

Chapter 5 Operating Theatre Planning and Scheduling

Erwin W. Hans, Peter T. Vanberkel

Center for Healthcare Operations Improvement & Research, dep. Operational Methods for Production & Logistics, University of Twente, Enschede

Abstract In this chapter we present a number of approaches to operating theatre planning and scheduling. We organize these approaches hierarchically which serves to illustrate the breadth of problems confronted by researchers. At each hierarchical planning level we describe common problems, solution approaches and results from studies at partner hospitals.

Acknowledgments The hospitals Erasmus MC and Netherlands Cancer Institute, and all co-authors involved in the various research projects described here: Boucherie RJ, Harten WH van, Houdenhoven M van, Hulshof PJH, Hurink JL, Kazemier G, Lans M van der, Lent WAM van, Oostrum JM van, Wullink G.

5.1 Introduction

Within the OR/OM healthcare literature, operating theatre planning and scheduling is one of the most popular topics. This is not surprising, as many patients in a hospital undergo surgical intervention in their care pathway. For a hospital, the operating theatre accounts for more than 40% of its revenues and a similar large part of its costs (HFMA 2005). An efficient operating theatre department thus significantly contributes to an efficient healthcare delivery system as a whole.

An extensive overview and taxonomy of the operating theatre planning and scheduling literature is given by Cardoen et al. (2010). They conclude that the majority of the research is directed at planning and scheduling of elective patients at an operational level of control, and take a deterministic approach. Furthermore, they observe that only half of the literature contributions consider up- or downstream hospital resources, and few papers report about implementation in practice. This appears to be a common problem in OR/OM healthcare literature (Brailsford et al. 2009). An up-to-date online bibliography of the operating theatre management literature is maintained by Dexter (2011), and a structured literature review of operations research in the management of operating theatres is given by Guerrero and Guido (2011).

In this chapter we address operating theatre planning problems on three hierarchical managerial levels: strategic, tactical and offline operational planning, as introduced in Section 5.2.

The remainder of this chapter addresses recent work in each of these three levels of control. Section 5.2 outlines the planning and control functions on the aforementioned hierarchical levels in an operating theatre department. Section 5.3 addresses the strategic problem of determining the target utilization of an operating theatre department. Section 5.4 addresses the strategic problem of determining the number of surgical teams required during the night to deal with emergency cases. Section 5.5 addresses the strategic decision whether to use emergency operating theatres. Section 5.6 addresses the tactical problem of determining a master surgery schedule (a day-to-day allocation of operating theatres to surgical specialties) that levels the workload in subsequent departments (wards). Section 5.7 addresses the offline operational problem of scheduling elective surgeries with stochastic durations, and sequencing them in order to reduce access time of emergency surgeries. We will use a wide array of OR techniques, including discrete event and Monte Carlo simulation, statistical modeling and meta-heuristics.

5.2 A hierarchy of resource planning and control in operating theatres

Competitive manufacturing companies make planning and control decisions in a hierarchical manner (Zijm 2000). For example, the long term decision of what products to manufacture is at the top of the hierarchy and the real time decision of whether to discard a specific part due to its quality is at the bottom of the hierarchy. In general the reliance of one decision on another defines the hierarchy. Many planning and control frameworks classify decisions into the three hierarchical levels strategic, tactical, and operational, as suggested by Anthony (1965). Similar hierarchical planning and control frameworks have been proposed for healthcare (see Hans et al. 2011). Hans et al. refine the classical hierarchy by splitting the operational level into an offline and online operational level, where the former is the *in advance* short term decision making, and the latter the monitoring and control of the process in *real time*. In the remainder of this section we outline the main operating theatre planning and control functions on these four hierarchical levels.

5.2.1 Strategic planning and control

To reach organizational goals, the strategic level addresses the dimensioning of core OT resources, such as the number of OTs, the amount of personnel, instruments (e.g. X-ray machines), etc. It also involves case mix planning, i.e. the selec-

tion of surgery types, and the determination of the desired patient type volumes (Vissers et al. 2001). Agreements are made with surgical services / specialties concerning their annual patient volumes and assigned OT time. The dimensioning of subsequent departments' resources (e.g. ward beds) is also done (Vanberkel and Blake 2007). Strategic planning is typically based on historical data and/or forecasts. The planning horizon is typically long term, e.g. a year or more.

5.2.2 Tactical planning and control

The tactical level addresses resource usage over a medium term, typically with a planning horizon of several weeks (Blake and Donald 2002, Wachtel and Dexter 2008). The actual aggregate patient demand (e.g. waiting lists, appointment requests for surgery) is used as input. In this stage, the weekly OT time is divided over specialties or surgeons, and patient types are assigned to days. For the division of OT time, two approaches exist (Denton et al. 2010). When a closed block planning approach is used, each specialty will receive a number of OT blocks (usually OT-days). In an (uncommon) open block planning approach, OT time is assigned following the arrival of requests for OT time by surgeons.

On the tactical level, the surgery sequence is usually not of concern. Instead, on this level it is verified whether the planned elective surgeries cause resource conflicts for the OT, for subsequent departments (ICU, wards), or for required instruments with limited availability (e.g. X-ray machines). The design of a Master Surgical Schedule is a tactical planning problem.

5.2.3 Operational planning and control (offline)

The offline operational level addresses scheduling of specific patients to resources (and as a consequence, the sequencing of activities) and typically involves a planning horizon of a week. It encompasses the rostering of OT-personnel, and reserving resources for add-on surgeries (Dexter et al. 1999). In addition, it addresses the sequencing of surgeries (Denton et al. 2007), to prevent critical resource conflicts, e.g. regarding X-ray machines, instrument sets, surgeons, etc. When there are no dedicated emergency OTs, the sequencing of the elective surgeries can also aid in spreading the planned starting times of elective surgeries (which are “break-in moments” for emergency surgeries) in order to reduce the emergency surgery waiting time (Wullink et al. 2007).

5.2.4 Operational planning and control (online)

The online operational level addresses the monitoring and control of the day-to-day activities in the OT. Obviously at this level of control, all uncertainty materializes and has to be dealt with. If necessary, surgeries are rescheduled, or even cancelled (Dexter et al. 2004, McIntosh et al. 2006). This is usually done by a day coordinator in the OT department. Emergency surgeries, which have to be dealt with as soon as possible, are scheduled, and emergency OT teams may have to be assembled and dispatched to the first available OT. If there are emergency OTs, these emergency surgeries are dispatched to these OTs. If there are no such OTs, they are scheduled within the elective surgical schedule.

In summary, strategic planning typically addresses capacity dimensioning decisions, considering a long planning horizon of multiple years. Tactical planning addresses the aggregate capacity allocation to patient types, on an intermediate horizon of weeks or months. Offline operational planning addresses the in-advance detailed capacity allocation to elective patients, with a short planning horizon of days and up to a few weeks. Online operational planning addresses the monitoring and control of the process during execution, and encompasses for example reacting to unforeseen events.

5.3 Strategic: the problem with using target OT utilization levels

Utilization of operating theatres is high on the agenda of hospital managers and researchers and is often used as a measure of efficiency, both introspectively as well as in benchmarking against other OT departments. As a result, much effort is spent trying to maximize OT utilization and sometimes, without understanding the factors affecting it. Using straightforward statistical analysis we show how the target OT utilization of a hospital depends on the patient mix and the hospital's willingness to accept overtime. This work is described in detail in Houdenhoven et al. (2007). Similarly, the erroneous nature of target *ward* occupancies is studied by Harper and Shahani (2002) and discussed by Green in Hall (2006).

5.3.1 General model

There are various ways to compute the utilization rate. We define the OT utilization as the expected total surgery duration (including changeover/cleaning time) divided by the amount of time allotted (see Figure 5.1):

$$\text{Expected OT utilization} = \frac{\text{expected total surgery duration}}{\text{allotted time}} \quad (5.1)$$

Our approach can be easily extended to deal with more extensive definitions of OT utilization.

