

---

## The Emergency Observation and Assessment Ward

Maartje E. Zonderland · Richard J. Boucherie ·  
Michael W. Carter · David A. Stanford

**Abstract** A recent development to reduce ED crowding and increase urgent patient admissions is the opening of an Emergency Observation and Assessment Ward (EOA Ward). At these wards urgent patients are temporarily hospitalized until they can be transferred to an inpatient bed. In this paper we present an overflow model to evaluate the effect of employing an EOA Ward on elective and urgent patient admissions.

### 1 Introduction

Over the last period, the Emergency Department (ED) has become more and more crowded, resulting in among others an increased length of stay and prolonged waiting times for patients. Also, ED crowding may result in increased mortality rates and lower quality of care [10]. These problems are not only caused by an aging population [17], a higher demand for

---

Maartje E. Zonderland  
Division I, Leiden University Medical Center  
Postbox 9600, 2300 RC Leiden, The Netherlands  
Tel: +31-71-5263108, Fax: +31-71-5266745  
E-mail: [m.e.zonderland@lumc.nl](mailto:m.e.zonderland@lumc.nl)

Maartje E. Zonderland · Richard J. Boucherie  
Stochastic Operations Research & Center for Healthcare Operations Improvement and Research, University of Twente  
Postbox 217, 7500 AE Enschede, The Netherlands  
E-mail: ([m.e.zonderland](mailto:m.e.zonderland@utwente.nl) · [r.j.boucherie](mailto:r.j.boucherie@utwente.nl))@utwente.nl

Michael W. Carter  
Centre for Research in Healthcare Engineering, University of Toronto  
5 King's College Road, Toronto ON, M5S 3G8, Canada  
E-mail: [carter@mie.utoronto.ca](mailto:carter@mie.utoronto.ca)

David A. Stanford  
Department of Statistical and Actuarial Sciences, University of Western Ontario  
1151 Richmond Street North, London ON, N6A 5B7, Canada  
E-mail: [stanford@stats.uwo.ca](mailto:stanford@stats.uwo.ca)

acute care [14], and the inability to transfer patients to inpatient beds [8, 14], but also by hospital restructuring leading to fewer inpatient beds and more ambulatory care [19].

There are several measures hospitals can take in order to improve ED patient flow [15]. A recent development is the creation of an Emergency Observation and Assessment Ward (EOA Ward). The definition and purpose of such wards varies across hospitals. Also, in literature a consistent definition seems to be lacking. The review papers [6] and [20] provide a comprehensive overview of definitions and concepts for EOA Wards. Patient types that can be admitted vary, for example sometimes only medical patients are considered [20]. Patients that need intensive care are usually excluded [6, 20]. At an EOA Ward, patients are temporarily (less than 24 hours, 24-48 hours) hospitalized until a bed at an inpatient ward becomes available. ED patients who have to wait for test results or require observation for a short period of time can also be admitted. Given the close monitoring, a staffed bed at this location is usually more expensive than a bed at a regular inpatient ward.

The success of an EOA Ward depends on the overlying organizational structures, together with clear agreements upon transfers to regular inpatient wards, a well-defined chain of command, and access to specialist consultations [6, 20]. Currently, little evidence is available that operating an EOA Ward reduces ED crowding (see [6, 10, 15, 20] and the references therein). This is not only related to the ambiguity in the terminology and definitions of the EOA Ward used in practice, but could also be caused by a lack of management information. This makes it very hard to measure the effects of opening an EOA Ward on the ED patient flow. Furthermore, there may be a publication bias since it is common to report only positive experiences [20]. This paper aims at filling this gap in literature by providing a clear model taking into account all relevant patient flows, which allows for a quantitative analysis of the benefits of an EOA Ward.

The patient flow between the ED and inpatient wards is in most hospitals organized as follows. Patients arrive at the ED, receive treatment and are then either discharged, admitted, or might die. Patients can be admitted at another hospital as where the ED is at. While some transfers to other hospitals are necessary, for example because the other hospital is specialized in the type of care the patient needs, other transfers are inevitable since there are no inpatient beds available at the current location. The process of finding a bed and subsequently waiting for transport can easily take several hours. During this time the patient usually occupies an ED room, resulting in fewer ED capacity and substantial delay for patients in the waiting room. Additionally, ED treatment is expensive compared to inpatient care. It is therefore also financially attractive for hospitals to continue the care process at one of their own inpatient wards, instead of transferring the patient to another hospital after ED treatment.

Since the inpatient wards admit elective patients as well, it might be difficult to set aside inpatient beds for urgent patients whose arrival is uncertain. At an EOA Ward this is avoided since only urgent and observation patients from the ED are admitted. The maximum LOS at the EOA Ward is usually short, with regular transfers to the inpatient wards. Transfer moments can be fixed (for example twice a day) or patients are transferred immediately when a bed becomes available. In addition to the elective and urgent patient flows, the inpatient wards also receive patients from the ICU.

