Journal of the Operational Research Society (2014), 1–16

www.palgrave-journals.com/jors/

Integral resource capacity planning for inpatient care services based on bed census predictions by hour

Nikky Kortbeek[1,2,3]*, Aleida Braaksma[1,2,3], Ferry HF Smeenk[1,2], Piet JM Bakker[2] and Richard J Boucherie[1,3]

[1]Center for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Enschede, The Netherlands; [2]Academic Medical Center, Amsterdam, The Netherlands; and [3]Stochastic Operations Research, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

The design and operations of inpatient care facilities are typically largely historically shaped. A better match with the changing environment is often possible, and even inevitable due to the pressure on hospital budgets. Effectively organizing inpatient care requires simultaneous consideration of several interrelated planning issues. Also, coordination with upstream departments like the operating theatre and the emergency department is much-needed. We present a generic analytical approach to predict bed census on nursing wards by hour, as a function of the Master Surgical Schedule and arrival patterns of emergency patients. Along these predictions, insight is gained on the impact of strategic (ie, case mix, care unit size, care unit partitioning), tactical (ie, allocation of operating room time, misplacement rules), and operational decisions (ie, time of admission/discharge). The method is used in the Academic Medical Center Amsterdam as a decision-support tool in a complete redesign of the inpatient care operations.

Journal of the Operational Research Society advance online publication, 23 July 2014; doi:10.1057/jors.2014.67

Keywords: probability; health service; hospitals; medical care units; bed occupancy; surgical scheduling

1. Introduction

Inpatient care facilities provide care to hospitalized patients by offering a room, a bed and board (PubMed, 2012). Societal developments and budget constraints demand hospitals to on the one hand increase quality of care and on the other hand efficiency (RVZ, 2012). This entails a strong incentive to reconsider the design and operations of inpatient care services. Since the 1950s, the application of operational research methods yields significant contributions in accomplishing essential efficiency gains in health-care delivery (Hulshof *et al*, 2012). In this paper, we present an exact method to assist hospital management in adequately organizing their inpatient care services.

Effectively designing inpatient care services requires simultaneous consideration of several interrelated strategic and tactical planning issues (Hulshof *et al*, 2012). Given service mix and case mix decisions, hospital management has to decide on care unit partitioning (which care units are created and which patient groups are assigned to these units) and care unit size (the number of staffed beds per care unit). Since the inpatient care facility is a downstream department, the outflow of the operating theatre 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 (Harper, 2002; Vanberkel *et al*, 2010a). While smoothing patient inflow prevents large differences between peak and off-peak periods, and so realizes a more efficient use of resources (Harper, 2002; Vissers *et al*, 2007; Adan *et al*, 2009), the authority of inpatient care facilities on their admission control is limited. 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 (Green and Nguyen, 2001). Anticipation for elective surgical patients is possible as well, by taking the surgical schedule into account (Green and Nguyen, 2001; Harper, 2002; Vissers *et al*, 2007; Adan *et al*, 2009). Hospitals typically allocate operating room capacity through a Master Surgical Schedule (MSS), a (cyclic) block schedule that allocates operating time capacity among patient groups (Van Oostrum *et al*, 2008; Fei *et al*, 2010; Guerriero and Guido, 2010). In this paper, we address these various patient flows and take the necessity of integral decision making into account.

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 (Harper *et al*, 2005; Villa *et al*, 2009). Performance measures are required that reflect efficiency and quality of care to assess the quality of logistical layout. Efficiency is often expressed in high bed occupancy, which is assumed to imply efficient use of staff and

*Correspondence: Nikky Kortbeek, Stochastic Operations Research, Department of Applied Mathematics, University of Twente, Drienerlolaan 5, Enschede 7500AE, The Netherlands.
E-mail: n.kortbeek@utwente.nl

other equipment (Ridge *et al*, 1998; Gorunescu *et al*, 2002). The drawback of high bed occupancy is that it may cause congestion, which manifests itself in two main consequences, both being a threat to the provided quality of care (Green and Nguyen, 2001; Goulding *et al*, 2012): (i) patients may have to be rejected for admission due to lack of bed capacity, so-called admission refusals or rejections, (ii) patients may (temporarily) be placed in less appropriate units, so-called misplacements (Harper and Shahani, 2002; Costa *et al*, 2003; Harrison *et al*, 2005). Owing to such misplacements, planning decisions regarding a specific care unit can affect the operations of other units (Akkerman and Knip, 2004; Cochran and Bharti, 2006; Li *et al*, 2009). Planning of the inpatient care facility should not only take into account the upstream departments, but also the interrelationship between care units.

Previous analytical studies have addressed partial resource capacity planning issues within the inpatient care chain, for example by dimensioning care units in isolation (eg Green and Nguyen, 2001; Gorunescu *et al*, 2002; Bekker and De Bruin, 2010), balancing bed utilization across multiple units (eg Akcali *et al*, 2006; Cochran and Bharti, 2006; Li *et al*, 2009), or focussing on improving the MSS to balance inpatient care demand (eg Van Oostrum *et al*, 2008; Adan *et al*, 2009; Beliën *et al*, 2009; Vanberkel *et al*, 2010b; Bekker and Koeleman, 2011). More integral approaches can be found in simulation studies (eg Harper and Shahani, 2002; Harper, 2002; Vanberkel and Blake, 2007; Troy and Rosenberg, 2009). The advantage of such approaches is their flexibility and therefore modelling power. However, the disadvantage is that the nature of such studies is typically context-specific, which limits the generalizability of application and findings.

We present a generic exact analytical approach to achieve the required integral and coordinated resource capacity planning decision making for inpatient care services. The method builds upon the approach presented in Vanberkel *et al* (2010b), 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 MSS. Based on a cyclic arrival pattern of emergency patients and an MSS block schedule of surgical patients, we derive demand predictions on an hourly level for several inpatient care units simultaneously for both acute and elective patients. (The method is also applicable for departments catering for non-surgical elective patients, as these can be incorporated in our model via fictitious OR blocks.) This hourly level of detail is required to adequately incorporate the time-dependent behaviour of the inpatient care process. 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. In addition, what-if questions can be addressed considering the impact of operational interventions such as shortening length of stay or changing the times of admissions and discharges.