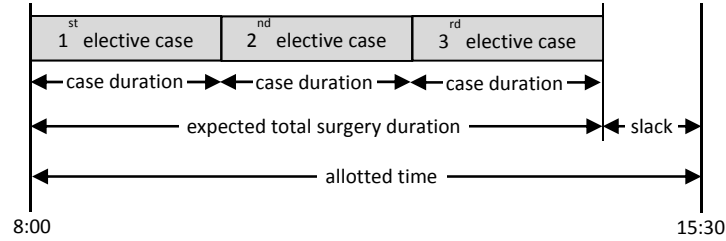


Fig. 5.1 Timeline for surgical cases

The amount of allotted time is computed as follows:

$$\text{Allotted time} = \text{expected total surgery duration} + \text{slack time} \quad (5.2)$$

where the slack time (reserved capacity) is determined in such a way that a certain frequency of overtime is achieved. This is a managerial choice: slack time reduces cancellations and/or costly overtime, but also reduces OT utilization. The frequency of overtime depends on the distribution of the total surgery duration and can be computed according to:

$$\begin{aligned} \text{Frequency of overtime} \\ = P(\text{total surgery duration} > \text{allotted time}) \end{aligned} \quad (5.3)$$

Now more formally, let μ_s and σ_s denote the average and standard deviation of elective surgical case durations of type s , and let n_s denote the number of cases completed in one block. A type s may correspond with the surgeries of for example a surgical specialty or a specific surgeon. Likewise, n_s^e , μ_s^e and σ_s^e denote the same for emergency cases. All these parameters are based on historical data. It follows that the total expected duration of all elective cases in one OT block is $n_s \times \mu_s$ and the standard deviation of the total duration of these n_s cases equals $\sqrt{n_s \times \sigma_s^2}$. Accordingly:

$$\text{Expected total surgery duration} = n_s \times \mu_s + n_s^e \times \mu_s^e \quad (5.4)$$

$$\text{Total surg. duration std. deviation} = \sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2} \quad (5.5)$$

The accepted risk (or frequency) of overtime is denoted by r_s . Now we can complete equation (5.2). The amount of allotted time required to achieve an overtime frequency of r_s can be computed as follows:

$$\text{Alloted time} = n_s \times \mu_s + n_s^e \times \mu_s^e + \alpha(r_s) \sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2} \quad (5.6)$$

where $\alpha(r_s)$ is a function yielding probability r_s . The outcome of this function depends on the distribution of the surgery duration. Using $\alpha(r_s)$ in this way allows the approach to be independent of the surgery duration distribution, i.e. function $\alpha(r_s)$ can be changed to reflect various distributions. Using equations (5.4) and (5.6) we can complete formula (5.1) for the expected OT utilization as a function of the frequency of overtime as follows:

$$\text{Expected OT utilization} = \frac{n_s \times \mu_s + n_s^e \times \mu_s^e}{n_s \times \mu_s + n_s^e \times \mu_s^e + \alpha(r_s) \sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2}} \quad (5.7)$$

5.3.2 General results

We use formula (5.7) for the expected OT utilization to illustrate the relationship between OT utilization, patient mix and overtime frequency. In a theoretical scenario where there is no surgery duration variability (i.e., $\sigma_s = \sigma_s^e = 0$), the expected OT utilization is obviously 100%.

As a case study we consider Erasmus Medical Center in Rotterdam, the Netherlands. OT management in this hospital accepts a 30% risk of overtime. For simplicity, they assume that the total surgery duration follows a normal distribution $\sim N(n_s \times \mu_s + n_s^e \times \mu_s^e, \sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2})$. Using straightforward statistical analysis we can show that $\alpha(r_s) = 0.5$ when the acceptable frequency of overtime is 30%. This is shown as follows. Let X be the total surgery duration, then:

$$\begin{aligned} 30\% &= P\left(X > n_s \times \mu_s + n_s^e \times \mu_s^e + \alpha(r_s) \sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2}\right) \Leftrightarrow \\ &P\left(X \leq n_s \times \mu_s + n_s^e \times \mu_s^e + \alpha(r_s) \sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2}\right) = 0.7 \Leftrightarrow \\ &P\left(\frac{X - n_s \times \mu_s - n_s^e \times \mu_s^e}{\sqrt{n_s \times \sigma_s^2 + n_s^e \times (\sigma_s^e)^2}} \leq \alpha(r_s)\right) = 0.7 \Leftrightarrow \\ &P(Z \leq \alpha(r_s)) = 0.7 \end{aligned}$$

where $Z \sim N(0,1)$. It follows that $\alpha(r_s) = 0.5$.

We use 2 years of historical data from the aforementioned hospital. We consider three different surgical specialties (i.e. three different patient mixes) and for each we show the trade-off between expected OT utilization and overtime frequency. This is illustrated in Figure 5.2.

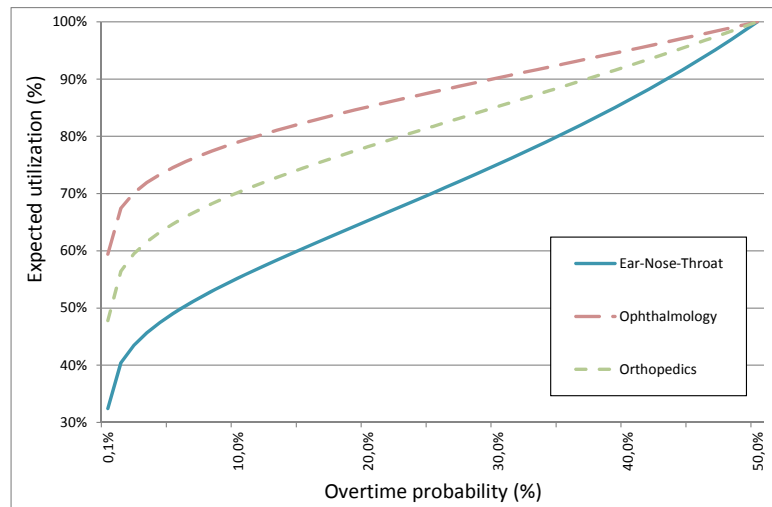


Fig. 5.2 Trade-off between overtime probability and expected utilization

The calculated expected OT utilization can also be regarded as a target utilization, or benchmark. Figure 5.2 shows that a single OT utilization target will result in different overtime frequencies for each specialty. For example, a target utilization of 80% will result in an overtime frequency of approximately 12% for ophthalmology but an overtime frequency of approximately 35% for ENT. In general, a low risk of overtime and a complex patient mix will result in a low utilization rate. If the accepted risk of overtime is higher and the patient mix less complex, then a higher utilization can be achieved. Given that overtime is expensive (and perhaps limited by collective bargaining agreements), this example illustrates the inadequacy of a single target OT utilization as a performance metric. It also illustrates the importance of taking case mix characteristics into account when comparing utilization figures between different OT departments.

5.4 Strategic: on-call or in-house nurses for overnight coverage for emergency cases?

Treating emergency patients is a common activity for most hospitals. Likewise, the operating theatre must be available to provide emergency operations 24 hours per day. The night shift (e.g., from 11:00 pm to 7:30 am) is typically the most expensive shift to staff due to collective labor agreements and the inconvenient hours. Determining minimum cost staffing levels that provide adequate coverage to meet emergency demand is a strategic problem. In this section we describe a

case study to determine appropriate night shift staffing levels at Erasmus Medical Center. The outcomes of the study were successfully implemented.

5.4.1 General problem formulation

Covering the night shift is usually accomplished by using in-hospital and on-call nurses. The in-hospital nurses are stationed in the hospital while waiting for emergency cases. The on-call nurses wait at their homes for emergency cases (typically there is a requirement that they can be present in the hospital within e.g. 30 minutes of being requested to do so) and are typically cheaper than in-hospital teams. In general, A single nurse can complete exactly one case at a time but can complete any number of cases in series until the end of the shift. The decision required in this problem is to determine how many in-hospital and on-call nurses are necessary to meet the demand for emergency cases. Timeliness is of the essence here, as these emergency cases may be very urgent.

