Available online at www.sciencedirect.com

ScienceDirect

European Journal of Operational Research 185 (2008) 998–1010

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

Managing the overflow of intensive care patients

Nelly Litvak [a,*], Marleen van Rijsbergen [b], Richard J. Boucherie [a], Mark van Houdenhoven [c]

[a] Faculty of Electrical Engineering, Mathematics and Computer Science, Department of Applied Mathematics, University of Twente, 7500 AE Enschede, The Netherlands

[b] ORTEC bv, Groningenweg 6-33, 2803 PV Gouda, The Netherlands

[c] Cluster 17, Erasmus Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Available online 16 October 2006

Abstract

Many hospitals in the Netherlands are confronted with capacity problems at their intensive care units (ICUs) resulting in cancelling operations, overloading the staff with extra patients, or rejecting emergency patients. In practice, the last option is a common choice because for legal reasons, as well as for hospital logistics, rejecting emergency patients has minimal consequences for the hospital. As a result, emergency patients occasionally have to be transported to hospitals far away. In this work, we propose a cooperative solution for the ICU capacity problem. In our model, several hospitals in a region jointly reserve a small number of beds for regional emergency patients. We present a mathematical method for computing the number of regional beds for any given acceptance rate. The analytic approach is inspired by overflow models in telecommunication systems with multiple streams of telephone calls. Simulation studies show that our model is quite accurate. We conclude that cooperation between hospitals helps to achieve a high acceptance level with a smaller number of beds resulting in improved service for all patients.

Keywords: OR in health services; Intensive care units; Overflow models; Equivalent random method

1. Introduction

"Each year, hundreds of patients die unnecessarily." This was announced in the Dutch current affairs program NOVA on November 6th, 2001, during the discussion on the capacity shortage at Intensive Care Units (ICUs) in Dutch hospitals [1]. The Dutch minister of Health, Welfare and Sports recognized the problems and initiated studies into the capacity problems of ICUs. A primary report [7] indicated that almost 10% of the severely ill patients were refused, 4% were admitted even though there was actually no space, and 3% were released earlier to make place for new patients. The most important reason for the refusal of a patient was the lack of operational (staffed) IC beds caused mostly by shortage of nurses. The ICU capacity problem

* Corresponding author.
E-mail addresses: N.Litvak@ewi.utwente.nl (N. Litvak), m.vanrijsbergen@alumnus.utwente.nl (M. van Rijsbergen), R.J.Boucherie@utwente.nl (R.J. Boucherie), m.vanhoudenhoven@erasmusmc.nl (M. van Houdenhoven).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.08.021

for emergency or trauma patients (victims of accidents) is strengthened by the complicated chain logistics of hospitals. In particular, cancellation of planned operations due to ICU capacity shortage is highly expensive. As a consequence, trauma patients are refused to accommodate these planned operations.

In the Netherlands, care for trauma patients is organized in a regional setting. In principle, each trauma patient has to be admitted to an ICU within the region. Only when all ICU beds in a region are occupied, a trauma patient may be transported to a hospital outside that region, with obvious degradation of the quality of health care due to, e.g., extended transportation times. In the current situation, where each ICU decides independently whether or not a trauma patient is admitted, it may be that a trauma patient is transferred outside the region due to simultaneous reservation of capacity at some of the ICUs, while other ICUs have actually reached their capacity. An initial capacity study [2] indicates that indeed in the Rijnmond Region sufficient ICU capacity seems to be available, and that lack of cooperation is a major cause for trauma patients to be transported outside the region.

This paper focuses on solutions for cooperation among ICUs so as to minimize the number of trauma patients transported outside the region while maintaining a sufficient amount of ICU beds for planned operations. We show that reservation of several ICU beds for regional trauma patients and sharing these beds among the hospitals (so-called *regional beds*) results in a higher acceptance rate for emergency patients with a smaller number of beds in the region, without serious degradation of the fraction of cancelled operations. This is mainly due to the more efficient use of ICU capacity. Cooperation among hospitals thus helps to achieve a high acceptance level with a smaller number of beds resulting in improved service for all patients.

This paper provides a mathematical model for regional capacity allocation at ICUs under constraints on the number of refused patients. The model includes regional ICU capacity for regional emergency patients, and contains a detailed description of patient classes admitted to ICUs, and of solutions to accommodate bed shortages. Typical solutions in case of bed shortage are: transferring a patient to another hospital/region; postponing a planned operation; and releasing another patient earlier. These solutions have serious drawbacks, and the solution also depends on the patient class. Patients arriving at an ICU are of three classes, which mainly differ in the decision for admittance to the ICU. An elective patient may require an ICU bed following a planned operation. A planned operation can start only when an ICU bed is available. When all ICU beds are occupied, the operation is cancelled. An internal trauma patient, due to e.g. an emergency at the ward, must always be admitted to the ICU. When all ICU beds are occupied, a so-called *over-bed* is created. An over-bed is an originally non-staffed bed which is forcefully brought into operation thus loading the staff with an extra patient. This results in a decreased level of care at the ICU. A regional trauma patient, due to e.g. an accident in the region, is accepted only when an ICU bed is available. Otherwise the patient is not admitted and sent to another ICU. From a mathematical perspective, a regional model for ICUs shows major similarities with queueing theoretical models developed for circuit switched telephone systems with overflow capacity. For such systems, the highly accurate Equivalent Random Method (ERM) allows us to approximate the fraction of blocked telephone calls [11]. Unfortunately, internal emergency patients placed in over-beds cannot be included in the ERM. Therefore, in this paper, we develop a generalisation of the ERM that also allows for these patients. A detailed simulation study indicates that our generalisation of the ERM accurately approximates the fraction of refused patients.