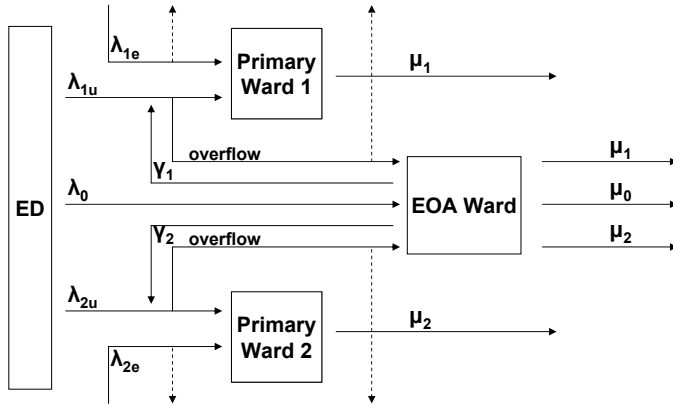
In this paper we present a queuing model (Section 2) that can be used to evaluate the effect of employing an EOA Ward. We analyze a small example in Section 3 and calculate performance measures such as the number of elective and urgent inpatient admissions. Many work in improving ED patient flow has been done using simulation techniques (see e.g., [1, 4, 7, 11]). Fewer examples use queuing theory [5, 9, 13, 16]. Even though the EOA Ward has

been subject of research quite often in the last decade, we were not able to find an analytical evaluation of its effect in terms of inpatient admissions as we present here.

## 2 Model

In this section we describe our mathematical model, which is based on Wilkinson's Equivalent Random Method (ERM) [21], a methodology developed to analyze overflow systems. The inpatient wards and EOA Ward can also be modeled as an overflow system, where the inpatient wards are the primary wards (i.e., the wards that generate the overflow of urgent patients) and the EOA Ward is the overflow ward where urgent patients are routed if the inpatient ward is full. We have  $I$  primary wards, with capacity  $c_i$ ,  $i = 1, \dots, I$ . We assume that the length of stay (LOS) at ward  $i$  is exponentially distributed with mean  $\mu_i^{-1}$ , where the LOS for elective and urgent patients at ward  $i$  is the same, but the LOS per ward may be different, so that  $\mu_i \neq \mu_j$  for  $i, j \in I$ ,  $i \neq j$ . Urgent patients arrive at primary ward  $i$  with rate  $\lambda_{iu}$ . If all beds at the primary ward are occupied, the urgent patient is routed to the EOA Ward. If the EOA Ward, which has a capacity of  $c_0$  staffed beds, is fully occupied as well, the patient is blocked (again) and leaves.

Fig. 1 ED – primary ward – EOA Ward patient flow; example with two primary wards



All urgent patients at the EOA Ward have the same exponential distributed LOS with rate  $\mu_{over}$ . Urgent patients who originated from primary ward  $i$  are transferred from the EOA Ward to primary ward  $i$  with rate  $\gamma_i$ . The LOS at primary ward  $i$  is for these patients again exponential with mean  $\mu_i^{-1}$ . Patients directly routed from the ED to the EOA Ward arrive with Poisson rate  $\lambda_0$  and have an exponentially distributed LOS with mean  $\mu_0^{-1}$ . Elective patients are blocked when the primary ward is full. The elective patient demand at the primary wards, which also incorporates patients from the ICU, is modeled with a Poisson process with rate  $\lambda_{ie}$ . Although elective arrivals are scheduled, random fluctuations in the number of scheduled arrivals make the Poisson assumption plausible [3]. Figure 1 summarizes our overflow system.

## 2.1 Global Balance Equation

We denote the number of elective and urgent patients present at primary ward  $i$  with  $n_{ie}$  and  $n_{iu}$  resp. The number of urgent patients from primary ward  $i$  present at the EOA Ward is given by  $n_{0i}$ , and the number of patients present directly routed to the EOA Ward is denoted by  $n_{00}$ . The state space for the overflow system in Figure 1 is given by:

$$\mathbf{S}: \{ \mathbf{n} = (n_{00}, n_{01}, \dots, n_{0I}, n_{1e}, \dots, n_{Ie}, n_{1u}, \dots, n_{Iu}); \quad n_{ie} + n_{iu} \leq c_i \quad \forall i; \\ \sum_{i=0}^I n_{0i} \leq c_0; \quad n_{ie}, n_{iu}, n_{0i}, n_{00} \geq 0 \quad \forall i \}. \quad (1)$$

Denote  $\pi(\mathbf{n})$  as the equilibrium probability that  $\mathbf{n}$  patients are present in the system. We obtain the following global balance equation:

