Bed Census Predictions and Nurse Staffing

Aleida Braaksma, Nikky Kortbeek, and Richard J. Boucherie

Abstract Workloads in nursing wards depend highly on patient arrivals and lengths of stay, both of which are inherently variable. Predicting these workloads and staffing nurses accordingly are essential for guaranteeing quality of care in a cost-effective manner. This chapter describes a stochastic method that uses hourly census predictions to derive efficient nurse staffing policies. The generic analytic approach minimizes staffing levels while satisfying so-called nurse-to-patient ratios. In particular, we explore the potential of flexible staffing policies that allow hospitals to dynamically respond to their fluctuating patient population by employing float nurses. The method is applied to a case study of the surgical inpatient clinic of the Academic Medical Center Amsterdam (AMC).

1 Introduction

Societal developments and budget constraints demand hospitals to on the one hand increase quality of care and on the other hand efficiency [41]. This entails a strong incentive to reconsider the design and operations of inpatient care services that provide care to hospitalized patients by offering a room, a bed, and board [40]. Since the 1950s, the application of operational research methods yields significant contributions in accomplishing essential efficiency gains in healthcare delivery [30].

A. Braaksma (✉) · R. J. Boucherie
Center for Healthcare Operations Improvement and Research, University of Twente, Enschede, The Netherlands
e-mail: a.braaksma@utwente.nl; r.j.boucherie@utwente.nl

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

Rhythm b.v., Amsterdam, The Netherlands
e-mail: Nikky.Kortbeek@rhythm.nl

M. E. Zonderland et al. (eds.), *Handbook of Healthcare Logistics*,
International Series in Operations Research & Management Science 302,
https://doi.org/10.1007/978-3-030-60212-3_9

This chapter, combining the results of [32, 33], presents an exact method to assist hospital management in adequately organizing their inpatient care services.

The challenge in decision-making for inpatient care delivery is to guarantee care from appropriately skilled nurses and required equipment to patients with specific diagnoses while making efficient use of scarce resources [28, 50]. Deploying adequate nurse staffing levels is one of the prime responsibilities of inpatient care facility managers. Nursing staff typically accounts for the majority of hospital budgets [54], which means that every incidence of overstaffing is scrutinized during times when cost containment efforts are required [34]. Performance measures are required that reflect efficiency and quality of care to assess the quality of the logistical layout. Efficiency is often expressed in high bed occupancy. The drawback of high bed occupancy is that it may cause congestion and a threat to the provided quality of care [21, 22]: (i) patients may have to be rejected for admission due to lack of bed capacity, so-called admission refusals or rejections, and (ii) patients may be placed in less appropriate units, so-called misplacements [13, 27, 29]. Due to such misplacements, planning decisions regarding a specific care unit can affect the operations of other units [4, 12, 36]. At the same time, maintaining appropriate staffing levels is crucial to be able to provide high-quality care. Planning of the inpatient care facility should not only take into account the upstream departments, such as the emergency department and the operating rooms, but also the interrelationship between care units. In this chapter, following [32, 33], we present an exact method to assist healthcare administrators in ensuring safe patient care while also maintaining an efficient and cost-effective nursing service.

We first present a generic exact analytical approach to achieve the required integral and coordinated resource capacity planning decision-making for inpatient care services, building upon the approach introduced in [49], which determines the workload placed on hospital departments by describing demand for elective inpatient care beds on a daily level as a function of the Master Surgical Schedule (MSS). The MSS is a (cyclic) block schedule that allocates operating time capacity among patient groups as typically used by hospitals to allocate operating room capacity [18, 24, 46]. Based on a cyclic arrival pattern of emergency patients and an MSS block schedule of surgical patients, we present demand predictions on an hourly level for several inpatient care units simultaneously for both acute and elective patients. Based on overflow rules, we translate the demand predictions to bed census predictions, since demand and census may differ due to rejections and misplacements. The combination of the hourly level perspective and the bed census conversion enables us to derive several performance measures, along which the effectiveness of different logistical configurations can be assessed.

Subsequently, following [32], we incorporate the tactical decision that is referred to as ‘staff-shift scheduling’ in [30] into this integrated modeling framework. For each working shift, during a given planning horizon, we determine the number of employees that should be assigned to each inpatient care unit. The predictable fluctuation in inpatient population due to the operating room schedule and other predictable variabilities in patient arrivals (e.g., seasonal, day of week, and time of day effects) can be taken into account when determining the staffing levels for 'dedicated nurses', which are nurses with a fixed assignment to a care unit. When only dedicated nurses are employed, the buffer capacity required to protect against random demand fluctuations can lead to regular overstaffing. When two or more care units cooperate by jointly appointing a flexible nurse pool, the variability of these random demand fluctuations balances out due to economies of scale, so that less buffer capacity is required. We explore the potential of flexible staffing policies that allow hospitals to dynamically respond to their fluctuating patient populations. This flexibility is achieved by employing a pool of cross-trained nurses, or 'float nurses' [19, 42], for whom assignments to specific care units are decided at the start of their shifts.

To illustrate its potential, the method is applied to a case study that involves the care units in the surgical inpatient clinic of the Dutch university hospital, the Academic Medical Center Amsterdam (AMC), which serve the specialties of traumatology, orthopedics, plastic surgery, urology, vascular surgery, and general surgery. Inspired by the quantitative results, the AMC decided that the method will be fully implemented as part of the global redesign of its inpatient care services.

This chapter is organized as follows: Sect. 2 provides a review of relevant literature; Sect. 3 presents the model to predict bed census; Sect. 4 presents the models for the fixed and the flexible staffing policies; Sect. 5 presents the numerical results for the case study; and Sect. 6 closes the chapter with a general discussion.

2 Literature

Effectively designing inpatient care services requires simultaneous consideration of several interrelated strategic and tactical planning issues [30]. The inpatient care facility is a downstream department. The outflow of the operating theater and the emergency department are main drivers behind its workload. Therefore, it is highly desirable to apply coordinated planning: considering the inpatient care facility in isolation yields suboptimal decision-making [25, 48]. Smoothing patient inflow prevents large differences between peak and off-peak periods and so realizes a more efficient use of resources [1, 25, 51]. Although the control on the inflow of patients from the emergency department is inherently very limited due to its nature, anticipation for emergency admissions is possible, by statistically predicting the arrival process of emergency patients that often follows a cyclic pattern [22]. Hospitals typically allocate operating room capacity through a Master Surgical Schedule (MSS). Anticipation for elective surgical patients is possible as well, by taking the surgical schedule into account [1, 22, 25, 51]. In this chapter, we address these various patient flows and take the necessity of integral decision-making into account.

Several analytical studies have addressed partial resource capacity planning within the inpatient care chain, for example, by dimensioning care units in isolation [5, 20, 22], balancing bed utilization across units [3, 12, 36], or improving the MSS to balance inpatient care demand [1, 6, 8, 46, 49]. More integral approaches can be found in simulation studies [25, 27, 44, 47]. The advantage of such approaches is their flexibility and therefore modeling power. However, the disadvantage is that the nature of such studies is typically context-specific, which limits the generalizability of application and findings.

Personnel scheduling in general and capacity planning for nursing staff in specific have received considerable attention from the operations research community)see the extensive literature review [45] and the survey and classification articles [9, 14, 17]). The nurse staffing process involves a set of hierarchical decisions over different time horizons with different levels of precision. The first strategic level of decision-making is the workforce dimensioning decision which concerns both the number of employees that must be employed and is often expressed as the number of full-time equivalents and the mix in terms of skill categories [26, 35, 37]. The second tactical level concerns staff-shift scheduling, which deals with the problem of selecting which shifts are to be worked and how many employees should be assigned to each shift to meet the patient demand [17, 31]. The third operational offline decision level concerns the creation of individual nurse timetables, designed with the objective to meet the required shift staffing levels set on the tactical level while satisfying a complex set of restrictions involving work regulations and employee preferences. This planning step is often referred to as 'nurse rostering' [9–11]. The fourth operational online decision level concerns the reconsideration of the staff schedule at the start of a shift. At this level, float nurses are assigned to specific care units [9, 42], and, based on the severity of need, on-call nurses, overtime, and voluntary absenteeism can be used to further align patient care supply and demand [23, 39]. The interdependence of the decision levels must be recognized to facilitate systematic improvements in nurse staffing. As expressed in the literature review by [39], each level is constrained by previous commitments made at higher levels, as well as by the degrees of flexibility conserved for later correction at lower levels. For a more elaborate exposition of the relevant decisions and considerations involved at each decision level and a detailed overview of relevant literature, see [30].

Tactical workforce decision-making in healthcare has received little attention. A spreadsheet approach is presented by [16], to retrospectively fit optimal shift staffing levels to historical census data. Simulation studies have shown to be successful in taking a more integral approach [23, 26]. Analytic yet deterministic approaches can be found in [7, 38, 52]. Stochastic approaches to determine shift staffing levels are available in [15, 54] and [55]. These references do not present an integral care chain approach, given that the demand distributions underlying the staffing decisions are not based on patient arrival patterns from the operating theaters and emergency departments.

Concerning the operational online assignment strategy to place a given number of available float nurses in care units at the start of their shifts, [43] indicate that formulating such an assignment strategy requires the consideration of three issues: (1) a method for measuring of the urgency of need for an additional nurse; (2) a prediction per care unit of that urgency of need for an upcoming shift; and (3) development of a technique for the allocation of the available float nurses to care units in order to meet this need. Whereas [43] focus on the third issue by developing a branch-and-bound algorithm, our assignment strategy involves the consideration of all three steps.

Nurse-to-patient ratios are commonly applied when determining staffing levels [2, 55]. These ratios indicate how many patients a registered nurse can care for during a shift, taking into account both direct and indirect patient care. Staffing according to nurse-to-patient ratios has received attention in the operations research literature [15, 54, 55]. In practice, setting the numerical values of the ratios seems to be more based on negotiation than on science [15, 54].

In this chapter, following [32, 33], we present an exact stochastic analytic approach to derive appropriate staffing levels, including the flexibility of float nurses, using nurse-to-patient ratios while taking an integrated care chain perspective.