A case study focusing on the Rijnmond Region indicates the capacity gain that may be achieved. In this region, the Erasmus Medical Centre (Erasmus MC) is appointed as one of the ten trauma centres in the Netherlands (see the National Atlas of Public Health [3]). According to a strategic analysis of cluster 17 of the Erasmus MC, responsible for Anesthesiology, ICUs, and Operating Theatres, the number of trauma patients sent to the ICU of the Erasmus MC has increased since its recognition as a trauma centre [2]. Some of the capacity problems at the ICU of the Erasmus MC are presumably caused by other hospitals in the region, which are not willing to cancel elective (planned) operations to allow for admission of emergency patients. As indicated in [2], it seems that the operational IC capacity in the Rijnmond Region reasonably approaches the demand for IC beds. At present however, emergency patients are occasionally sent outside the Rijnmond Region because no operational bed can be found in the region. If all hospitals in the region allocate several IC beds as emergency beds, the region can most likely take care of most of the emergency patients in the Rijnmond Region [2]. This does imply that sometimes hospitals might have to cancel elective operations while having an empty operational bed. Our case study indicates that the increased ICU bed occupation due to regional cooperation indeed reduces the fraction of regional trauma patients not accepted at an ICU in the region, but that this is *not* at the cost of an increased fraction of cancelled planned operations.

The paper is organized as follows. In the next section we describe the structure of the ICU, and available data. In Section 3, we present an overflow model of an ICU inspired by closely related models in telecommunications systems. In Section 4 we carry out the analysis of the model and provide the method for computing the fraction of rejected patients. Section 5 provides a simulation study to indicate the accuracy of our approximation, and is devoted to the case study for the Rijnmond Region. Conclusions and recommendations are given in the final Section 6.

2. Patient flows in the ICU

Intensive Care is specific medical treatment and nursing to severely ill patients who require intensive monitoring, mostly elaborate pharmacological treatment and in many cases support with artificial ventilation. The admission and release of a patient in the ICU is subject to a number of rules [5]. There are, however, no unambiguous agreements on how to deal with an arriving patient when no operational IC bed is available. An IC bed is operational when sufficient staff is available.

In practice, one can roughly distinguish three patient types: elective patients, internal emergency patients and external/regional emergency patients. Elective patients arrive from the operating theatre after undergoing a planned operation. If no operational IC bed is available, the operation is cancelled. An exception is made for operations that involve many people (staff and patients), for example a liver transplantation with a living donor. For such patients, beds are reserved that will not be taken by another patient.

Emergency patients arrive unexpectedly and require immediate care. Internal emergency patients arrive from a nursing ward. Regional emergency patients arrive through the emergency room, mostly brought by ambulance. The ambulance nurse does not have information on the availability of IC beds. If there is no bed available for an emergency patient, an attempt is made to create a place. For instance, another patient may be predischarged from the ICU but only if the discharge of the patient was already imminent. Also, a patient who came from a different hospital for some special procedure may be sent back if the special procedure is finished. If none of these options is available, the solution depends on the type of patient.

An internal emergency patient should be kept in the hospital mostly because it is not desirable to transport a critically ill patient, but also because legally, a patient can only be transferred if it is beneficial for the patient. Therefore, for an internal emergency patient an over-bed is created, which is an IC bed that was not staffed. The drawback of the over-bed is that physicians and nurses have to work harder as they have an extra patient to take care of, which requires flexible staff and negatively affects the quality of care. As soon as a patient is discharged, the over-bed is cancelled. For regional patients an over-bed is generally not an option because the hospitals tend to give priority to already admitted patients, and legally, a patient not yet admitted to the hospital can be sent to another hospital. Thus, for a regional emergency patient, generally an operational bed in another hospital is sought, and sometimes an available bed can be found only outside the region.

Fig. 1 schematically depicts the patient flows for two ICUs. Flow 1 reflects the regional emergency patients, who are transferred to another hospital/region if all the beds are occupied. Flow 2 is the flow of elective patients. If no operational bed is available at their arrival, they are sent home to return later. Flow 3 corresponds to the flow of internal emergency patients who are not transferred in case of a full ICU, but are placed in an over-bed. Flow 4 depicts the patients whose discharge is imminent and who can be predischarged in case of an incoming emergency. We do not take this flow into account. Flow 5 is the flow of patients who leave the ICU (because of recovery or mortality). The Overflow block denotes the patients who are rejected at the ICU.

The order of magnitude of the number of arriving patients, and the length of stay (LOS) are required for the selection of a proper approximation. According to the data presented in [9] for the Erasmus MC, the average inter-arrival time is 0.18 days. More detailed data on different patients types are given in Table 1. As the elective patients never arrive at weekends, we also provide the mean and standard deviation of the interarrival times for elective patients leaving out the weekend days.

The total mean LOS given in [9] for the Erasmus MC is 6.93 days. Table 2 contains the mean LOS for the three types of patients. The mean LOS of the elective patients differs significantly from the two types of emergency patients. The LOS is measured in whole days and includes the arrival and the release days not taking