During the upcoming years the presented method will be applied in a Dutch teaching hospital, the Academic Medical Center (AMC) in Amsterdam in supporting the intended complete redesign of the inpatient care facility. As part of the total redesign, in the case study of the present article we restrict ourselves to a set of interrelated (with respect to capacity planning) specialties: traumatology, orthopaedics, plastic surgery, urology, vascular surgery, and general surgery. By means of this case study we illustrate the practical potential of our analytical approach for logistical redesign of inpatient care services.

This paper is organized as follows. First, we describe the model consisting of demand predictions, bed census predictions, and performance measures. The next section introduces the case study at the AMC and describes the numerical results. The paper closes with a discussion of our findings and opportunities for further research.

2. Predictive model

In this section, the model is described that predicts the workload at several care units of an inpatient care facility on a time scale of hours, due to patients originating from the operating theatre and emergency department. The basis for the operating room outflow prediction is the MSS. The basis for the emergency department outflow prediction is a cyclic random arrival process which we define as the Acute Admission Cycle (AAC). Schematically, the approach is as follows. First, the impact of the MSS and the AAC are separately determined and then combined to obtain the overall steady-state impact of the repeating cycles. Second, the obtained demand distributions are translated to bed census distributions. Finally, 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, which in principle can be regarded as hours (but could also resemble for example 2- or 4-h time intervals). 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 (eg weekly) non-homogeneous Poisson arrival process corresponding to the unpredictable nature of emergency arrivals. 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.

For the demand predictions, for both elective and acute patients, three steps are performed. First, 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. Then, since the MSS and AAC are cyclical, the predictions from the second step are overlapped to find the overall steady-state impact of the repeating cycles. The workload predictions for elective and acute patients are combined to find the probability distributions of the number of recovering patients at the inpatient care facility on each unique day in the cycle which we denote as the Inpatient Facility Cycle (IFC). The length of the IFC is the least common multiple of the lengths of the MSS and the AAC.

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.

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 $t=0,\ldots,T$, 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, that is, $n=0$ indicates the day of surgery.

MSS utilization. For each day $s\in\{1,\ldots,S\}$, a (sub)specialty j can be assigned to an available operating room i, $i\in\{1,\ldots,I\}$. 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, that is, $c^j(k)$ is the probability of k surgeries performed in one block, $k\in\{0,1,\ldots,C^j\}$. 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 modelled 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 $t=\vartheta_j+1,\ldots,T-1$.

Discharges. $P^j(n)$ is the probability that a type j patient stays n days after surgery, $n\in\{0,\ldots,L^j\}$. 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, that is, $m^j_{0,t}=0$ for $t=0,\ldots,\vartheta_j$.

Single surgery block. In this first step we consider a single specialty j operating in a single OR block. We compute 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. 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 $n=1,\ldots,L^j$. Therefore, we calculate $h^j_{n,t}(x)$ as follows:

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

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 derivations of $a^j_{n,t}$ and $d^j_{n,t}$ are presented in Appendix A 'Single surgery block'.

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 ($m\in\{0,1,2,\ldots,S,S+1,S+2,\ldots\}$) resulting from a single MSS cycle (see Appendix A 'Single MSS cycle').

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 $H^{\mathrm{SS}}_{s,t}$, the steady-state probability distributions of the number of recovering patients at time t on day s of the cycle ($s\in\{1,\ldots,S\}$) (see Appendix A 'Steady state').

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 $t\leqslant\theta$.

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 presented in Appendix B 'Single patient type'.

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 (see Appendix B 'Single cycle').

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 ($r \in \{1, \ldots, R\}$) (see Appendix B 'Steady state').

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, that is, $S \neq R$. 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:

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

where $\otimes$ denotes the discrete convolution function. 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 A.1 and V^k in Equation (B.1).

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

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

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:

$$\phi(\mathbf{x}) = \begin{cases} (x_1, x_2) & ,x_1 \leqslant M_1, x_2 \leqslant M_2, \\ (M_1, \min\{x_2 + (x_1 - M_1), M_2\}) & ,x_1 > M_1, x_2 \leqslant M_2, \\ (\min\{x_1 + (x_2 - M_2), M_1\}, M_2) & ,x_1 \leqslant M_1, x_2 > M_2, \\ (M_1, M_2) & ,x_1 > M_1, x_2 > M_2. \end{cases} \quad (1)$$

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.

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 the next section 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:

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

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

(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:

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

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

Admission rate

Patient admissions may increase the nursing workload. Let $\Lambda_{q,t}^k$ 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 $\Lambda_{q,t}^k$, we first determine $\overline{a}_{n,t}^j$, the distribution of the number of elective type j arrivals during time interval $[t, t+1)$ on day n ($n \in \{-1, 0\}$):

$$\overline{a}_{n,t}^j(x) = \sum_{y=0}^{C^j} c^j(y) \overline{a}_{n,t}^j(x \mid y) \quad \text{, with} \quad \overline{a}_{n,t}^j(x \mid y)$$

$$= \binom{y}{x} \left(e_n^j w_{n,t}^j \right)^x \left(1 - e_n^j w_{n,t}^j \right)^{y-x}.$$

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

$$\Lambda_{q,t}^k = \left\{ \otimes_{i=1}^{I} \left\{ \otimes_{j \in W^k:\, j \in b_{i,s'}} \overline{a}_{-1,t}^j \right\} \otimes \left\{ \otimes_{j \in W^k:\, j \in b_{i,s''}} \overline{a}_{0,t}^j \right\} \right\}$$

$$\otimes \left\{ \otimes_{j \in V^k:\, r = r'} \tilde{a}_t^j \right\}. \tag{2}$$

where $s' = 1 + q \bmod S$, $s'' = q \bmod S + S \cdot 1_{(q \bmod S = 0)}$, $r' = q \bmod R + R \cdot 1_{(q \bmod R = 0)}$, and $\otimes_{x \in X} f_x$ denotes the discrete convolution over the probability distributions $f_x, x \in \mathcal{X}$. The first term in the right-hand side of (2) 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 (2) 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 $\rho_{q,t}^k, \rho_q^k, \rho^k$ 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:

$$\rho_{q,t}^k = \frac{1}{M^k} \sum_{x=0}^{M^k} x \cdot \hat{Z}_{q,t}^k(x), \qquad \rho_q^k = \frac{1}{T} \sum_{t=0}^{T-1} \rho_{q,t}^k,$$

$$\rho^k = \frac{1}{Q} \sum_{q=1}^{Q} \rho_q^k.$$

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 $R_{q,t}^{\phi,k}$: the probability of such an admission rejection at time t on day q. $R^{\phi,k}$ is then calculated as follows:

$$R^{\phi,k} = \frac{1}{\sum_{q,t} E\left[\Lambda_{q,t}^k\right]} \sum_{q,t} E\left[\Lambda_{q,t}^k\right] R_{q,t}^{\phi,k}.$$

