

# Combining optimization with simulation to obtain credible models for intensive care units

Fermín Mallor · Cristina Azcárate

Published online: 16 December 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** In this paper we develop a simulation model to study bed occupancy levels in an Intensive Care Unit (ICU). The main contributions of this study are: (1) A proposal for generalized regression models to fully capture the high variability of patients' length of stay; (2) Proof that a simulation model that does not incorporate the management decisions by clinical staff cannot be considered valid; (3) The development of a mathematical model to represent these management decisions, and (4) A proposal for a method combining optimization with simulation to estimate the model parameters.

This provides a valid simulation model that includes the physician management of an ICU. Validation is accomplished by comparing distribution patterns in daily bed occupancy records against simulated bed occupancy data.

The methodology is tested using data provided by the Hospital of Navarre in Spain.

**Keywords** Simulation · Statistical modeling · Validation of simulation models · Optimization with simulation · Intensive care units

## 1 Introduction

The Intensive Care Unit (ICU) is a key area within hospitals caring for critically ill patients. The beds and the specialized staff of an ICU are costly resources, subject to budget constraints. High quality care provision involves the analysis of a number of important ICU sizing and capacity issues and resource management policy decisions (see, for example, Griffiths et al. 2005).

This paper describes the construction of a credible ICU simulation model designed in the purpose of convincing managers of its reliability as a tool for sizing and capacity analysis. A simulation model and its results have *credibility* if they are accepted as *correct* by the decision-maker and other key project personnel (Law 2006). Credibility is therefore one of

---

F. Mallor (✉) · C. Azcárate  
Department of Statistics and Operations Research, Public University of Navarre, Pamplona, Spain  
e-mail: [mallor@unavarra.es](mailto:mallor@unavarra.es)

the attributes we need to build into our simulation model in order to provide a meaningful guide for ICU management decisions. Our specific task is to develop an ICU simulation model for the Hospital of Navarre (Spain). This hospital is the main center of referral for a city population over 200,000, and a complementary center for the whole province of Navarre (population over 500,000) and other bordering regions (for specialties such as Oncology, Neurosurgery, etc.). Managerial decisions include determining the number of beds required to manage an increased workload due to the opening of new operating theaters and increasing population growth. These decisions have an important impact on both the quality and cost of the health service offered (the estimated cost of a bed in the ICU is around 2000 euro per day). Complete confidence in the simulation tool is essential for it to be of use in guiding this decision-making process.

Simulation has been widely used to tackle health-care system management and operation problems, which are characterized by an uncertain working environment and limited human and material resources. Since the introduction of simulation modeling for hospital facilities (Fetter and Thompson 1965), this operations research technique has been used to address a wide range of problems, such as patient flow modeling, bed capacity, waiting list management, health care center design, emergency facilities etc. Reviews and discussion papers dealing with the application of simulation modeling in health care can be found in Brailsford et al. (2009), Eldabi et al. (2007), Fone et al. (2003), Günal and Pidd (2010), Jun et al. (1999) and (Katsaliaki and Mustafee 2010). Many studies use simulation to analyze hospital capacity and bed allocation, but only a few deal specifically with ICUs. In Kim et al. (1999), ICU admission and discharge processes are analyzed through simulation; in Kim et al. (2000), the same authors evaluate several rules for bed allocation in a two-criteria problem; and Litvack et al. (2008) and Ridge et al. (1998) analyze the problem of ICU capacity. In (Kolker 2009) the author uses simulation to model ICU patient flow, to establish a quantitative link between the daily load leveling of elective surgeries and ICU diversion.

A key issue in the building of valid ICU simulation models is the determination of appropriate statistical models for the stochastic elements involved: arrival patterns, patient's personal and medical characteristics and length of stay (LoS) in the ICU. An overview of length of stay and patient flow modeling techniques can be found in Marshall et al. (2005).

Many studies have found a high percentage of *extreme* values in LoS data. That is, values that are far enough from the mean value to be considered outliers (see Vasilakis and Marshall 2005). In these cases, distributions commonly used to represent service times in the health context are not suitable to fit real LoS data. This problem of lack of fit to the original data has been addressed in the literature in different ways, sometimes erroneously, as when patients with long LoS values are considered as outliers and not included in the statistical analysis. Some studies compensate for withdrawn data by reducing the number of beds by the same percentage as the percentage of outliers in the data. This approach is also erroneous, however, because a given percentage of outlier patients will not always use the same percentage of resources. This study also finds heavy-tailed distributions for the LoS data and a lack of fit to the usual distributions. We address this statistical fitting problem by developing non-normal regression models, including variables with the power to explain some of the LoS variability.

