

Implementing Algorithms to Reduce Ward Occupancy Fluctuation Through Advanced Planning



Peter T. Vanberkel, Richard J. Boucherie, Erwin W. Hans, Johann L. Hurink, Wineke A. M. van Lent, and Wim H. van Harten

Abstract As with many hospitals, NKI-AVL is eager to improve patient access through intelligent capacity investments. To this end, the hospital expanded its operating capacity from five to six operating rooms (ORs) and redesigned their master surgical schedule (MSS) in an effort to improve utilization and decrease hospital-wide congestion; an MSS is a cyclical schedule specifying when surgical specialties operate. Designing an efficient MSS is a complex task, requiring commitment and concessions on the part of competing stakeholders. There are many restrictions which need to be adhered to, including limited specialized equipment and physician availability. These restrictions are, for the most part, known in advance. The relationship between the MSS and the ward, however, is not known in advance and is plagued with uncertainties. For example, it may be known which patient type will be admitted to the ward after surgery; however, the number of patients changes from week to week, and it is not known with certainty how long each patient will stay in the hospital. Inpatient wards, furthermore, are one of the most expensive hospital resources and can be a major source of hospital congestion, as many departments rely on inpatient wards to receive and treat their patients prior to discharge from the hospital (e.g., the emergency department). This congestion leads to long waiting times for patients, patients receiving the wrong level of care, and extended lengths of stay for patients. Well-designed surgical schedules which take into account inpatient ward resources lead to reduced cancellations and higher and balanced utilization. We observed that peaks in the ward occupancy

P. T. Vanberkel (✉)

Department of Industrial Engineering, Dalhousie University, Halifax, NS, Canada
e-mail: peter.vanberkel@dal.ca

R. J. Boucherie · E. W. Hans · J. L. Hurink

Center for Healthcare Operations Improvement and Research, University of Twente, Enschede, The Netherlands

W. A. M. van Lent · W. H. van Harten

Netherlands Cancer Institute – Antoni van Leeuwenhoek (NKI-AVL) Hospital, Amsterdam, The Netherlands

© Springer Nature Switzerland AG 2021

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_8

129

are particularly dependent on the MSS, and, as a result, ward occupancies can be leveled through careful MSS design. Avoiding peaks and leveling ward occupancy across weekdays makes staff scheduling easier and limits the risk of exceeding capacity, which causes congestion and perpetuates inefficiencies throughout the hospital. Working with NKI-AVL we developed an operations research model to support the redesign of their MSS. The redesigned MSS improved the use of existing ward resources, thereby allowing an additional operating room to be built without additional investments in ward capacity. A post implementation review of bed-use statistics validated our model's projections. The success of the project served as proof-of-concept for our model, which has since been applied in several other hospitals.

1 Introduction

Driven by an aging population, public opinion, increased health expenditures, and long waiting lists, a flood of changes in the health-care system have been set into motion. Health care constitutes the largest industry in many developed countries [9], and managing it is a complex task due to its importance to society and the often politically charged atmosphere within which it exists. Furthermore, the nature of health-care delivery does not allow the direct copying of success stories from the manufacturing industry, where logistical optimization has a long history. Health-care processes and supply chains show considerable differences, such as a high degree of uncertainty, the medical autonomy of clinicians, and the fact that patients cannot be treated as products.

As this research began in the Netherlands, we begin by discussing the state of the Dutch health-care system. Like most countries, the Dutch health-care system has struggled with poor quality and wasteful expenditures. This came to a head in 2006, when the country passed a new Health Insurance Act. The Act reformed the structure of health-care delivery with the intent of using competition to breed efficiencies and improve value for money. To ensure that all Dutch citizens have the same basic level of health insurance regardless of ability to pay, a number of regulatory elements were introduced. Significant competition was created at two different levels. Competition exists between health insurance companies, which vie for enrollees, and also health-care providers (new and existing), which vie for contracts with health insurance companies. Insurance companies compete mainly by offering extended coverage packages (e.g., additional dentistry, eyewear, cosmetic surgery, alternative medicine, etc.) at lower prices. Health-care providers compete mainly on the remuneration amounts (paid by insurers to providers) and quality of care (e.g., access times, treatment options) [35].

The crucial underpinning of this system is to use competitive markets and insurance companies to increase performance and create a more cost-effective health-care system. The extent to which this has worked can be debated; however, the concept has been generally lauded [26]. It has been our experience that this new

competition has applied significant pressure on health-care providers, which has, as a result, significantly increased the use of (and demand for) operations research. Financial reforms are just one example of the many efforts by developed nations to eliminate poor quality and wasteful expenditures in health care. Perhaps not surprising, given that value for money was a guiding mandate in the reforms, they have acted as a catalyst for making operations research commonplace in many Dutch hospitals. This has led to an enormous increase in research questions motivated by health-care providers. Research results are influencing national health-care policies and changing the way health care is delivered across the country. Although the financial reforms are unique to the Netherlands, the operations research it has motivated has broader appeal and can support improvement efforts around the world.

In this chapter we discuss a project motivated by a Dutch hospital which has broad applicability. The challenges and opportunities related to surgical scheduling are discussed below, but for now it is sufficient to say that it is a topic of significant academic study (see reviews [5] from 1997 and [8] from 2009) and is a challenge faced by health-care providers in many parts of the world. There has been limited reported success in terms of implementation and impact on health-care operations [8], and hence there is a need to develop solutions which can be readily applied and generalized for applicability in various hospital settings.

The reported research project evolved from an operations research model to an application at the Netherlands Cancer Institute-Antoni van Leeuwenhoek (NKI-AVL) to a standard practice. The structure of the chapter reflects this. The formal model is discussed in Sect. 2, immediately following the problem description (Sect. 1.1). The application is discussed in Sect. 3 and includes the post implementation analysis (Sect. 3.1.1) used to validate the model. Finally Sect. 3.2 provides concluding remarks and briefly discusses applications of the model at other hospitals. Much of this research has been reported before in [14, 30, 32–34]. Accordingly, the intent of this chapter is to summarize this research with a specific focus on how it has impacted NKI-AVL and generally how it has impacted other hospitals.

1.1 Problem Description

No other single hospital department influences the workload of the hospital more than the Department of Surgery, in particular the activities in the operating room (OR) [20]. As such, its activities (or lack thereof) cause a ripple effect throughout the hospital. Upstream processes are less sensitive to changes, as there is often a waiting list for surgical operations which acts as a buffer, dampening the effect. For downstream processes, this is different as a buffer of post-surgery patients waiting to be admitted to a ward cannot exist. Since post-surgical activities are sensitive to the activities in the OR, it is important to derive one in terms of the other. As described in this section, the workload for downstream departments can be modelled as a function of the master surgical schedule (MSS).