3 Hourly Bed Census Predictions

This section presents a general model to predict the workload at several care units due to patients arriving and departing according to a statistically predicted inflow and outflow pattern. Following [33], we will focus on the workload at an inpatient care facility on a timescale of hours, due to patients originating from the operating theater and emergency department. The basis for the operating room outflow prediction is the Master Surgical Schedule (MSS). The basis for the emergency department outflow prediction is a cyclic random arrival process, the Acute Admission Cycle (AAC). The cycles are combined into the Inpatient Facility Cycle (IFC), with length the least common multiple of the lengths of the MSS and the AAC. For the demand predictions, for both elective and acute patients, the impact of a single patient type in a single cycle (MSS or AAC) is determined, by which in the second step the impact of all patient types within a single cycle can be calculated. Since the IFC is cyclical, the predictions from the second step are combined to find the probability distributions of the number of recovering patients at the inpatient care facility on each time interval in the IFC. The resulting demand distributions are translated to bed census distributions, and performance measures are formulated based on the demand and census distributions.

The operation of the inpatient care facility is as follows. Each day is divided in time intervals (hours). Patient admissions are assumed to take place independently at the start of a time interval. Elective patients are admitted to a care unit either on the day before or on the day of surgery. For acute patients we assume a cyclic (e.g., weekly) non-homogeneous Poisson arrival process corresponding to the unpredictable nature of emergency arrivals (see, e.g., [53]). Discharges take place independently at the end of a time interval. For elective patients we assume the length of stay to depend only on the type of patient and to be independent of the day of admission and the day of discharge. For acute patients the length of stay and time of discharge are dependent on the day and time of arrival, in particular to account for possible disruptions in diagnostics and treatment during nights and weekends.

Patient admission requests may have to be rejected due to a shortage of beds, or patients may (temporarily) be placed in less appropriate units. As a consequence, demand predictions and bed census predictions do not coincide. Therefore, an additional step is required to translate the demand distributions into census distributions. This translation is performed by assuming that after a misplacement the patient is transferred to his preferred care unit when a bed becomes available, where we assume a fixed patient-to-ward allocation policy, which prescribes the prioritization of such transfers.

3.1 Demand Predictions for Elective Patients

Model input

Time. An MSS is a repeating blueprint for the surgical schedule of S days. Each day is divided in T time intervals. Therefore, we have time points $t = 0, \dots, T$, in which $t = T$ corresponds to $t = 0$ of the next day. For each single patient, day n counts the number of days before or after surgery, i.e., $n = 0$ indicates the day of surgery.

MSS utilization. For each day $s \in \{1, \dots, S\}$, a (sub)specialty j can be assigned to an available operating room i, $i \in \{1, \dots, I\}$. The OR block at operating room i on day s is denoted by $b_{i,s}$ and is possibly divided in a morning block $b^m_{i,s}$ and an afternoon block $b^a_{i,s}$, if an OR day is shared. The discrete distributions c^j represent how specialty j utilizes an OR block, i.e., $c^j(k)$ is the probability of k surgeries performed in one block, $k \in \{0, 1, \dots, C^j\}$. If an OR block is divided in a morning OR block and an afternoon OR block, c^j_M and c^j_A represent the utilization probability distributions, respectively. For brevity, we do not include shared OR blocks in our formulation, since these can be modeled as two separate (fictitious) operating rooms.

Admissions. With probability e^j_n, $n \in \{-1, 0\}$, a patient of type j is admitted on day n. Given that a patient is admitted on day n, the time of admission is described by the probability distribution $w^j_{n,t}$. We assume that a patient who is admitted on the day of surgery is always admitted before or at time ϑ_j; therefore, we have $w^j_{0,t} = 0$ for $t = \vartheta_j + 1, \dots, T - 1$.

Discharges. $P^j(n)$ is the probability that a type j patient stays n days after surgery, $n \in \{0, \dots, L^j\}$. Given that a patient is discharged on day n, the probability of being discharged in time interval $[t, t + 1)$ is given by $m^j_{n,t}$. We assume that a patient who is discharged on the day of surgery is discharged after time ϑ_j, i.e., $m^j_{0,t} = 0$ for $t = 0, \dots, \vartheta_j$.

Single surgery block In this first step, we consider a single specialty j operating in a single OR block. Note that admissions can take place during day $n = -1$ and during day $n = 0$ until time $t = \vartheta_j$. Discharges can take place during day $n = 0$ from time $t = \vartheta_j + 1$ and during days $n = 1, \dots, L^j$. Therefore, the probability $h^j_{n,t}(x)$ that n days after carrying out a block of specialty j, at time t, x patients of the block are still in recovery is

$$h^j_{n,t}(x) = \begin{cases} a^j_{n,t}(x) & , n = -1 \text{ and } n = 0, t \leq \vartheta_j, \\ d^j_{n,t}(x) & , n = 0, t > \vartheta_j \text{ and } n = 1, \dots, L^j, \end{cases}$$

where $a^j_{n,t}(x)$ represents the probability that x patients are admitted until time t on day n and $d^j_{n,t}(x)$ is the probability that x patients are still in recovery at time t on day n. The derivation of $a^j_{n,t}$ is presented below and that of $d^j_{n,t}$ is by analogy and is presented in [33].

Observe that

$$a^j_{n,t}(x) = \sum_{y=x}^{C^j} a^j_{n,t}(x|y)c^j(y),$$

where $a^j_{n,t}(x|y)$ is the probability that x patients are admitted until time t on day n, given that y admissions take place in total:

$$a^j_{n,t}(x|y) = \begin{cases} \binom{y}{x}(v^j_{n,t})^x(1-v^j_{n,t})^{y-x} & , n = -1, t = -0, \\ \sum_{g=0}^{x}\binom{y-g}{x-g}(v^j_{n,t})^{x-g}(1-v^j_{n,t})^{y-x}a^j_{n-1,T-1}(g|y) & , n=0, t=0, \\ \sum_{g=0}^{x}\binom{y-g}{x-g}(v^j_{n,t})^{xs-g}(1-v^j_{n,t})^{y-x}a^j_{n,t-1}(g|y) & , n = -1, t=1, \dots, T-1 \text{ and} \\ & \quad n=0, t=1, \dots, \vartheta_j - 1, \\ 0 & , n=0, t \geq \vartheta_j, \end{cases}$$

where $v^j_{n,t}$ is the probability for a type j patient to be admitted at time t, given that he/she will be admitted at day n and is not yet admitted before t:

$$v^j_{n,t} = \frac{w^j_{n,t}e^j_n}{e^j_n \sum_{k=t}^{T-1} w^j_{n,k} + e^j_0 \cdot \mathbb{1}_{(n=-1)}}.$$

Single MSS cycle Now, we consider a single MSS in isolation. From the distributions $h^j_{n,t}$, we can determine the distributions $H_{m,t}$, the discrete distributions for the total number of recovering patients at time t on day m ($m \in \{0, 1, 2, \dots, S, S+1, S+2, \dots\}$) resulting from a single MSS cycle. We determine the overall probability distribution of the number of patients in recovery resulting from a single MSS, using discrete convolutions. If specialty j is assigned to OR block $b_{i,s}$, then the distribution $\bar{h}^{i,s}_{m,t}$ for the number of recovering patients of block $b_{i,s}$ present at time t on day m ($m \in \{0, 1, 2, \ldots, S, S+1, S+2, \ldots\}$) is given by

$$\bar{h}^{i,s}_{m,t} = \begin{cases} \mathbf{0} & , m < s-1, \\ h^{j}_{m-s,t} & , m \geq s-1, \end{cases}$$

where $\mathbf{0}$ means $\bar{h}^{i,s}_{m,t}(0) = 1$ and all other probabilities $\bar{h}^{i,s}_{m,t}(x), x > 0$ are 0. Then, $H_{m,t}$ is computed by

$$H_{m,t} = \bar{h}^{1,1}_{m,t} \otimes \bar{h}^{1,2}_{m,t} \otimes \ldots \otimes \bar{h}^{1,S}_{m,t} \otimes \bar{h}^{2,1}_{m,t} \otimes \ldots \otimes \bar{h}^{I,S}_{m,t}, \tag{1}$$

where $\otimes$ denotes the discrete convolution.

Steady state In this step, the complete impact of the repeating MSS is considered. The distributions $H_{m,t}$ are used to determine the distributions $H^{SS}_{s,t}$, the steady-state probability distributions of the number of recovering patients at time t on day s of the cycle ($s \in \{1, \ldots, S\}$).

Since the cyclic structure of the MSS implies that the recovery of patients receiving surgery during one cycle may overlap with patients from the next cycle, the distributions $H_{m,t}$ have to be overlapped in the correct manner. $H^{SS}_{s,t}$ can be computed as follows:

$$H^{SS}_{s,t} = \begin{cases} H_{s,t} \otimes H_{s+S,t} \otimes \ldots \otimes H_{s+\lceil M/S \rceil S,t} & , s = 1, \ldots, S-1, \\ H_{0,t} \otimes H_{S,t} \otimes \ldots \otimes H_{\lceil M/S \rceil S,t} & , s = S, \end{cases}$$

where $M = \max\{m \mid \exists t, x \text{ with } H_{m,t}(x) > 0\}$.

3.2 Demand Predictions for Acute Patients

Model input

Time. The AAC is the repeating cyclic arrival pattern of acute patients with a length of R days. For each single patient, day n counts the number of days after arrival.

Admissions. An acute patient type is characterized by patient group p, $p = 1, \ldots, P$, arrival day r, and arrival time θ, which is for notational convenience denoted by type $j = (p, r, \theta)$. The Poisson arrival process of patient type j has arrival rate λ^{j}.

Discharges. $P^j(n)$ is the probability that a type j patient stays n days, $n \in \{0, \ldots, L^j\}$. Given that a patient is discharged at day n, the probability of being discharged in time interval $[t, t+1)$ is given by $\tilde{m}^j_{n,t}$. By definition, $\tilde{m}^j_{0,t} = 0$ for $t \leq \theta$.

Single patient type In this first step we consider a single patient type j. We compute the probability $g^j_{n,t}(x)$ that on day n at time t, x patients are still in recovery. Admissions can take place during time interval $[\theta, \theta + 1)$ on day $n = 0$ and discharges during day $n = 0$ after time θ and during days $n = 1, \ldots, L^j$. Therefore, we calculate $g^j_{n,t}(x)$ as follows:

$$g^j_{n,t}(x) = \begin{cases} \tilde{a}^j_t(x) & , n = 0, t = \theta, \\ \tilde{d}^j_{n,t}(x) & , n = 0, t > \theta \text{ and } n = 1, \ldots, L^j, \end{cases}$$

where $\tilde{a}^j_t(x)$ represents the probability that x patients are admitted in time interval $[t, t+1)$ on day $n = 0$ and $\tilde{d}^j_{n,t}(x)$ is the probability that x patients are still in recovery at time t on day n. The derivations of $\tilde{a}^j_t$ and $\tilde{d}^j_{n,t}$ are by analogy with those probabilities for elective patients and may be found in [33].