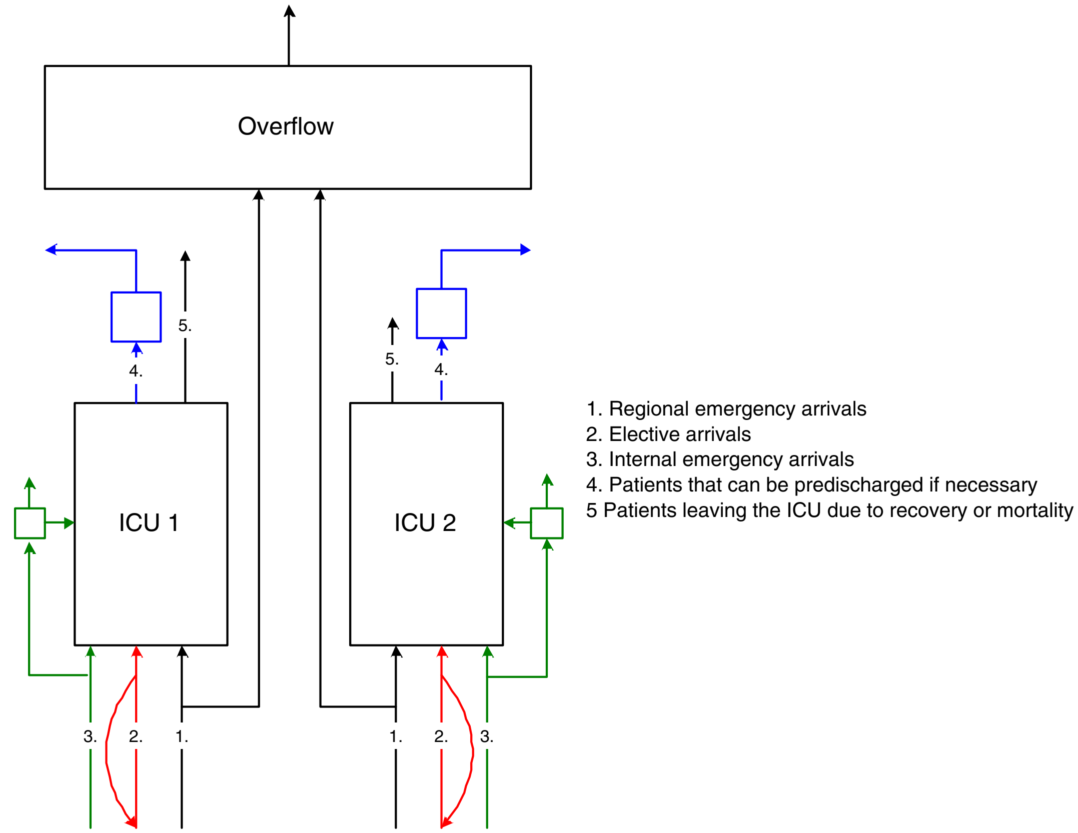

Fig. 1. Overview of the ERM including all patient streams.

Table 1
Interarrival times in days for Erasmus MC

Type of arrival	Mean	Standard deviation
Elective	0.58	0.92
Elective excluding weekend days	0.42	0.79
Internal emergency	0.62	0.74
Regional emergency	0.46	0.60
Total	0.18	0.39

Table 2
Mean length of stay in days for Erasmus MC

Type of arrival	Mean	Standard deviation
Elective	3.88	6.44
Internal emergency	8.15	12.69
Regional emergency	7.95	13.78
Total	6.93	11.90

into account the time of release/arrival. It is also shown in [9] that the data on the LOS fits a Log Normal distribution.

The number of operational IC beds ranges from 5 for small hospitals to 40 for larger hospitals. Typically, the occupation degree of IC beds is above 80%.

3. Overflow model for regional ICU capacity

Consider a region containing multiple ICUs that jointly reserve beds (regional beds) for regional emergency patients only. In this case, the overflow block in Fig. 1 depicts this regional emergency capacity, consisting of an extra ICU that is intended for regional emergency patients who are refused at an original ICU. In practice, these beds will be distributed over the ICUs in the region, but will be reserved for regional emergency patients, thus creating a virtual ICU. Our goal is to compute the fraction of rejected patients (rejection probabilities).

We assume that all hospitals have similar patient stream structure. Assume that patients arrive to the hospital according to a Poisson flow. For the emergency arrivals this assumption is reasonable and is supported by statistical data [9]. The elective arrivals, however, are scheduled and therefore most likely do not constitute a Poisson flow. However, a surgeon is not aware of the occupation of the ICU when planning operations. As only a fraction of 5% of operated patients require Intensive Care after the operation, the assumption of Poisson arrivals is reasonable. In our model, ICUs may have a different mean arrival rate, reflecting the size of the area immediately surrounding a hospital from which patients are sent to the ICU. Let λ_i denote the total arrival rate (average number of patients arriving per time unit) at ICU i. The fraction of regional emergency patients, elective patients, and internal emergency patients is denoted as $p_{1,i}$, $p_{2,i}$, and $p_{3,i}$, respectively, with $p_{1,i} + p_{2,i} + p_{3,i} = 1$. The return of elective patients after a cancelled operation is modelled as a new arrival.

For analytical tractability, we assume that the LOS is exponentially distributed. A large class of queueing loss models is insensitive to the distribution of the service time. In Section 5 we present simulation results that support this kind of insensitivity in our model, and justify the assumption of exponential LOS.

To simplify notation, we do not discriminate between the mean LOS of different patient types. The mean LOS for patients at ICU i is denoted as μ_i^{-1}. Data indicate that the LOS is indeed similar for internal and regional emergency patients [9]. For elective patients the LOS is generally smaller and less variable. Nevertheless, the model with equal mean LOS provides a good approximation and can be readily extended to the case of different mean LOS for different patient types.

4. Analysis

From a mathematical perspective, the behaviour of ICUs with shared regional capacity closely resembles that of a circuit switched telephone system with common overflow. In that system, a telephone call occupies a circuit during its call-length, and a call generated when all circuits are occupied is blocked and clear. To see the resemblance, identify patients with calls, beds with circuits (servers), and LOS with call-length. By capacity of the system we mean the number of servers, and we use these terms interchangeably throughout the paper. The computation of call blocking probabilities in such systems is an important research question that has received considerable attention in the literature. In the simplest case of one telephone switch with one incoming flow and c circuits, the system is referred to as the Erlang loss system, and the blocking probability can be computed using the famous Erlang loss formula [6]:

$$\text{Blocking probability} = B(c, \rho) = \frac{\rho^c/c!}{\sum_{k=0}^{c} \rho^k/k!},$$

where $\rho = \lambda\mu^{-1}$ is the load, with λ the call arrival rate, and μ^{-1} the mean call length.