For a hospital, the OR accounts for more than 40% of its revenues and a similarly large portion of its costs [14]. An efficient OR thus significantly contributes to an efficient health-care delivery system as a whole. The planning and scheduling of OR time is discussed by many authors [2, 5, 6, 8, 17, 36] and is often described as a multiple stage process.

The multiple stage process used by many hospitals starts with the long-term allocation of OR time to surgical specialties, e.g., the number of surgery hours per year. This allocation, referred to as Stage 1, is a strategic decision that reflects patient demand patterns and the priorities defined by hospital management. From this strategic decision, an MSS is developed for a shorter time horizon which divides OR time (aggregated into blocks) among the specialties, known as Stage 2. The specific assignment of patients to OR blocks within the MSS is commonly referred to as Stage 3. A fourth stage “addresses the monitoring and control of the OR activities” [23] on the day of surgery. In this chapter, we focus on the development of an MSS in Stage 2.

The MSS is often specialty specific [2], meaning OR time is dedicated to a surgical specialty. In these MSSs, the decision of which patients (and consequently which surgeries) to schedule within each OR block is determined by the surgical specialty through consultation with the OR manager. Other MSSs are more specific with OR blocks being allotted to specific surgical procedures [18, 24]. Instead of using the term MSS, other authors refer to the distribution of OR time among surgical specialties as a surgical block schedule [25] and a timetable of operations [16].

The development of an MSS is often a complex balancing act. Since the OR is one of the hospital’s most expensive resources, hospitals wish to maximize its performance through high resource utilization, minimal overtime, minimal cancellations, and the elimination of conflicting equipment needs between rooms. Many authors describe methods for developing the MSS that take into account various resources within the OR such as staff, equipment, and instrument trays. For a review see [8]. Furthermore, the OR is often described as the engine that drives the hospital [20], implying many other departments are impacted by the MSS. The effect of the MSS on ward occupancy [1–4, 10, 11, 16, 18, 19, 24, 28, 29, 31], critical care resources [3, 10, 12, 18, 22, 24], and waiting lists [25, 31] has notably been studied. Three of these papers represent the relationship with deterministic models, while the remaining consider at least one variable as stochastic. The stochastic models are either simulation models, mathematical programming models, or queueing theory models.

The analytical model presented in this chapter most closely resembles a queueing model. The model quantifies the effect of an MSS on admission/discharge rates, ward occupancy rates, and the workload of all departments treating inpatients. The robustness of this model and, as illustrated later, its ease of implementation are the main contributions of the model to surgical scheduling literature.

Using our model, downstream workload distributions can be computed as a function of the MSS for all departments that provide care for recovering surgical patients. Specifically, the model computes the ward occupancy distributions, the

patient admission/discharge distributions, and the distributions for the ongoing interventions/treatments required by recovering patients. Furthermore, the cumulative influence of multiple MSS cycles is considered. Since the MSS is identical from cycle to cycle, the overlapping of patients from one cycle to the next can be anticipated. A single MSS design is expected to remain in place for a long period of time leading to “steady-state” workload distributions for each day of the MSS cycle.

2 Methodology

This section describes a model to determine the workload placed on hospital departments by recovering surgical patients. In the same way an MSS describes resource demands within the OR, we show how the resources of other departments can be seen as a function of the MSS. The method relies only on data which are easily extractable from typical patient management systems.

The model is most easily described from a queueing theory perspective. The basic component of the model is a single OR block and its expected impact on the arrival rate to the hospital wards. The number of cases scheduled in such a block varies per specialty and is modelled as a specialty-specific random variable. This variable also represents the number of patients arriving to the ward (batch size). At the ward, each patient directly occupies a bed for a certain period of time. In the queueing model, the ward is seen as an infinite server system in which the patients occupy a server (ward bed) without delay. The time spent occupying a bed (length of stay, LOS) is the service time, which is modelled as a random variable. Again, this random variable is specific to the surgical specialty. Since patients occupying a server do not interfere with each other during their recovery, the aggregate number of patients for all OR blocks can be computed by adding the individual effects of all OR blocks. Finally, since the MSS is cyclical, the cumulative number of patients from recurring MSS cycles can be computed.

The main output of the model is the distribution for the number of patients anticipated in the system on each day of the MSS. The model used for these calculations is explained in the following subsection. The three subsequent subsections explain how the model can be modified to obtain the distributions for (1) ward occupancies, (2) admissions/discharges, and (3) the number of patients in a specific day of their recovery. The time scale in the model is days; therefore, all metrics are considered on a daily basis.

2.1 Model Inputs

An 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 ORs can 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 OR block $b_{i,q}$ where $i \in \{1, 2, \dots, I\}$. Using this notation, an empty MSS (i.e., before specialties have been assigned to OR blocks) is shown in Fig. 1 where each cell represents an OR block. It is possible for multiple blocks to be assigned to a single specialty on the same day.

The way specialty j fills in an OR block is described by two parameters, c^j and d_n^j . Parameter c^j is a discrete distribution for the number of surgeries carried out in one block, i.e., $\mathbb{P}(c^j = k)$ is the probability of k surgeries, $k \in \{0, 1, \dots, C^j\}$, where C^j is the maximum number of surgeries of specialty j that can be completed in one block. Specialties independently decide which patients to schedule during each block, meaning that the number of surgeries completed in one block does not influence the number of surgeries completed in another. The second parameter d_n^j is the probability that a patient, who is still in the ward on day n , is to be discharged that day ($n \in \{0, 1, \dots, L^j\}$, where L^j is the maximum LOS for specialty j ; a finite LOS is used for numerical purposes). Note that d_0^j is the probability that the patient is discharged on the same day as surgery (i.e., an outpatient surgery or day-case surgery) and $d_{L^j}^j = 1$. The parameter d_n^j is computed by dividing the probability that a patient's total stay is exactly n days by the probability that the patient was not yet discharged before day n . Let $P^j(n)$ be the probability that the LOS of a patient from specialty j is exactly n days long; then formally d_n^j is computed as follows:

$$d_n^j = \frac{P^j(n)}{\sum_{k=n}^{L^j} P^j(k)}. \tag{1}$$

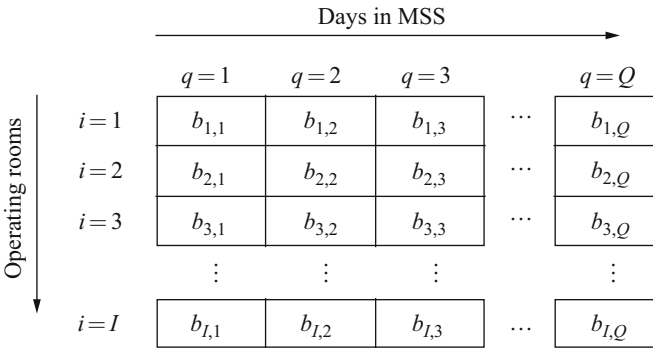


Fig. 1 Empty MSS illustrating model notation

2.2 Recovering Patients in the Hospital

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 is computed in three steps. Step 1 computes the distribution of recovering patients from a single OR block of a specialty j , i.e., we essentially pre-calculate the distribution of recovering patients expected from an OR 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 of the MSS.