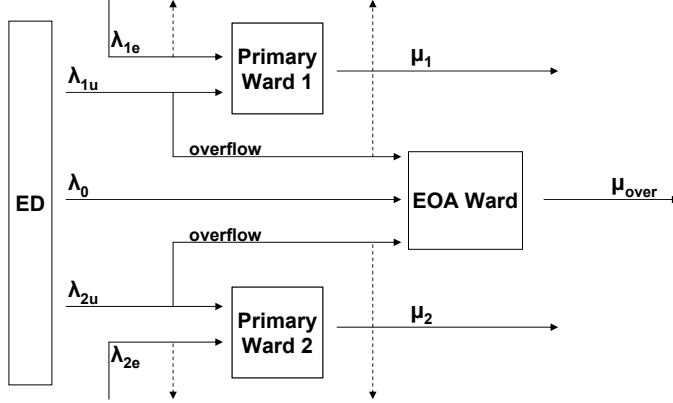
$$\begin{aligned} \pi(\mathbf{n}) & \left[ \sum_{i=1}^I \lambda_{i,e} \mathbb{1}_{n_{ie}+n_{iu} < c_i} + \sum_{i=1}^I \lambda_{i,u} \left( \mathbb{1}_{n_{ie}+n_{iu} < c_i} + \mathbb{1}_{n_{ie}+n_{iu}=c_i, \sum_{i=0}^I n_{0i} < c_0} \right) \right. \\ & + \lambda_0 \mathbb{1}_{\sum_{i=0}^I n_{0i} < c_0} + \sum_{i=1}^I (n_{ie} + n_{iu}) \mu_i + \sum_{i=1}^I n_{0i} \mu_{over} \\ & \left. + n_{00} \mu_0 + \sum_{i=1}^I n_{0i} \gamma_i \mathbb{1}_{n_{ie}+n_{iu} < c_i} \right] \\ & = \sum_{i=1}^I \lambda_{i,e} \pi(\mathbf{n} - e_{ie}) \mathbb{1}_{n_{ie} > 0} \\ & + \sum_{i=1}^I \lambda_{i,u} \left( \pi(\mathbf{n} - e_{iu}) \mathbb{1}_{n_{iu} > 0} + \pi(\mathbf{n} - e_{0i}) \mathbb{1}_{n_{0i} > 0, n_{ie}+n_{iu}=c_i} \right) \\ & + \lambda_0 \pi(\mathbf{n} - e_{00}) \mathbb{1}_{n_{00} > 0} + \sum_{i=1}^I (n_{ie} + 1) \mu_i \pi(\mathbf{n} + e_{ie}) \mathbb{1}_{n_{ie}+1+n_{iu} \leq c_i} \\ & + \sum_{i=1}^I (n_{iu} + 1) \mu_i \pi(\mathbf{n} + e_{iu}) \mathbb{1}_{n_{ie}+n_{iu}+1 \leq c_i} \\ & + \sum_{i=1}^I (n_{0i} + 1) \mu_{over} \pi(\mathbf{n} + e_{0i}) \mathbb{1}_{n_{0i}+1 \leq c_0} + (n_{00} + 1) \mu_0 \pi(\mathbf{n} + e_{00}) \mathbb{1}_{n_{00}+1 \leq c_0} \\ & + \sum_{i=1}^I (n_{0i} + 1) \gamma_i \pi(\mathbf{n} + e_{0i} - e_{iu}) \mathbb{1}_{n_{ie}+n_{iu}+1 \leq c_i}. \quad (2) \end{aligned}$$

This equation can be solved explicitly only for specific values of the system parameters [2]. We therefore apply the method as discussed in [12], which requires that  $\mu_i = \mu_{over}$ . We adapt it such that  $\mu_i \neq \mu_{over}$ . Our model shows strong similarities with the model of Schehrer [18], but we add an extra flow to represent the elective patients arriving at the primary ward. We first analyze the model without transfers from the EOA Ward to the primary wards. Subsequently we introduce transfers and apply the approach presented in [2] to determine the number of patients present at each ward.

## 2.2 No Transfers from the EOA Ward to the Primary Wards

We first analyze the situation where patients at the EOA Ward are not transferred to the primary wards (see Figure 2).

**Fig. 2** No transfers to the primary wards; example with two primary wards



We do so by adapting the global balance equation (2) by setting  $\gamma_i = 0 \forall i$ . Now we can solve the global balance equations explicitly. In line with [12] we use a probability generating function approach to determine the mean,  $\mathbb{E}_i$ , and variance,  $\mathbb{V}_i$ , of the overflow of urgent patients from primary ward  $i$  at the EOA Ward in case of infinite EOA Ward capacity. Since  $c_0 = \infty$ , and since the overflow processes from the primary wards are independent,  $\mathbb{E}_i$  and  $\mathbb{V}_i$  can be determined for each primary ward  $i$  in isolation. We only want to calculate the blocking probability at the overflow, and thus it is not required to know whether a patient residing at primary ward  $i$  is of the urgent or elective type. Let  $n_i = n_{ie} + n_{iu}$  denote the number of patients at primary ward  $i$ , and let  $\lambda_i = \lambda_{ie} + \lambda_{iu}$  denote the total arrival rate at primary ward  $i$ . The global balance equation simplifies to:

$$\begin{aligned}
 & \pi(n_{0i}, n_i) (\lambda_i + n_i \mu_i + n_{0i} \mu_{over}) \\
 &= \lambda_i \pi(n_{0i}, n_i - 1) + (n_{0i} + 1) \mu_{over} \pi(n_{0i} + 1, n_i) + (n_i + 1) \mu_i \pi(n_{0i}, n_i + 1) \\
 & \text{for } n_i < c_i, \\
 & \pi(n_{0i}, n_i) (\lambda_{iu} + n_i \mu_i + n_{0i} \mu_{over}) \\
 &= \lambda_i \pi(n_{0i}, n_i - 1) + (n_{0i} + 1) \mu_{over} \pi(n_{0i} + 1, n_i) + \lambda_{iu} \pi(n_{0i} - 1, n_i) \\
 & \text{for } n_i = c_i.
 \end{aligned} \tag{3}$$

We define the probability generating function of the number of urgent patients from ward  $i$  present at the EOA Ward,  $G_{i,n_i}(z)$ , as:

$$G_{i,n_i}(z) = \sum_{n_{0i}=0}^{\infty} \pi(n_{0i}, n_i) z^{n_{0i}}. \tag{4}$$

Multiplication of (3) with  $z^{n_0i}$  and the summation of the result over  $n_{0i} = 0, \dots, \infty$  yields

$$\begin{aligned} & [\lambda_i + n_i \mu_i] G_{i,n_i}(z) + \mu_{over}(z-1) \frac{d}{dz} G_{i,n_i}(z) \\ & = \lambda_i G_{i,n_i-1}(z) + (n_i + 1) \mu_i G_{i,n_i+1}(z) \quad \text{for } 0 \leq n_i < c_i, \\ & [\lambda_{iu}(1-z) + n_i \mu_i] G_{i,n_i}(z) + \mu_{over}(z-1) \frac{d}{dz} G_{i,n_i}(z) \\ & = \lambda_i G_{i,n_i-1}(z) \quad \text{for } n_i = c_i. \end{aligned} \quad (5)$$

Now  $\mathbb{E}_i$  and  $\mathbb{V}_i$  can be derived from:

$$\begin{aligned} \mathbb{E}_i &= \sum_{n_i=0}^{c_i} \frac{d}{dz} G_{i,n_i}(z)|_{z=1} \\ \mathbb{V}_i &= \sum_{n_i=0}^{c_i} \frac{d^2}{dz^2} G_{i,n_i}(z)|_{z=1} + \mathbb{E}_i - (\mathbb{E}_i)^2. \end{aligned} \quad (6)$$

Taking the first derivative of (5) and evaluating at  $z = 1$ , gives:

$$\begin{aligned} & (\lambda_i + n_i \mu_i + \mu_{over}) g_i[n_i] \\ & = \lambda_i g_i[n_i - 1] + (n_i + 1) \mu_i g_i[n_i + 1] \quad \text{for } 0 \leq n_i < c_i, \\ & (\mu_{over} + n_i \mu_i) g_i[n_i] - \lambda_{iu} \mathbb{P}_i(c_i) \\ & = \lambda_i g_i[n_i - 1] \quad \text{for } n_i = c_i, \end{aligned} \quad (7)$$

where  $g_i[n_i] = \frac{d}{dz} G_{i,n_i}(z)|_{z=1}$  and  $\mathbb{P}_i(c_i) = \text{Erl}\left(\frac{\lambda_i}{\mu_i}, c_i\right)$ . Then  $\mathbb{E}_i$  is obtained by the summation of (7) for  $n_i = 0, \dots, c_i$ :

$$\mathbb{E}_i = \frac{\lambda_{iu}}{\mu_{over}} \mathbb{P}_i(c_i). \quad (8)$$

The variance can be calculated accordingly by taking the second derivative of (5) and evaluating at  $z = 1$ :

$$\begin{aligned} & (\lambda_i + n_i \mu_i + 2\mu_{over}) h_i[n_i] \\ & = \lambda_i h_i[n_i - 1] + (n_i + 1) \mu_i h_i[n_i + 1] \quad \text{for } 0 \leq n_i < c_i, \\ & (n_i \mu_i + 2\mu_{over}) h_i[n_i] - 2\lambda_{iu} g_i[n_i] \\ & = \lambda_i h_i[n_i - 1] \quad \text{for } n_i = c_i, \end{aligned} \quad (9)$$

where  $h_i[n_i] = \frac{d^2}{dz^2} G_{i,n_i}(z)|_{z=1}$ . The summation of the result for  $n_i = 0, \dots, c_i$  yields:

$$\mathbb{V}_i = \frac{\lambda_{iu}}{\mu_{over}} g_i[c_i] + \mathbb{E}_i - (\mathbb{E}_i)^2, \quad (10)$$

where  $g_i[c_i]$  can be determined recursively from (7) with  $g_i[-1] = 0$ . The direct patient flow arriving at the EOA Ward can be represented by an  $M/M/\infty$  queue. The mean  $\mathbb{E}_0$  and variance  $\mathbb{V}_0$  of this stream is therefore given by:

$$\begin{aligned} \mathbb{E}_0 &= \frac{\lambda_0}{\mu_0} \\ \mathbb{V}_0 &= \frac{\lambda_0}{\mu_0}. \end{aligned} \quad (11)$$