Real-life systems, however, require analysis that is far beyond this basic model. For instance, the problem becomes much more complex when several multi-server units share a common overflow. To approximate the blocking probabilities in this model, the Equivalent Random Method (ERM) introduced by Wilkinson [11] can be efficiently applied. The idea of the classical ERM and its numerous modifications is to replace several multi-server units by one Equivalent Random unit that generates the same expectation and variance of the overflow as in the original system. Then the Erlang loss formula can be applied as for a classical loss system with equivalent random load ρ and capacity $c + r$, where r is the capacity of the overflow buffer, and c is the capacity of the Equivalent Random unit (see Fig. 2).

More formally, consider a system of I multi-server units. In order to apply the ERM we need to find an equivalent random load ρ and capacity c. To this end, first consider an overflow with unlimited capacity.

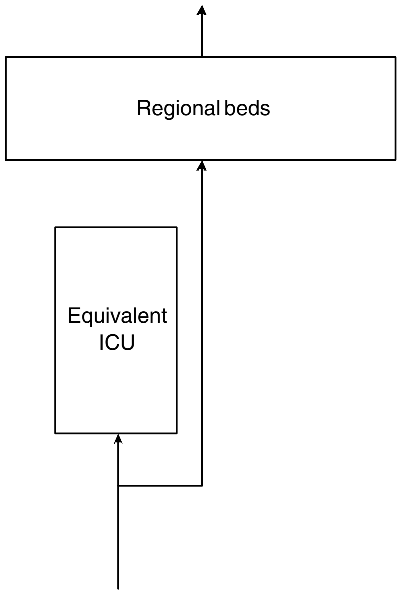

Fig. 2. Equivalent Random IC with regional emergency patients.

The mean and variance of the number of calls in the overflow from unit $i = 1, \ldots, I$, with load $\rho_i = \lambda_i/\mu_i$, and capacity c_i, are [6]

$$E_i = \rho_i B(c_i, \rho_i), \quad V_i = E_i\left(1 - E_i + \frac{\rho_i}{c_i + 1 + E_i - \rho_i}\right).$$

The mean and variance of the *total* number of calls in the overflow buffer, assuming that the latter has an unlimited capacity, is

$$E = \sum_{i=1}^{I} E_i, \quad V = \sum_{i=1}^{I} V_i. \tag{1}$$

The Equivalent Random system is the Erlang loss queue with capacity c and load ρ that satisfy [6]

$$E = \rho B(c, \rho), \quad V = E\left(1 - E + \frac{\rho}{c + 1 + E - \rho}\right). \tag{2}$$

System (2) can readily be solved numerically. We can also find a solution using analytic approximations such as the equations given by Rapp [10]:

$$\rho = V + 3\frac{V}{E}\left(\frac{V}{E} - 1\right), \quad c = \frac{\rho\left(E + \frac{V}{E}\right)}{E} + \frac{V}{E - 1} - E - 1. \tag{3}$$

Cooper [6] states that these estimates of ρ and c are generally on the high side of the exact values. Rounding c down to an integer $\lfloor c \rfloor$ and then finding ρ by

$$\rho = \frac{(\lfloor c \rfloor + E + 1)\left(E + \frac{V}{E} - 1\right)}{E + \frac{V}{E}}, \tag{4}$$

gives a better approximation.

Let r be the capacity of the overflow determined above. Once ρ and c for the Equivalent Random unit are defined, we can compute the approximate average number $\overline{E}$ that is rejected at the overflow

$$\overline{E} = \rho B(c+r,\rho) = \rho \frac{\rho^{c+r}/(c+r)!}{\sum_{k=0}^{c+r}\rho^k/k!}. \tag{5}$$

Now consider a region with regional ICU. Compared to the known versions of ERM, our model is different because (i) internal emergency patients cannot be rejected, and (ii) elective patients are never sent to the overflow. In order to apply the ERM, we have to be able to compute the mean E and variance V of the overflow for our model with three patient flows and the possibility of over-beds.

From (1), it is sufficient to find E_i and V_i for the ith ICU. For $j, k \geqslant 0$, let $P_i(j,k)$ be the steady-state probability that there are j patients at ICU i and k patients at the overflow, $j \geqslant 0$, $k \geqslant 0$. Setting $P_i(-1,k) = 0$, $k \geqslant 0$, we can write the global balance equations that uniquely determine these probabilities as follows:

$$\begin{aligned}&(\lambda_i + j\mu_i + k\mu_i)P_i(j,k) = \lambda_i P_i(j-1,k) \\ &\quad + (j+1)\mu_i P_i(j+1,k) + (k+1)\mu_i P_i(j,k+1), \quad j < c_i;\ k \geqslant 0,\end{aligned} \tag{6}$$

$$\begin{aligned}&((p_{1,i}+p_{3,i})\lambda_i + c_i\mu_i + k\mu_i)P_i(c_i,k) = \lambda_i P_i(c_i-1,k) + p_{1,i}\lambda_i P_i(c_i,k-1) \\ &\quad + (c_i+1)\mu_i P_i(c_i+1,k) + (k+1)\mu_i P_i(c_i,k+1), \quad k \geqslant 0,\end{aligned} \tag{7}$$

$$\begin{aligned}&((p_{1,i}+p_{3,i})\lambda_i + j\mu_i + k\mu_i)P_i(j,k) = p_{3,i}\lambda_i P_i(j-1,k) + p_{1,i}\lambda_i P_i(j,k-1) \\ &\quad + (j+1)\mu_i P_i(j+1,k) + (k+1)\mu_i P_i(j,k+1), \quad j > c_i;\ k \geqslant 0.\end{aligned} \tag{8}$$

The left-hand side of (6) represents the probability flow out of state (j,k) due to arrivals at rate λ_i, departures of patients from the ICU at rate $j\mu_i$, and departures from the overflow at rate $k\mu_i$. The right-hand side represents the probability flow into state (j,k) due to arrivals from state $(j-1,k)$, due to departures from the ICU from state $(j+1,k)$, and due to departures from the overflow from state $(j,k+1)$. The other equations have a similar interpretation.