Single cycle Now, we consider a single AAC in isolation. From the distributions $g^j_{n,t}(x)$, we can determine the distributions $G_{w,t}$, the distributions for the total number of recovering patients at time t on day w ($w \in \{1, \ldots, R, R+1, R+2, \ldots\}$) resulting from a single AAC by analogy with those probabilities for elective patients (see [33]).

Steady state In this step, the complete impact of the repeating AAC is considered. The distributions $G_{w,t}$ are used to determine the distributions $G^{SS}_{r,t}$, the steady-state probability distributions of the number of recovering patients at time t on day r of the cycle ($r \in \{1, \ldots, R\}$) by analogy with those probabilities for elective patients (see [33]).

3.3 Demand Predictions Per Care Unit

To determine the complete demand distribution of both elective and acute patients, we need to combine the steady-state distributions $H^{SS}_{s,t}$ and $G^{SS}_{r,t}$. In general, the MSS cycle and AAC are not equal in length, i.e., $S \neq R$. This has to be taken into account when combining the two steady-state distributions. Therefore, we define the new IFC length $Q = LCM(S, R)$, where the function LCM stands for *least* *common multiple*. Let $Z_{q,t}$ be the probability distribution of the total number of patients recovering at time t on day q during a time cycle of length Q:

$$Z_{q,t} = H^{SS}_{q \bmod S + S\cdot\mathbb{1}_{(q \bmod S=0)},t} \otimes G^{SS}_{q \bmod R + R\cdot\mathbb{1}_{(q \bmod R=0)},t} \ .$$

Let W^k be the set of specialties j whose operated patients are (preferably) admitted to unit k ($k \in \{1, \ldots, K\}$) and V^k the set of acute patient types j that are (preferably) admitted to unit k. Then, the demand distribution for unit k, $Z^k_{q,t}$, can be calculated by only considering the patients in W^k in equation (1), and V^k may be obtained by analogy (see [33]).

3.4 Bed Census Predictions

We translate the demand distributions $Z^k_{q,t}$, $k = 1, \ldots, K$, into bed census distributions $\hat{Z}_{q,t}$, the distributions of the number of patients present in each unit k at time t on day q. To this end, we require an allocation policy ϕ that uniquely specifies from a demand vector $\mathbf{x} = (x_1, \ldots, x_K)$ a bed census vector $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_K)$, in which x_k and $\hat{x}_k$ denote the demand for unit k and the bed census at unit k, respectively. Let $\phi(\cdot)$ be the function that executes allocation policy ϕ. Let $\hat{Z}^k_{q,t}$ denote the marginal distribution of the census at unit k given by distribution $\hat{Z}_{q,t}$. With M_k, the capacity of unit k in number of beds, we obtain:

$$\hat{Z}_{q,t}(\hat{\mathbf{x}}) = \left(\hat{Z}^1_{q,t}(\hat{x}_1), \ldots, \hat{Z}^K_{q,t}(\hat{x}_K)\right) = \sum_{\{\mathbf{x}|\hat{\mathbf{x}}=\phi(\mathbf{x})\}} \left\{\prod_{k=1}^{K} Z^k_{q,t}(x_k)\right\}.$$

We do not impose restrictions on the allocation policy ϕ other than specifying a unique relation between demand $\mathbf{x}$ and census configuration $\hat{\mathbf{x}}$. Recall that the underlying assumption is that a patient is transferred to his preferred unit when a bed becomes available. The policy ϕ also reflects the priority rules that are applied for such transfers. As an illustration, we present an example for an inpatient care facility with two care units of capacity M_1 and M_2, respectively:

$$\phi(\mathbf{x}) = \begin{cases} (x_1, x_2) & , x_1 \le M_1, x_2 \le M_2, \\ (M_1, \min\{x_2 + (x_1 - M_1), M_2\}) & , x_1 > M_1, x_2 \le M_2, \\ (\min\{x_1 + (x_2 - M_2), M_1\}, M_2) & , x_1 \le M_1, x_2 > M_2, \\ (M_1, M_2) & , x_1 > M_1, x_2 > M_2. \end{cases} \qquad (2)$$

Under this policy patients are assigned to their bed of preference if available and are otherwise misplaced to the other unit if beds are available there.

3.5 Performance Indicators

Based on the demand distributions $Z^k_{q,t}$ and the census distributions $\hat{Z}^k_{q,t}$ we are able to formulate a variety of performance indicators. We present a selection of such performance indicators, which will be used in Sect. 5 to evaluate the impact of different scenarios and interventions.

Demand and bed census percentiles Let $D^k_{q,t}(\alpha)$ and $\hat{D}^k_{q,t}(\alpha)$ be the α-th percentile of, respectively, demand and bed census at time t on day q:

$$D^k_{q,t}(\alpha) = \arg\min_x \left\{ \sum_{i=0}^{x} Z^k_{q,t}(i) \geq \alpha \right\}, \quad \hat{D}^k_{q,t}(\alpha) = \arg\min_x \left\{ \sum_{i=0}^{x} \hat{Z}^k_{q,t}(i) \geq \alpha \right\}.$$

(Off-)Peak demand Reducing peaks and drops in demand will balance bed occupancy and therefore allows more efficient use of available staff and beds. Define $\overline{P}^k_q(\alpha)$ ($\underline{P}^k_q(\alpha)$) and $\overline{P}^k(\alpha)$ ($\underline{P}^k(\alpha)$) to be the maximum (minimum) α-th demand percentile per day and over the complete cycle, respectively:

$$\overline{P}^k_q(\alpha) = \max_t \left\{ D^k_{q,t}(\alpha) \right\}, \qquad \overline{P}^k(\alpha) = \max_q \left\{ \overline{P}^k_q(\alpha) \right\},$$

$$\underline{P}^k_q(\alpha) = \min_t \left\{ D^k_{q,t}(\alpha) \right\}, \qquad \underline{P}^k(\alpha) = \min_q \left\{ \underline{P}^k_q(\alpha) \right\}.$$

Admission rate Patient admissions may increase the nursing workload. Let $\Lambda^k_{q,t}$ be the distribution of the number of arriving patients during time interval $[t, t+1)$ on day q who are preferably admitted to care unit k. To obtain $\Lambda^k_{q,t}$, we first determine $\bar{a}^j_{n,t}$, the distribution of the number of elective type j arrivals during time interval $[t, t+1)$ on day n ($n \in \{-1, 0\}$):

$$\bar{a}^j_{n,t}(x) = \sum_{y=0}^{C^j} c^j(y)\bar{a}^j_{n,t}(x|y) \quad , \text{ with } \quad \bar{a}^j_{n,t}(x|y) = \binom{y}{x}(e^j_n w^j_{n,t})^x(1-e^j_n w^j_{n,t})^{y-x}.$$

$\Lambda^k_{q,t}$ is then determined by taking the discrete convolution over all relevant arrival distributions of both elective and acute patient types:

$$\Lambda^k_{q,t} = \left\{ \bigotimes_{i=1}^{I} \left\{ \bigotimes_{j\in W^k: j\in b_{i,s'}} \bar{a}^j_{-1,t} \right\} \otimes \left\{ \bigotimes_{j\in W^k: j\in b_{i,s''}} \bar{a}^j_{0,t} \right\} \right\} \otimes \left\{ \bigotimes_{j\in V^k: r=r'} \tilde{a}^j_t \right\}, \tag{3}$$

where $s' = 1 + q \bmod S$, $s'' = q \bmod S + S \cdot \mathbb{1}_{(q \bmod S = 0)}$, $r' = q \bmod R + R \cdot \mathbb{1}_{(q \bmod R=0)}$, and $\bigotimes_{x \in \mathscr{X}} f_x$ denotes the discrete convolution over the probability distributions $f_x, x \in \mathscr{X}$. The first term in the right-hand side of (3) represents the elective patients who claim a bed at unit k ($j \in W^k$), who are operated in any OR, and who are admitted on the day $s' - 1$ before surgery or on the day s'' of surgery. The second term in the right-hand side of (3) represents the acute patients who claim a bed at unit k ($j \in V^k$) and who arrive on the corresponding day r' in the AAC.

Average bed occupancy Let $\rho^k_{q,t}$, ρ^k_q, and ρ^k be the average bed utilization rate at care unit k respectively at time t on day q, on day q, and over the complete cycle:

$$\rho^k_{q,t} = \frac{1}{M^k} \sum_{x=0}^{M^k} x \cdot \hat{Z}^k_{q,t}(x), \qquad \rho^k_q = \frac{1}{T} \sum_{t=0}^{T-1} \rho^k_{q,t}, \qquad \rho^k = \frac{1}{Q} \sum_{q=1}^{Q} \rho^k_q.$$

Rejection probability Let $R^{\phi,k}$ denote the probability that under allocation policy ϕ an admission request of an arriving patient for unit k has to be rejected, because all beds at unit k are already occupied and none of the alternative beds (prescribed by ϕ) are available. To determine $R^{\phi,k}$, we first determine $R^{\phi,k}_{q,t}$: the probability of such an admission rejection at time t on day q. $R^{\phi,k}$ is then calculated as follows:

$$R^{\phi,k} = \frac{1}{\sum_{q,t} E[\Lambda^k_{q,t}]} \sum_{q,t} E[\Lambda^k_{q,t}] R^{\phi,k}_{q,t}.$$

Let n indicate the number of arriving patients who are preferably admitted to unit k and $\mathbf{x} = (x_1, \ldots, x_K)$ the demand for each unit (in which these arrivals are already incorporated). Introduce $\mathscr{R}^{\phi,k}(\mathbf{x}, n)$, the number of rejected patients under allocation policy ϕ of the n arriving patients to unit k, and $Z^k_{q,t}(x_k|n)$ the probability that at time t on day q in total x_k patients demand a bed at unit k and n of them have just arrived. Then, $R^{\phi,k}_{q,t}$ is calculated by

$$\begin{aligned} R^{\phi,k}_{q,t} &= \frac{E[\#\text{ rejections at unit } k \text{ on time } (q,t)]}{E[\#\text{ arrivals to unit } k \text{ on time } (q,t)]} \\ &= \frac{1}{E[\Lambda^k_{q,t}]} \sum_{\mathbf{x}} \prod_{\ell \neq k} Z^\ell_{q,t}(x_\ell) \sum_n \mathscr{R}^{\phi,k}(\mathbf{x}, n) \Lambda^k_{q,t}(n) Z^k_{q,t}(x_k|n). \end{aligned} \tag{4}$$