We are now able to define an equivalent primary ward with service rate  $\mu_{over}$  which generates the same traffic as the  $i$  overflow (urgent) and direct streams together. Since only urgent patients are routed to the EOA Ward, the elective patients who do not cause overflow do not need to be incorporated in the equivalent primary ward. The equivalent primary ward has load  $a$  and capacity  $C$  such that [21]:

$$\begin{aligned} aErl(a, C) &= \mathbb{E} \\ \mathbb{E} \left( 1 - \mathbb{E} + \frac{a}{C + 1 + \mathbb{E} - a} \right) &= \mathbb{V}, \end{aligned} \quad (12)$$

where, since the overflow processes from the primary wards and the direct arrival process are independent, the expectation and variation of the aggregated overflow,  $\mathbb{E}$  and  $\mathbb{V}$ , are given by:

$$\begin{aligned} \mathbb{E} &= \sum_{i=0}^I \mathbb{E}_i \\ \mathbb{V} &= \sum_{i=0}^I \mathbb{V}_i. \end{aligned} \quad (13)$$

The blocking probability for patients from ward  $i$  and the direct arriving patients ( $i = 0$ ) is given by the Katz approximation [22], which takes the peakedness,  $\zeta_i = \frac{\mathbb{V}_i}{\mathbb{E}_i}$ , of the separate flows into account:

$$K_i = \frac{aErl(a, C + c_0)}{aErl(a, C)} \left( v(C, c_0)^{-1} + \frac{\zeta_i - 1}{\zeta_i - 1} (1 - v(C, c_0)^{-1}) \right), \quad (14)$$

with  $\zeta = \frac{\mathbb{V}}{\mathbb{E}}$  and  $v(C, c_0)$  can be determined recursively from

$$\begin{aligned} v(C, j) &= \frac{aj}{aErl(a, C) (C + j - a - aErl(a, C)v(C, j - 1))}, \quad j = 1, 2, \dots \\ v(C, 0) &= 1. \end{aligned} \quad (15)$$

The Katz approximation is exact in this case since the direct arrival stream is Poisson [22].

The mean number of urgent patients from primary ward  $i$  present at the EOA Ward,  $\mathbb{E}[N_{0i}]$ , is given by:

$$\mathbb{E}[N_{0i}] = \frac{\lambda_{iu}}{\mu_{over}} Erl \left( \frac{\lambda_i}{\mu_i}, c_i \right) (1 - K_i). \quad (16)$$

The mean number of patients who directly arrived at the EOA Ward,  $\mathbb{E}[N_{00}]$ , equals:

$$\mathbb{E}[N_{00}] = \frac{\lambda_0}{\mu_0} (1 - K_0). \quad (17)$$

### 2.3 Transfers from the EOA Ward to the Primary Wards

Finally, we allow patient transfers from the EOA Ward back to the primary wards. It was assumed that  $\mu_{over}$  was the same for all patients, and therefore we define  $\gamma_i$  such that  $\mu_{over} = \mu_i + \gamma_i \forall i$ . In this scenario we obviously have that  $\mu_{over} \geq \mu_i$ , and this is exactly the model as depicted in Figure 1. We approximate the arrival rate at primary ward  $i$ ,  $v_i$ , by the sum

of the arrivals of elective and urgent patients ( $\lambda_{ie}$  and  $\lambda_{iu}$  resp.) and the patients transferred from the EOA Ward,  $\gamma_i \mathbb{E}[N_{0i}]$ , [2]:

$$v_i = \lambda_{ie} + \lambda_{iu} + \gamma_i \mathbb{E}[N_{0i}]. \quad (18)$$

A fraction of this stream,  $\kappa_i$ , is routed to the EOA Ward when all beds at primary ward  $i$  are occupied:

$$\kappa_i = \lambda_{iu} + \gamma_i \mathbb{E}[N_{0i}]. \quad (19)$$

To analyze the model for  $\gamma_i > 0$ , we replace  $\lambda_{iu}$  by  $\kappa_i$  and  $\lambda_i$  by  $v_i$  in (7 – 10). We then obtain a system of equations, which can be solved for  $\mathbb{E}[N_{0i}]$  using fixed point iteration with initial value  $\mathbb{E}[N_{0i}] = 0$  [2].

The mean of the total number of patients present at the EOA Ward,  $\mathbb{E}[N_0]$ , and the mean of the total number of patients present at primary ward  $i$ ,  $\mathbb{E}[N_i]$ , are given by:

$$\begin{aligned} \mathbb{E}[N_0] &= \sum_{i=0}^I \mathbb{E}[N_{0i}] \\ \mathbb{E}[N_i] &= \frac{v_i}{\mu_i} \left( 1 - \text{Erl} \left( \frac{v_i}{\mu_i}, c_i \right) \right). \end{aligned} \quad (20)$$

The mean occupation,  $\rho_i$ , of the EOA Ward and the primary wards, is given by:

$$\rho_i = \frac{E[N_i]}{c_i}. \quad (21)$$

The mean number of patients blocked or admitted at the wards can be calculated accordingly.