Step 1: Distribution of recovering patients from specialty j following from a single OR block In Step 1 we ignore the MSS and consider a single specialty j operating in a single OR block. The patient flow process is as follows: During the OR block patients receive surgery. The number of patients who undergo surgery in one OR block is given by the random variable c^j . After surgery each patient still on the ward on day n has the probability d_n^j of being discharged that day. In the following, we compute the probability $\mathbb{P}(h_n^j = x)$ that n days after scheduling a block of specialty j , x patients of the block are still in recovery. Note that $n \in \{0, 1, \dots, L^j\}$ and $x \in \{0, 1, \dots, C^j\}$ and that, for example, $\mathbb{P}(h_3^j = 5) = 0.25$ means that 3 days after surgery, there is a 25% chance that five patients are still recovering in the hospital.

Day $n = 0$ is defined as the day of surgery, and it is assumed that patients occupy a bed all day on the day of surgery even though they may physically be in the OR. This is consistent with practice where patients have a recovery bed reserved for them before surgery. As such, the number of patients in recovery from specialty j on day $n = 0$ is by definition the number of surgeries performed that day by specialty j . It follows that the distribution for the number of recovering patients on day $n = 0$ is $h_0^j = c^j$.

Note that on day n , each patient still in the hospital has a probability d_n^j of being discharged that day and $(1 - d_n^j)$ of staying. If there are k patients in recovery on day n , then the probability of s patients in recovery (where $s \leq k$) on day $n + 1$ is computed using the binomial distribution, $\binom{k}{s} (d_n^j)^{k-s} (1 - d_n^j)^s$. Since we know the probability distribution for the number of patients at the end of day $n = 0$, we can iteratively use this formula to compute the probability of k patients at the end of all days $n > 0$. Summarizing, the distribution for the number of recovering patients on day n is recursively computed by

$$\mathbb{P}(h_n^j = x) = \begin{cases} \mathbb{P}(c^j = x) & \text{when } n = 0 \\ \sum_{k=x}^{c^j} \binom{k}{x} (d_{n-1}^j)^{k-x} (1 - d_{n-1}^j)^x \mathbb{P}(h_{n-1}^j = k) & \text{otherwise.} \end{cases} \quad (2)$$

Step 2: Aggregate distribution of recovering patients following from a single MSS cycle In this step we consider the previously computed probability distribution h_n^j and a given MSS as input. Although the MSS is cyclical and repeats after Q days, in this subsection, we consider only a single MSS cycle in isolation. The MSS defines when each specialty is assigned an OR block and thus the days on which patients of specialty j arrive to the ward. Based on these, we compute the total number of patients in recovery by means of discrete convolutions.

To calculate the overall distribution of recovering patients, we first identify for each block $b_{i,q}$ the impact that this block has on the number of recovering patients in the hospital on days $(q, q + 1, \dots)$. If z denotes the specialty assigned to block $b_{i,q}$ which follows from the MSS, then the distribution $\bar{h}_m^{i,q}$ for the number of recovering patients of block $b_{i,q}$ on day m ($m \in \{1, 2, \dots, Q, Q + 1, Q + 2, \dots\}$) is given by

$$\bar{h}_m^{i,q} = \begin{cases} h_{m-q}^z & \text{if } q \leq m < L^z + q, \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{3}$$

where $\mathbf{0}$ means $\mathbb{P}(\bar{h}_m^{i,q} > 0) = 0$. Note that specialties index j is no longer needed as specialties are accounted for by their designated OR block(s).

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 blocks impacting day m to get H_m . Adding two independent discrete distributions is done by discrete convolutions which we indicate by $*$. Let A and B be two independent discrete distributions. Then $C = A * B$ is computed by

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

where τ is equal to the largest x value with a positive probability that can result from $A * B$. Using this notation, H_m is computed by

$$H_m = \bar{h}_m^{1,1} * \bar{h}_m^{1,2} * \dots * \bar{h}_m^{1,Q} * \bar{h}_m^{2,1} * \dots * \bar{h}_m^{I,Q}. \tag{4}$$

Step 3: Steady-state distribution of recovering patients In Step 3 we 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 , for example, patients from many different cycles can overlap.

In Step 2 we have computed H_m for a single cycle of the MSS in isolation. Let M be the last day where there is still a positive probability that a recovering patient is present as computed by H_m . Thus $M = \max_j \{L^j + x^j\}$ (where x^j is the latest day q of a block assigned to specialty j) indicates the range of the MSS. To calculate the overall distribution of recovering patients when the MSS is repeatedly

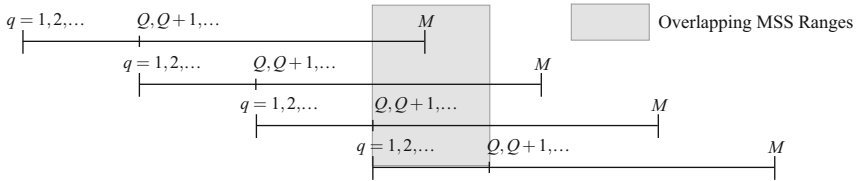


Fig. 2 Consecutive MSS cycles illustrating overlapping recovering patients

executed, we must take into account $\lceil M/Q \rceil$ consecutive cycles of the MSS (see Fig. 2). Let H_q^{SS} denote the probability distribution of recovering patients on day q of the MSS cycle, resulting from $\lceil M/Q \rceil$ consecutive MSS cycles. Since the MSS does not change from cycle to cycle, H_q^{SS} is the same for all MSS cycles. Using discrete convolutions, H_q^{SS} is computed by

$$H_q^{SS} = H_q * H_{q+Q} * H_{q+2Q} * \dots * H_{q+\lceil M/Q \rceil Q}. \tag{5}$$

The relationship between the distribution H_q^{SS} and the workload associated with recovering patients is discussed in detail in the following three subsections.