When the first emergency case presents for surgery during a night shift, in-house nurses respond. Depending on the hospital policy and the total number of in-house nurses, an on-call nurse may be called in. In other words, some hospitals may wait until 1, 2, ... or all in-house nurses are busy before calling in a nurse from home, while other hospitals may wait until all in-house nurses are busy and an emergency case is present. For each subsequent emergency case this process is repeated. Note that nurses are available to complete multiple surgeries per night and are available again after completing a surgery. Finally surgeries cannot be preempted. In this subsection we assume the hospital's policy for calling an on-call nurse is fixed, although determining this policy is, in and of itself, and interesting research question.

There are generally two types of emergency cases: those that need to be started immediately and those that can be delayed before being started. The former we refer to as emergent cases and the latter as urgent cases. The acceptable delay, or safety interval, for starting an urgent case varies: "for example a facility may consider it imperative for a patient with a ruptured abdominal aortic aneurysm to be operated on within 30 min of arrival, while a patient with an amputated finger should be operated on within 90 min of arrival, and a patient with a perforated gastric ulcer should be operated on within 3 h of arrival" (Oostrum et al. 2008).

By incorporating the acceptable delays for urgent cases it is possible to postpone urgent case demand to a later cheaper shift, and/or postpone the case until busy in-house nurses are free. To examine these possibilities in detail, Oostrum et al. (2008) use a discrete-event simulation and a case study at Erasmus Medical Center Rotterdam (Erasmus MC). To illustrate the benefits of postponing surgeries, the authors compare results with surgery postponements with the approach of Dexter and O'Neill (2001) where surgery postponements are not used. In the following subsection we provide an overview of the results.

5.4.2 General results

Current practice at Erasmus MC had a team composition of 9 in-house nurses and 2 on-call nurses. Using the approach of Dexter and O'Neill, a team of 8 in-house nurses and 2 on-call nurses was determined to be appropriate. A number of other team compositions were considered, ranging from a total of 11 nurses to a total of 6. Each team composition represented a what-if scenario in the simulation model. The simulation model was used to determine the number of surgeries cases starting later than required.

To compute the cost of each team composition, observe that – since the number of working hours does not depend on the team composition – we only need to look at the cost of idle staff. Under Dutch law, the costs for nurses who wait during the night shift are 107.5% of the regular hourly daytime wage for in-house nurses, and 106% for nurses on call. We thus compute the cost of waiting nurses in each team composition as follows:

$$\text{Cost of waiting} = (\text{number of in-house nurses} \times 1.075 + \text{number of on-call nurses} \times 1.06) \times \text{hourly wage}$$

Figure 5.3 displays the cost of waiting and percentage of surgeries starting late for the considered team compositions, where we assumed for simplicity that a regular hour's wage is 1. It shows that current practice of 9 in-house and 2 on-call nurses performs the best. However, the waiting cost can be decreased by approximately 18.5% by switching to a team composition of 5 in-house and 4 on-call nurses, at the expense of a 2% increase of late starts.

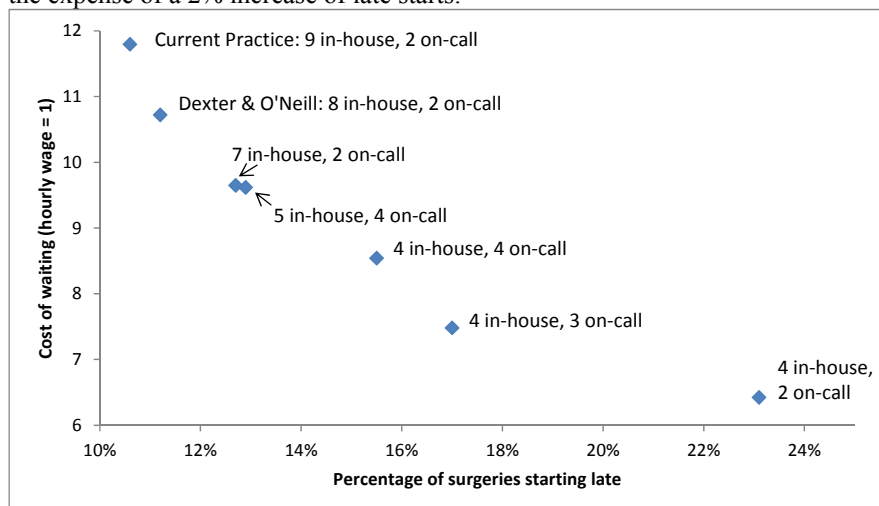


Fig. 5.3 Cost of waiting and late surgery starts for various team compositions

For policy making, managers can use results like these to see the relative performance cost associated with each staff assignment. The decision autonomy remains with the policy makers and they are left to determine if cheaper staffing levels justify the a decrease in performance.

For more extensive results, we refer to (Oostrum et al. 2008), where the authors present the distribution of cases starting later than required, surgical specialty specific results, results for multiple nurse types, and an extensive sensitivity analyses. The sensitivity analyses showed that the approach can be generalized for use in other centers.

Oostrum et al. (2008) report that heavy involvement of clinical staff in this project was essential for the following reasons. Staff assessed the safety intervals for urgent patients to ensure changes did not negatively affect patient's safety. They validated the discrete event simulation model, and suggested various scenarios for sensitivity analyses. The visualizations provided by the computer simulation aided to convince them of the final conclusions. As a result, despite the negative impact on their salary, the staff accepted the adjustment of the team composition to 5 in-house and 4 on-call nurses. For Erasmus MC this intervention resulted in an annual cost saving of 275,000 Euro.

5.5 Strategic: Emergency operating theatres or not?

During regular working hours, most hospitals either perform emergency operations in dedicated emergency OTs, or in regular elective patient OTs. For the second option a certain amount of slack is scheduled in order to fit in emergency cases without causing excessive cancellations of elective cases. The choice to use Policy 1 (reserving capacity in dedicated emergency OTs) or Policy 2 (reserving capacity in multiple regular emergency OTs) is the strategic decision addressed in this section. The difference between these two policies is illustrated graphically in Figure 5.4.

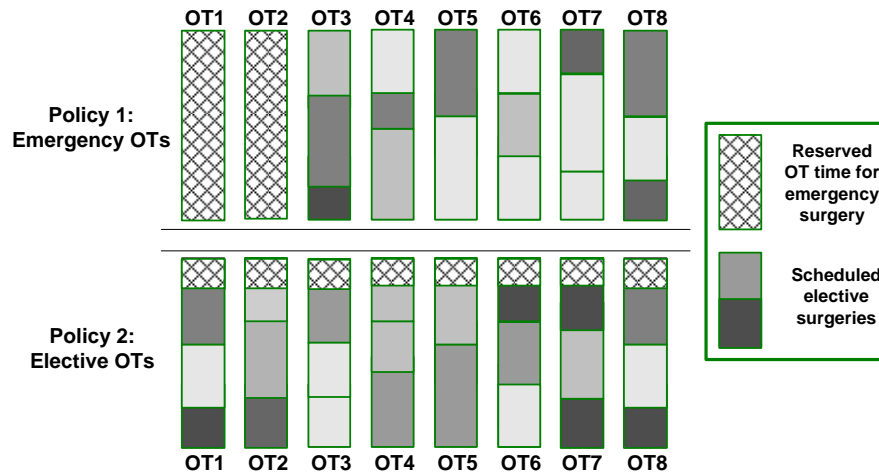


Fig. 5.4 Cost of waiting and late surgery starts for various team compositions

The flow of patients is summarized as follows: “Emergency patients arriving at a hospital that has adopted the first policy, will be operated immediately if the dedicated OT is empty and will have to queue otherwise, whereas patients arriving at a hospital that has adopted the second policy can be operated once one of the ongoing elective cases has ended. Other planned cases will then be postponed to allow the emergency operation” (Wullink et al. 2007).