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 $\mathcal{R}^{\phi,k}(\mathbf{x}, n)$, the number of rejected patients under allocation policy ϕ of the n arriving patients to unit k, and $Z_{q,t}^k(x_k \mid n)$ 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, $R_{q,t}^{\phi,k}$ is calculated by:

$$R_{q,t}^{\phi,k} = \frac{E[\#\text{rejections at unit } k \text{ on time } (q,t)]}{E[\#\text{arrivals to unit } k \text{ on time } (q,t)]}$$

$$= \frac{1}{E[\Lambda_{q,t}^k]} \sum_{\mathbf{x}} \prod_{\ell \neq k} Z_{q,t}^{\ell}(x_\ell) \sum_{n} \mathcal{R}^{\phi,k}(\mathbf{x}, n)$$

$$\Lambda_{q,t}^k(n) Z_{q,t}^k(x_k \mid n). \tag{3}$$

The derivation of $Z_{q,t}^k(x_k \mid n)$ is presented in Appendix C. $\mathcal{R}^{\phi,k}(\mathbf{x}, n)$ is uniquely determined by allocation policy ϕ. For example, for the case with $K = 2$ presented in (1), we have for unit $k = 1$:

$$\mathcal{R}^{\phi,1}(\mathbf{x}, n) =$$

$$\begin{cases} \min\{n, x_1 - M_1\} & , x_1 \geqslant M_1, x_2 \geqslant M_2, \\ \max\{0, (x_1 - M_1) - (M_2 - x_2)\} & , x_1 \geqslant M_1, x_2 < M_2, n \geqslant (x_1 - M_1), \\ n - \max\{0, \min\{n, (M_2 - x_2 - [x_1 - M_1 - n])\}\} & , x_1 \geqslant M_1, x_2 < M_2, n < (x_1 - M_1), \\ 0 & , \text{otherwise}. \end{cases}$$

(4)

In (4), 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 (3), $\mathcal{R}^{\phi,k}(\mathbf{x},n)$ has to be replaced by $\mathcal{M}^{\phi}(\mathrm{x},n)$, 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 (1), we have:

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

Productivity

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

$$\mathcal{P}^{\mathcal{K}} = \frac{365}{Q} \frac{1}{\sum_{k\in\mathcal{K}} M^k} \sum_{k\in\mathcal{K}} \sum_{q,t} \left(1 - R^{\phi,k}_{q,t}\right) E\left[\Lambda^k_{q,t}\right]. \qquad (5)$$

Remark 1 (Approximation) Observe that the calculations of misplacements and rejections 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 health care.

Remark 2 (Numerical evaluation) Recall that to compute all performance measures formulated above it is only required to specify the input parameters that were specified under the headers 'model input' for the elective and the acute patients.

3. Quantitative results

The case study entails the university hospital AMC, which has 20 operating rooms and 30 inpatient departments with in total 1000 beds. Owing to both economic and medical developments, the AMC is forced to reorganize the operations of the inpatient services during the upcoming years. This section describes the exercise we performed together with the managers of four care units to explore the potential of the presented method to direct these reorganizations. These managers are responsible for the care units that in the current situation together house six surgical specialties. The structure of this section is as follows. We first provide some details on the care units under study. Second, we show that the prediction model is able to generate a valid representation of this practical setting. Third, we present the results on numerical experiments that were designed in close cooperation with the health-care professionals. Based on the outcomes of this exercise, the quantitative method that we presented in the previous section is embraced by the hospital as a valuable instrument to support the resource capacity planning of its inpatient care services. Outside the scope of this paper are the actual decision-making process on which interventions to apply in practice, and the subsequent implementation phase. These will take place during the upcoming years embedded in a hospital-wide improvement programme.

Case study description

We take the following specialties into account: traumatology (TRA), orthopaedics (ORT), plastic surgery (PLA), urology (URO), vascular surgery (VAS), and general surgery (GEN). In the present setting, the patients of the mentioned specialties are admitted in four different inpatient care departments. Care unit A houses GEN and URO, unit B VAS and PLA, unit C TRA, and unit D ORT. The physical building is such that units A and B are physically adjacent (Floor I), so 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). Currently, no cyclical MSS is applied. Each time, roughly six weeks in advance the MSS is determined for a period of four weeks. The capacities of units A, B, C, and D are 32, 24, 24, and 24 beds, respectively. However, it often happens that not all beds are available, due to personnel shortages. The utilizations over 2009–2010 were 53.2, 55.6, 54.4, and 60.6% (which includes some patients of other than the given specialties that were placed in these care units). These utilizations reflect administrative bed census, 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 theatre or at the intensive care department. Unfortunately, no confident data was available on rejections and misplacements.

Validation

We validate the model on the 'base case scenario', the situation that closely resembles current practice. The base case takes the current bed capacities, and misplacements take place between care units A and B (Floor I) and between units C and D (Floor II).

Table 1 Overview historical data 2009–2010

Specialty	*Acronym*	*Care unit*	*Elective admissions*	*Acute admissions*	*Average LOS (in days)*	*Load (# 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
Orthopaedics	ORT	D	836	845	6.23	14.38

The main difference between current practice and the base case scenario is that the model assumes the available beds to be always open, so no *ad hoc* closings are allowed.

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 specialities who stayed at departments A–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 one 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 $n \geqslant 2$.

In validating the bed census predictions resulting from running the model, we have to address the following challenge. Since no cyclical MSS was applied in the AMC, each day in the time horizon is unique; in fact, each time slot in the considered 2-year period is unique. Thus, by applying the predictive model, for each care unit we estimate 104 (weeks) × 7 (days) × 24 (h) = 17,472 census distributions. As a consequence, for each predicted census distribution, we have only one observation available in historical data to compare against. To address this challenge, the validation below consists of two components, which are both adapted from existing techniques.