2.3 Ward Occupancy

Perhaps the most common measure of inpatient workload is ward occupancy. Ward occupancy, the distribution of the number of inpatients on a ward, follows easily from the basic model where we compute the distribution of all recovering patients. In practice patients tend to be segregated into different wards depending on the type of surgery they receive. 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 OR blocks assigned to a specialty in W_k .

2.4 Admission Rate/Discharge Rate

Bed 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. As such, in this subsection, we explain how to derive the probability distribution for daily admissions and discharges.

The admission rate is the rate of arriving patients which we previously defined as the number of surgeries completed on day $n = 0$. For this metric we are only interested in a patient on the day of admission and wish to ignore them afterward. To modify the model to reflect this new purpose replace (3) with

$$h_n^j = \begin{cases} c^j & \text{when } n = 0 \\ \mathbf{0} & \text{otherwise.} \end{cases} \quad (6)$$

With this modification, the resulting H_m represents the distribution for daily admission for each day q of the MSS. To have ward-specific results, we again can restrict this to blocks belonging to specialties of the specific ward.

The discharge rate is the rate at which patients leave the ward and can be computed by adding an additional calculation in Step 1. The number of patients in recovery on day n is distributed according to h_n^j ; see (3). On day n , each patient has the probability d_n^j of being discharged and the probability $(1 - d_n^j)$ of staying. Let D_n^j be a discrete distribution for the number of discharges from specialty j on day n . Given h_n^j and d_n^j , D_n^j can be computed with a binomial distribution as follows:

$$\mathbb{P}(D_n^j = x) = \sum_{k=x}^{c^j} \binom{k}{x} (d_n^j)^x (1 - d_n^j)^{k-x} \mathbb{P}(h_n^j = k). \quad (7)$$

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} represents the distribution for daily discharges for each day q of the MSS. As with admissions, ward-specific results can also be obtained.

2.5 Patients in Day n of Their Recovery

The final workload metric we consider is the distribution of patients in day n of their recovery. This is relevant for predicting workload for the many hospital departments who treat recovering patients. For example, a patient recovering from hip surgery may need to see a physiotherapist every other day during their recovery. This metric states the frequency of such visits. The analogy holds true for all types of services that take place on specific intervals during a patient's recovery (e.g., chemotherapy, diagnostics, social work, discharge planning). In particular, this metric can help dimension capacity for clinical pathways patients whose recovery is intended to follow a strict regime.

The metric requires substantial modifications to the original model, since we now must carry an index (n) for the "day of recovery" throughout the three steps. Let $\bar{h}_{m,n}^{i,q}$ be a discrete distribution for the number of recovering patients from block $b_{i,q}$ on day m in day n of their recovery. To compute $\bar{h}_{m,n}^{i,q}$, replace (3) with the following:

Table 1 Example results for the frequency of inpatient chemotherapy treatments

Example results	Interpretation
$\mathbb{P}(H_{1,3}^{SS} = 2) = 0.3 \quad n = 3, q = 1$	30% probability that exactly two treatments are required on the first day of the MSS cycle
$\mathbb{P}(H_{1,3}^{SS} = 3) = 0.5 \quad n = 3, q = 1$	50% probability that exactly three treatments are required on the first day of the MSS cycle
$\mathbb{P}(H_{1,3}^{SS} = 4) = 0.2 \quad n = 3, q = 1$	20% probability that exactly four treatments are required on the first day of the MSS cycle
$\mathbb{P}(H_{2,3}^{SS} = 2) = 0.4 \quad n = 3, q = 2$	40% probability that exactly two treatments are required on the second day of the MSS cycle
$\mathbb{P}(H_{2,3}^{SS} = 3) = 0.4 \quad n = 3, q = 2$	40% probability that exactly three treatments are required on the second day of the MSS cycle
$\mathbb{P}(H_{2,3}^{SS} = 4) = 0.2 \quad n = 3, q = 2$	20% probability that exactly four treatments are required on the second day of the MSS cycle

$$\bar{h}_{m,n}^{i,q} = \begin{cases} \mathbf{0} & \text{if } m - q \neq n \\ h_{m-q}^z & \text{otherwise,} \end{cases} \tag{8}$$

and we replace (4) with the following:

$$H_{m,n} = \bar{h}_{m,n}^{1,1} * \bar{h}_{m,n}^{1,2} * \dots * \bar{h}_{m,n}^{1,Q} * \bar{h}_{m,n}^{2,1} * \dots * \bar{h}_{m,n}^{I,Q} \tag{9}$$

where $H_{m,n}$ now denotes the number of patients from a single MSS on day m in day n of their recovery.

This alteration to the model eliminates the need for convolutions in Step 3. Since patients are indexed by their recovery day, patients from one MSS cycle are not aggregated with patients from the next. As such we need to replace (5) with

$$H_{q,n}^{SS} = H_{q+Q\lfloor n/(Q+1)\rfloor,n} \tag{10}$$

To help to interpret this metric, consider the following fictitious example for patients who require chemotherapy treatment on day three of their recovery. The Chemotherapy Department would like to know how frequently they need to provide this service. Example results for $H_{q,n}^{SS}$ are illustrated in Table 1.

2.6 Assumptions

Inherent to the model are a number of assumptions which are discussed in this subsection. One assumption resulting from the use of the infinite server system is that there is always a bed available for a patient after surgery. This implies that surgeries are never cancelled due to bed shortages. In practice this means that there is not a physical bed restriction and that additional staff can be called in when demand

is higher than expected. The frequency of this occurring follows from the model. For example, if a hospital staffs 50 beds, then the probability of an additional staffed bed being needed on day q is $\mathbb{P}(H_q^{SS} = 51)$.

In the current formulation, the model ignores seasonality. Of course, at certain times of the year, surgical blocks are cancelled to accommodate vacations and slowdowns, representing a change in supply. In this case, a modified MSS is temporarily used in breaking down the assumption that the same MSS repeats every Q days. However, given that the modifications to the MSS are typically cancellations of certain OR blocks, then the original result can act as an upper bound.

Only elective surgeries are considered. To incorporate non-elective surgeries, it is possible to convolute a historic bed occupancy distribution for non-elective patients. Alternatively, it is possible to incorporate a virtual OR block into the model that represents emergency admissions.