Policy 1 has the advantage that the first emergency case of the day can begin without delay, but all following cases may be subject to delay. Furthermore this policy means only the emergency OTs needed to be equipped to deal with emergency cases. Finally, as a result of emergency surgeries elective surgeries will experience no delay (Bhattacharyya et al. 2006, Ferrand et al. 2010) and elective OTs will experience no overtime (Wixted et al. 2008).

Policy 2 cannot guarantee any emergency case will begin without delay, but since emergency cases can be completed in more OTs, an opening (i.e. a case finishing) for the subsequent cases may happen sooner than in policy 1. The benefits from this policy essentially result from flexibility. To ensure this flexibility (and the corresponding benefits) multiple (or all) of the OTs must be equipped to deal with emergency cases.

5.5.1 General problem description

The decision that is required is to determine how to reserve OT capacity for emergency cases, i.e. according to policy 1 or policy 2. There are advantages and disadvantages of both policies introduced above. Due to the stochastic nature of emergency cases (arrivals and surgery durations) choosing the best policy is not

immediately obvious. To compare the policies we suggest evaluating the following metrics:

- *emergency surgery waiting time*: the total delay, or the delay past what is allowed to receive emergency surgery.
- *elective surgery waiting time*: the difference between the planned and actual starting time of an elective surgery.
- *OT overtime*: the time used for surgical procedures after the regular block time has ended.
- *OT utilization*: the ratio between the total used operating time for elective procedures and the available regular time.

The following instance parameters are taken into account: elective surgery volume and duration characteristics, emergency surgery arrival and duration characteristics.

5.5.2 General results

We summarize a case study (presented in detail in Wullink et al. 2007) where discrete event simulation was used to prospectively evaluate both policies. The case study was used to support decision making at Erasmus MC. When applying Policy 2, the hospital decided that all of their 12 OTs would be equipped to handle emergency cases. In policy 1, with emergency capacity allocated to 1 dedicated emergency OT, the remaining free OT time is allocated exclusively to elective OTs. In policy 2, with emergency time allocated to each elective OT, the reserved OT time is distributed evenly over all elective OTs. Figure 5.5 and Table 5.1 summarize the results from the discrete event simulation.

Table 5.1 Summary of simulation results for policy 1 and 2

	Policy 1	Policy 2
Total overtime per day	10.6	8.4
Mean number of OTs with overtime per day	3.6	3.8
Mean emergency patient's waiting time	74 (± 4.4)	8 (± 0.5)
OT utilization	74%	77%

From Table 5.1 it is clear that policy 2 outperforms policy 1 on all given outcomes.

Under Policy 1, all emergency patients were operated on within 7 hours with a mean waiting time of 74 (± 4.4) min. Under Policy 2, all emergency patients were operated upon within 80 min with a mean waiting time of 8 (± 0.5) min. OT utilization for policy 1 was 74% and 77% for policy 2. Policy 1 resulted in 10.6 hours of overtime on average per day and policy 2 resulted in 8.4. Policy 2, with emergency capacity allocated to all elective OTs, thus substantially outperforms policy 1, on all outcomes measured.

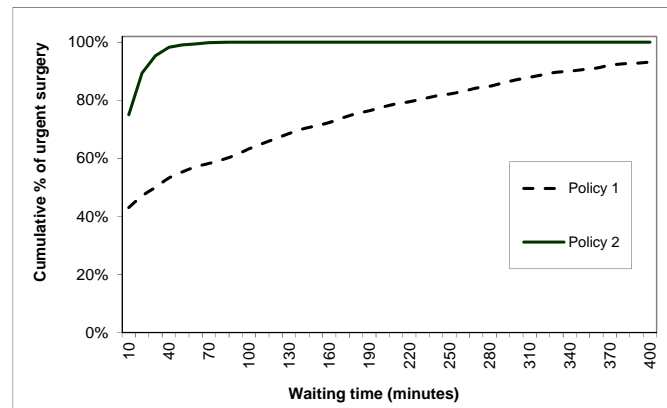


Fig. 5.5 Cumulative percentage of emergency patients in policy 1 and 2 (simulation results)

Table 5.2 summarizes the results of additional simulation experiments in which we vary the number of emergency OTs (0, 1, 2, or 3) as well as the number of elective OTs used for emergency surgeries (0, 5, 10, or 15). We use the case mix of the previous experiment, but resize the problem to 15 elective OTs (instead of 12). Furthermore, approximately 10% of the surgeries are emergency surgeries. The results show that policy 2 (dealing with emergencies in (some) elective OTs) results in improved emergency waiting performances, at the expense of increased waiting time of the elective surgeries. A mixed policy combines the advantages of both policies – the table can be used as a guideline to make a trade-off.

Table 5.2 Simulation results: (1) average and (2) maximum emergency surgery waiting time (minutes), (3) percentage of emergency surgeries that has to wait, (4) average elective surgery waiting time (minutes)

Elective OTs used for emergency	Number of emergency OTs			
	0	1	2	3
0	-	21.9	2.4	0.5
	-	3026	949	292
	-	4.4%	2.4%	1.1%
	-	12.6	12.6	12.6
5	1.3	0.6	0.3	0.1
	204.9	152.7	113.1	83.3
	4.5%	2.9%	1.5%	0.7%
	32.3	21.2	14.2	11.4
10	0.5	0.3	0.1	0.1

	94.2	76.3	63.8	50.2
	4.2%	2.6%	1.3%	0.6%
	22.2	16.0	12.1	10.3
15	0.3	0.2	0.1	0.0
	60.3	52.3	43.3	36.0
	4.0%	2.5%	1.2%	0.5%
	18.6	14.9	11.6	10.0

5.6 Tactical: designing a master surgical schedule to level ward usage

Managers are inclined to solve problems at the moment they occur (i.e., on the operational level). In Hans et al. (2011) we refer to this phenomenon as the “real-time hype” of managers. For healthcare managers, while inundated with operational problems, the universal panacea for all productivity related problems is “more capacity”. It is thereby often overlooked to tactically allocate and reorganize the available resources, which may turn out to be even more effective, and will certainly be cheaper. However, due to its longer (intermediate) planning horizon, tactical planning is less tangible and inherently more abstract than operational planning. In the majority of our healthcare process optimization research projects we find that the tactical planning level is typically not formalized and overlooked. Tactical planning decisions are rather a result of historical development (“This year’s tactical plan is last year’s tactical plan”), than a result of periodic planning. We also find they have often evolved to hard constraints for operational planning (“We don’t do orthopedic patients on Wednesday afternoons. Why? Well, we just don’t!”).

This is also typical for the tactical planning of OTs, the block planning or Master Surgical Scheduling (MSS) problem, which concerns the weekly allocation of OT-days to surgical specialties (or surgeons). To a surgeon: “operating theatre 6 on Monday is *her/his* OT”. Re-allocating OT-days may however lead to a more stable workload in subsequent departments (wards, ICU), and even reduce the required capacity of these departments. In this section we present a model to analyze and improve the impact of the MSS on the resource usage in subsequent departments.

5.6.1 General problem description

Tactical OT planning, typically involves the assignment of OT capacity to aggregate patient groups (i.e. patient cohorts) for a fixed planning horizon. This assignment should reflect the strategic goals of the hospital. For example, consider a

planning horizon of one month and a hospital with a strategic goal to complete 1000 orthopedic surgeries over the next six months, then the orthopedic surgical specialty should be assigned enough OT capacity to complete $1000/6 \approx 167$ surgeries per month.

The tactical plan is used to organize capacity over an intermediate planning horizon such that long term goals are met and to create a structure from which operational level planning can be based. In the OT this is usually accomplished with a MSS. The MSS defines which surgical specialties operate on which days during the planning horizon. Such a schedule allows the surgical specialties to plan their other functions, such as outpatient clinics, education, etc. The schedule also allows the OT department to make their own planning decisions, such as how many pieces of equipment are needed each day, what are the staffing levels, when can OT maintenance happen, etc.

When an MSS is being developed, not all of the details about the planning horizon are known, i.e. which patients will show up, which doctors will be available, etc. What is known is that a certain volume of patients with a certain case mix is expected. Hence the assignment of patient cohorts (not individual patients) to OT time is the primary factor considered when designing a MSS.