The impact of workload on service time in health care has recently been considered by Kc and Terwiesch (2009), who show that hospital resources are sensitive to load levels and that service workers can adapt to system needs by increasing the service rate. Nevertheless, all previous ICU simulation studies assume patient length of stay to be independent of the ICU workload, which is why our first model also incorporated this assumption. However, the validation process demonstrated that this premise does not always hold and that the length

of stay of some patients can be influenced by the ICU bed occupancy level. This finding was confirmed by the ICU medical staff, who reported that some of their decisions depend on the number of occupied beds.

Specifically, doctors may discharge patients earlier when the number of occupied beds reaches the point where the medical care of incoming patients may be compromised. Conversely, when the ICU workload is low, patients may be allowed to continue in intensive care to ensure a complete recovery. There is no written protocol or structured methodology for these decisions. We therefore accommodate them in the model by defining a set of rules dependent on a set of parameters. We then obtain the parameter values, by solving an optimization problem using a combination of optimization with simulation techniques. The set of rules with their estimated optimal parameters are included in the simulation model. The validation process is then repeated and presented to the medical staff. The result is a valid and credible simulation model to guide the medical staff's resource allocation decisions. Our review of the literature revealed no previous studies where the modeling of human decision rules had been validated in a simulation model.

The combination of simulation and optimization techniques is novel in the analysis of ICUs, although there have been some recent proposals for its use in the health care context. Simulation and goal programming are used in Oddoye et al. (2009) to solve a planning problem in the medical assessment unit of a general hospital. An optimization simulation model is proposed in Ahmed and Alkhamis (2009) to obtain the optimal staff distribution for a hospital emergency unit, and, in Azcárate et al. (2008), to solve a hospital sizing problem in a multiple criteria context. In de Angelis et al. (2003), the authors combine a simulation model with nonlinear programming and neuronal networks to determine the optimal configuration of a transfusion center. In Brailsford et al. (2007), a discrete-event simulation model is embedded in an ant colony optimization model for the optimal choice of screening policies for diabetic retinopathy.

The paper is organized as follows. Section 2 describes the ICU, the available data and the statistical analysis of patient arrival and length of stay. Section 3 describes the simulation model, its validation process and the combination of simulation and optimization to model ICU decision rules. Finally, Sect. 4 contains the conclusions and a discussion about the main point of the paper that is the modeling of the management decisions made by the physicians of the ICU. We also include our work in progress and some suggestions for future research.

## 2 Input modeling

### 2.1 Recorded data at the ICU

The Hospital of Navarre is a general public hospital in the Autonomous Community of Navarre, which is a reference in a number of specialties (Neurosurgery, cardiac surgery, vascular surgery, thoracic surgery, oncology, radiotherapy, nuclear medicine, infectious diseases, endocrinology, dermatology, rheumatology and bone marrow transplant). It has 483 beds, 2015 members of staff and 10 surgery rooms. The ICU of this hospital has 20 beds and 86 physicians and nurses. It receives patients from three sources (emergency, operating theater and ward). The ICU classifies the patients into eight pathology types:

- Group 1 (G1): patients from elective surgery, i.e., scheduled surgery. These patients would never be admitted over the weekend.
- Group 2 (G2): patients referred from emergency units.
- Group 3 (G3): patients admitted due to deterioration post surgery.

**Table 1** Initial summary of the data: mean (standard deviation)

Pathology type	Patients percentage	Daily arrival rate	APACHE II	% died (exitus = 1)	Length of stay
G1	39%	0.744 (0.75)	11.157 (5.641)	4.32%	4.379 (8.739)
G2	14%	0.27 (0.528)	20.758 (9.885)	31.19%	8.032 (15.276)
G3	7%	0.135 (0.372)	19.116 (8.274)	27.27%	9.927 (14.174)
G4	12%	0.226 (0.466)	21.550 (9.257)	36.81%	9.820 (18.694)
G5	11%	0.205 (0.449)	17.7 (7.895)	31.73%	10.307 (12.932)
G6	10%	0.137 (0.382)	15.665 (8.963)	20.08%	10.552 (12.442)
G7	5%	0.096 (0.309)	15.555 (9.135)	24.84%	10.677 (15.029)
G8	2%	0.054 (0.243)	21.9 (9.249)	34.25%	5.2707 (9.5886)

Group 4 (G4): patients transferred from nursing wards.

Group 5 (G5): patients from (unscheduled) emergency surgery.

Group 6 (G6): multiple trauma patients having required no surgery.