The inherent assumption of using the binomial distribution in this model is that all patients (experiments) have equal probability of each outcome and that the outcome is independent of other patients, i.e., it is assumed that the patients are independent and identically distributed. The independence assumption is natural as it 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. The identically distributed requirement means that we must compute the number of beds needed tomorrow (and the number of case completed in one OR 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 (for a discussion on defining statistically equivalent patient cohorts, see [15]). In our model we aggregate patient such that each surgical specialty is a patient cohort. It follows then that patients within each surgical specialty should be identically distributed.

If a heterogeneous population is grouped together, this causes the ward census distribution to have longer tails (although the mean remains the same) and will overestimate the bed requirements when staffing for a certain percentile of demand. On the other hand, however, less aggregation (such as dividing a specialty by short- and long-stay patients) decreases the sample size from which to derive the parameters which in turn reduces the statistical confidence of the estimated parameters. In our case study that follows, we aggregate the data by specialty which allows for enough data to have a sufficient sample size and results in relatively homogeneous patient cohorts.

In cases where patients of a surgical specialty are not identically distributed and cannot be aggregated into a single cohort, the model can still be used. First the heterogeneous specialty has to be divided into multiple homogeneous cohorts, and then these cohorts can be treated as if they were assigned their own OR block. Using this, the binomial distribution is applied as described above to determine the bed requirements of each cohort. Again, using the independence assumptions, these cohorts can be added (with discrete convolutions) to determine the total bed requirements for the complete surgical specialty.

3 Application

The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital (NKI-AVL) is a comprehensive cancer center, which provides hospital care and research and is located in Amsterdam, The Netherlands. The hospital has 150 inpatient beds and sees about 24,000 new patients every year, making it approximately the size of a mid-sized general hospital. As with many Dutch hospitals, NKI-AVL is eager to improve access and increase capacity. To this end, the hospital has expanded its operating capacity from five to six operating rooms (ORs). The hospital welcomed this expansion as an opportunity to develop a new MSS.

NKI-AVL distributes its surgical capacity to its six surgical specialties in the typical manner described previously. The yearly amount of operating time is first allotted to each specialty reflecting patient demand and hospital priorities. To implement this allotment, and to make the surgical department manageable, NKI-AVL divides OR time into OR blocks over a 1-week planning horizon. One OR block represents a full day of operating room time. The assignment of the surgical specialties to each OR block represents the MSS. An example MSS for five operating rooms is shown in Fig. 3.

As previously discussed, the use of operations research in surgical scheduling is not new; however, the rate of implementation for this type of study is low [8]. To overcome this, Cardoen et al. [8] “encourage the provision of additional information on the behavioral factors that coincide with the actual implementation. Identifying the causes of failure or the reasons that lead to success, may be of great value to the research community.” In this section we describe the process of developing a new MSS for NKI-AVL and results observed after its implementation. The development process, which combined an operational research model and staff input, led to an MSS which was agreeable to staff from both the wards and the OR. Staff selected and implemented an MSS which the model predicted would result in a balanced ward occupancy.

The development of the new MSS was completed over 3 months in an iterative manner. A team was formed consisting of a team leader from the surgical department, a team leader from the inpatient wards, the manager of both groups,

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
OR 1	Surg. 1	Surg. 3	Surg. 3	HIPEC	Surg. 2		
OR 2	ENT 1	Surg. 4	Urol.	Urol.	Urol.		
OR 3	ENT 2	P. Surg.	ENT 1	ENT 2	P. Surg.		
OR 4	Surg. 2	Gyne.	ENT 2	P. Surg.	Gyne.		
OR 5	RT	Surg. 5	RT	RT	Urol.		

Fig. 3 Sample MSS for five operating rooms

and two of this chapter's authors. The team members from the surgical department ensured MSS proposals did not cause conflicts within the OR, such as with physician schedules and available equipment. The projected impact that each MSS proposal would have on the wards was evaluated with the model described previously. Each new MSS proposal represented a new scenario to be evaluated by the model. From the model output, staff decided whether the MSS was acceptable or if further modifications to the MSS were necessary.

The original MSS was roughly developed as follows. Based on production targets, the number of OR blocks to be assigned to each specialty during the 1-week MSS cycle was determined. Next, the physicians' commitments elsewhere in the hospital were determined, and their preferred operating days were considered. Potential equipment and resources conflicts were addressed, for example, it would be problematic to assign two specialties to the same operating day when both routinely require the same specialized OR. Considering these restrictions, OR staff proposed the *original* MSS.

To determine how the original MSS impacted the wards, the model was used. As illustrated in the "Results" section, the original MSS results in an unbalanced ward occupancy (the motivation for this metric is also provided in the "Results" section). As such, the team decided the original MSS was not acceptable.

Next, modifications to the original MSS were made, and a new MSS proposal was put forth. Given that in this project we were not asking surgical specialties to change how they operate (i.e., the number of surgeries they perform in an OR block and/or the invasiveness of their surgeries which can dictate length of stay), modifications to the MSS were limited to changes in the assignment of surgery specialties to OR blocks. Essentially, modifications consisted of swapping a specialty operating on one day with a specialty operating on a different day. Deciding which blocks to swap followed, first, from OR staff knowledge of what was possible within the constraints of the OR and, second, by intuition gained from seeing results from several MSS proposals. See Fig. 4 for an illustration of the type of modifications made.

A number of MSSs were proposed, and the impact that each would have on the ward was evaluated by the model. This process of modifying and evaluating MSSs continued for several weeks until an MSS was found that satisfied staff from both the OR and the wards. A schematic overview for this process is displayed in Fig. 5.

The MSS chosen by the team was implemented concurrently with the opening of the new OR in March of 2009. The new OR was phased in over several months, and once it became fully utilized, ward occupancy statistics were collected. The data, observed over a 33-week period when all six ORs were being regularly scheduled, was compared with what was projected from the model. The purpose, to ensure a more balanced ward occupancy, was indeed being achieved with the implemented MSS. In this way we could validate the model output and confirm the implemented MSS is resulting in the desired ward occupancy.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
OR 1	Surg. 1	Surg. 3	Surg. 3	HIPEC	Surg. 2		
OR 2	ENT 1	Surg. 4	Urol.	Urol.	Urol.		
OR 3	ENT 2	P. Surg.	ENT 1	ENT 2	P. Surg.		
OR 4	Surg. 2	Gyne.	ENT 2	P. Surg.	Gyne.		
OR 5	RT	Surg. 5	RT	RT	Urol.		