### 3 Results

We now use the model from Subsection 2.3 to analyze a simple example for a hospital with two primary wards. Primary ward 1 has a capacity of  $c_1 = 200$  beds and admits only medical patients, whose mean LOS is five days (so  $\mu_1 = \frac{1}{5}$ ). The elective patient arrival rate  $\lambda_{1e}$  equals 26 patients per day, and the urgent patient arrival rate  $\lambda_{1u}$  is 14 patients per day, so that the total patient arrival rate at ward 1,  $\lambda_1 = 40$ . Primary ward 2 admits only surgical patients, with  $c_2 = 200$ , a mean LOS of four days ( $\mu_2 = \frac{1}{4}$ ),  $\lambda_{2e} = 37$ ,  $\lambda_{2u} = 13$ , and  $\lambda_2 = 50$ . Adding capacity by creating so-called overbeds is not allowed.

Patients arrive directly at the EOA Ward with rate  $\lambda_0 = 2$  and have a service rate of  $\mu_0 = 4$ . With this flow we represent patients who only require observation for a short period of time (on average six hours in this case). The mean length of stay for urgent patients at the EOA Ward is set on 36 hours, so that  $\mu_{over} = \frac{2}{3}$ . Consequently,  $\gamma_1 = \mu_{over} - \mu_1 = \frac{7}{15}$  and  $\gamma_2 = \mu_{over} - \mu_2 = \frac{5}{12}$ . This implies that patients with a longer LOS should be transferred back sooner to the primary ward than patients with a shorter LOS, in order to keep the LOS at the EOA Ward the same for all urgent patients. Table 1 summarizes the parameter values.



**Table 1** Parameter values for EOA Ward example

Parameter	Value	Parameter	Value
$c_1$	200	$\mu_0$	4
$c_2$	200	$\mu_{over}$	$\frac{2}{3}$
$\lambda_{1e}$	26	$\mu_1$	$\frac{1}{5}$
$\lambda_{1u}$	14	$\mu_2$	$\frac{1}{4}$
$\lambda_{2e}$	37	$\gamma_1$	$\frac{7}{15}$
$\lambda_{2u}$	13	$\gamma_2$	$\frac{5}{12}$

### 3.1 Opening the EOA Ward

Suppose the hospital considers opening an EOA Ward. We first analyze the situation where only urgent patients are admitted at this ward (so for now, we set  $\lambda_0 = 0$ ). In Table 2 for  $c_0 = 0$  (no EOA Ward, i.e., the old situation), and  $c_0 = 4, 6, 8, 12$  the blocking probabilities for elective,  $\mathbb{P}(B_{ie})$ , and urgent patients,  $\mathbb{P}(B_{iu})$ , are given. The number of rejected elective,  $B_{ie}$ , and urgent patients,  $B_{iu}$ , is given per ward per day, and the number of admitted elective,  $EP/y$ , and admitted urgent,  $UP/y$ , patients per year are given.

**Table 2** Results for opening an EOA Ward

$c_0$	$\mathbb{P}(B_{ie})$	$B_{1e}$	$\mathbb{P}(B_{1u})$	$B_{1u}$	$\mathbb{P}(B_{2e})$	$B_{2e}$	$\mathbb{P}(B_{2u})$	$B_{2u}$	$EP/y$	$UP/y$
0	5.4%	1.4144	5.4%	0.7616	5.4%	2.0128	5.4%	0.7072	21,753	9,323
4	6.0%	1.5547	2.4%	0.3313	5.8%	2.1469	2.1%	0.2733	21,642	9,633
6	6.2%	1.5998	1.5%	0.2024	5.9%	2.1885	1.1%	0.1468	21,610	9,726
8	6.3%	1.6352	0.1%	0.1035	6.0%	2.2110	0.1%	0.0796	21,587	9,845
12	6.4%	1.6652	0.2%	0.0215	6.0%	2.2322	0.1%	0.0169	21,577	9,840

What we see is that the blocking probability for urgent patients decreases, which was expected since we added capacity for these patients. However, since the hospital is now able to admit more urgent patients, ultimately there is less capacity available at the primary wards for the elective patients which results in repression of elective patients. An EOA Ward with four beds results in a total of 310 more (9,633 vs. 9,323) urgent patients admitted per year, but at the same time 111 less elective patients are admitted per year (21,642 vs. 21,753). It follows that the opening of the EOA Ward also affects the elective patient flow.

### 3.2 Admitting Observation Patients

Following the opening of the EOA Ward, the hospital decides that patients from the ED requiring observation should also be admitted at the EOA Ward (thus we set  $\lambda_0 = 2$ ). It is obvious that more beds are required to maintain the decreased blocking probabilities for urgent patients (Table 3), but the blocking probabilities for elective patients remain about the same.