First, we look at the validity of the expectations of the predicted bed census distributions. To compare the model census expectations with the practical averages obtained from the AMC observations, we calculate two measures well-known from statistical forecasting to describe prediction errors. Common quantities used to measure how close predictions are to the actual realizations are the mean absolute error[1] (MAE) and the mean absolute percentage error[2] (MAPE). To judge whether the model expectations follow the practical averages, we apply the MAE and MAPE measures as follows. Let us denote the observations obtained from the historical data of the number of patients present at each point in time by $Obs^k_{q,t}$ ($k \in \{A, B, C, D\}$, $q \in \{1, \ldots, 728\}$, $t \in \{0, \ldots, 23\}$). Based on these observations, we determine the average realized bed census over each specific time in different weeks. Thus, we calculate for each day of the week ($d = 1, \ldots, 7$, corresponding to Monday till Sunday) and hour of the day ($t = 0, \ldots, 23$, corresponding to 0:00 h till 23:00 h) the average realized bed census, denoted by $\overline{Obs}^k_{d,t}$,

$$\overline{Obs}^k_{d,t} = \frac{1}{104} \sum_{i=0}^{103} Obs^k_{7i+d,t}.$$

In addition, we calculate the corresponding average model expectation, denoted by $\overline{Mod}^k_{d,t}$,

$$\overline{Mod}^k_{d,t} = \frac{1}{104} \sum_{i=0}^{103} M^k \cdot \rho^k_{7i+d,t}.$$

Figure 1 displays the model results for the expected census average against the historical data. It illustrates that the model predictions closely follow historical data. Slight differences can be observed for (1) the elective patients on Sunday afternoon, since in practice Sunday-admission times differ from weekdays, where we assume the same admission time distributions for all days, and (2) the elective patients on Friday afternoon, since in practice more patients are discharged just before the weekend, where we assume the length-of-stay distributions to be independent of the day of surgery. The similarity between the historical and model averages is quantified by the MAE and MAPE scores. The MAE is 0.32 for ward A, 0.19 for ward B, 0.11 for ward C, and 0.35 for ward D; and the MAPE is 2.02% for ward A, 1.51% for ward B, 0.86% for ward C, and 2.54% for ward D.

Second, we look at the validity of the shape of the predicted bed census distributions. To compare the shape of the census distributions resulting from the model against the observed variation, we judge how the predicted bed census percentiles relate to the historical data set. We judge the quality of the bed census percentile predictions $\hat{D}^k_{q,t}(\alpha)$, by comparing each historical census observation to the demand percentile of the corresponding predicted bed census distribution. Our approach is adapted from known methods to compare a data sample to the

[1]The mean absolute error (MAE) is given by $MAE = (1)/(n)\sum_{i=1}^{n} |a_i - y_i|$, where a_i is the actual value and y_i the predicted value.

[2]The mean absolute percentage error (MAPE) is given by $MAPE = (100\%)/(n)\sum_{i=1}^{n} |(a_i - y_i)/a_i|$, where a_i is the actual value and y_i the predicted value.

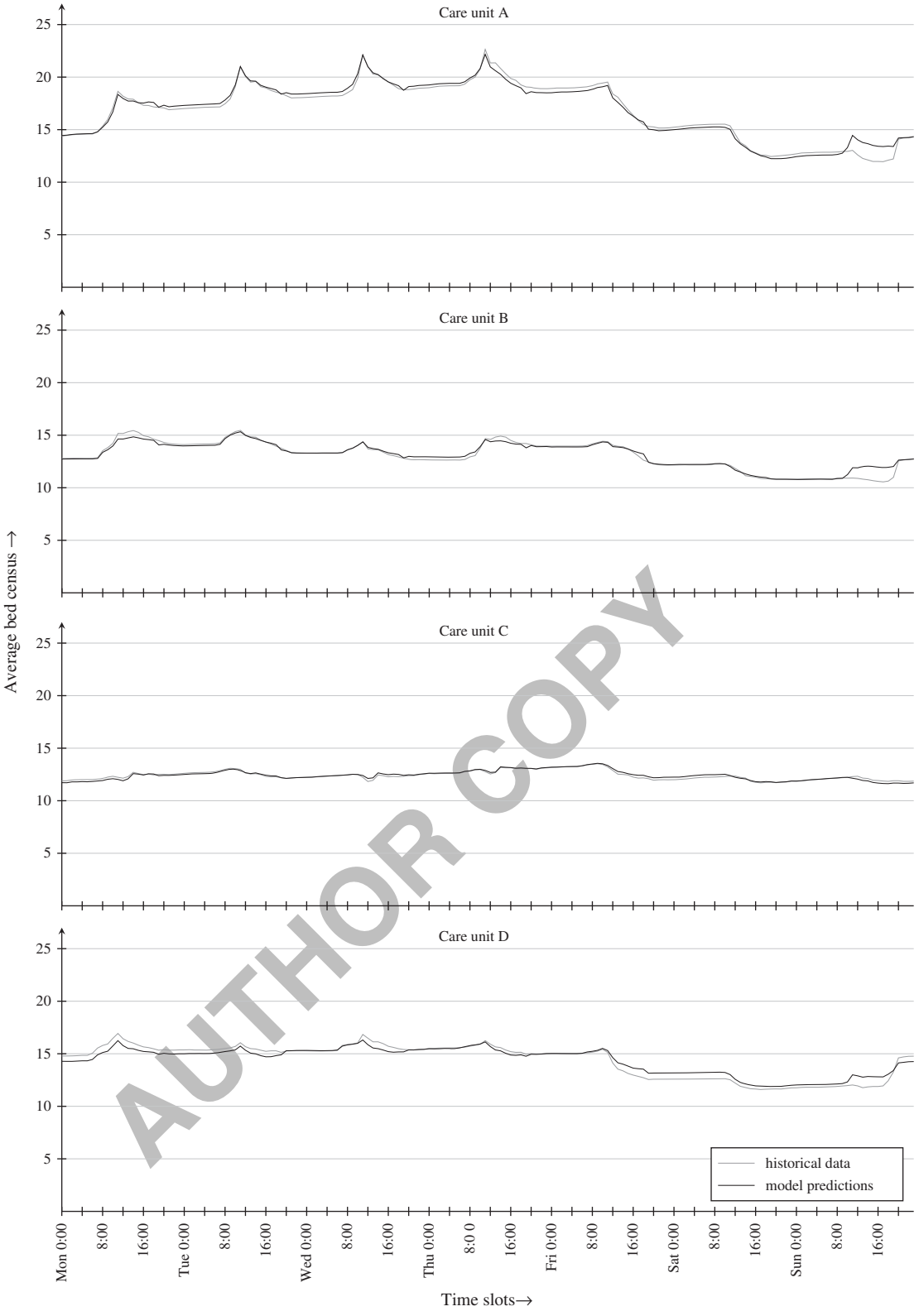

Figure 1 Comparison of the model results against historical data on average bed census per time slot.