We are interested in the mean and variance. To obtain expressions for these measures, let $G_{i,j}(z)$ be the marginal generating function

$$G_{i,j}(z) = \sum_{k=0}^{\infty} P_i(j,k)z^k, \quad |z| \leqslant 1.$$

Multiplying the balance Eqs. (6)–(8) by z^k, $|z| \leqslant 1$, and summing both sides of the equations over k, we obtain the following relations:

$$\begin{aligned}&(\lambda_i + j\mu_i)G_{i,j}(z) = \lambda_i G_{i,j-1}(z) + (j+1)\mu_i G_{i,j+1}(z) \\ &\qquad + \mu_i(1-z)\frac{d}{dz}G_{i,j}(z), \quad j < c_i,\end{aligned} \tag{9}$$

$$\begin{aligned}&((p_{1,i}(1-z) + p_{3,i})\lambda_i + c_i\mu_i)G_{i,c_i}(z) = \lambda_i G_{i,c_i-1}(z) \\ &\qquad + (c_i+1)\mu_i G_{i,c_i+1}(z) + \mu_i(1-z)\frac{d}{dz}G_{i,c_i}(z),\end{aligned} \tag{10}$$

$$\begin{aligned}&((p_{1,i}(1-z) + p_{3,i})\lambda_i + j\mu_i)G_{i,j}(z) = p_{3,i}\lambda_i G_{i,j-1}(z) + \\ &\qquad (j+1)\mu_i G_{i,j+1}(z) + \mu_i(1-z)\frac{d}{dz}G_{i,j}(z), \quad j > c_i.\end{aligned} \tag{11}$$

The expectation and variance of the overflow can be now calculated by using first and second order derivatives of $G_{i,j}(z)$ with respect to z as follows:

$$E_i = \sum_{j=0}^{\infty} \frac{\partial}{\partial z} G_{i,j}(z)\bigg|_{z=1},$$

$$V_i = \sum_{j=0}^{\infty} \frac{\partial^2}{\partial z^2} G_{i,j}(z)\bigg|_{z=1} + E_i - (E_i)^2.$$

Therefore, differentiating both sides of Eqs. (9)–(11) with respect to z and substituting $z = 1$ and denoting $E_i(j) = \frac{\partial}{\partial z}G_{i,j}(z)|_{z=1}$ we obtain

$$(\lambda_i + (j+1)\mu_i)E_i(j) = \lambda_i E_i(j-1) + (j+1)\mu_i E_i(j+1), \quad 0 \leqslant j < c_i, \tag{12}$$
$$(p_{3,i}\lambda_i + (c_i+1)\mu_i)E_i(c_i) = \lambda_i E_i(c_i-1) + (c_i+1)\mu_i E_i(c_i+1) + p_{1,i}\lambda_i P_i(c_i), \tag{13}$$
$$(p_{3,i}\lambda_i + (j+1)\mu_i)E_i(j) = p_{3,i}\lambda_i E_i(j-1) + (j+1)\mu_i E_i(j+1) + p_{1,i}\lambda_i P_i(j), \quad j > c_i, \tag{14}$$

where $P_i(j) = G_{i,j}(1)$ is the probability that there are j patients at ICU i, and $E_i(-1) = 0$, $i = 1, \ldots, I$. To obtain E_i, we sum the above equations over j. It now follows that

$$E_i = p_{1,i}\rho_i \sum_{j=c_i}^{\infty} P_i(j). \tag{15}$$

To find the variance of the number of patients in the overflow, we take the second order derivatives in (9)–(11) with respect to z and put $z = 1$. Summing up the resulting equations, we obtain the second factorial moment of the number of patients in the overflow, which leads to the following expression for the variance:

$$V_i = p_{1,i}\rho_i \sum_{j=c_i}^{\infty} E_i(j) + E_i - (E_i)^2. \tag{16}$$

It now remains to find $\sum_{j=c_i}^{\infty} P_i(j)$ and $\sum_{j=c_i}^{\infty} E_i(j)$. The probabilities $P_i(j)$, $j \geqslant 0$, can be found by iteratively solving (9)–(11) with $z = 1$. This gives

$$P_i(j) = \begin{cases} \frac{1}{j!}(\rho_i)^j P_i(0), & 0 \leqslant j \leqslant c_i; \\ \frac{1}{j!}(p_{3,i})^{j-c_i}(\rho_i)^j P_i(0), & j > c_i. \end{cases} \tag{17}$$

From (17) and the normalizing condition $\sum_{j=0}^{\infty} P_i(j) = 1$ we obtain $P_i(0)$:

$$P_i(0) = \left[\sum_{j=0}^{c_i} \frac{(\rho_i)^j}{j!} + \sum_{j=c_i+1}^{\infty} \frac{(\rho_i)^j}{j!}(p_{3,i})^{j-c_i}\right]^{-1}. \tag{18}$$

It now follows from (17) and (18) that

$$\sum_{j=c_i}^{\infty} P_i(j) = 1 - \sum_{j=0}^{c_i-1} \frac{(\rho_i)^j}{j!} \cdot \left[\sum_{j=0}^{c_i} \frac{(\rho_i)^j}{j!}(1 - (p_{3,i})^{j-c_i}) + (p_{3,i})^{-c} e^{p_{3,i}\rho_i}\right]^{-1}. \tag{19}$$

The expression for $\sum_{j=c_i}^{\infty} E_i(j)$ can be found similarly. Iterating (12)–(14), we can express $E_i(j)$ via $E_i(0)$ for all $j > 0$. Then $E_i(0)$ can be found from the normalisation $E_i = \sum_{j=0}^{\infty} E_i(j)$. This will give the values of $E_i(j)$ for all $j > 0$ and thus we can compute the sum on the right-hand side of (16). The resulting analytical expression is, however, cumbersome and will not be given here. Instead, we suggest a simpler computational approach. Since $E_i(j)$ decreases quickly with j, we may iterate (12)–(14) only up to some sufficiently large value of $j = M$. Then $E_i(0)$ can be found from $E_i = \sum_{j=0}^{M} E_i(j)$, and the required expression will be $E_i = \sum_{j=c_1}^{M} E_i(j)$. This approach reflects reality, for instance, if M is the number of constructional beds. On the other hand, if we want an accurate solution of the proposed model with the unlimited over-bed capacity, we can choose M large enough so that the resulting values of $E_i(0)$ are sufficiently close for M and $M - 1$, and $E_i(M)$ is close to zero.