### 3.3 Increasing Urgent Admissions

As mentioned in the Introduction, one of the reasons to open an EOA Ward is to increase the number of urgent patient admissions through the ED. Table 4 shows for various rates of

**Table 3** Results for admitting observation patients

$c_0$	$\mathbb{P}(B_{ie})$	$B_{1e}$	$\mathbb{P}(B_{1u})$	$B_{1u}$	$\mathbb{P}(B_{2e})$	$B_{2e}$	$\mathbb{P}(B_{2u})$	$B_{2u}$	$\mathbb{P}(B_0)$	$B_0$
4	5.9%	1.5409	2.7%	0.3716	5.8%	2.1377	2.3%	0.3019	2.2%	0.4452
12	6.4%	1.6647	0.2%	0.0228	6.0%	2.2319	0.1%	0.0176	1.1%	0.0218

increase,  $f_u$ , in the arrival rate of urgent patients,  $\lambda_{iu}$ , the required size of the EOA Ward for which  $\mathbb{P}(B_{iu}) \approx 1\%$ . Note that  $\lambda_0 = 0$ .

**Table 4** Results for increasing urgent admissions

$f_u$	$c_0$	$\mathbb{P}(B_{ie})$	$B_{1e}$	$\mathbb{P}(B_{1u})$	$B_{1u}$	$\mathbb{P}(B_{2e})$	$B_{2e}$	$\mathbb{P}(B_{2u})$	$B_{2u}$	$EP/y$	$UP/y$
5%	8	7.6%	1.9716	1.2%	0.1758	6.9%	2.5465	0.8%	0.1176	21,342	10,262
10%	10	9.1%	2.3664	1.0%	0.1601	7.9%	2.9208	0.7%	0.0954	21,065	10,748
20%	12	12.2%	3.1753	1.3%	0.2112	9.9%	3.6731	1.0%	0.1502	20,500	11,768
50%	22	22.2%	5.7827	1.3%	0.2816	16.6%	6.1434	1.0%	0.1955	18,646	14,612

We see that an increase in the number of urgent patient admissions has a tremendous effect on the number of elective patient admissions. For example in the case of a 10% increase, the number of elective patient admissions decreases from 21,753 to 21,065 per year.

### 3.4 Maintaining the Number of Elective Patient Admissions

If the hospital wishes to maintain the number of elective patient admissions, it has two options: increase the number of beds at the primary wards or stop transferring patients from the EOA Ward back to the primary wards. The latter option would transform the EOA Ward to a long stay ward for urgent patients, and therefore we only analyze the first option. Table 5 gives for each value of  $f_u$  the required number of beds at the primary wards,  $c_1$  and  $c_2$ , and the extra number of beds required at the primary wards,  $c_+$ , compared to the initial situation where  $c_1 = c_2 = 200$ , such that the elective patient blocking probability is maintained at its initial value of  $\approx 5\%$ . Again, note that  $\lambda_0 = 0$ .

**Table 5** Results for maintaining the number of elective patient admissions

$f_u$	$c_0$	$c_1$	$c_2$	$c_+$
0%	6	203	202	5
5%	6	207	205	12
10%	6	210	208	18
20%	6	215	215	30
50%	8	238	229	67

Since the number of inpatient beds increases, more urgent patients can be admitted directly at the primary wards and thus less EOA Ward capacity is required to keep  $\mathbb{P}(B_{iu}) \approx 1\%$ .

## 4 Discussion

Inpatient wards designed to improve the urgent patient flow have gained increased popularity during the last decade. In this paper we have developed a queuing model, based on Wilkinson's Equivalent Random Method [21], that allows for a quantification of the effects of these EOA Wards in terms of elective and urgent patient admissions. In addition to the extensions to the ERM made in [12] and [18], our model enables the analysis of an overflow system for which the service rate at the primary wards (or cells in the ERM terminology) is not equal to the service rate at the overflow ward (or cell). Furthermore we have added an extra arrival stream to the primary wards which is blocked when all beds are occupied. For analytical tractability, we assumed the LOS at the primary wards and EOA Ward is exponentially distributed. In [12] the authors show for a similar system that the outcomes are insensitive to the service time distribution, as is the case for many loss models.

With a simple example of a hospital with two primary wards, we have shown that opening an EOA Ward results in an increase of urgent patient admissions, but at the same time in a decrease in the number of elective patient admissions. The elective patients are repressed by the extra admitted urgent patients who return to the primary ward from the EOA Ward. We assumed that the urgent patient flow remained constant over time. In reality, the added capacity may attract extra urgent patients, which will in turn result in even less capacity for elective patients. To overcome this effect, next to the EOA Ward capacity created, additional inpatient beds should be added. This in turn results in a decrease of the number of EOA Ward beds required, which makes the EOA Ward a small ward that is possibly difficult to staff. In the example we incorporated only two, very large inpatient wards. In case of more wards with smaller capacities than in our example, the blocking probabilities at these wards are more sensitive to an increase in patient arrivals and thus the repression effect will remain and possibly even worsen. Another important factor to consider is the maximum LOS at the EOA Ward. In our example we used a maximum LOS of 36 hours. When this is shortened to 24 hours, the number of urgent admissions will increase and thus the repression effect will get worse. When determining the maximum LOS this phenomenon should be taken into account as well. The model allows for an easy evaluation of the repression effect in case of changes in the EOA Ward LOS.