For the derivation of $Z^k_{q,t}(x_k|n)$, let us first introduce the concept *cohort*. A cohort is a group of patients originating from a single instance of an OR block (electives) or admission time interval (acute patients). Then

$$Z^k_{q,t}(x_k|n) =$$

$$\frac{P[\text{Demand } x_k \text{ patients for unit } k \text{ on time } t \text{ on day } q \text{ of which } n \text{ are arriving in } [t, t+1)]}{P[n \text{ arrivals for unit } k \text{ on day } q \text{ in } [t, t+1)]}$$

$$= \frac{1}{\Lambda^k_{q,t}(n)} \sum_{\substack{y_{\sigma(1)},\dots,y_{\sigma(\Omega)},\\ n_{\sigma(1)},\dots,n_{\sigma(\omega)}:\\ \sum_i y_i = x_k, \sum_j n_j = n}} \left\{ \prod_{i=\omega+1}^{\Omega} f^{\sigma(i)}_{q,t}(y_{\sigma(i)}) \right\}$$
$$\left\{ \prod_{j=1}^{\omega} \alpha^{\sigma(j)}_{q,t}(y_{\sigma(j)}) \check{a}^{\sigma(j)}_{q,t}(n_{\sigma(j)}|y_{\sigma(j)}) \right\},$$

where Ω is the total number of cohorts, ω the number of cohorts that do generate arrivals during time interval $[t, t+1)$ on day q, and the permutation σ is such that the patient types $\sigma(1), \dots, \sigma(\omega)$ are the types that can generate those arrivals. Further, for notational convenience we introduce the function $f^i_{q,t}$ as $f^i_{q,t} = h^i_{q,t}$ for the elective patients and $f^i_{q,t} = g^i_{q,t}$ for acute patient types. Also, we introduce $\alpha^j_{q,t}$ as $\alpha^j_{q,t} = a^j_{q,t}$ for the elective patient types and $\alpha^j_{q,t} = \tilde{a}_t^{(p,q \bmod R + R \cdot \mathbb{1}_{q \bmod R = 0}, t)}$ for the acute patient types. It remains to define $\check{a}^j_{q,t}(n_j|y_j)$, the probability that for an arriving cohort, from the y_j patients present in total, n_j arrivals occur during time interval $[t, t+1)$:

$$\check{a}^j_{q,t}(n_j|y_j) = \binom{y_j}{n_j} (\nu^j_{n,t})^{n_j} (1 - \nu^j_{n,t})^{y_j - n_j},$$

where for elective patient types $\nu^j_{n,t} = \frac{w^j_{n,t} e^j_n}{e^j_n \sum_{k=0}^{t} w^j_{n,k} + e^j_{-1} \cdot \mathbb{1}_{(n=0)}}$ and for acute patient types $\nu^j_{n,t} = 1$.

$\mathscr{R}^{\phi,k}(\mathbf{x}, n)$ is uniquely determined by allocation policy ϕ. For example, for the case with $K = 2$ presented in (2), we have for unit $k = 1$:

$$\mathscr{R}^{\phi,1}(\mathbf{x}, n) = \begin{cases} \min\{n, x_1 - M_1\} & , x_1 \geq M_1, x_2 \geq M_2, \\ \max\{0, (x_1 - M_1) - (M_2 - x_2)\} & , x_1 \geq M_1, x_2 < M_2, n \geq (x_1 - M_1), \\ n - \max\{0, \min\{n, (M_2 - x_2 - [x_1 - M_1 - n])\}\} & , x_1 \geq M_1, x_2 < M_2, n < (x_1 - M_1), \\ 0 & , \text{otherwise}. \end{cases} \tag{5}$$

In (5), the first case reflects the situation in which all beds at care unit 2 are occupied so that all arriving patients who do not fit in unit 1 have to be rejected. The second and third cases reflect the situation that (some of) the arriving patients can be misplaced to unit 2 so that only a part of the arriving patients have to be rejected.

In the second case, the $(x_1 - M_1)$ patients that do not fit at unit 1 are all arriving patients. In the third case, some of the $(x_1 - M_1)$ patients were already present so that not all $(M_2 - x_2)$ beds at unit 2 can be used to misplace arriving patients.

Misplacement probability Let $M^{\phi,k}$ denote the probability that under allocation policy ϕ a patient who is preferably admitted to care unit k is admitted to another unit. The derivation of $M^{\phi,k}$ is equivalent to that of $R^{\phi,k}$. In (4), $\mathscr{R}^{\phi,k}(\mathbf{x}, n)$ has to be replaced by $\mathscr{M}^{\phi,k}(\mathbf{x}, n)$, which gives the number of misplaced patients under allocation policy ϕ of the n arriving patients to unit k and which is again uniquely determined by ϕ. Observe that for the two unit example presented in (2), we have:

$$\mathscr{M}^{\phi,1}(\mathbf{x}, n) = \begin{cases} \min\{x_1 - M_1, M_2 - x_2\} & , x_1 > M_1, x_2 < M_2, n \geq (x_1 - M_1), \\ \max\{0, \min\{n, (M_2 - x_2 - [x_1 - M_1 - n])\}\} & , x_1 > M_1, x_2 < M_2, n < (x_1 - M_1), \\ 0 & , \text{otherwise.} \end{cases}$$

Productivity Let $\mathscr{K}$ be a set of cooperating care units, i.e., units that mutually allow misplacements. Let $\mathscr{P}^{\mathscr{K}}$ reflect the productivity of the available capacity at care units $k \in \mathscr{K}$, defined as the number of patients that is treated per bed per year:

$$\mathscr{P}^{\mathscr{K}} = \frac{365}{Q} \frac{1}{\sum_{k \in \mathscr{K}} M^k} \sum_{k \in \mathscr{K}} \sum_{q,t} (1 - R_{q,t}^{\phi,k}) E[\Lambda_{q,t}^k] \tag{6}$$

Remark 1 (Approximation) Observe that the misplacement and rejection probabilities are an abstract approximation of complex reality. In our model, we count each time interval how many of the arriving patients have to be misplaced or rejected. Since we do not remove rejected patients from the demand distribution, it is likely that we overestimate the rejection and misplacement probabilities. However, also in reality strict rejections are often avoided: by postponing elective admissions, predischarging another patient, or letting acute patients wait at the emergency department. These are all undesired degradations of provided quality of care. Therefore, our method provides a secure way of organizing inpatient care services. It is applicable to evaluate performance for care unit capacities that give low rejection probabilities, thus when high service levels are desired, which is typically the case in healthcare.

4 Flexible Nurse Staffing

This section reviews two staffing models based on the bed census predictions described above as introduced in [32]. First, we discuss the requirements that need to be satisfied in setting appropriate staffing levels. Then we present the fixed staffing model and subsequently a model to find optimal staffing levels when float nurse pools are applied: the flexible staffing model.

4.1 Staffing Requirements

Recall that we consider a planning horizon of Q days ($q = 1, \ldots, Q$), during which each day is divided in T time intervals ($t = 0, 1, \ldots, T - 1$). The set of working shifts is denoted by $\mathscr{T}$, where a shift τ is characterized by its start time b_τ and its length ℓ_τ. Within the time horizon, (q, t) is a unique time interval and (q, τ) a unique shift. For notational convenience, $t \geq T$ indicates a time interval on a later day, e.g., $(q, T + 5) = (q + 1, 5)$. For each of K inpatient care units, with the capacity of unit k being M^k beds, staffing levels have to be determined for each shift (q, τ).

We consider two types of staffing policies: 'fixed' and 'flexible' staffing. Under fixed staffing, the number of nurses working in unit k during shift (q, τ), denoted by $s^k_{q,\tau}$, is completely determined in advance. In the flexible case, 'dedicated' staffing levels $d^k_{q,\tau}$ per unit are determined, together with the number of nurses $f_{q,\tau}$ available in a flex pool. The decision regarding the particular units to which the float nurses are assigned is delayed until the start of the execution of a shift. We assign float nurses to one and the same care unit for a complete working shift, to avoid frequent handovers, which increase the risk of medical errors. Thus, we obtain staffing levels $s^k_{q,\tau} = d^k_{q,\tau} + f^k_{q,\tau}$, $k = 1, \ldots, K$, where $f^k_{q,\tau}$ denotes the number of float nurses assigned to unit k from the available $f_{q,\tau}$. Taking into account the current bed census and the predictions on patient admissions and discharges, the allocation of the float nurses to care units at the start of a shift is decided according to a predetermined assignment procedure. We denote such an assignment procedure by π. For both staffing policies, we assume shifts to be non-overlapping, and for the flexible policy, we assume shifts to be equivalent for each care unit.

Our goal is to determine the most cost-efficient staffing levels such that certain quality-of-care constraints are satisfied. Because float nurses are required to be cross-trained, it is likely that these staff members are more expensive to employ. To be able to differentiate such costs, we therefore consider staffing costs ω_d for each dedicated nurse who is staffed for one shift and ω_f for each flexible nurse. Next, the nurse-to-patient ratio targets during shift (q, τ) are reflected by $r^k_{q,\tau}$, indicating the number of patients a nurse can be responsible for at any point in time. To keep track of the compliance to these targets, we define the concept 'nurse-to-patient coverage', or shortly 'coverage'. With x^k_t the number of patients present at unit k at a certain time (q, t), $b_\tau \leq t < b_\tau + \ell_\tau$, the coverage at that time is given by $r^k_{q,\tau} \cdot s^k_{q,\tau} / x^k_t$. Thus, a coverage of one or higher corresponds to a preferred situation.

Starting from the following quality-of-care requirements as prerequisites, we will formulate the fixed and flexible staffing models by which the most cost-effective staffing levels can be found:

(i) **Staffing minimum.** For safety reasons, at least S^k nurses have to be present at care unit k at any time.
(ii) **Coverage minimum.** The coverage at care unit k may never drop below β^k.
(iii) **Coverage compliance.** The long-run fraction of time that the coverage at care unit k is one or higher is at least α^k. We denote the expectation of the coverage

compliance at care unit k during shift (q, τ) by $c^k_{q,\tau}(\cdot)$; the arguments of this function depend on which staffing policy is considered. (Note that 'coverage compliance' is a measure defined for a shift, based on the measure 'coverage' that is defined for the time periods within that shift).

(iv) **Flexibility ratio.** To ensure continuity of care, at any time, the fraction of nurses at care unit k that are dedicated nurses has to be at least γ^k.

(v) **Fair float nurse assignment.** The policy π, according to which the allocation of the available float nurses to care units at the start of a shift is done, has to be 'fair'. Fairness is defined as assigning each next float nurse to the care unit where the expected coverage compliance during the upcoming shift is the lowest.