Having computed the mean E and the variance V of the overflow, we can use (2) or (3) and (4) to define ρ and c of the Equivalent Random ICU. An expression for the number of patients rejected at the overflow of capacity r is then given in (5). The loss probability $B^{(r)}$ for regional emergency patients is

$$B^{(r)} = \frac{\overline{E}}{\sum_{i=1}^{n} \rho_i p_{1,i}}, \tag{20}$$

where $\rho_i p_{1,i} = \lambda_i p_{1,i}/\mu_i$ is the load of regional emergency patients at ICU i. The blocking probability $B_i^{(r)}$ for an emergency patient arriving at ICU i can be approximated as follows. For $i = 1, \ldots, I$, let

$$B_i = \sum_{j=c_i}^{\infty} P_i(j)$$

be the probability that ICU i is full. According to the PASTA property this is the probability of rejection of a regional patient at the original ICU. Thus, the probability that an emergency patient attempts to access a regional bed equals $\sum_{i=1}^{I} B_i(p_{1,i}\lambda_1/\lambda_{1\bullet})$, where $\lambda_{1\bullet} = p_{1,1}\lambda_1 + \cdots + p_{1,I}\lambda_I$ is the total arrival rate of regional emergency patients, and $p_{1,i}\lambda_i/\lambda_{1\bullet}$ is the probability that an emergency patient claiming a regional bed comes from the ICU i. Assume that the rejection probability at the regional ICU, B_0 say, is the same for patients originating from any ICU. Hence, using the total probability formula for the blocking probability $B^{(r)}$, for any $i = 1, \ldots I$, $r \geqslant 0$, we write:

$$B^{(r)} \approx \left[\sum_{i=1}^{I} B_i(p_{1,i}\lambda_i/\lambda_{1\bullet})\right] B_0,$$

so that

$$B_i^{(r)} \approx B_i B_0 \approx \frac{\lambda_{1\bullet} B_i B^{(r)}}{\sum_{i=1}^{I} B_i p_{1,i}\lambda_i}. \tag{21}$$

Note that the equivalent random load ρ and capacity c are defined from the mean and variance of the overflow which consists only of emergency patients. Thus, the Equivalent Random ICU has load and capacity related only to the regional emergency flow. However, the blocking probability is also related only to the regional emergency patients. Therefore, one may hope that $B^{(r)}$ in (20) provides a good approximation for the real percentage of rejected regional patients. The numerical results in the next section show that this is indeed the case.

5. Simulation model and numerical results

This section contains both a simulation study of patient flows to investigate the accuracy of the ERM approximation, and a case study for the Rijnmond Region in the Netherlands. Data for the simulation study are obtained from a database of the Erasmus MC containing detailed information on patients, operations, and LOS for the years 1994–2004.

5.1. Accuracy of the ERM approximation

To investigate the accuracy of the ERM approximation, a simulation model has been developed in eM-Plant, version 7.0.2. eM-Plant is software for object-oriented, graphical modelling for simulating and visualizing systems and business processes [4]. Our simulation model is generic in the sense that the number of ICUs in the region, the number of beds per ICU, the arrival times and Length Of Stay (LOS) can all be adjusted. The simulation study includes detailed acceptance rules, and closely mimics the actual patient flows in ICUs including general LOS. The aim of the simulation study is to (i) investigate the influence of the distribution of the LOS, and (ii) investigate the accuracy of the ERM approximation.

The main frame of the simulation model represents the region which contains several ICUs and a unit with a number of regional beds. The three types of patients arrive at an ICU according to a Poisson process, each with its own rate. Elective patients do not arrive on weekends. If a bed is available the patient is treated at this ICU. The length of stay of the patient is modelled through a LogNormal distribution, each patient type having a different mean LOS. In the case when no beds are available and an internal emergency patient arrives, an over-bed is created for this patient. When no bed is available upon arrival of an elective patient, the patient is deleted from the system. When a regional emergency patient arrives and no bed is available, the patient is sent to a regional bed (in the frame of the region). Fig. 3 shows the basic patient streams in the simulation model with one ICU.

In the simulation model, we used the actual data from the IC department of the Erasmus MC and the estimated data from three other hospitals in the Rijnmond Region, the Albert Schweizer hospital, hospital Dirksland, and the Sint-Franciscus hospital. Further hospitals in the Rijnmond Region have not been taken into account. We have made 20 simulation runs for each test. Results of the first 10 runs are used to determine the warm-up period. This is necessary because the system starts empty, which is not the case in reality. Results

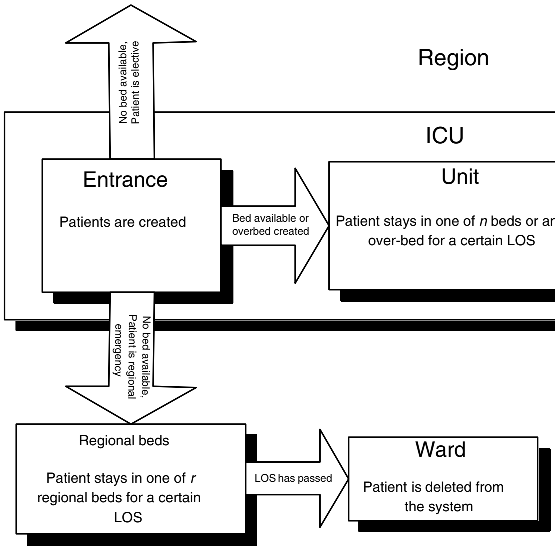

Fig. 3. The basic patient streams in the simulation model.

of the second 10 runs are used to calculate the 95% confidence intervals with at least 10% relative precision by the Replication/Deletion Method [8].