A MSS represents a repetitive pattern over a certain number of days (say Q). For each day $q \in \{1, 2, \dots, Q\}$ in the MSS each of the I available OTs has to be assigned to one of the available surgical specialties. More precisely, the MSS is described by the assignment of a surgical specialty j to each operating theatre block b_{iq} where $i \in \{1, 2, \dots, I\}$ and $q \in \{1, 2, \dots, Q\}$. Using this notation, an empty MSS (i.e. before specialties have been assigned operating theatre blocks) is shown in Figure 5.6 where each cell represents an operating theatre block. It is common that multiple blocks are assigned to a single specialty on the same day.

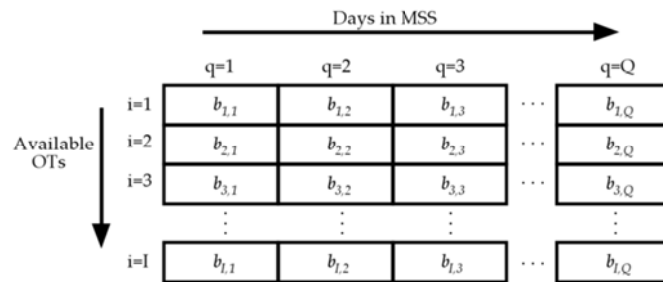


Fig. 5.6 Example empty Master Surgery Schedule (MSS)

The MSS is defined for period Q and executed repeatedly. Let M be the maximum LOS of any patient. Figure 5.7 displays how the multiple MSS cycles repeat and how patients overlap.

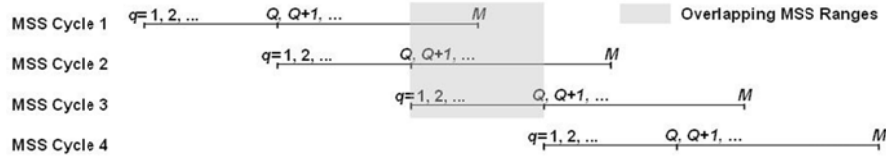


Fig. 5.7 Illustration of the overlap between multiple MSS cycles

In this section we describe a model by Vanberkel et al. (2011). The objective of the model is to make a cyclical assignment of OT time to patient cohorts for an intermediate term planning horizon, such that strategic “production levels” and performance goals are achieved. We allow for a stochastic LOS and, for a given MSS, compute a number of workload metrics associated with recovering surgical inpatients. This approach does not find the optimal MSS but rather evaluates MSS proposals. Adopting this approach to be used in conjunction with a search heuristic to find the best MSS proposal is of course a natural and very plausible extension.

The aim is to determine the number of patients in recovery as a function of the MSS. We do this by modeling the recovering patient cohorts with binomial distributions. We then add these discrete distributions (with convolutions) to determine the number of patients recovering. Once we know the number of patients recovering, we predict a number of workload metrics including admissions, ongoing inpatient care, discharges and specialized inpatient care.

Consider a single patient who is recovering from surgery and each day has the option of staying or being discharged. From historical data we can compute the probability of being discharged and conversely the probability of staying for each day of the patient’s LOS. Now consider a cohort of patients with similar discharge probabilities. From probability theory it is known that when multiple experiments have two (and only two) outcomes, then the probability for the number of experiments resulting in each outcome can be computed with a binomial distribution (assuming experiments are independent and identically distributed). Thus on any day and for each patient cohort, we can compute the number of discharges and consequently we know the number of patients who will remain until tomorrow. For a given MSS and using historic records, we can compute the number of admission to the ward (i.e. the number of completed inpatient surgeries each day) for each patient cohort. Thus for each patient cohort we have a distribution for the admission rate and using a binomial distribution we have a distribution for the discharge rate and we can easily compute the ward occupancy distribution. Finally, using discrete convolutions we can compute the overall ward occupancy. The formal model description follows.

Assume that each surgical specialty represents a single patient cohort. Let the MSS be defined such that b_{iq} is an operating theatre block where $i \in \{1, 2, \dots, I\}$ indexes the OTs and $q \in \{1, 2, \dots, Q\}$ the days in a cycle. Let each surgical specialty j be characterized by two parameters c^j and d_n^j , where c^j is a discrete distribution for the number of surgeries carried out in one OT block and d_n^j

the probability that a patient, who is still in the ward on day n , is to be discharged that day ($n = 0, 1, \dots, L^j$, where L^j denotes the maximum LOS for specialty j).

Using c^j and d_n^j as model inputs, for a given MSS the probability distribution for the number of recovering patients on each day q can be computed. The required number of beds is computed with the following three steps. *Step 1* computes the distribution of recovering patients from a single OT block of a specialty j ; i.e. we essentially pre-calculate the distribution of recovering patients expected from an OT block of a specialty. In *Step 2*, we consider a given MSS and use the result from *Step 1* to compute the distribution of recovering patients given a single cycle of the MSS. Finally in *Step 3* we incorporate recurring MSSs and compute the probability distribution of recovering patients on each day q .

Step 1: For each specialty j we use the binomial distribution to compute the number of beds required from the day of surgery $n = 1$ until $n = L^j$. Since we know the probability distribution for the number of patients having surgery (c_j), which equates to the number of beds needed on day $n = 0$, we can use the binomial distribution to iteratively compute the probability of needing beds on all days $n > 0$. Formally, the distribution for the number of recovering patients on day n is recursively computed by:

$$h_n^j(x) = \begin{cases} c^j(x) & \text{when } n = 0 \\ \sum_{k=x}^{c^j} \binom{k}{x} (d_{n-1}^j)^{k-x} (1 - d_{n-1}^j)^x h_{n-1}^j(k) & \text{otherwise.} \end{cases}$$

Step 2: We calculate for each OT block b_{iq} the impact this OT block has on the number of recovering patients in the hospital on days $q, q+1, \dots$. If j denotes the specialty assigned to OT block b_{iq} , then let $\bar{h}_m^{i,q}$ be the distribution for the number of recovering patients of OT block b_{iq} on day $m = 1, 2, \dots, Q, Q+1, \dots$. It follows that:

$$\bar{h}_m^{i,q} = \begin{cases} h_{m-q}^j & \text{if } q \leq m < L^j + q \\ \mathbf{0} & \text{otherwise} \end{cases}$$

where $\mathbf{0}$ means $\bar{h}_m^{i,q}(0) = 1$. Let H_m be a discrete distribution for the total number of recovering patients on day m resulting from a single MSS cycle. Since recovering patients do not interfere with each other we can simply iteratively add the distributions of all the OT blocks corresponding to the day m to get H_m . Adding two independent discrete distributions is done by discrete convolutions which we indicated by “*”. For example, let A and B be two independent discrete distributions. Then $C = A * B$, which is computed by:

$$C(x) = \sum_{k=0}^x A(k)B(x-k)$$

where τ is equal to the largest x value with a positive probability that can result from $A * B$ (e.g. if the maximum value of A is 3 and the maximum value of B is 4, then when convoluted the maximum value of the resulting distribution is 7, therefore in this example $\tau = 7$). Using this notation, H_m is computed by:

$$H_m(x) = \bar{h}_m^{1,1} * \bar{h}_m^{1,2} * \dots * \bar{h}_m^{1,Q} * \bar{h}_m^{2,1} * \dots * \bar{h}_m^{l,Q}$$

Step 3: We now consider a series of MSSs to compute the steady-state probability distribution of recovering patients. The cyclic structure of the MSS implies that patients receiving surgery during one cycle may overlap with patients from the next cycle. In the case of a small Q patients from many different cycles may overlap.