4.2 Fixed Staffing

When only dedicated staffing is allowed, there is no interaction between care units. Therefore, the staffing problem decomposes in the following separate decision problems for each care unit k and each shift (q, τ):

$$\min \quad z_F = \omega_d s^k_{q,\tau} \tag{7}$$

$$\text{s.t.} \quad s^k_{q,\tau} \geq S^k \tag{8}$$

$$s^k_{q,\tau} \geq \left\lceil \beta^k \cdot M^k / r^k_{q,\tau} \right\rceil \tag{9}$$

$$c^k_{q,\tau}\left(s^k_{q,\tau}, r^k_{q,\tau}\right) \geq \alpha^k \tag{10}$$

The constraints (8), (9), and (10) reflect requirements (i), (ii), and (iii), respectively. Let $X^k_{q,t}$ be the random variable with bed census distribution $\hat{Z}^k_{q,t}$ counting the number of patients present on care unit k at time (q, t). Then, the coverage compliance in (10) can be calculated as follows:

$$c^k_{q,\tau}\left(s^k_{q,\tau}, r^k_{q,\tau}\right) = \mathbb{E}\left[\frac{1}{\ell_\tau} \sum_{t=b_\tau}^{b_\tau+\ell_\tau-1} \mathbb{1}\left(X^k_{q,t} \leq s^k_{q,\tau} \cdot r^k_{q,\tau}\right)\right]$$
$$= \frac{1}{\ell_\tau} \sum_{t=b_\tau}^{b_\tau+\ell_\tau-1} \sum_{x=0}^{s^k_{q,\tau} \cdot r^k_{q,\tau}} \hat{Z}^k_{q,t}(x).$$

Observe that $\sum_{x=0}^{s^k_{q,\tau} \cdot r^k_{q,\tau}} \hat{Z}^k_{q,t}(x)$ reflects the probability that with staffing level $s^k_{q,\tau}$ and under ratio $r^k_{q,\tau}$ the nurse-to-patient ratio target is satisfied during time interval

$[t, t+1)$. The optimum of (7) is found by choosing the minimum $s^k_{q,\tau}$ satisfying constraints (8) and (9) and increasing it until constraint (10) is satisfied.

4.3 Flexible Staffing

The next step is to formulate the flexible staffing model. Note that for requirements (i) and (ii), the constraints are similar to those for fixed staffing. Under the assumption $\omega_d \leq \omega_f$, we can replace $s^k_{q,\tau}$ by $d^k_{q,\tau}$ in (8) and (9). Due to the presence of a flex pool, the care units cannot be considered in isolation anymore. Hence, constraint (10) has to be replaced. An assignment procedure has to be formulated that fulfills requirement (v), and this assignment procedure influences the formulation of the constraint for requirement (iii). In addition, a constraint needs to be added for requirement (iv).

For an assignment procedure π that allocates the float nurses to care units at the start of a shift (q, τ), let $g^{\pi}_{q,\tau}(\boldsymbol{d}, f, \boldsymbol{y}) = (g^{1,\pi}_{q,\tau}(\boldsymbol{d}, f, \boldsymbol{y}), \ldots, g^{K,\pi}_{q,\tau}(\boldsymbol{d}, f, \boldsymbol{y}))$ be the vector denoting the number of float nurses assigned to each care unit, when f flex nurses are available to allocate, the number of staffed dedicated nurses equals $\boldsymbol{d} = (d^1, \ldots, d^K)$, and the census at the different care units at time (q, b_τ) equals $\boldsymbol{y} = (y^1, \ldots, y^K)$. A vector of the type $\boldsymbol{y}$ reflects what we will call a *census configuration.*

Let π^* denote the assignment procedure that ensures constraint (v). The assignment procedure π^* depends on $\boldsymbol{d}_{q,\tau}$, $f_{q,\tau}$, and $r^k_{q,\tau}$, $k = 1, \ldots, K$ and therefore the coverage as well. Hence, requirement (v) gives a constraint of the form $c^k_{q,\tau}(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, r^k_{q,\tau}) \geq \alpha^k$. However, assignment procedure π^* depends on the census configuration $\boldsymbol{y}$ at time (q, b_τ), so calculation of the coverage compliance first requires the computation of $c^k_{q,\tau}(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, r^k_{q,\tau}; \boldsymbol{y})$, which describes the coverage compliance, given that at the start of shift (q, τ) census configuration $\boldsymbol{y}$ is observed. Then, the coverage compliance is given by

$$c^k_{q,\tau}\big(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, r^k_{q,\tau}\big) = \sum_{\boldsymbol{y}} \Big\{ c^k_{q,\tau}\big(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, r^k_{q,\tau}; \boldsymbol{y}\big) \prod_{w=1}^{K} \hat{Z}^w_{q,b_\tau}(y^w) \Big\}.$$

Using $c^k_{q,\tau}(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, r^k_{q,\tau}; \boldsymbol{y})$, the assignment policy π^* satisfying requirement (v) is the one that satisfies

$$g^{\pi^*}_{q,\tau}(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, \boldsymbol{y}) = \underset{\left\{(f^1_{q,\tau}, \ldots, f^K_{q,\tau}) : \sum_k f^k_{q,\tau} = f_{q,\tau}\right\}}{\arg\max} \; \min_k \; c^k_{q,\tau}\big(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, r^k_{q,\tau}; \boldsymbol{y}\big). \tag{11}$$

Applying policy π^* provides $s^k_{q,\tau}(\boldsymbol{y})$, the number of nurses staffed at care unit k if census configuration $\boldsymbol{y}$ is observed at the start of shift (q, τ). Hence, the flexible model for each shift (q, τ) is the following, where constraints (13)–(17) reflect (i)–(v), respectively: :

$$\min \quad z_E = \omega_f f_{q,\tau} + \omega_d \sum\nolimits_k d^k_{q,\tau} \tag{12}$$

$$\text{s.t.} \quad d^k_{q,\tau} \geq S^k, \qquad \text{for all } k, \tag{13}$$

$$d^k_{q,\tau} \geq \left\lceil \beta^k \cdot M^k / r^k_{q,\tau} \right\rceil, \qquad \text{for all } k, \tag{14}$$

$$c^k_{q,\tau}\left(\boldsymbol{d}_{q,t}, f_{q,\tau}, r^k_{q,\tau}\right) \geq \alpha^k, \qquad \text{for all } k, \tag{15}$$

$$d^k_{q,\tau} \geq \gamma^k \cdot s^k_{q,\tau}(\boldsymbol{y}), \qquad \text{for all } k, \boldsymbol{y}, \tag{16}$$

$$s^k_{q,\tau}(\boldsymbol{y}) = d^k_{q,\tau} + g^{k,\pi^*}_{q,\tau}\left(\boldsymbol{d}_{q,\tau}, f_{q,\tau}, \boldsymbol{y}\right), \qquad \text{for all } k, \boldsymbol{y}. \tag{17}$$

Finding the optimum for (12) requires the computation of $c^k_{q,\tau}(\boldsymbol{d}, f_{q,\tau}, r^k_{q,\tau}; \boldsymbol{y})$ by considering every sample path of census configurations during a shift. For realistic instances, it is computationally expensive to find the optimal solution for $d^1_{q,\tau}, \ldots, d^K_{q,\tau}, f_{q,\tau}$ (see [32]).

5 Quantitative Results

This section presents the experimental results. We present a selection of results taken from [32, 33].

5.1 *Case Study Description*

The case study entails six surgical specialties of the university hospital AMC, which together have 104 beds in operation. The entire hospital has 20 operating rooms and 30 inpatient departments, with a total of 1000 beds. The following specialties are taken into account: traumatology (TRA), orthopedics (ORT), plastic surgery (PLA), urology (URO), vascular surgery (VAS), and general surgery (GEN). In the present setting, the patients of the abovementioned specialties are admitted to four different inpatient care departments. On Floor I, care unit A houses GEN and URO and unit B VAS and PLA. On Floor II, care unit C houses TRA and unit D ORT.

The physical building is such that units A and B are physically adjacent (Floor I), as are units C and D (Floor II). For these specialties, we have historical data available over 2009–2010 on 3498 (5025) elective (acute) admissions, with an average length of stay (LOS) of 4.85 days (see Table 1). At the time of the original study, no cyclic MSS was applied. Each time, roughly 6 weeks in advance the MSS was determined for a period of 4 weeks. The capacities of units A, B, C, and D are 32, 24, 24, and 24 beds, respectively. The utilizations over 2009–2010 were 53.2%, 55.6%, 54.4%, and 60.6%, respectively (which includes some patients from other specialties that were placed in these care units). These utilizations reflect administrative bed census,

Table 1 Overview of historical data 2009–2010

Specialty	Acronym	Care unit	Elective admissions	Acute admissions	Average LOS (in days)	Load[a] (# patients)
General surgery	GEN	A	611	901	3.31	6.88
Urology	URO	A	818	1157	3.68	9.99
Vascular surgery	VAS	B	257	634	8.30	10.16
Plastic surgery	PLA	B	639	288	2.29	2.91
Traumatology	TRA	C	337	1200	5.88	12.41
Orthopedics	ORT	D	836	845	6.23	14.38

[a] Load = Expected number of patient arrivals per day $*$ Average LOS

which means the percentage of time that a patient physically occupies a bed or keeps it reserved during the time the patient is at the operating theater or at the intensive care department. Unfortunately, no confident data was available on rejections and misplacements.

Working days are divided in three shifts: the day shift (8:00–15:00), the evening shift (15:00–23:00), and the night shift (23:00–8:00). These time intervals indicate the times that nurses are responsible for direct patient care. Around these time intervals, the working shifts also incorporate time for patient handovers, indirect patient care, and professional development. At all times, there should be at least two nurses present at each care unit. According to agreements on working conditions for nurses in all university hospitals in the Netherlands, the contractual number of annual working hours per full-time equivalent (FTE) is 1872. The number of hours that one FTE can be employed for direct nursing care, after deduction of time reserved for professional development, holiday hours, and sick leave, is 1525.7 on average (also see [16]). The yearly cost per FTE, including all costs and bonuses, is roughly €53,000.

The nurse-to-patient ratio targets prescribed by the board of the AMC for the care units of interest are 1:4 during the day shifts, 1:6 during the evening shifts, and 1:10 during the night shifts. At the time of the original study, the current staffing practice was based on the number of beds in service, independent of whether they were occupied, and no float nurse pools were employed. Thus, for example, for a care unit size of 24 beds and staffing ratio of 1:4, the number of dedicated nurses to staff was always 6. A scarcity of nursing capacity frequently leads to the expensive hiring of temporary nurses from external agencies, as well as to undesirable ad hoc bed closings. Also, the prescribed staffing levels cannot always be realized in practice. As a result, the inpatient care units experienced a lack of consistency in the delivered quality of nursing care.