The first aim in the simulation study is to compare the results for LogNormal and exponential length of stay. As was mentioned in Section 2, the LogNormal distribution fits the real-life data, whereas the exponential distribution is used for analytical tractability. Table 3 gives 95% confidence intervals for the mean bed occupation, the proportion of cancelled operations, the proportion of refused emergency patients and the average number of over-beds, for exponentially and LogNormally distributed LOS, as well as the confidence interval for the difference, based on Common Random Numbers (CRN) [8]. For all performance measures except mean bed occupancy, 0 is contained in the confidence interval based on CRN for the difference between performance measures under LogNormal and exponential LOS. This result supports the use of the exponential distribution in our analytical model to approximate the performance measures for a region with cooperating ICUs. Insensitivity to the LOS distribution is a common property of the Erlang loss model, the mathematical model underlying our approximation, which suggests that the accuracy of our results will not change when compared to results obtained using the empirical distribution of the LOS.

The second aim of the simulation study is to verify the quality of the ERM approximation of the blocking probability $B^{(r)}$ for regional emergency patients. Table 4 contains the results for different numbers of regional beds. As can be seen from these results, the ERM provides an engineering approximation (roughly 10% accuracy) of the fraction of rejected regional emergency patients. ERM overestimates the loss probability for a small number of beds and underestimates the loss probability for a large number of beds. The reason may be that ERM smooths the discrepancy between distinct ICUs. We conclude that ERM captures the loss probability and the required number of regional beds with good precision.

We have also used simulation to verify analytical formulae for blocking probabilities at each hospital separately, with and without cooperation. For the case with cooperation, we used the formula (21) to determine

Table 3
Exponential versus LogNormal LOS, based on the Erasmus MC data

Confidence intervals	Exponential	LogNormal	Difference (CRN)
Mean bed occupation	0.90 ± 0.002	0.89 ± 0.001	0.003 ± 0.002
Proportion of cancelled operations	0.26 ± 0.006	0.26 ± 0.005	0.003 ± 0.009
Proportion of refused regional patients	0.18 ± 0.005	0.18 ± 0.004	-0.002 ± 0.007
Average number of over-beds	0.08 ± 0.004	0.08 ± 0.004	-0.002 ± 0.008

Table 4
Blocking probability regional emergency patients in the region with cooperation

Number of regional beds	ERM blocking probability	Proportion refused regionals in simulation
0	0.255	0.232 ± 0.006
1	0.215	0.195 ± 0.003
2	0.177	0.162 ± 0.006
3	0.142	0.134 ± 0.005
4	0.112	0.107 ± 0.004
5	0.085	0.083 ± 0.004
6	0.063	0.065 ± 0.002
7	0.045	0.049 ± 0.001
8	0.030	0.036 ± 0.001
9	0.020	0.026 ± 0.002
10	0.013	0.018 ± 0.001
11	0.008	0.011 ± 0.001

the probability that an emergency patient arriving at ICU i eventually has to be sent outside the region. In case without cooperation, we assumed that a hospital reserves several emergency beds, and we used ERM involving only one unit in order to compute the blocking probability. The error in the analytical approximation in these two cases turns out to be of a similar order as in Table 4. In the case study below we use the analytical approximation.

5.2. Case study for Rijnmond region

The objective of this case study is to investigate the advantage of cooperation between the hospitals. For that, we used the ERM to compute the blocking probabilities in each hospital separately assuming that they handle the emergency patients on their own, without the regional beds capacity. We will illustrate the advantage of cooperation within the region by means of the following example.

The goal of the management of the ICUs in the region is that at most 1% of the regional patients are rejected and transferred to an ICU outside the region. Table 4 indicates that 11 regional beds are required to achieve this goal. Table 5 provides the fraction of rejected regionals per hospital, where the approximate blocking probabilities computed using formula (21). The row with 11 beds indicates that this results in a blocking probability of approximately 0.6% for regional patients arriving at the Erasmus MC, approximately 2% for the Albert Schweizer Hospital, 2% for the Sint Franciscus Gasthuis, and approximately no rejected regionals for the Dirksland Hospital. Notice that these numbers at the Albert Schweizer and Sint Franciscus hospitals exceed those of the Erasmus MC. As the total number of regionals arriving at the ICU of the Erasmus MC is considerably larger than that number at the other hospitals, the total rejection probability is 0.8%.

Table 5
Blocking probability regional emergency patients for each hospital with cooperation

Nr of regional beds	Erasmus MC	Albert Schweizer	Dirksland	Sint Franciscus
0	0.207	0.689	0.004	0.715
1	0.174	0.478	0.003	0.602
2	0.144	0.496	0.003	0.478
3	0.116	0.385	0.002	0.399
4	0.091	0.302	0.002	0.313
5	0.069	0.230	0.001	0.239
6	0.051	0.169	0.001	0.176
7	0.036	0.120	0.001	0.125
8	0.025	0.082	0.000	0.085
9	0.016	0.054	–	0.056
10	0.010	0.034	–	0.035
11	0.006	0.020	–	0.021

Table 6
Blocking probability of regional emergency patients for each hospital, without cooperation

Nr reserved beds	Erasmus MC	Albert Schweizer	Dirksland	Sint Franciscus
0	0.207	0.732	0.016	0.742
1	0.168	0.230	0.001	0.357
2	0.133	0.049	0.000	0.135
3	0.102	0.007	–	0.039
4	0.077	0.001	–	0.009
5	0.056	0.001	–	0.002
6	0.039	0.000	–	0.000
7	0.026	–	–	–
8	0.017	–	–	–
9	0.011	–	–	–
10	0.006	–	–	–
11	0.004	–	–	–

Furthermore, note that the Albert Schweizer and Sint Franciscus hospitals seem to benefit more than the Erasmus MC from the introduction of regional beds. This is due to the fact that the initial rejection rate at these hospitals is much higher than at the Erasmus MC.