Table 2 Comparison of the model results against historical data on bed census percentiles (for the definition of $\Psi^k(\alpha)$ see Equation (6))

		α	0.500	0.600	0.700	0.800	0.900	0.950	0.975
Care unit	A	$\Psi^A(\alpha)$	0.496	0.603	0.705	0.810	0.909	0.960	0.987
	B	$\Psi^B(\alpha)$	0.536	0.648	0.754	0.853	0.944	0.968	0.985
	C	$\Psi^C(\alpha)$	0.469	0.587	0.693	0.812	0.932	0.974	0.987
	D	$\Psi^D(\alpha)$	0.486	0.573	0.669	0.789	0.921	0.973	0.992

percentiles of a single nonparametric distribution (eg, see Chapter 14 of Bain and Engelhardt, 1992). To this end, we introduce the measure

$$\Psi^k(\alpha) = \frac{1}{17\,472}\sum_{q=1}^{728}\sum_{t=0}^{23} 1_{\left(Obs^k_{q,t} \leqslant \hat{D}^k_{q,t}(\alpha)\right)}, \qquad (6)$$

representing for level α the fraction of data points in the historical data set that is less than or equal to the model's αth bed census percentile. By the law of large numbers, if the data points $Obs^k_{q,t}$ would have been sampled from the census distributions $\hat{Z}^k_{q,t}$, for a growing number of observations, $\Psi(\alpha)$ converges to α. Therefore, the closer the measures $\Psi(\alpha)$ are to α for different values of α, the greater the similarity between the model predictions and historical data. The results presented in Table 2 support the validity of our prediction model.

Based on the results of these two validation exercises, we conclude that the model presented in this paper provides a valid representation of the AMC practice.

Analysis

We present several interventions that improve the efficiency of the inpatient care service operations. The goal is to maximize the number of patients that can be treated per bed per year, while maintaining a high standard on rejections and misplacements to guarantee accessibility and quality of care. The interventions were formulated in close cooperation with the involved health-care professionals. A range of interventions were designed and tested and each time the outcomes were discussed with the care unit managers. For the purpose of this paper, we choose to present the interventions that we believe are the most insightful with respect to numerical outcomes and that best illustrate the manner in which the predictive model can be exploited to quantitatively evaluate the impact of suggested practical interventions.

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 (as described in the validation section). In all experiments, no *ad hoc* closings are allowed, like 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 3–5:

(1) *Rationalize bed requirements*. The current numbers of beds are a result of historical development. Given particular service requirements, which are to be specified by the hospital management, 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 with for instance a stationary Poisson arrival process. The hospital recognizes that simultaneously prohibiting bed closings on an *ad hoc* basis and downsizing the total number of beds is more effective in realizing a consistent quality-of-service level, while it is also more efficient (reflected by the clear increase in the productivity measure, ie, the number of patients that can be treated per bed per day).

(2) *No misplacements*. As it would medically be desirable and would reduce the complexity of operations, the care unit managers wondered what would happen if the practice of misplacing patients when the preferred bed is not available would be abandoned. To gain this insight, in this intervention we explore what would happen if misplacements are not allowed. In the model, we changed allocation policy ϕ so that no misplacements take place. The numerical results demonstrate the benefits of capacity pooling when overflow between care units is allowed. These benefits are due to the so-called portfolio effect which induces that the relative variability in demand is reduced by economies of scale. It can be concluded that in the case under study, care

Table 3 The numerical results for the base case, intervention 1, and intervention 2 (with the productivity-Δ% relative to the base case)

Intervention	*Unit*	*Capacity* (# *beds*)	*Rejection* (%)	*Misplace* (%)	*Utilization* (%)	*Floor*	*Capacity* (# beds)	*Productivity*	
								Equation (5)	(Δ%)
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. *No misplacements*									
Rejection < 5%	A	30	4.22	—	60.5	AB	52	51.7	+3.5
	B	22	3.67	—	61.5				
	C	20	4.93	—	66.1	CD	44	36.7	+4.4
	D	24	3.78	—	61.5				
Rejection < 2.5%	A	32	2.00	—	56.8	AB	55	49.9	−0.2
	B	23	2.22	—	58.9				
	C	22	1.67	—	60.3	CD	47	35.2	+0.1
	D	25	2.42	—	59.1				
Rejection < 1%	A	34	0.86	—	53.5	AB	59	47.1	−5.7
	B	25	0.73	—	54.2				
	C	23	0.91	—	57.8	CD	50	33.4	−4.8
	D	27	0.90	—	54.8				

units in the order of size 20–30 beds are too small to operate efficiently in isolation.

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

Compared with 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.

(4) *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 constructively created a cyclical MSS with a length of four 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.

As an illustration, Figure 2 displays the average bed utilization per day of the week for care unit A (rejection probability < 1%) before and after balancing the MSS. From this figure it is clear that both the midweek peak and the weekend dip can be cleared to a large extent, which results in distinct efficiency gains (see Table 4). We have reason to believe that even larger gains can be achieved. First, by developing a

Table 4 The numerical results for interventions 2, 3, and 4 (with the productivity-Δ% relative to the base case)

Intervention	*Unit*	*Capacity* (# *beds*)	*Rejection* (%)	*Misplace* (%)	*Utilization* (%)	*Floor*	*Capacity* (# beds)	*Productivity*	
								Equation (5)	(Δ%)
3. *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				
4. *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				
5. *Combination* (1), (3), *and* (4)									
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				

structured method to optimize the MSS instead of manual optimization. Second, the lack of detail in the available historical MSS data resulted in high variation in the input probability distributions of the number of cases per OR block and the length-of-stay distributions. When more information would be available on the content of MSS blocks, for instance on the level of subspecialty or even surgery type, the census predictions would show lower variability, resulting in lower bed requirements.

(5) *Combinations* 1, 3, *and* 4. This intervention combines interventions (1), (3), and (4). 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.