We have estimated the input parameters for our model based on historical data of 2009–2010 from the hospital's electronic databases. The event logs of the operating room and inpatient care databases had to be matched. Since the data contained many errors, extensive cleaning was required. Patients of other specialties who stayed at departments A to D have been deleted. No cyclical MSS was applied in practice; therefore, in our model we set the MSS length at 2 years, following the surgery blocks as occurred in practice during 2009–2010. Elective surgery blocks are only executed on weekdays. For the elective patient types, the distributions for the number of surgeries and for the admission/discharge processes are estimated per specialty. We set the length of the AAC at 1 week. For the acute patients, the discharge distributions are estimated per specialty, and to have enough measurements, via the following clustering: admission time intervals 0–8, 8–18, and 18–24. Furthermore, for all patient types, the discharge distributions during a day are assumed to be equal for the days $n \geq 2$.

5.2 Case Study Results: Bed Census

To illustrate the potential of the presented staffing methodology for the case study, we present a selection of the interventions presented in [33]. For the interventions that are based on the current MSS, we run the model for the estimated 2-year MSS, and we calculate the performance measures only over the second year, to account for warm-up effects. To assess the effects of the interventions, we first evaluate the performance of the base case scenario. In all experiments, no ad hoc closings are allowed, as decided by the hospital board to be the near-future policy. Note that the calculated rejection and misplacement percentages are therefore most likely an underestimation of current practice (of which no reliable data is available). The productivity measure is calculated per floor, since the misplacement policy implies that capacity is 'shared' per floor. The following interventions are considered, of which the results are displayed in Tables 2 and 3:

(0) Base case. To assess the effects of the interventions, we first evaluated the performance of a base case scenario, which is the situation that most closely resembles current practice. The base case involved the current bed capacities and misplacements between care units A and B (Floor I) and between units C and D (Floor II).

(1) Rationalize bed requirements. The numbers of beds in the base case are a result of historical development. Given particular service requirements, we determine whether the number of beds can be reduced to achieve a higher bed utilization while a certain quality level is guaranteed. We consider rejection probabilities not exceeding 5%, 2.5%, and 1%. Often, there are different bed configurations with the same total number of beds per floor, satisfying a given maximum rejection probability. Per floor, from the available configurations, the one is chosen that gives the lowest maximum misplacement probability.

It can be seen that a significant reduction in the number of beds is possible. However, the overall bed utilizations are still modest, because demand drops during weekend days when no elective surgeries take place. In addition, there is a correlation between moments of higher census and moments that patients arrive, which leads to higher rejection probabilities compared to, for instance, a stationary Poisson arrival process.

Table 2 The numerical results for the base case, intervention 1, and intervention 2 (with the productivity-$\Delta\%$ relative to the base case)

Intervention	Unit	Capacity (# beds)	Rejection (%)	Misplace (%)	Utilization (%)	Floor	Capacity (# beds)	Productivity (eq.(6))	Productivity ($\Delta\%$)
Base case	A	32	0.14	1.85	56.9	AB	56	50.0	–
	B	24	0.08	1.22	56.5				
	C	24	0.03	0.45	55.6	CD	48	35.1	–
	D	24	0.10	3.68	61.5				
1. Rationalize bed requirements									
Rejection <5%	A	27	4.92	6.07	67.7	AB	45	59.3	+18.6
	B	18	4.59	14.35	74.3				
	C	18	3.42	8.90	74.0	CD	38	42.5	+21.1
	D	20	4.92	11.72	73.3				
Rejection <2.5%	A	28	2.31	5.86	65.0	AB	48	57.2	+14.4
	B	20	1.67	7.30	67.7				
	C	18	2.02	10.30	73.3	CD	40	41.3	+17.5
	D	22	2.27	6.14	67.5				
Rejection <1%	A	29	0.94	5.00	62.6	AB	51	54.5	+9.1
	B	22	0.52	3.15	61.8				
	C	20	0.54	4.39	66.5	CD	43	39.0	+11.0
	D	23	0.79	4.93	64.3				
2. Change operational process									
Rejection <5%	A	24	4.51	9.24	66.4	AB	43	62.5	+25.2
	B	19	3.03	6.53	66.1				
	C	17	3.65	11.21	74.3	CD	37	43.6	+24.2
	D	20	5.00	9.12	69.7				
Rejection <2.5%	A	26	2.31	5.22	61.7	AB	45	60.9	+21.8
	B	19	2.03	7.54	65.7				
	C	17	2.11	12.74	73.8	CD	39	42.3	+20.5
	D	22	2.28	4.62	64.0				
Rejection <1%	A	27	0.94	4.44	59.3	AB	48	57.9	+15.8
	B	21	0.64	3.26	59.7				
	C	19	0.58	5.59	66.8	CD	42	39.9	+13.6
	D	23	0.83	3.78	60.7				

(2) Change operational process. Hospital management proposes to admit all elective patients on the day of surgery, since admitting patients the day before surgery is often induced by logistical reasons and not by medical necessity. Second, to reduce census peaks during the middle of the day, management proposes to aim for discharges to happen before noon. To predict the potential impact of these changes in the operational process, we mimic the changes as follows: we adjust the admission distributions of elective patients, so that admissions on the day before surgery are postponed to time $t = 8$ on the day of surgery (which impacts 81.9% of the elective patients), and we adjust

Table 3 The numerical results for interventions 3 and 4 (with the productivity-$\Delta\%$ relative to the base case)

		Capacity	Rejection	Misplace	Utilization		Capacity	Productivity	
Intervention	Unit	(# beds)	(%)	(%)	(%)	Floor	(# beds)	(eq.(6))	($\Delta\%$)
3. Balance MSS									
Rejection <5%	A	25	4.85	8.43	74.5	AB	44	62.5	+25.0
	B	19	3.93	8.73	74.4				
	C	18	3.24	8.84	74.6	CD	38	43.5	+23.7
	D	20	3.99	10.03	75.6				
Rejection <2.5%	A	27	2.25	4.29	69.5	AB	46	61.1	+22.3
	B	19	2.41	10.25	73.9				
	C	19	1.46	6.21	70.8	CD	40	42.1	+19.9
	D	21	1.86	7.50	72.2				
Rejection <1%	A	28	0.83	3.57	66.7	AB	49	58.3	+16.6
	B	21	0.66	4.32	67.4				
	C	20	0.60	4.05	67.3	CD	42	40.5	+15.3
	D	22	0.79	5.21	69.0				
4. Combination (1), (2), and (3)									
Rejection <5%	A	23	4.92	9.17	70.9	AB	42	65.5	+31.1
	B	19	3.47	5.56	68.9				
	C	17	3.77	11.04	74.9	CD	37	44.5	+26.5
	D	20	4.21	7.34	71.7				
Rejection <2.5%	A	25	2.28	4.72	65.7	AB	44	64.0	+28.0
	B	19	2.18	6.85	68.4				
	C	18	1.74	7.87	71.0	CD	39	43.1	+22.7
	D	21	2.02	5.54	68.2				
Rejection <1%	A	26	0.82	3.90	63.1	AB	47	60.8	+21.7
	B	21	0.57	2.75	62.2				
	C	19	0.74	5.21	67.5	CD	41	41.4	+18.0
	D	22	0.89	3.87	65.1				

the discharge distributions of days $n \geq 1$, so that discharges later than time $t = 11$ are moved forward to $t = 11$ (which impacts 51.8% of the total patient population).

Compared to intervention 1, the number of beds can be further decreased. Also, the results indicate that the care unit managers of these departments should not only focus on achieving high bed utilizations: although somewhat lower utilization is achieved, productivity is significantly increased.

(3) Balance MSS. The outcomes of the previous experiments showed that the MSS that was realized in practice created artificial demand variability. This intervention estimates the potential of a cyclical MSS that is designed with the purpose to balance bed census. We created a cyclical MSS with a length of 4 weeks. First, for each specialty, an integer number of OR blocks is chosen so that an output is achieved similar to the original MSS; due to this integrality

average demand is slightly increased. Second, these blocks have been manually divided over the days in the MSS, and by trial and error, a more balanced outflow was realized.

(4) Combination (1), (2), and (3). This intervention combines interventions (1), (2), and (3). Hospital management agreed upon a service level norm of rejection probabilities <2.5%. Under this requirement, it is possible to reduce the number of beds by 20% (from 104 to 83) and increase productivity by roughly 25%. Considering that the AMC has 30 inpatient departments, the savings potential for the entire hospital seems substantial.

5.3 Case Study Results: Nurse Staffing

In this section, we present the results for the case study on the interventions described in Sect. 5.2. We investigate both the value of aligning staffing levels with bed census predictions and of employing float nurses, by comparing the results of the fixed and flexible staffing models with the current staffing policy, which we refer to as 'full staffing'. With a care unit capacity of M^k beds at unit k, under the full staffing policy, $\lceil M^k/r^k_{q,\tau}\rceil$ nurses are required at all times.

The intended AMC practice will be that registered nurses will alternately be rostered as a dedicated or float nurse. Therefore, we consider the case in which dedicated and float nurses are equally expensive, i.e., $\omega_d = \omega_f$. In addition to the fixed input $S^k = 2$, and staffing cost dedicated nurse = 1, board of the AMC has chosen to deploy the following quality of care requirements: nurse-to-patient ratios $r^k_{q,1} = 4$, $r^k_{q,2} = 6$, $r^k_{q,3} = 10$, minimum coverage $\beta^k = 0.70$, coverage compliance $\alpha^k = 0.90$, and at least two out of three nurses should be dedicated nurses, i.e., $\gamma^k = 0.67$.