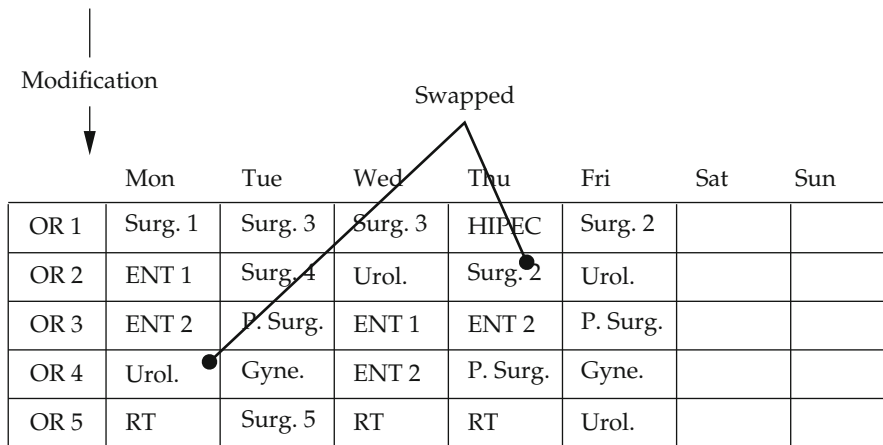


Fig. 4 Illustration of MSS modification

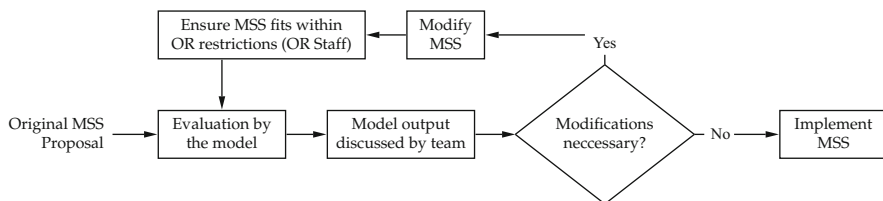


Fig. 5 MSS development schematic

3.1 Results

This section is divided into two subsections. The first subsection discusses ward occupancy projected by the model during the MSS development process. We show ward occupancy projections from the *original* MSS proposal and from the MSS proposal that staff chose to implement (which we refer to as the *implemented* MSS). In the second subsection, we compare the ward occupancy projected by the

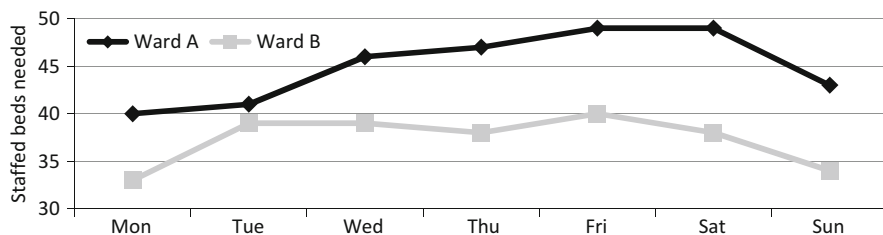


Fig. 6 90th percentile of demand projected by the model for each day of the MSS cycle (original MSS proposal)

model for the implemented MSS with the ward occupancy observed after it was implemented.

3.1.1 Projected Results

NKI-AVL has two wards for treating surgical patients, Ward A and Ward B, with a combined physical capacity of 100 beds. Management strives to staff enough beds such that for 90% of the days, there is sufficient coverage. In other words, they staff for the 90th percentile of demand; their accepted risk for needing to call in additional staff is thus 10%. Figure 6 illustrates the 90th percentile demand for staffed beds on each of the wards, resulting from the original MSS proposal. As is clear from the figure, the staffing requirements are relatively balanced across the weekdays (Monday to Friday) for Ward B. This is not the case for Ward A. On Ward A the occupancy is relatively low on Monday and Tuesday and relatively high on Thursday, Friday, and Saturday.

This projected demand for staffed beds concerned the ward manager, as such an unbalanced demand profile 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 leading to significant problems, particularly as the wards approach peak capacity. For example, when inpatient 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 anxiety for patients. When a bed is 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 preferred 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 not possible without an exorbitant amount of resources, sound planning ahead of time may help to minimize occurrences.

After discussing the model output, all participating staff agreed that the original MSS, although appropriate for the OR, was not ideal for the wards. The discussion

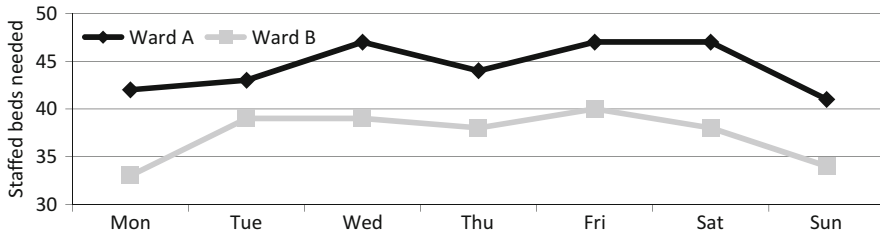


Fig. 7 90th percentile of demand projected by the model for each day of the MSS cycle (implemented MSS)

then moved to how to correct the imbalance across the weekdays by changing the assignment of OR blocks to specialties. Modifications to the original MSS were made by considering what changes were possible within the restrictions of the OR (e.g., physician schedules and equipment availability).

Eventually, after considering several MSS proposals, the process led to an MSS (the implemented MSS) which was acceptable to all staff members. The implemented MSS fit within the restrictions of the OR and, as illustrated in Fig. 7, resulted in a more balanced ward occupancy. Comparing the implemented MSS with the original MSS, the implemented MSS dampened the fluctuation on Ward A by lowering occupancy on Thursday, Friday, and Saturday and increasing it on Monday and Tuesday. With the implemented MSS, the model predicted that no days would require more than 47 staffed beds, which reduced the maximum from 49 (predicted for the original MSS). Furthermore, the implemented MSS ensured the staffing requirements remained relatively balanced across the working days for both wards.

3.1.2 Observed Results

The ward occupancy was observed over a 33-week period after the new OR was fully operational. From these data, probability distributions of beds used for each day of the MSS cycle were derived. Using chi-square goodness-of-fit tests [21], these observed distributions were compared to those projected by the model. For Ward B, six of the seven distributions (one for each day of the MSS cycle) were found to be statistically equivalent at a level $\alpha = 0.05$, while the seventh day was statistically equivalent at a level $\alpha = 0.2$. For Ward A, the tests revealed statistical equivalence at levels $\alpha = 0.15$ (for three days), 0.25 (for two days), and 0.35 (for two days). At these alpha levels, we conclude that the observed ward occupancy is well predicted by the model. Explanations for the poorer fit of Ward A data are discussed in the following paragraphs where the 90th percentiles (desired staffing level) are compared for the observed and projected results.

