York.
2. Borgman, N. J. (2017). Managing urgent care in hospitals. *PhD thesis, University of Twente, Enschede, The Netherlands*.
3. Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.
4. Denton, B. T., Miller, A. J., Balasubramanian, H. J., and Huschka, T. R. (2010). Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, 58(4-part-1):802–816.
5. Ferrand, Y. B., Magazine, M. J., and Rao, U. S. (2014). Managing operating room efficiency and responsiveness for emergency and elective surgeries, a literature survey. *IIE Transactions on Healthcare Systems Engineering*, 4(1):49–64.
6. Guda, H., Dawande, M., Janakiraman, G., and Jung, K. S. (2016). Optimal policy for a stochastic scheduling problem with applications to surgical scheduling. *Production and Operations Management*, 25(7):1194–1202.
7. Gupta, D. (2007). Surgical suites' operations management. *Production and Operations Management*, 16(6):689–700.
8. Hans, E. W. and Vanberkel, P. T. (2012). *Operating Theatre Planning and Scheduling*, pages 105–130. Springer US, Boston, MA.
9. Hulshof, P. J. H., Kortbeek, N., Boucherie, R. J., Hans, E. W., and Bakker, P. J. M. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems*, 1(2):129–175.
10. Lamiri, M., Dreo, J., and Xie, X. (2007). Operating room planning with random surgery times. In *2007 IEEE International Conference on Automation Science and Engineering, Scottsdale, AZ, USA*, pages 521–526.
11. Lamiri, M., Grimaud, F., and Xie, X. (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics*, 120(2):400–410.
12. Lamiri, M., Xie, X., and Zhang, S. (2008). Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions*, 40(9):838–852.
13. Landa, P., Aringhieri, R., Soriano, P., Tánfani, E., and Testi, A. (2016). A hybrid optimization algorithm for surgeries scheduling. *Operations Research for Health Care*, 8:103–114.
14. Leeftink, G. and Hans, E. W. (2018). Case mix classification and a benchmark set for surgery scheduling. *Journal of Scheduling*, 21(1):17–33.
15. Min, D. and Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206(3):642–652.
16. Molina-Pariente, J. M., Hans, E. W., and Framinan, J. M. (2016). A stochastic approach for solving the operating room scheduling problem. *Flexible Services and Manufacturing Journal*, 30(1-2):224–251.
17. Otten, J. W. M. and Boucherie, R. J. (2018). Minimizing Earliness/Tardiness costs on multiple machines with an application to surgery scheduling. *Submitted*.
18. Razmi, J., Yousefi, M., and Barati, M. (2015). A stochastic model for operating room unique equipment planning under uncertainty. *IFAC-PapersOnLine*, 48(3):1796–1801.
19. Spieksma, F. C. R., Woeginger, G. J., and Yu, Z. (1995). Scheduling with safety distances. *Annals of Operations Research*, 57(1):251–264.
20. Van Der Lans, M., Hans, E. W., Hurink, J. L., Wullink, G., van Houdenhoven, M., and Kazemier, G. (2005). *Anticipating urgent surgery in operating room departments*. Number WP-158 in Beta working papers. BETA Research School for Operations Management and Logistics.
21. Van Essen, J. T., Hans, E. W., Hurink, J. L., and Oversberg, A. (2012). Minimizing the waiting time for emergency surgery. *Operations Research for Health Care*, 1(2-3):34–44.

22. Van Houdenhoven, M., Hans, E. W., Klein, J., Wullink, G., and Kazemier, G. (2007). A norm utilisation for scarce hospital resources: Evidence from operating rooms in a Dutch university hospital. *Journal of Medical Systems*, 31(4):231–236.
23. Van Riet, C. and Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, 7:52–69.
24. Van Veen-Berkx, E., Elkhuizen, S., Kuijper, B., and Kazemier, G. (2016). Dedicated operating room for emergency surgery generates more utilization, less overtime, and less cancellations. *American Journal of Surgery*, 211(1):122–128.
25. Weiss, E. N. (1990). Models for determining estimated start times and case orderings in hospital operating rooms. *IIE Transactions*, 22(2):143–150.
26. Wullink, G., Van Houdenhoven, M., Hans, E. W., van Oostrum, J. M., van der Lans, M., and Kazemier, G. (2007). Closing emergency operating rooms improves efficiency. *Journal of Medical Systems*, 31(6):543–546.
27. Zonderland, M., Boucherie, R., Litvak, N., and Vleggeert-Lankamp, C. (2010). Planning and scheduling of semi-urgent surgeries. *Health care management science*, 13(3):256–267. 10.1007/s10729-010-9127-6.