The detailed results are displayed in Tables 4 and 5. Table 6 provides an overview of the results for the various interventions and includes the calculation of the

Table 4 The numerical results for the base case (Floor I: 56 beds, 56.7% utilization; Floor II: 48 beds, 58.6% utilization; with the FTE-Δ% relative to full staffing)

		Full staffing	*Fixed staffing*			*Flexible staffing*			
		FTE	Average	FTE		Error	Average	FTE (float)	
Intervention	*Floor*	(#)	coverage	(#)	(Δ%)	bound (%)	coverage	(#)	(Δ%)
Base case									
$\alpha = 0.85$	I	57.7	0.96	44.8	−22.2	0.4	0.96	44.7 (1.7)	−22.4
	II	48.3	0.96	38.9	−19.5	0.0	0.95	38.8 (2.0)	−19.7
$\alpha = 0.90$	I	57.7	0.98	46.0	−20.3	0.8	0.97	45.7 (2.7)	−20.8
	II	48.3	0.97	40.0	−17.3	0.1	0.97	39.6 (2.8)	−18.0
$\alpha = 0.95$	I	57.7	0.99	47.9	−16.9	1.4	0.99	47.4 (4.6)	−17.8
	II	48.3	0.99	42.5	−12.1	0.4	0.99	41.1 (4.3)	−14.9

Table 5 The numerical results for the various interventions (with the FTE-$\Delta\%$ relative to full staffing)

			Full staffing	*Fixed staffing*			*Flexible staffing*		
	Capacity	*Utilization*	FTE	Average	FTE		Average	FTE (float)	
Intervention	(# beds)	(%)	(#)	coverage	(#)	($\Delta\%$)	coverage	(#)	($\Delta\%$)
1. Rationalize bed requirements									
Floor I	48	66.1	48.1	0.99	43.8	−8.9	0.98	43.3 (6.2)	−9.9
Floor II	40	70.1	42.6	0.99	39.3	−7.8	0.98	38.7 (5.2)	−9.1
2. Change operational process									
Floor I	45	63.4	48.1	0.98	41.8	−13.0	0.98	41.6 (4.4)	−13.5
Floor II	39	68.3	42.6	0.98	38.4	−9.9	0.98	37.2 (6.9)	−12.7
3. Balance MSS									
Floor I	46	71.3	48.1	0.99	45.7	−5.0	0.99	44.9 (7.8)	−6.7
Floor II	40	71.5	44.5	0.98	40.9	−8.2	0.98	39.6 (6.1)	−11.0
4. Combination (1), (2), and (3)									
Floor I	44	66.9	48.1	0.98	42.4	−11.7	0.98	41.8 (6.4)	−13.1
Floor II	39	69.5	42.6	0.98	38.8	−8.8	0.98	38.1 (4.6)	−10.6

Table 6 FTE and productivity results for all interventions (with both the FTE-$\Delta\%$ and the productivity-$\Delta\%$ relative to full staffing in the base case)

	Full staffing				*Fixed staffing*				*Flexible staffing*			
	FTE		Productivity		FTE		Productivity		FTE		Productivity	
Intervention	(#)	($\Delta\%$)	(#/yr)	($\Delta\%$)	(#)	($\Delta\%$)	(#/yr)	($\Delta\%$)	(#)	($\Delta\%$)	(#/yr)	($\Delta\%$)
Base case	106.0	–	42.3	–	85.9	−18.9	52.2	+23.3	85.3	−19.5	52.6	+24.2
(1)	90.7	−14.4	48.5	+14.5	83.1	−21.6	52.9	+25.0	82.1	−22.6	53.5	+26.5
(2)	90.7	−14.4	48.4	+14.4	80.2	−24.3	54.7	+29.4	78.7	−25.7	55.8	+31.8
(3)	92.6	−12.6	48.6	+14.8	86.5	−18.4	52.0	+22.8	84.5	−20.3	53.2	+25.8
(4)	90.7	−14.4	49.6	+17.2	81.3	−23.3	55.3	+30.7	79.8	−24.7	56.3	+33.0

Productivity: number of patients treated per employed FTE per year

productivity measure of the number of patients treated per employed FTE per year.

(0) Base case. First, we evaluate the performance of the base case scenario (see Table 4). In the flexible staffing policy, two flex pools are installed, one on each floor; we therefore present the results per floor. For the base case, we show three values for the coverage compliance threshold ($\alpha^k = \{0.85, 0.90, 0.95\}$) to illustrate the effect of this quality-of-care constraint on required nursing capacity.

The number of FTEs required is calculated by summing the total number of staffed nurse hours and dividing by the 1525.7 direct nursing hours that one FTE has available. Note that in this calculation we do not include scheduling restrictions that might be involved when assigning individual nurses to working shifts. Therefore, at a particular inpatient clinic, the number of FTEs to hire might need to be larger than the displayed number of FTEs required, depending on the local labor regulations and nurse rostering practice.

For both the fixed and the flexible staffing models, it turns out that the realized coverage compliance is, on average, much higher than the minimum requirement. This result occurs because when the coverage compliance constraint is slightly violated, an additional nurse needs to be staffed, which significantly increases the coverage compliance because this nurse can care for $r^k_{q,\tau}$ patients. Although full staffing ensures a coverage compliance of 100%, it frequently overstaffs care units. It is clear that the acceptance of slight coverage reductions (still realizing average coverage compliances higher than 95%) allows managers to better match care supply and demand, thereby realizing efficiency gains of 12–22%. The largest gain is achieved by the staffing based on census predictions (see results of the fixed model). The additional value of employing float nurses is case dependent, and in most cases, the value is higher with increasing α^k due to the increasing gap with the minimum coverage requirement set by β^k.

(1) Rationalize bed requirements. Intervention (1) rationalizes the care unit dimensions. Table 5 shows that fixed staffing with $\alpha^k = 0.90$ reduces nursing capacity requirements by 8–9% compared to full staffing and flexible staffing yields an additional 1% reduction. Table 6 indicates the gain against current practice: 22.6% reduction in FTE requirements, with a simultaneous increase of staff productivity by 26.5%.

(2) Change operational process. Intervention (2) focuses on changes in the operational process that shorten the average lengths of stay. The reduction of demand and its variability lowered the number of beds required. Here, we see that our staffing methodology also translates this into significantly lower staff requirements, as well as higher productivity.

(3) Balance MSS. Intervention (3) intends to decrease the artificial demand variability by designing a balanced cyclic MSS. Note that due to the integrality of the number of scheduled operating room blocks, the resulting MSS has slightly increased patient demand. Therefore, its impact on staffing requirements is not directly evident.

(4) Combination (1), (2), and (3). Intervention (4) outperforms all previous configurations on the productivity measure. As an illustration, the effect of staffing levels following bed census demand patterns, including the differences between fixed and flexible staffing therein, is visualized in Fig. 1. Also, in this figure, average demand is displayed for day shifts in a 4-week period as the average bed census divided by the applied nurse-to-patient ratios. It signals that the high variability in bed census implies that the number of nurses to be staffed, to guarantee the coverage compliance on the nurse-to-patient ratios, is considerably higher than average demand. It is a clear indication of the savings potential of increasing the predictability of demand for nursing staff by balancing bed census.

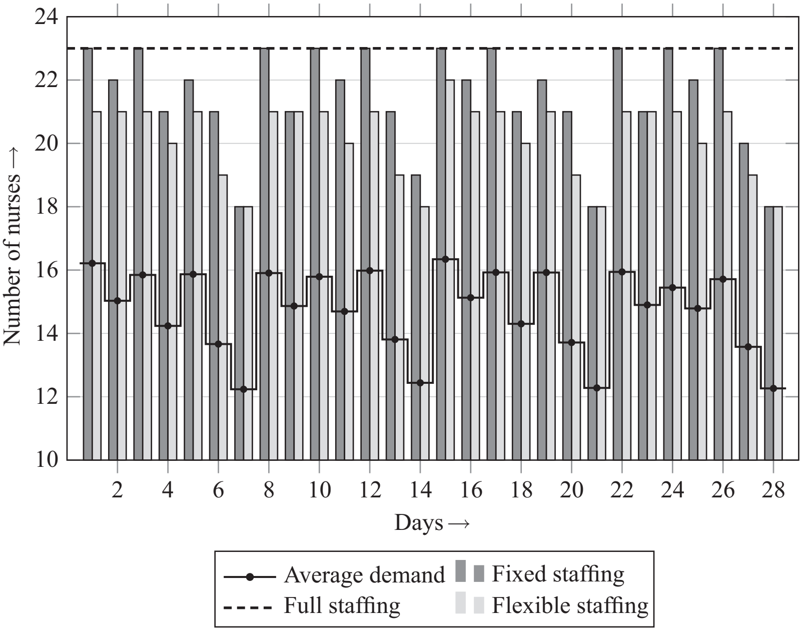

Fig. 1 Total staffing levels for day shifts during the 4-week period starting on Monday January 25 (the average demand pattern shows the average census divided by ratios $r^k_{q,1}$)

Finally, let us state two general insights. First, note that under the old (full) staffing policy, a reduction in the number of beds not always translates into a reduction in staffing requirements. This is the case when the number of beds does not decrease to a capacity level such that it crosses a level that is a multiple of one of the nurse-to-patient ratios. Second, based on our results, we cannot deduce general rules of thumb for the potential of float nurses. The outcomes for each particular care unit are a complex interplay between care unit sizes, nurse-to-patient ratios, and the shapes of the bed census distributions.

6 Discussion

Rising healthcare costs and increasing nurse shortages make cost-effective nurse staffing of utmost importance. In many hospitals, staffing levels are a result of historical development, given that hospital managers lack the tools to base current staffing decisions on information about future patient demand. In this chapter, combining the results of [32, 33], we have presented a generic analytical method that can quantitatively support decision-making about required staffing levels in inpatient care facilities. We have demonstrated its potential with a case study of the AMC, for which we have shown that, by achieving coherence between patient demand and staffing supply, simultaneous cost reductions and quality of care improvements are possible.

The combined application of the bed census prediction model and the staffing models from the present chapter enables hospital administrators to gain insight into the value of integrated decision-making. The interrelation between decisions, such as case mix, care unit partitioning, care unit size, and admission/discharge times, is made explicit. Because the demand prediction model incorporates the operating room block schedule and the patient arrival pattern from the emergency department, the presented methodology also facilitates alignment between the design and operations of the inpatient care facility and its surrounding departments. With this integrated framework, staffing effectiveness can be attained in three steps. First, the method can help to reduce artificial variability of bed occupancies, for example, by adjusting the operating room schedule. Second, by predicting the bed census distributions and determining staffing levels for dedicated nurses accordingly, the predictive part of the remaining variability can be anticipated. Third, to be able to effectively respond to random variability, adequately sized float nurse pools can be created.

The case study of the AMC provides an example of how the methodology can be applied in practice. Nurse staffing is high on the agenda because staffing costs account for 66% of the total expenses in the AMC. We have applied our staffing models to data from several care units, and we presented results from four of them in this chapter. The formulations of all interventions and the eventual parameter settings are the results of close cooperation between operations researchers and hospital managers from different levels within the organization. This collaboration resulted in the joint conclusion that substantial efficiency gains are possible while improving upon the adherence to nurse-to-patient ratio targets.