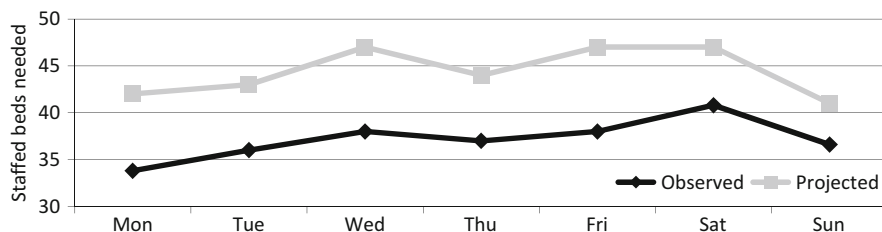


Fig. 8 Comparison of projected and observed ward occupancies (90th percentile) on Ward A

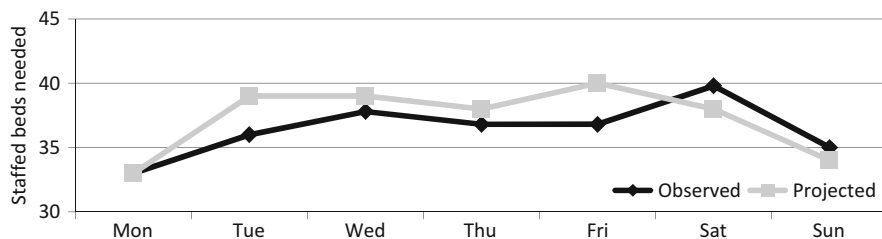


Fig. 9 Comparison of projected and observed ward occupancies (90th percentile) on Ward B

Figures 8 and 9 compare the projected ward occupancy with the observed ward occupancy during the 33-week period. Figure 8 displays results for Ward A and Fig. 9 for Ward B.

As is observable in Figs. 8 and 9, the data indicates that both wards have balanced ward occupancies across the week days. However, it is also observable that our model overestimated the number of beds required in Ward A by approximately 16%. The overestimate is due to an unexpected increase in short-stay patients during the period of measurement. Had this change in patient mix been expected at the time the projections were made (and model input altered to reflect it), such an overestimate would likely not have been observed, and we would expect to have similarly accurate results as those for Ward B.

As a final note on the model results, consider if hospital management decided to staff only for the average number of beds projected to be needed for six ORs. In this case, 32 beds would be assigned to Ward A and 29 beds to Ward B. This would have led to a bed shortage on 51% of the days, illustrating the importance of considering probability distributions in hospital planning.

3.2 Discussion

With the approach discussed in this section, a new MSS was developed for NKI-AVL which reduced the fluctuations in the daily ward census, creating a more balanced workload on the wards. The roll out of the new MSS corresponded with

the opening of an additional OR which was expected to overwhelm the wards. By using the described process to develop an MSS that accounted for the inpatient wards, peaks in ward occupancy were reduced. As such capacity is used more efficiently, and the hospital has the means to support the additional OR without a major expansion in the wards.

The main benefit of the model was the ability to quantify the concerns of ward staff, thereby providing a platform 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 the model output intuitively. For example, on Thursdays and Fridays, the wards tended to be crowded with patients. To remedy this, specialties that completed many cases per OR block were removed from Thursday and Friday OR blocks and assigned to OR blocks earlier in the week. To accommodate these changes, specialties which complete a relatively small number of cases per OR block were moved to Thursday and Friday. Once staff could foresee the impact of swapping one surgical OR block assignment with another, the MSS which was eventually implemented came quickly.

In the NKI-AVL application, we treated the equipment and physician schedule restrictions as unchangeable. It is possible that further improvements in the ward occupancy 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.

At NKI-AVL, our model was used to solve the *tactical* surgical scheduling problem – a medium-term planning horizon with patients aggregated by surgical specialty. Alternatively, the same model can support decisions at an *operational* level – a shorter-term planning horizon without patient aggregation (for a discussion on levels of planning and control in health care, see [13]). Instead of computing the expected patients in recovery, the actual patients in recovery can be used as input. By combining this with the expected new arrivals from the OR, real-time workload projections can be used to identify upcoming staffing needs.

4 Conclusion

Many good research projects conclude after the implementation of results with team members satisfied that anticipated improvements have been realized. This project, in some ways, merely began at this point. Variants of this model have been developed (and applied) at the request of three other Dutch hospitals, and the model forms the basis of a similar application at a German hospital. Below we discuss these other applications of our model.

Leiden University Medical Center (LUMC), the Netherlands LUMC is one of eight University Medical Centers in the Netherlands and employs approximately 7000 professionals. Highly variable ward occupancy was proving problematic for staffing the inpatient wards at LUMC. Utilizing our model, staff proposed and evaluated a number of MSS proposals in order to reduce occupancy fluctuations. They found that with very little disruption to the current MSS (only four swaps of OR blocks), the maximum bed occupancy would reduce from 74 to 71. Additional reductions in the maximum bed occupancy were found only to be possible when additional OR time was made available [27].

Haga Hospital, The Hague, the Netherlands Haga Hospital is a top clinical teaching hospital in the Netherlands with 245 specialists, 729 beds, and 35,571 admissions in 2010. At Haga Hospital, management wanted to develop a new structured scheduling procedure in order to increase OR utilization and balance ward occupancy. To achieve this, the scheduling procedure was redesigned, and a formal process was put into practice supported by a decision support system. To model stochastic length of stays and to integrate bed leveling into this software, the model described in this chapter was used [7].

Technical University of Munich, Germany To make the model appropriate for a German hospital, modifications were made in collaboration with researchers at the Technical University of Munich. The first modification involved increasing the scope to include the Intensive Care Unit (ICU), a unit in which most acute surgical patients are admitted to receive one-on-one nursing care. This modification essentially amounted to changing the model from a single queue to a network of two queues, one for the ICU and one for the Ward. The second modification involves developing heuristics for determining good MSS proposals. A number of objectives are considered, including minimizing costs and determining the best improvement with the smallest change (disruption) of the existing MSS.

The buy-in of other hospitals and the dissemination of our approach can be credited to the way in which the model was incorporated *into* the decision-making process and does not *replace* it. Supporting and not replacing the process allows for a less complex model and more staff engagement. This approach to problem-solving proved crucial for implementation and for making a meaningful impact.