Now consider the hospitals without regional cooperation. Table 6 presents the fraction of rejected regionals for each hospital. To achieve at most 1% of rejected regionals without cooperation, the Erasmus MC needs 10 emergency beds, the Albert Schweizer hospital 3 beds, the Sint Franciscus Gasthuis 4 beds, and the Dirksland Hospital one bed, resulting in 18 beds in total. Only a slightly higher fraction of rejected regionals will be guaranteed with 9 beds at the Erasmus MC and 0 beds at the Dirksland Hospital. Further decreasing of the number of reserved beds results in a much higher rejection rate. Thus, at least 16 reserved beds are required in the region without cooperation. Accordingly, cooperation between the hospitals can save at least 5 beds (31%). We note that simulation results (that we do not present here) come down to the same numbers.

In the case study reported above, the reservation of regional beds does not influence rejection of elective patients or the use of over-beds. In practice, an elective operation sometimes will have to be cancelled although there is an empty regional bed available. However, since the LOS of elective patients is more predictable and their arrivals can be controlled, exact knowledge about how many emergency patients can be present in the ICU may help to decrease the number of cancellations by better planning of elective arrivals and thus result in a smaller number of cancellations. Cooperation helps to decrease the total capacity required for regional emergency patients, which eventually is advantageous for elective patients, too.

6. Conclusions and further research

Hospitals in the Netherlands are responsible for their own budget. In contrast, efficient care for patients within a region covered by multiple hospitals requires coordination among hospitals. A strong basis for coordination is provided by proper insight into the benefits and drawbacks of cooperation. To this end, this paper has investigated the effect of regional Intensive Care capacity on the quality of patient care, in particular focusing on the fraction of regional emergency patients not admitted to an ICU in the region, and the fraction of cancelled operations. Reserving IC beds for regional emergency patients seems to increase the number of cancelled operations. As is demonstrated in a case study for the Rijnmond Region in the Netherlands, cooperation may both lead to a reduction of the fraction of rejected regionals, and a reduction in the fraction of cancelled operations. Both reductions are due to the more efficient use of IC capacity.

Establishing an IC bed is extremely costly. Therefore, making the trade-off between regional and local IC capacity requires an adequate tool to quantify the number of required IC beds for each hospital in various scenarios taking into account aspects including the expected number of patients, the division of beds over hospitals, but also the fraction of cancelled operations and rejected regional patients allowed by the management, by health insurers, or by the government. Based on mathematical methods developed for circuit switched telephone systems, this paper has developed an extension of the Equivalent Random Method that allows us to quantify both the local (for each hospital) and regional fractions of rejected patients. The advantage of the

Equivalent Random Method over simulation is that the ERM provides insight into the nature of the regional overflow problem, and that the ERM allows for a fast evaluation of all different combinations of the number of beds at each hospital and the number of regional beds. This allows for optimisation of the distribution of beds over hospitals. The model may also be applicable to other departments such as Radiology, or the wards.

There is room for improvements. Our results for the blocking probabilities seem to be too high. In part, this is due to the data provided by the hospitals. In particular the length of stay is on average one day too long for the Erasmus MC, since both the day of arrival and the day of departure are included. Furthermore, we have assumed that patients at peripheral hospitals have the same LOS. As the Erasmus MC is an academic hospital that also serves as a regional trauma centre, the LOS for other hospitals seems to be overestimated. A detailed data analysis, including data for hospitals in the region is beyond the scope of the current paper, and is among our aims for further research. The aim of the current paper is to show that the developed mathematical model provides an adequate predictions for required capacity in the given setting.

A second improvement may be to include non-Poissonian arrivals of elective patients. Although this seems to be an important improvement, in practice the assumption of Poisson arrivals may be reasonable, since only 5% of patients from the operating theatre require an IC bed. Therefore, the arrival process of elective patients to the ICU is more variable than the scheduled arrival of patients to the operating theatre. From a mathematical perspective, however, the generalisation is very interesting.

References

[1] www.novatv.nl (in Dutch).
[2] *Schakels in een keten*, Strategische visie 2004–2008 van Cluster 17 (in Dutch).
[3] Rijksinstituut voor Volksgezondheid en Milieu, *Nationale Atlas Volksgezondheid*. Available from: <http://www.rivm.nl/vtv/generate/objv2atl_atlas_0.htm> (in Dutch).
[4] Technomatrix, Tutorial version 5.5, eM-Plant, 2002.
[5] J. Bakker, Criteria voor opname en ontslag van Intensive Care afdelingen in Nederland, Nederlandse Vereniging voor Intensive Care, 2001 (in Dutch).
[6] R.B. Cooper, Introduction to Queueing Theory, Elsevier North Holland, New York, 1981.
[7] J.L.A. Hautvast, J. Bakker, N. Boekema-Bakker, J.A.J. Faber, D.E. Grobbee, A.J.P. Schrijvers, Plaats in de herberg, een studie naar determinanten van opname- en ontslagproblemen in IC-afdelingen in Nederland, Julius Centrum voor Huisartsgeneeskunde en Patiëntgebonden Onderzoek, Utrecht, 2001 (in Dutch).
[8] A.M. Law, W.D. Kelton, Simulation Modeling and Analysis, McGraw-Hill Higher Education, 2000.
[9] R. Lesscher, E. Meutstege, J. Rouhof, Data Analyse van de Intensive Care, Universiteit Twente, 2004 (in Dutch).
[10] Y. Rapp, Planning of junction networks in a multiexchange area, Ericsson Technics 20 (1) (1964) 77–130.
[11] R.I. Wilkinson, Theories for toll traffic engineering in the USA, The Bell System Technical Journal 35 (1956) 421–514.