In Step 2 we have computed H_m for a single MSS in isolation. Let M be the last day where there is still a positive probability that a recovering patient is present in H_m . To calculate the overall distribution of recovering patients when the MSS is repeatedly executed we must take into account $[M/Q]$ consecutive MSSs. Let H_q^{SS} denote the probability distribution of recovering patients on day q of the MSS cycle, resulting from the consecutive MSSs. Since the MSS does not change from cycle to cycle, H_q^{SS} is the same for all MSS cycles. Such a result, where the probabilities of various states remain constant over time, is referred to as a steady-state result. Using discrete convolutions, H_q^{SS} is computed by:

$$H_q^{SS}(x) = H_q * H_{q+Q} * H_{q+2Q} * \dots * H_{q+[M/Q]Q}$$

From this result a number of workload metrics can be derived. To determine the demand for ward beds from the variable H_q^{SS} consider the following example. Let the staffing policy of the hospital be such that they staff for the 90th percentile of demand and let D_q denote the 90th percentile of demand on day q . It follows that D_q is also the number of staffed beds needed on day q , and is computed from H_q^{SS} as follows:

$$D_q = \max\{x | H_q^{SS}(x) \leq 0.9\}$$

In practice, patients tend to be segregated into different wards depending on the type of surgery they received. To incorporate this segregation into the model and to consequently have recovering patient distributions for each ward, a minor modification needs to be made to the model. Let W_k be the set of specialties j whose patients are admitted to ward k . Then in Step 2 we only have to consider those OT blocks assigned to a specialty in W_k and continue with the calculations.

Ward occupancy alone does not fully account for the workload associated with caring for recovering patients. During patient admissions and discharges the nursing workload can increase. From the model the probability distribution for daily admissions and discharges can be computed. To compute the admission rate,

set $d_1^j = 1$ for all j and repeat the steps above. The resulting H_q^{SS} will denote the admissions on day q .

The discharge rate is the rate at which patients leave the ward and can be computed by adding an additional calculation in Step 1. Let D_n^j be a discrete distribution for the number of discharges from specialty j on day n which is computed as follows:

$$\mathbb{P}(D_n^j = x) = \sum_{k=x}^{c^j} \binom{k}{x} (d_n^j)^x (1 - d_n^j)^{k-x} \mathbb{P}(h_n^j = k)$$

Finally, after computing D_n^j , one can set $h_n^j = D_n^j$ and continue with Step 2. By doing so, the resulting H_q^{SS} will denote the distribution for daily discharges for each day q of the MSS.

The inherent assumption of the described method is that all patients with a patient cohort have equal probability of being discharged and that it is independent of other patients, i.e. it is assumed that patients are identically distributed and independent. The independence assumption implies that the amount of time one patient is in the hospital does not influence the amount of time another patient is in the hospital. This seems like a natural assumption in most cases and appropriate so long as surgeries are rarely cancelled due to a bed shortage (cancellations due to bed shortages creates a dependency). The identically distributed requirement means that we must compute the number of beds needed tomorrow (and the number of case completed in one OT block), for all identically distributed cohorts of patients separately. In other words, the parameters of the binomial distribution must reflect all of the patients in a given cohort.

5.6.2 General results

The model was applied at the Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital (NKI-AVL) to support the design of a new MSS. Selected results from Vanberkel et al. (2011-2) are summarized in this subsection.

Management at NKI-AVL strive to staff enough beds such that for 90% of the week days there is sufficient coverage. This implies that on 10% of the days they will be required to call in additional staff. Using the model a number of MSS proposals were evaluated and eventually staff choose an MSS that the model predicted would lead to a balanced ward occupancy.

An unbalanced ward occupancy makes staff scheduling, and ward operations, difficult. Early in the week, beds would be underutilized whereas later in the week, beds would become highly utilized and the risk of a shortage would increase. Such peaks and valleys represent variation in the system which possibly could be eliminated with a different MSS. This variation leads to significant problems, particularly as the wards approach peak capacity. For example, when inpa-

tient wards reach their peak capacity and a patient admission is pending, staff often scramble to try and make a bed available. If one cannot be made available, additional staff are called in (or in rare cases when additional staff cannot be found, the elective surgery is cancelled), which causes extra work for OR planners, wasted time for surgeons and extra anxiety for patients. When a bed was made available, it often means a patient was transferred from one ward to another (often to a ward capable of caring for the patient but not the designated one) or discharged. Either way, extra work is required by ward staff and there is a disruption in patient care. Although completely eliminating such problems is likely not possible without an exorbitant amount of resources, sound planning ahead of time may help to minimize occurrences.

After implementing the new MSS, the ward occupancy was observed over a 33 week period. From these observations, probability distributions of beds used for each day of the MSS cycle were derived. Using Chi-square goodness-of-fit tests, these observed distributions were compared to those projected by the model. Six of the seven projected distributions (one for each day of the MSS cycle) were found to be a good fit for the observed data at a level $\alpha=0.05$, while for the seventh day, this was true at a level $\alpha=0.2$. Figure 5.8 compares the projected ward occupancy with the observed ward occupancy during the 33 week period.

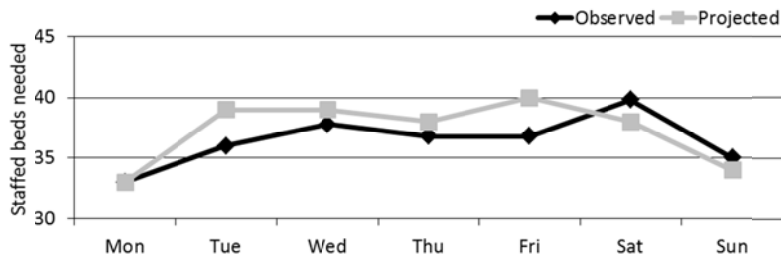


Fig. 5.8 Comparison of the projected and observed (90th percentile) ward occupancies

5.6.3 Discussion

The main benefit of using the model was to be able to quantify the concerns of ward staff, thereby providing a platform from which they could begin to negotiate a solution. Staff was quick to embrace the model output, particularly after seeing several modifications to the original MSS, at which point they were able to roughly predict model output for a given modification.

In this project we treated the equipment and physician schedule restrictions as unchangeable. It is possible that further improvements in the ward oc-

cupancy could have been achieved if these restrictions were relaxed. In this way the model can be used to illustrate the benefits of buying an extra piece of equipment or of changing physicians' schedules. An additional restriction, which if relaxed may have allowed further improvements, is the assignment of wards to surgical specialties. In other words, in addition to changing when a specialty operates, it may prove advantageous to change which ward the patients are admitted to after surgery. Finally, we chose the best MSS from those created through swapping OR block and surgical specialty assignments. It is possible that a search heuristic may have found a better MSS, although it would have required the many surgical department restrictions to be modeled and the more complex model may not have garnered the same level of staff understanding and support.

Oostrum et al. (2008-2) propose another approach, where the MSS is planned in more detail: here it comprises of a cyclical schedule of frequently occurring elective surgery *types*. The resulting combinatorial optimization problem is to determine a MSS that balances OT utilization and ward occupancy. By scheduling surgery types, the surgeon/surgical specialty can assign a patient's name at a later time, without affecting the performance of the MSS. The model considers stochastic OT capacity constraints and empirical LOS distributions. As the resulting problem is NP-hard, heuristics are provided. For a review on the suitability and managerial implication of this particular MSS approach see Oostrum et al. (2010).

5.7 Operational: elective surgery scheduling and sequencing

Operational planning and scheduling of operating theatres is arguably one of the most popular topics in the healthcare operations research literature. The literature reviews of Cardoen et al. (2010) and Guerriero and Guido (2011) outline many contributions regarding the elective surgery scheduling and sequencing literature. Cardoen et al. (2010-2) also propose a classification scheme for operating theatre planning and scheduling problems, which contains four descriptive fields $\langle \alpha | \beta | \gamma | \delta \rangle$. Here, α holds the patient characteristics, β the delineation of the decision, γ the extent to which uncertainty is incorporated, and δ the performance measures.