(6) *Separation elective and acute*. Clinicians and managers in the AMC discuss the desirability to split elective and acute patient flows. This intervention illustrates the capability of the model to provide quantitative support in decision making on care unit partitioning. Intervention 6a

Table 5 The numerical results for intervention 6 (with the productivity-Δ% relative to 6a)

Intervention	*Unit*	*Capacity* (# *beds*)	*Rejection* (%)	*Misplace* (%)	*Utilization* (%)	*Floor*	*Capacity* (# beds)	*Productivity*	
								Equation (5)	(Δ%)
6a. Separation elective and acute									
Rejection < 5%	A	21	4.36	10.34	68.9	AB	42	45.0	—
	B	21	4.00	7.65	68.9				
	C	21	3.92	7.82	70.3	CD	42	57.6	—
	D	21	3.89	11.12	76.0				
Rejection < 2.5%	A	22	2.40	8.31	66.0	AB	44	43.7	—
	B	22	2.30	6.04	65.9				
	C	22	2.03	6.00	67.0	CD	44	56.1	—
	D	22	1.99	8.41	72.9				
Rejection < 1%	A	24	0.80	4.43	60.7	AB	47	41.6	—
	B	23	0.95	4.87	62.9				
	C	23	0.98	4.33	64.0	CD	46	54.2	—
	D	23	0.95	6.02	69.9				
6b. Combination (6a) *and balance MSS*									
Rejection < 5%	A	21	3.35	6.70	73.6	AB	41	48.9	+8.7
	B	20	3.30	8.55	75.7				
	C	21	3.92	7.82	70.3	CD	42	57.6	0.0
	D	21	3.89	11.12	76.0				
Rejection < 2.5%	A	22	2.07	4.33	70.6	AB	42	48.2	+10.3
	B	20	2.44	9.41	75.5				
	C	22	2.03	6.00	67.0	CD	44	56.1	0.0
	D	22	1.99	8.41	72.9				
Rejection < 1%	A	23	0.65	3.25	67.2	AB	45	45.8	+10.2
	B	22	0.59	3.90	69.2				
	C	23	0.98	4.33	64.0	CD	46	54.4	0.0
	D	23	0.95	6.02	69.9				
6c. Combination (6b) *and change operational process*									
Rejection < 5%	A	19	4.00	7.01	68.7	AB	39	51.3	+14.1
	B	20	3.07	4.46	66.4				
	C	20	4.73	9.59	71.8	CD	41	58.8	+2.0
	D	21	3.74	9.29	75.2				
Rejection < 2.5%	A	20	2.46	4.59	65.5	AB	40	50.6	+15.6
	B	20	2.38	5.16	66.2				
	C	22	2.10	4.62	65.9	CD	43	57.3	+2.2
	D	21	2.20	10.82	74.8				
Rejection < 1%	A	21	0.77	3.55	62.4	AB	43	47.9	+15.2
	B	22	0.56	2.04	60.3				
	C	23	0.78	3.60	62.6	CD	46	54.4	+0.2
	D	23	0.67	5.21	68.9				

is formulated such that all elective patients are treated at Floor I (unit A: GEN, URO, VAS; unit B: PLA, TRA, ORT) and all acute patients at Floor II (unit C: GEN, URO, VAS, PLA; unit D: TRA, ORT). In intervention 6b splitting electives and acute patients is combined with creating a balanced MSS, and intervention 6c extends this by including the changes in the operational process from intervention 3. Table 5 shows that the logistical performance is similar to the previous care unit configuration. We conclude therefore that whether or not to separate elective and acute patients in the studied case, should mainly be decided based on medical arguments.

4. Discussion

The design and operations of inpatient care facilities are typically to a large extent historically shaped. Accomplishing a better match with the changing environment is often possible, and even inevitable due to the pressure on hospital budgets. As an illustration, Dutch hospitals observe a shift from inpatient to outpatient care as a result of technological developments and increased medical knowledge. Consequently, many of these hospitals are organized in many care units that slowly decrease in size. Low bed utilizations occur, while at the same time a national shortage of nursing staff is observed. Therefore, the

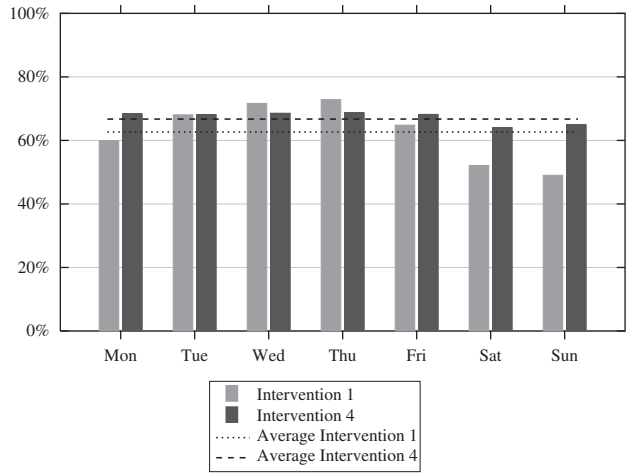

Figure 2 Average bed utilization per day of the week (care unit A, rejections < 1%).

majority of Dutch hospitals is reorganizing its inpatient clinic. In this paper, we have presented a generic analytical method that can support logistical decision making for inpatient care services, by quantitatively predicting the impact of different scenarios and interventions.

We are able to assist decision making on various planning levels. Insight can be gained on the impact of strategic (ie capacity dimensioning, case mix), tactical (ie the allocation of operating room time, misplacement rules), and operational decisions (ie time of admission/discharge). For these decisions, rules-of-thumb can be established. For example, we have shown the economies-of-scale effect: larger facilities can operate under a higher occupancy level than smaller ones in trying to achieve a given patient service level, since randomness balances out. In addition, by allowing overflow and setting appropriate rules, the benefits of bed capacity pooling are utilized, while the placement of patients on the preferred ward is maximized. Also, by adjusting the surgical schedule, extremely busy and quiet periods can be avoided. Once such basic rules are obtained, explicit interventions can be formulated of which the effect can be predicted. This combination between basic insights and quantifications is highly valuable to hospital managers.

The method is currently being used in the AMC in redesigning its inpatient care services, of which the improvement potential is substantial (as numerically illustrated in this paper). We illustrated in this paper how the method can be employed to predict the consequences of suggested improvement actions, and, as such, make recommendations to hospital managers so that they are eventually able to make better logistical decisions. The follow-up of choosing and implementing the preferred interventions takes place in a hospital-wide improvement programme that will be performed during the upcoming years. Such a process of drastically changing an existing health-care environment is often highly political. Our experience in the AMC is that the benefit of quantitative analysis in this 'negotiation' process is that it rationalizes the process of realizing a good trade-off between interests of clinicians and patients. We observe that quantification helps to formulate robust organizational plans, for instance also anticipating the expected increase of acute admissions due to a changing nature of the emergency department. Finally, in the AMC we observe that applying the method and discussing the results triggers the discussion to also focus on other potential gains like a more efficient use of the operating theatre.