To fully exploit the potential of the forecasting and staffing method, based on the bed census prediction and staffing models presented in this chapter, in cooperation with Rhythm, we developed a user-friendly decision support tool (DSS). The prediction model, and thus the software tool, relies on data which is easily extractable from typical hospital management information systems. This makes it possible to automate the process of collecting the required input parameters to run the model. The DSS allows for visualization of the results and the possibility to run what-if scenarios. Building on these preconditions, the example of the AMC has been taken forward by various Dutch hospitals. Each hospital has its own configuration and its own application of the tool. As a consequence, the methods presented in this chapter have shown to be powerful in different hospital settings. Some hospitals have been able to achieve more balanced bed occupancies, thereby reducing admission rejections, misplacements, and costs. Others also improved their nurse staffing, thereby significantly reducing under- and overstaffing. Integration with hospital management systems and integration with nurse rostering software will be the next development steps. In addition, pilots to implement our DSS outside the Netherlands have started.

References

1. Adan, I., Bekkers, J., Dellaert, N., Vissers, J., and Yu, X. (2009). Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning. *Health Care Management Science*, 12(2):129–141.
2. Aiken, L., Sermeus, W., van den Heede, K., Sloane, D., Busse, R., McKee, M., Bruyneel, L., Rafferty, A., Griffiths, P., Moreno-Casbas, M., Tishelman, C., Scott, A., Brzostek, T., Kinnunen, J., Schwendimann, R., Heinen, M., Zikos, D., Sjetne, I. S., Smith, H., and Kutney-Lee, A. (2012). Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. *British Medical Journal*, 344(3):1–14.
3. Akcali, E., Coté, M., and Lin, C. (2006). A network flow approach to optimizing hospital bed capacity decisions. *Health care management science*, 9(4):391–404.
4. Akkerman, R. and Knip, M. (2004). Reallocation of beds to reduce waiting time for cardiac surgery. *Health Care Management Science*, 7(2):119–126.
5. Bekker, R. and De Bruin, A. (2010). Time-dependent analysis for refused admissions in clinical wards. *Annals of Operations Research*, 178(1):45–65.
6. Bekker, R. and Koeleman, P. (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Management Science*, pages 1–13.
7. Beliën, J. and Demeulemeester, E. (2008). A branch-and-price approach for integrating nurse and surgery scheduling. *European Journal of Operational Research*, 189(3):652–668.
8. Beliën, J., Demeulemeester, E., and Cardoen, B. (2009). A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling*, 12(2):147–161.
9. Burke, E., de Causmaecker, P., vanden Berghe, G., and van Landeghem, H. (2004). The state of the art of nurse rostering. *Journal of Scheduling*, 7(6):441–499.
10. Cheang, B., Li, H., Lim, A., and Rodrigues, B. (2003). Nurse rostering problems–a bibliographic survey. *European Journal of Operational Research*, 151(3):447–460.
11. Chiaramonte, M. and Chiaramonte, L. (2008). An agent-based nurse rostering system under minimal staffing conditions. *International Journal of Production Economics*, 114(2):697–713.
12. Cochran, J. and Bharti, A. (2006). Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science*, 9(1):31–45.
13. Costa, A., Ridley, S., Shahani, A., Harper, P., De Senna, V., and Nielsen, M. (2003). Mathematical modelling and simulation for planning critical care capacity. *Anaesthesia*, 58(4):320–327.
14. de Causmaecker, P. and vanden Berghe, G. (2011). A categorisation of nurse rostering problems. *Journal of Scheduling*, 14(1):3–16.
15. de Véricourt, F. and Jennings, O. (2011). Nurse staffing in medical units: a queueing perspective. *Operations Research*, 59(6):1320–1331.
16. Elkhuizen, S., Bor, G., Smeenk, M., Klazinga, N., and Bakker, P. (2007). Capacity management of nursing staff as a vehicle for organizational improvement. *BMC health services research*, 7(1):196–205.
17. Ernst, A., Jiang, H., Krishnamoorthy, M., and Sier, D. (2004). Staff scheduling and rostering: a review of applications, methods and models. *European Journal of Operational Research*, 153(1):3–27.
18. Fei, H., Meskens, N., and Chu, C. (2010). A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering*, 58(2):221–230.
19. Gnanlet, A. and Gilland, W. (2009). Sequential and simultaneous decision making for optimizing health care resource flexibilities. *Decision Sciences*, 40(2):295–326.
20. Gorunescu, F., McClean, S., and Millard, P. (2002). A queueing model for bed-occupancy management and planning of hospitals. *Journal of the Operational Research Society*, 53(1):19–24.

21. Goulding, L., Adamson, J., Watt, I., and Wright, J. (2012). Patient safety in patients who occupy beds on clinically inappropriate wards: a qualitative interview study with nhs staff. *BMJ Quality & Safety*, 21(3):218–224.
22. Green, L. and Nguyen, V. (2001). Strategies for cutting hospital beds: the impact on patient service. *Health Services Research*, 36(2):421–442.
23. Griffiths, J., Price-Lloyd, N., Smithies, M., and Williams, J. (2005). Modelling the requirement for supplementary nurses in an intensive care unit. *Journal of the Operational Research Society*, 56(2):126–133.
24. Guerriero, F. and Guido, R. (2010). Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, 14(1):1–26.
25. Harper, P. (2002). A framework for operational modelling of hospital resources. *Health Care Management Science*, 5(3):165–173.
26. Harper, P., Powell, N., and Williams, J. (2010). Modelling the size and skill-mix of hospital nursing teams. *Journal of the Operational Research Society*, 61(5):768–779.
27. Harper, P. and Shahani, A. (2002). Modelling for the planning and management of bed capacities in hospitals. *The Journal of the Operational Research Society*, 53(1):11–18.
28. Harper, P., Shahani, A., Gallagher, J., and Bowie, C. (2005). Planning health services with explicit geographical considerations: a stochastic location-allocation approach. *Omega*, 33(2):141–152.
29. Harrison, G., Shafer, A., and Mackay, M. (2005). Modelling variability in hospital bed occupancy. *Health Care Management Science*, 8(4):325–334.
30. Hulshof, P., Kortbeek, N., Boucherie, R., Hans, E., and Bakker, P. (2012). Taxonomic classification of planning decisions in health care: a review of the state of the art in OR/MS. *Health Systems*, 1(2):129–175.
31. Kellogg, D. and Walczak, S. (2007). Nurse scheduling: from academia to implementation or not? *Interfaces*, 37(4):355.
32. Kortbeek, N., Braaksma, A., Burger, C., Bakker, P., and Boucherie, R. (2015). Flexible nurse staffing based on hourly bed census predictions. *International Journal of Production Economics*, 161:167–180.
33. Kortbeek, N., Braaksma, A., Smeenk, F., Bakker, P., and Boucherie, R. (2014). Integral resource capacity planning for inpatient care services. *Journal of the Operational Research Society*, Published online 23 July 2014:–.
34. Lang, T., Hodge, M., Olson, V., Romano, P., and Kravitz, R. (2004). Nurse-patient ratios: a systematic review on the effects of nurse staffing on patient, nurse employee, and hospital outcomes. *Journal of Nursing Administration*, 34(7-8):326–337.
35. Lavieri, M. and Puterman, M. (2009). Optimizing nursing human resource planning in British Columbia. *Health Care Management Science*, 12(2):119–128.
36. Li, X., Beullens, P., Jones, D., and Tamiz, M. (2009). An integrated queuing and multi-objective bed allocation model with application to a hospital in China. *Journal of the Operational Research Society*, 60(3):330–338.
37. Oddoye, J., Jones, D., Tamiz, M., and Schmidt, P. (2009). Combining simulation and goal programming for healthcare planning in a medical assessment unit. *European Journal of Operational Research*, 193(1):250–261.
38. Oddoye, J., Yaghoobi, M., Tamiz, M., Jones, D., and Schmidt, P. (2007). A multi-objective model to determine efficient resource levels in a medical assessment unit. *Journal of the Operational Research Society*, 58(12):1563–1573.
39. Pierskalla, W. and Brailer, D. (1994). Applications of operations research in health care delivery. In Pollock, S., Rothkopf, M., and Barnett, A., editors, *research and the public sector*, volume 6 of *Handbooks in OR & MS*, pages 469–505. North-Holland, Amsterdam, The Netherlands.
40. PubMed (2011). *Retrieved June 19, 2012, from:* http://www.s.gov/.
41. RVZ (2012). Council for Public Health and Health Care [Raad voor Volksgezondheid & Zorg]. Medisch-specialistische zorg in 2020 (In Dutch). *Retrieved August 1, 2012, from:* http://www.rvz.net/.

42. Smith-Daniels, V., Schweikhart, S., and Smith-Daniels, D. (1988). Capacity management in health care services: review and future research directions. *Decision Sciences*, 19(4):889–919.
43. Trivedi, V. and Warner, D. (1976). A branch and bound algorithm for optimum allocation of float nurses. *Management Science*, 22(9):972–981.
44. Troy, P. and Rosenberg, L. (2009). Using simulation to determine the need for ICU beds for surgery patients. *Surgery*, 146(4):608–620.
45. Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., and De Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3):367–385.
46. Van Oostrum, J., Van Houdenhoven, M., Hurink, J., Hans, E., Wullink, G., and Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR spectrum*, 30(2):355–374.
47. Vanberkel, P. and Blake, J. (2007). A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Management Science*, 10(4):373–385.
48. Vanberkel, P., Boucherie, R., Hans, E., Hurink, J., and Litvak, N. (2010a). A survey of health care models that encompass multiple departments. *International Journal of Health Management and Information*, 1(1):37–69.
49. Vanberkel, P., Boucherie, R., Hans, E., Hurink, J., van Lent, W., and van Harten, W. (2010b). An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society*, 62(10):1851–1860.
50. Villa, S., Barbieri, M., and Lega, F. (2009). Restructuring patient flow logistics around patient care needs: implications and practicalities from three critical cases. *Health Care Management Science*, 12(2):155–165.
51. Vissers, J., Adan, I., and Dellaert, N. (2007). Developing a platform for comparison of hospital admission systems: An illustration. *European Journal of Operational Research*, 180(3):1290–1301.
52. Walts, L. and Kapadia, A. (1996). Patient classification system: an optimization approach. *Health Care Management Review*, 21(4):75.
53. Whitt, W. and Zhang, X. (2017). A data-driven model of an emergency department. *Operations Research for Health Care*, 12:1–15.
54. Wright, P., Bretthauer, K., and Côté, M. (2006). Reexamining the Nurse Scheduling Problem: Staffing Ratios and Nursing Shortages. *Decision Sciences*, 37(1):39–70.
55. Yankovic, N. and Green, L. (2011). Identifying good nursing levels: a queuing approach. *Operations Research*, 59(4):942–955.