In previous work (Hans et al. 2008) we demonstrated that by combining advanced optimization techniques with extensive historical statistical records on surgery durations, the OT department utilization can be improved significantly. We demonstrated that, if slack time is reserved in OTs according to the method described in Section 5.3.1 (particularly equation 5.6), the portfolio effect can be exploited in a local search meta-heuristic as follows. By swapping surgeries between OTs (1-swap or 2-swap), the total slack time of both involved OTs is affected. By clustering surgeries with similar duration variability characteristics, the total slack time is reduced due to the portfolio effect. This principle can be used in a local search heuristic to minimize the total slack time, and thus free OT time. A

result of the portfolio optimization is that the fragmentation of the free OT time is minimized. In fact, OTs in resulting solutions are either filled to a great extent with surgery and slack time, or are empty. As a result, OTs can be closed, or time is freed to perform more surgeries.

In this section we discuss the optimization of the elective surgery schedule, in order to minimize emergency surgery waiting time. This problem follows from policy 2 outlined in Section 5.5 (i.e., emergency surgeries are dealt with in elective OTs).

5.7.1 General problem description

Emergency surgery waiting time increases a patient's risk of postoperative complications and morbidity. When dealing with emergency patients according to policy 2 (Section 5.5), waiting time will occur when all elective OTs are busy. Typically, at the beginning of the regular working day all OTs will be busy with long procedures, as surgeries are often scheduled according to the LPT rule. As a result, emergency surgeries that arrive just after the start of the elective program may have to wait a long time, as surgeries cannot be preempted. This pleads for scheduling a short surgery at the beginning of the day, to obtain a so-called "Break-in-Moment" (BIM), at an early time for emergency surgeries. Extending on this idea, we may sequence the elective surgeries within their assigned OTs in such a way, that their expected completion times, which are BIMs for emergency surgery, are spread as evenly as possible. We do not re-assign surgeries to other OTs, but instead only re-sequence elective surgeries within their assigned OT. This is illustrated in Figure 5.9: the BIMs are clearly spread more evenly after re-sequencing the surgeries.

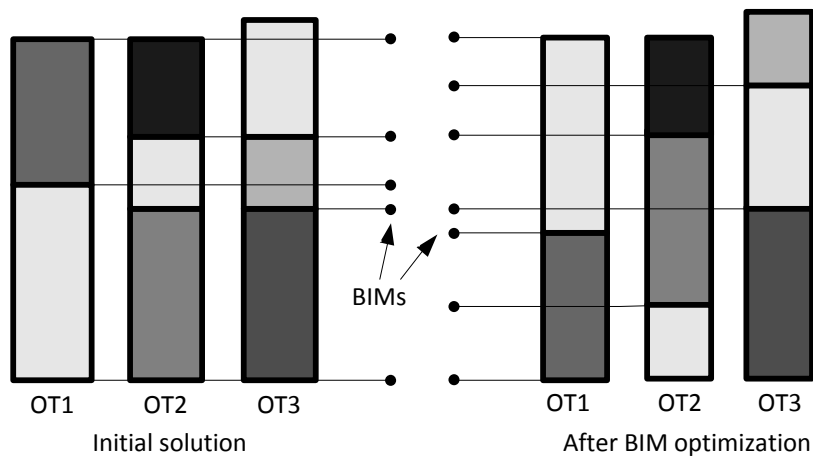


Fig. 5.9 Example BIM optimization

The problem of sequencing elective surgeries in such a way that the BIMs are spread as evenly as possible (or, alternatively, the break-in-intervals/BIIs are minimized) is in fact a new type of scheduling problem. This innovative idea was a result of a MSc thesis project at Erasmus MC (Lans et al. 2006), where it was proven that the problem is NP-hard by reduction from 3-partition.

We assume, as illustrated in Figure 5.9, that surgeries are executed directly after another, i.e., there is no planned slack between surgeries. The planning horizon is within a day, and starts on the first moment when all OTs are scheduled to have elective surgeries. If all OTs start at the same time, then this time marks the start of the planning horizon. It ends on the first moment when there is an OT without a scheduled surgery, since after this moment there are infinitely many BIMs. The objective is to lexicographically minimize the largest break-in-intervals (BII). In other words, we minimize the largest BII, then the second largest (without affecting the largest), etc. The reason that we do not only minimize the largest BII is that the expected duration of the shortest surgery is a lower bound to the longest BII. This can be seen as follows: assuming all OTs start at the same time, placing the shortest surgery at the beginning of its OT gives a BII that cannot be shortened.

In forthcoming work we will propose various exact and heuristic approaches for the BIM/BII optimization problem. Here we give the results of a Simulated Annealing (SA) local search heuristic, which iteratively swaps surgeries within their sequence. The SA method uses the following parameters: start temperature 0.2, final temperature 0.0001, Markov chain length 150, decrease factor 0.8. We fix the shortest surgery at the beginning of its OT.

5.7.2 General results

We generate instances with the case mix of Section 5.5 (academic hospital Erasmus MC), scaled to fill 4, 8 or 12 OTs. Surgeries are scheduled “First Fit” (Hans et al. 2008). First Fit assigns surgeries from the top of the list to the first available OT plus an amount of slack (Section 5.3.1, equation 5.6) to achieve a 30% probability of overtime caused by surgery duration variability, until no surgery can be found anymore that fits in the remaining OT capacity. Each instance has two variants: with *full* flexibility (all surgery sequences are allowed), and with *reduced* flexibility (randomly, $\frac{1}{12}$ of the first surgeries are fixed on this position in their OT, and $\frac{1}{12}$ of the last surgeries are fixed on this position in their OT). For example, surgeries on children are typically done first, and surgeries after which extensive OT cleaning is required are typically done last.

Table 5.3 presents the results for the SA algorithm. It compares the solutions found by SA to the initial First Fit solution (which does not aim to optimize BIM/BII). Particularly, it shows the frequency of the BII of size >15, >30, ...,

>90 minutes. Each number is an average over 260 instances (52 weeks of 5 working days). SA solves each instance in less than 2 seconds. Clearly, the large intervals are eliminated to a great extent, the more so when there are more OTs (and thus more BIMs).

Table 5.3 Avg. frequency of break-in-interval size (initial solution → SA solution; 260 instances per parameter setting)

# OTs	Reduced flexibility	>90 min.	>75 min.	>60 min.	>45 min.	>30 min.	>15 min.
4	No	1.01→0.29	1.51→0.67	2.01→1.50	2.72→2.84	4.09→5.52	5.51→7.11
		-71.3%	-55.6%	-25.4%	4.4%	35%	29%
4	Yes	1.01→0.30	1.51→0.74	2.00→1.59	2.72→2.85	4.10→5.47	5.55→7.01
		-70.3%	-51%	-20.5%	4.8%	33.4%	26.3%
8	No	0.48→0.00	0.82→0.01	1.21→0.09	1.97→0.46	3.84→3.75	6.89→10.22
		-100%	-98.8%	-92.6%	-76.6%	-2.3%	48.3%
8	Yes	0.47→0.00	0.82→0.02	1.23→0.11	1.94→0.56	3.82→3.91	6.88→10.02
		-100%	-97.6%	-91.1%	-71.1%	2.4%	45.6%
12	No	0.33→0.00	0.69→0.00	0.95→0.02	1.47→0.11	3.14→1.35	6.97→10.49
		-100%	-100%	-97.9%	-92.5%	-57%	50.5%
12	Yes	0.36→0.00	0.70→0.00	0.95→0.02	1.46→0.13	3.15→1.58	6.92→10.26
		-100%	-100%	-97.9%	-91.1%	-49.8%	48.3%

The question now is what impact these optimized BIMs/BIIs have on emergency surgery waiting time, particularly given the fact that elective surgery durations are stochastic, and the BIMs are *expected* surgery completion times. Table 5.4 presents the results of a Monte Carlo simulation of 260 instances with 12 OTs and *reduced* sequencing flexibility. The elective surgeries are assumed to have a lognormal distribution. The emergency surgeries arrive according to a Poisson process (on average 5.1 arrivals per day), and are served on a FCFS basis. Elective surgeries are not preempted.