EOA Wards definitely have advantages, such as that the admission of urgent patients is centrally organized, or that admissions during the night for regular inpatient wards are avoided. Given the results presented in this paper, we can also perceive the EOA Ward as an instrument to control the elective/urgent patient ratio. However, the effect on elective patient admissions should not be neglected and studied before the decision to open an EOA Ward is made. Other possibilities to improve urgent patient flow are likely to have less adverse effects and should also be taken into consideration.

## References

1. Bagust A, Place M, Posnett JW (1999) *Dynamics of bed use in accommodating emergency admissions: stochastic simulation model*. British Medical Journal 319(7203): 155-158
2. Borst S, Boucherie RJ, Boxma OJ (1999) *ERMR: a generalised Equivalent Random Method for overflow systems with repacking*. Proceedings of the 16th International Teletraffic Congress
3. de Bruin AM, Bekker R, van Zanten L, Koole GM (2010) *Dimensioning hospital wards using the Erlang loss model*. Annals of Operations Research 178(1):23-43
4. Coats TJ, Michalis S (2001) *Mathematical modelling of patients flow through an accident and emergency department*. Emergency Medicine Journal 18(3): 190-192

5. Cochran JK, Roche KT (2009) *A multi-class queuing network analysis methodology for improving hospital emergency department performance*. Computers & Operations Research 36(5):1497-1512
6. Cooke MW, Higgins J, Kidd P (2003) *Use of emergency observation and assessment wards: a systematic literature review*. Emergency Medicine Journal 20(2):138-142
7. Duguay C, Chetouane F (2007) *Modeling and improving emergency department systems using discrete event simulation*. Simulation 83(4):311-320
8. Forster AJ, Stiehl I, Wells G, Lee AJ, van Walraven C (2003) *The effect of hospital occupancy on emergency department length of stay and patient disposition*. Academic Emergency Medicine 10(2):127-133
9. Green LV, Soares J, Giglio JF, Green RA (2006) *Using queueing theory to increase the effectiveness of emergency department provider staffing*. Academic Emergency Medicine 13(1):61-68
10. Hoot NR, Aronsky D (2008) *Systematic review of emergency department crowding: causes, effects, and solutions*. Annals of Emergency Medicine 52(2):126-136
11. Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, Aronsky D (2008) *Forecasting emergency department crowding: a discrete event simulation*. Annals of Emergency Medicine 52(2):116-125
12. Litvak N, van Rijnsbergen M, Boucherie RJ, van Houdenhoven M (2008) *Managing the overflow of intensive care patients*. European Journal of Operational Research 185(3):998-1010
13. Mayhew L, Smith D (2008) *Using queueing theory to analyse the government's 4-h completion time target in accident and emergency departments*. Health Care Management Science 11(1):11-21
14. Moskop JC, Sklar DP, Geiderman JM, Schears RM, Bookman KJ (2009) *Emergency department crowding, part 1 - concept, causes, and moral consequences*. Annals of Emergency Medicine 53(5):605-611
15. Moskop JC, Sklar DP, Geiderman JM, Schears RM, Bookman KJ (2009) *Emergency department crowding, part 2 - barriers to reform and strategies to overcome them*. Annals of Emergency Medicine 53(5):612-617
16. Ramakrishnan M, Sier D, Taylor PG (2005) *A two-time-scale model for hospital patient flow*. IMA Journal of Management Mathematics 16(3):197-215
17. Roberts DC, McKay MP, Shaffer A (2008) *Increasing rates of emergency department visits for elderly patients in the United States, 1993 to 2003*. Annals of Emergency Medicine 51(6):769-774
18. Schehrer RG (1997) *A two moment method for overflow systems with different mean holding times*. Proceedings of the 15th International Teletraffic Congress
19. Schull MJ, Szalai J, Schwartz B, Redelmeier DA (2001) *Emergency department overcrowding following systematic hospital restructuring: trends at twenty hospitals of ten years*. Academic Emergency Medicine 8(11):1037-1043
20. Scott I, Vaughan L, Bell D (2009) *Effectiveness of acute medical units in hospitals: a systematic review*. International Journal for Quality in Health Care 21(6):397-407
21. Wilkinson RI (1956) *Theories for toll traffic engineering in the U.S.A.* The Bell System Technical Journal 35:421-514
22. Wolff RW (1989) *Stochastic modeling and the theory of queues*. Prentice Hall, Englewood Cliffs, NJ, USA