References

1. Adan, I.J.B.F., Vissers, J.M.H.: Patient mix optimisation in hospital admission planning: a case study. *International Journal of Operations and Production Management* **22**(4), 445–461 (2002)
2. Beliën, J., Demeulemeester, E.: Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research* **176**(2), 1185–1204 (2007)
3. Beliën, J., Demeulemeester, E., Cardoen, B.: Visualizing the demand for various resources as a function of the master surgery schedule: A case study. *Journal of Medical Systems* **30**(5), 343–350 (2006)

4. Beliën, J., Demeulemeester, E., Cardoen, B.: A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling* **12**(2), 147–161 (2009)
5. Blake, J.T., Carter, M.W.: Surgical process scheduling: a structured review. *Journal of the Society for Health Systems* **5**(3), 17–30 (1997)
6. Blake, J.T., Donald, J.: Mount Sinai Hospital uses integer programming to allocate operating room time. *Interfaces* **32**(2), 63–73 (2002)
7. Bosch, J.: Surgery scheduling. Master's thesis, University of Twente (2011)
8. Cardoen, B., Demeulemeester, E., Beliën, J.: Operating room planning and scheduling: A literature review. *European Journal of Operational Research* **201**(3), 921–932 (2010)
9. Carter, M.W.: Diagnosis: Mismanagement of resources. *OR/MS Today* **29**(2), 26–32 (2002)
10. Gallivan, S., Utley, M., Treasure, T., Valencia, O.: Booked inpatient admissions and hospital capacity: mathematical modelling study. *BMJ* **324**(7332), 280–282 (2002)
11. Gorunescu, F., McClean, S.I., Millard, P.H.: A queueing model for bed-occupancy management and planning of hospitals. *Journal of the Operational Research Society* **53**, 19–24 (2002)
12. Griffiths, J.D., Price-Lloyd, N., Smithies, M., Williams, J.E.: Modelling the requirement for supplementary nurses in an intensive care unit. *Journal of the Operational Research Society* **56**(2), 126–133 (2005)
13. Hans, E.W., van Houdenhoven, M., Hulshof, P.J.H.: Handbook of Health Care Systems Scheduling, chap. 12: A Framework for Health Care Planning and Control, pp. 1–18. Springer International Series in Operations Research & Management Science, Vol. 168 (2011)
14. Hans, E.W., Vanberkel, P.T.: Handbook of Health Care Systems Scheduling, chap. 5: Operating Theatre Planning and Scheduling, pp. 1–27. Springer International Series in Operations Research & Management Science, Vol. 168 (2011)
15. Harper, P.R.: A framework for operational modelling of hospital resources. *Health Care Management Science* **5**(3), 165–173 (2002)
16. Harris, R.A.: Hospital bed requirements planning. *European Journal of Operational Research* **25**(1), 121–126 (1986)
17. van Houdenhoven, M., Hans, E., Klein, J., Wullink, G., Kazemier, G.: A norm utilisation for scarce hospital resources: Evidence from operating rooms in a Dutch university hospital. *Journal of Medical Systems* **31**(4), 231–236 (2007)
18. van Houdenhoven, M., van Oostrum, J.M., Wullink, G., Hans, E., Hurink, J.L., Bakker, J., Kazemier, G.: Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. *Journal of Critical Care* **23**(2), 222–226 (2008)
19. Huang, X.M.: A planning model for requirement of emergency beds. *Mathematical Medicine and Biology* **12**(3-4), 345–353 (1995). doi:10.1093/imammb/12.3-4.345
20. Litvak, E., Long, M.C.: Cost and quality under managed care: Irreconcilable differences. *American Journal of Managed Care* **6**(3), 305–312 (2000)
21. McDonald, J.H.: Handbook of Biological Statistics, 2nd ed. Sparky House Publishing, Baltimore, Maryland. (2009)
22. McManus, M.L., Long, M.C., Cooper, A., Mandell, J., Berwick, D.M., Pagano, M., Litvak, E.: Variability in surgical caseload and access to intensive care services. *Anesthesiology* **98**(6), 1491–1496 (2003)
23. van Oostrum, J.M., Bredenhoff, E., Hans, E.W.: Suitability and managerial implications of a master surgical scheduling approach. *Annals of Operations Research* **178**(1), 91–104 (2010)
24. van Oostrum, J.M., van Houdenhoven, M., Hurink, J.L., Hans, E.W., Wullink, G., Kazemier, G.: A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum* **30**(2), 355–374 (2008)
25. Santibáñez, P., Begen, M., Atkins, D.: Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Management Science* **10**(3), 269–282 (2007)
26. Seddon, N.: Is the future Dutch? *The Lancet* **372**, 103–104 (2008)
27. Tjoonk, J.: Workload leveling of nursing wards in Leiden University Medical Center. Master's thesis, University of Twente (2010)

28. Utley, M., Gallivan, S., Davis, K., Daniel, P., Reeves, P., Worrall, J.: Estimating bed requirements for an intermediate care facility. *European Journal of Operational Research* **150**(1), 92–100 (2003)
29. Utley, M., Gallivan, S., Treasure, T., Valencia, O.: Analytical methods for calculating the capacity required to operate an effective booked admissions policy for elective inpatient services. *Health Care Management Science* **6**(2), 97–104 (2003)
30. Vanberkel, P.T.: Interacting hospital departments and uncertain patient flows: theoretical models and applications. Ph.D. thesis, University of Twente, Enschede, the Netherlands (2011). URL <http://doc.utwente.nl/77176/>
31. Vanberkel, P.T., Blake, J.T.: A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Management Science* **10**(4), 373–385 (2007)
32. Vanberkel, P.T., Boucherie, R.J., Hans, E.W., Hurink, J.L., van Lent, W.A.M., van Harten, W.H.: Accounting for inpatient wards when developing master surgical schedules. *Anesthesia & Analgesia* **112**, 1472–1479 (2011)
33. Vanberkel, P.T., Boucherie, R.J., Hans, E.W., Hurink, J.L., Lent, W.A.M., Harten, W.H.: An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society* **62**, 1851–1860 (2011)
34. Vanberkel, P.T., Boucherie, R.J., Hans, E.W., Hurink, J.L., van Lent, W.A.M., van Harten, W.H.: Een planningsmethode voor reductie van de fluctuaties in de belasting van verpleegafdelingen. *STAtOR* (forthcoming) (2012)
35. Vanberkel, P.T., Zonderland, M.E., Boucherie, R.J., Hans, E.W.: Commentary – leaving a mark on healthcare delivery with operations analysis. *M&SOM* (forthcoming) pp. 1–7 (2012)
36. Wachtel, R.E., Dexter, F.: Tactical increases in operating room block time for capacity planning should not be based on utilization. *Anesthesia & Analgesia* **106**(1), 215–226 (2008)