Table 5.4 Waiting time for the 1st, 2nd and 3rd arriving emergency patients (12 OTs, run length 780 days, max. relative error 10%, min. confidence level 90%)

Waiting time (minutes)	1 st emergency surgery		2 nd emergency surgery		3 rd emergency surgery	
	Initial solution	SA solution	Initial solution	SA solution	Initial solution	SA solution
< 10	28.8%	48.6%	34.9%	44.9%	40.4%	46.2%
< 20	53.0%	75.8%	56.9%	73.6%	63.0%	69.8%
< 30	70.5%	90.9%	71.8%	87.2%	76.3%	86.7%

We observe that the BIM/BII optimization by SA, despite the reduced flexibility, has a significant impact on emergency surgery waiting. For example, the relative number of first emergency patients that wait at most 10 minutes increases by 69% from 28.8% to 48.6%. The improvement decreases with every next arriving emergency patient of the day. This may be expected, as these emergency patients increasingly distort the original schedule.

5.7.3 Discussion

BIM/BII optimization has a big impact on emergency surgery waiting. More research is required into exact solution approaches, and perhaps applications of BII/BIM optimization in other sectors. For healthcare, it is easy to implement: it only requires re-sequencing of elective surgeries. As a first step, managers are advised to plan the shortest surgery at the beginning of the regular working day.

5.8 Future Directions

The operating theatre department offers challenging planning and control problems on all hierarchical levels of control. While operational planning and control has received a lot of attention from the OR/OM in healthcare research community, tactical planning is less exposed, and research has had less of an impact in practice due to its inherent complexity. In our experience, decision support software tools mostly focus on the operational planning level, whereas tools for the tactical planning level are scarce and are too simplified or limited in scope to deal with tactical decision making. Future research therefore has to focus on the tactical level, to a greater extent. This raises opportunities to expand the scope beyond the operating theatre department. From our survey of healthcare models that encompass multiple departments we concluded that researchers often model hospitals in a way that reflects the limited/departmental view of healthcare managers (Vanberkel et al. 2010). The research scope should particularly include the polyclinics, where new patients are taken in, and the wards, which are typically managed to follow the OT department but whose workloads may be leveled significantly by tactically optimizing the OT's master surgery schedule. Ultimately, we should aim to optimize the entire patient care pathway.

5.9 References

Anthony RN (1965) Planning and control systems: a framework for analysis. Harvard Business School Division of Research, Boston

- Bhattacharyya T, Vrahas MS, Morrison SM, Kim E, Wiklund RA, Smith RM, Rubash HE, The Value of the Dedicated Orthopaedic Trauma Operating Room. *J Trauma* 60:1336–1341
- Blake JT, Donald J (2002) Using Integer Programming to Allocate Operating Room Time at Mount Sinai Hospital. *Interfaces* 32(2):63–73
- Brailsford SC, Harper PR, Patel B, Pitt M (2009) An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation* 3:130–140
- Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. *European Journal of Operational Research* 201:921–932
- Cardoen B, Demeulemeester E, Beliën J (2010-2) Operating room planning and scheduling problems: a classification scheme. *International Journal of Health Management and Information* 1(1):71–83
- Denton BT, Miller A, Balasubramanian H, Huschka T (2010) Optimal Allocation of Surgery Blocks to Operating Rooms Under Uncertainty. *Operations Research* 58(4):802–816
- Denton BT, Viapiano J, Vogl A (2007) Stochastic Optimization of Surgery Sequencing and Start Time Scheduling Decisions. *Health Care Management Science* 10(1):13–24
- Dexter F, Macario A, Traub RD (1999) Which algorithm for scheduling add-on elective cases maximizes operating room utilization? *Anesthesiology* 91:1491–1500
- Dexter F (2011) Website: http://www.franklindexter.com/bibliography_TOC.htm
- Dexter F, O'Neill L (2001) Weekend operating room on-call staffing requirements. *AORN Journal* 74:666–671
- Dexter F, Epstein RH, Traub RD, Xiao Y (2004) Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. *Anesthesiology* 101:1444–1453
- Ferrand Y, Magazine M, Rao U (2010) Comparing two operating-room-allocation policies for elective and emergency surgeries. *Proceedings of the 2010 Winter Simulation Conference*. Johansson B, Jain S, Montoya-Torres J, Hagan J, Yücesan E (eds)
- Guerrero F, Guido R (2011) Operational research in the management of the operating theatre: a survey. *Health Care Management Science* 14:89–114
- HFMA (2005) Achieving operating room efficiency through process integration. Technical report, Health Care Financial Management Association
- Hall R (2006) *Patient Flow: Reducing Delay in Healthcare Delivery*, Springer
- Hans EW, Wullink G, Houdenhoven M van, Kazemier G (2008) Robust Surgery Loading. *European Journal of Operational Research* 185:1038–1050
- Hans EW, Houdenhoven M van, Hulshof PJH (2011) A framework for health care planning and control. Memorandum 1938, Department of Applied Mathematics, University of Twente, Enschede
- Harper PR and Shahani AK (2002) Modelling for the Planning and Management of Bed Capacities in Hospitals. *Journal of the Operational Research Society* 53(1):11–18
- Houdenhoven M van, Hans EW, Klein J, Wullink G, 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
- Lans M van der, Hans EW, Hurink JL, Wullink G (2006) Anticipating Urgent Surgery in Operating Room Departments. BETA working paper WP-158, ISSN: 1386–9213, University of Twente, 2006
- McIntosh C, Dexter F, Epstein RH (2006) The impact of service specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an Australian hospital. *Anesthesia & Analgesia* 103:1499–1516
- Oostrum JM van, Houdenhoven M van, Vrielink MMJ, Klein J, Hans EW, Klimek M, Wullink G, Steyerberg EW, Kazemier G (2008) A Simulation Model for Determining the Optimal Size of Emergency Teams on Call in the Operating Room at Night. *Anesthesia & Analgesia* 107:1655–1662
- Oostrum JM van, Bredenhoff E, Hans EW (2010) Suitability and managerial implications of a Master Surgical Scheduling approach. *Annals of Operations Research* 178(1):91–104

- Oostrum JM van, Houdenhoven M van, Hurink JL, Hans EW, Wullink G, Kazemier G (2008-2) A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, 30(2):355–374
- Vanberkel PT, Boucherie RJ, Hans EW, Hurink JL, Lent WAM van, Harten WH van (2011) An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society* (forthcoming)
- Vanberkel PT, Boucherie RJ, Hans EW, Hurink JL, Lent WAM van, Harten WH van (2011-2) Accounting for Inpatient Wards when developing Master Surgical Schedules. *Anesthesia & Analgesia* (forthcoming)
- Vanberkel PT, Blake JT (2007) A Comprehensive Simulation for Wait Time Reduction and Capacity Planning Applied in General Surgery. *Health Care Management Science*, 10 (4) 373–385, Special Issue on Simulation in Health Care. Anderson JG, Merode GG van (eds)
- Vanberkel PT, Boucherie RJ, Hans EW, Hurink JL, Litvak N (2010) A Survey of Health Care Models that Encompass Multiple Departments. *International Journal of Health Management and Information* 1(1):37–69
- Visser JMH, Bertrand JWM, De Vries G (2001) A framework for production control in healthcare organizations. *Production Planning & Control* 12:591–604
- Wachtel RE, Dexter F (2008) Tactical increases in operating room block time for capacity planning should not be based on utilization. *Anesthesia & Analgesia* 106(1):215–22
- Wixted JJ, Reed M, Eskander MS, Millar B, Anderson RC, Bagchi K, Kaur S, Franklin P, Leclair W (2008) The Effect of an Orthopedic Trauma Room on After-Hours Surgery at a Level One Trauma Center. *J Orthop Trauma* 22(4):234–236
- Wullink G, Houdenhoven M van, Hans EW, Oostrum JM van, Lans M van der, Kazemier G (2007) Closing emergency operating rooms improves efficiency. *Journal of Medical Systems* 31(6):543–546
- Zijm WHM (2000) Towards intelligent manufacturing planning and control systems. *OR Spectrum* 22(3):313–345