The development of a user-friendly decision support tool (DSS) based on our method will be a next step in achieving practical impact. Our model relies on data which is easily extractable from typical hospital management systems. This makes it possible to automate the process of collecting the required input parameters to run the model. Integration with the hospital management system, visualization of the results, and the possibility to run what-if scenarios will be desired specifications of the DSS. Note that in such a DSS it would be desirable to not have to rely on formulating and calculating the impact of different interventions by trial-and-error. In future work, we will therefore focus on designing a formal automated approach to optimize resource capacity planning of inpatient care services based on the predictive model, translating the presented work from a predictive method to a true prescriptive method.

In future research, we will also focus on the following three directions. First, by combining the current inpatient census with the settled upcoming MSS, the model could be exploited to support last minute decision making like whether or not to hire temporary staff. Second, we will focus on incorporating the possibility of intermediate intensive care unit stays for patients who have undergone a complex surgery. Finally, the hourly level of the model will provide the basis for a formal approach along which effective and efficient nurse staffing can be achieved.

Acknowledgements—This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs.The authors would like to acknowledge Peter Vanberkel and Erwin Hans for the discussion that contributed to this study.

References

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.

Akcali E, Côté M and Lin C (2006). A network flow approach to optimizing hospital bed capacity decisions. *Health Care Management Science* **9**(4): 391–404.

Akkerman R and Knip M (2004). Reallocation of beds to reduce waiting time for cardiac surgery. *Health Care Management Science* **7**(2): 119–126.

Bain LJ and Engelhardt M (1992). *Introduction to Probability and Mathematical Statistics*. Vol. 2. Duxbury Press: Belmont, CA, USA.

Bekker R and De Bruin A (2010). Time-dependent analysis for refused admissions in clinical wards. *Annals of Operations Research* **178**(1): 45–65.

Bekker R and Koeleman P (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Management Science* **14**(3): 237–249.

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.

Cochran J and Bharti A (2006). Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science* **9**(1): 31–45.

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.

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.

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.

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.

Green L and Nguyen V (2001). Strategies for cutting hospital beds: the impact on patient service. *Health Services Research* **36**(2): 421–442.

Guerriero F and Guido R (2010). Operational research in the management of the operating theatre: A survey. *Health Care Management Science* **14**(1): 89–114.

Harper P (2002). A framework for operational modelling of hospital resources. *Health Care Management Science* **5**(3): 165–173.

Harper P and Shahani A (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society* **53**(1): 11–18.

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.

Harrison G, Shafer A and Mackay M (2005). Modelling variability in hospital bed occupancy. *Health Care Management Science* **8**(4): 325–334.

Hulshof P, Kortbeek N, Boucherie R, Hans E and Bakker P (2012). Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. *Health Systems* **1**: 129–175.

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.

PubMed (2012). http://www.pubmed.gov/, accessed 1 August 2012.

Ridge J, Jones S, Nielsen M and Shahani A (1998). Capacity planning for intensive care units. *European Journal of Operational Research* **105**(2): 346–355.

RVZ (2012). Council for Public Health and Health Care [Raad voor de Volksgezondheid & Zorg]. Medisch-specialistische zorg in 2020 (In Dutch). http://www.rvz.net/, accessed 1 August 2012.

Troy P and Rosenberg L (2009). Using simulation to determine the need for ICU beds for surgery patients. *Surgery* **146**(4): 608–620.

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.

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.

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.

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.

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.

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.

Appendix A

In the appendix, the derivations are presented that were omitted in the main article for reasons of readability. Note that the exposition is such that it is supplementary to the main text, and is therefore not intended to be comprehensible in isolation.

Demand predictions for elective patients

Single surgery block

To calculate $a^j_{n,t}(x)$, we first determine the admission process under a given number of performed surgeries y. Define $a^j_{n,t}(x\,|\,y)$ as the probability that x patients are admitted until time t on day n, given that y admissions take place in total. Then:

$$a^j_{n,t}(x \mid y) = \begin{cases} \binom{y}{x}\left(v^j_{n,t}\right)^x\left(1-v^j_{n,t}\right)^{y-x} & ,n=-1,t=0,\\ \sum_{g=0}^{x}\binom{y-g}{x-g}\left(v^j_{n,t}\right)^{x-g}\left(1-v^j_{n,t}\right)^{y-x}a^j_{n-1,T-1}(g \mid y) & ,n=0,t=0,\\ \sum_{g=0}^{x}\binom{y-g}{x-g}\left(v^j_{n,t}\right)^{x-g}\left(1-v^j_{n,t}\right)^{y-x}a^j_{n,t-1}(g \mid y) & ,n=-1,t=1,\ \ldots\ ,T-1,\\ & n=0,t=1,\ \ldots\ ,\vartheta_j-1,\\ 0 & ,n=0,t\geqslant\vartheta_j, \end{cases}$$

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

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

Finally,

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

To calculate $d^j_{n,t}(x)$, we first determine $d^j_n(x)$, for day 0 the probability that x patients are present at the start of the discharge

process ($t=\vartheta_j$) and for days $n>0$ the probability that x patients are present at the start of the day:

$$d_n^j(x)=\begin{cases} c^j(x) & ,n=0,\\ \sum_{g=x}^{C^j}\binom{g}{x}\left(s_{n-1}^j\right)^{g-x}\left(1-s_{n-1}^j\right)^x d_{n-1}^j(g) & ,n=1,\ldots,L^j,\end{cases}$$

where s_n^j is the probability that a type j patient who is still present at the begin of day n is discharged on day n:

$$s_n^j=\frac{P^j(n)}{\prod_{m=0}^{n-1}\left(1-s_m^j\right)}.$$

Starting from $d_n^j(x)$, we determine the day process:

$$d_{n,t}^j(x)=$$

$$\begin{cases} 0 & ,n=0,t<\vartheta_j,\\ d_n^j(x) & ,n=0,t=\vartheta_j \text{ and } n>0,t=0,\\ \sum_{k=x}^{C^j}\binom{k}{x}\left(z_{n,t-1}^j\right)^{k-x}\left(1-z_{n,t-1}^j\right)^x d_{n,t-1}^j(k) & ,n=0,t>\vartheta_j \text{ and } n>0,t>0,\end{cases}$$

where $z_{n,t}^j$ is the probability of a type j patient to be discharged during time interval $[t,t+1)$ on day n, given this patient is still present at time t:

$$z_{n,t}^j=\frac{m_{n,t}^j P^j(n)}{P^j(n)\sum_{i=t}^{T-1}m_{n,i}^j+\sum_{k=n+1}^{L^j}P^j(k)}.$$

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 $\overline{h}_{m,t}^{i,s}$ 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:

$$\overline{h}_{m,t}^{i,s}=\begin{cases}0 & ,m<s-1,\\ h_{m-s,t}^j & ,m\geqslant s-1,\end{cases}$$

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

$$H_{m,t}=\overline{h}_{m,t}^{1,1}\otimes\overline{h}_{m,t}^{1,2}\otimes\cdots\otimes\overline{h}_{m,t}^{1,S}\otimes\overline{h}_{m,t}^{2,1}\otimes\cdots\otimes\overline{h}_{m,t}^{I,S}. \quad \text{(A1.1)}$$

Steady state

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. $H_{s,t}^{SS}$ can be computed as follows:

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

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

Appendix B

Demand predictions for acute patient types

Single patient type

For patient type $j=(p,r,\theta)$, the admission process $\tilde{a}_t^j$ is determined by a non-homogeneous Poisson process:

$$\tilde{a}_t^j(x)=\frac{\left(\lambda^j\right)^x e^{-\lambda^j}}{x!},\quad t=\theta.$$

To calculate $\tilde{d}_{n,t}^j(x)$, we first determine $\tilde{d}_n^j(x)$, for day 0 the probability that x patients are present at the start of the discharge process ($t=\theta+1$) and for days $n>0$ the probability that x patients are present at the start of the day:

$$\tilde{d}_n^j(x)=$$

$$\begin{cases}\tilde{a}_\theta^j(x) & ,n=0,\\ \sum_{g=x}^{\infty}\binom{g}{x}\left(\tilde{s}_{n-1}^j\right)^{g-x}\left(1-\tilde{s}_{n-1}^j\right)^x\tilde{d}_{n-1}^j(g) & ,n=1,\ldots,L^j,\end{cases}$$

where $\tilde{s}_n^j$ is the probability that a type j patient who is still present at the begin of day n is discharged during day n:

$$\tilde{s}_n^j=\frac{P^j(n)}{\prod_{m=0}^{n-1}\left(1-\tilde{s}_m^j\right)}.$$

Starting from $\tilde{d}_n^j$, we determine the day process:

$$\tilde{d}_{n,t}^j(x)=$$

$$\begin{cases}0 & ,n=0,t\leqslant\theta,\\ \tilde{d}_n^j(x) & ,n=0,t=\theta+1 \text{ and } n>0,t=0,\\ \sum_{k=x}^{\infty}\binom{k}{x}\left(\tilde{z}_{n,t-1}^j\right)^{k-x}\left(1-\tilde{z}_{n,t-1}^j\right)^x\tilde{d}_{n,t-1}^j & ,n=0,t>\theta+1 \text{ and } n>0,t>0,\end{cases}$$

where $\tilde{z}_{n,t}^j$ is the probability of a type j patient to be discharged during time interval $[t,t+1)$ on day n, given this patient is still

present at time t:

$$\tilde{z}^j_{n,t} = \frac{\tilde{m}^j_{n,t} P^j(n)}{P^j(n) \sum_{i=t}^{T-1} \tilde{m}^j_{n,i} + \sum_{k=n+1}^{L^j} P^j(k)}.$$

Single cycle

Appendix C

Performance indicators

In this appendix, the derivation of $Z^k_{q,t}(x_k \mid n)$ is presented. To this end, 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,

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

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

$$\left\{ \prod_{j=1}^{\omega} \alpha^{\sigma(j)}_{q,t}\left(y_{\sigma(j)}\right) \breve{a}^{\sigma(j)}_{q,t}\left(n_{\sigma(j)} \mid y_{\sigma(j)}\right) \right\},$$

To determine the overall probability distribution of the number of patients in recovery resulting from a single AAC, define $\overline{g}^j_{w,t}$ as the probability distribution of the number of recovering patients of type j present at time interval t on day $w (w \in \{0, 1, 2, \ldots, R, R+1, R+2, \ldots\})$. The distribution $\overline{g}^j_{w,t}$ is given by:

$$\overline{g}^j_{w,t} = \overline{g}^{p,r,\theta}_{w,t} = \begin{cases} 0 & , w < r, \\ g^j_{w-r,t} & , w \geqslant r. \end{cases}$$

Then, $G_{w,t}$ is computed by:

$$G_{w,t} = \overline{g}^{1,1,0}_{w,t} \otimes \cdots \otimes \overline{g}^{1,1,T-1}_{w,t} \otimes \overline{g}^{1,2,0}_{w,t} \otimes \cdots \otimes \overline{g}^{1,2,T-1}_{w,t} \otimes \overline{g}^{2,1,0}_{w,t} \otimes \cdots \otimes \overline{g}^{P,R,T-1}_{w,t}. \quad \text{(B2.1)}$$

Steady state

$G^{SS}_{r,t}$ can be computed as follows:

$$G^{SS}_{r,t} = G_{r,t} \otimes G_{r+R,t} \otimes G_{r+2R,t} \otimes \cdots \otimes G_{r+\lceil W/R \rceil R,t},$$

where $W = \max\{r \mid \exists\, t, x \text{ with } G_{r,t}(x) > 0\}$.

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 $\sigma(1), \ldots, \sigma(\omega)$ 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 $\alpha^j_{q,t} = \alpha^j_{q,t}$ for the elective patient types and $\alpha^j_{q,t} = \tilde{a}^{(p,\, q \bmod R + R \cdot 1_{q \bmod R = 0},\, t)}_t$ for the acute patient types. It remains to define $\breve{a}^j_{q,t}(n_j \mid y_j)$, the probability that for an arriving cohort, from the y_j patients present in total, n_j arrivals occur during time interval $[t, t+1)$:

$$\breve{a}^j_{q,t}\left(n_j \mid y_j\right) = \binom{y_j}{n_j} \left(\nu^j_{n,t}\right)^{n_j} \left(1 - \nu^j_{n,t}\right)^{y_j - n_j},$$

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

Received 6 September 2012;
accepted 2 June 2014 after one revision