Group 7 (G7): multiple trauma patients post surgery.

Group 8 (G8): other patients.

The Hospital of Navarre provided us with a large data set collected by the ICU administration staff over the period 1/1/2000 to 31/12/2008. The data occupy three files:

The first file contains records of 6,414 patients treated in the ICU during the nine-year period. There is a file entry for each patient showing *age*, *arrival and discharge dates*, *infections* in the ICU, *APACHE II scores* and *exitus*. APACHE II score (“Acute Physiology and Chronic Health Evaluation II”) is a severity-of-disease classification system: after admission of a patient to an intensive care unit, an integer score is computed based on several measurements (Knaus et al. 1985). Higher scores imply a more severe condition and a higher mortality risk. *Exitus* is a binary-coded variable that represents whether the patient leaves the ICU following recovery or mortality. Patient length of stay (number of days in the ICU) was computed from the arrival and discharge date records.

The second file records the daily total of arrivals to the ICU. The third is a daily record of bed occupancy taken at four pm each day.

We began by making a descriptive statistical analysis of these data in order to familiarize ourselves with them. An initial summary of the data for the eight patient groups considered in the model is shown in Table 1. This table shows the percentage of patients for each group, means and standard deviations of the daily arrival rate, Apache II score, percentage of exitus and length of stay (in days).

## 2.2 Arrival analysis

Many studies (see for example, Kim et al. 2000; Litvack et al. 2008; Oddoye et al. 2009 and Ridge et al. 1998) use a Poisson Process as a statistical model for patient arrival to a health care center. The following characterization of a Poisson Process illustrates the model’s suitability for representing the patient arrival pattern.

Let us denote by  $N_t$  the number of patients that arrive up to time  $t$ . The stochastic arrival process of patients  $\{N_t, t \geq 0\}$  is a *Poisson Process* if:

1. Patients arrive one at a time.

**Table 2** Arrival distributions. POIS() stands for the Poisson distribution and EMPDIST() stands for empirical distribution

Pathology type	Arrival distributions	$p$ -values of $\chi^2$ tests
G2	POIS(0.270)	0.393
G3	POIS(0.135)	0.459
G4	POIS(0.226)	0.204
G5	POIS(0.205)	0.773
G6	POIS(0.187)	0.088
G7	POIS(0.096)	0.474

Pathology type	Arrival distributions
G1	Weekends: 0 Monday-Friday: EMPDIST (0: $p = 0.1519$ ; 1: $p = 0.5869$ ; 2: $p = 0.2454$ ; 3: $p = 0.0148$ ; 4: $p = 0.001$ ) July: EMPDIST (0: $p = 0.4301$ ; 1: $p = 0.4514$ ; 2: $p = 0.1185$ )
G8	EMPDIST (0: $p = 0.9489$ ; 1: $p = 0.0487$ ; 2: $p = 0.0021$ ; 4: $p = 0.0003$ )

- The number of arrivals in the time interval  $(t, t + s]$ ,  $N_{t+s} - N_t$ , is independent of the number and times of arrivals from 0 to time  $t$ . That is, it is independent of the variable set  $\{N_u, 0 \leq u \leq t\}$ .
- The distribution of  $N_{t+s} - N_t$  is independent of  $t$  for all  $t, s \geq 0$ .

Properties 1 and 2 can be interpreted as follows. Patients arrive at the ICU on an individual basis, their arrival times are not influenced by prior patient arrivals and they are not coordinated according to any pre-established plan. Condition 3 sets the *homogeneity* of the process through time.

However, the Poisson Process fits the daily arrival data for Saturdays and Sundays, but not for weekdays. Closer analysis of the weekday data showed that they fit a *translated Poisson distribution*. This type of distribution, which is related to the Poisson distribution, is obtained by the sum of a Poisson and an independent, approximately normally-distributed, integer-valued random variable (see Röllin 2005). The difference between the weekday and weekend arrival patterns reflects the fact that there is no scheduled surgery at the weekend. Consequently, group G1 patients are never admitted at weekends. Therefore, from the original data files, we had to derive new ones containing the number of arrivals per pathology type per day.

The arrival data analysis for each group reveals that the arrival patterns of groups G2 to G7 fit a Poisson Process ( $p$ -values of the  $\chi^2$ -tests are shown in Table 2). However, the hypothesis of Poisson arrivals is rejected for both the pre-scheduled patients (G1) and the small percentage of patients in group G8 (others). Furthermore, the elective surgery group (G1) presents a different arrival pattern for July, due to this being a holiday period.

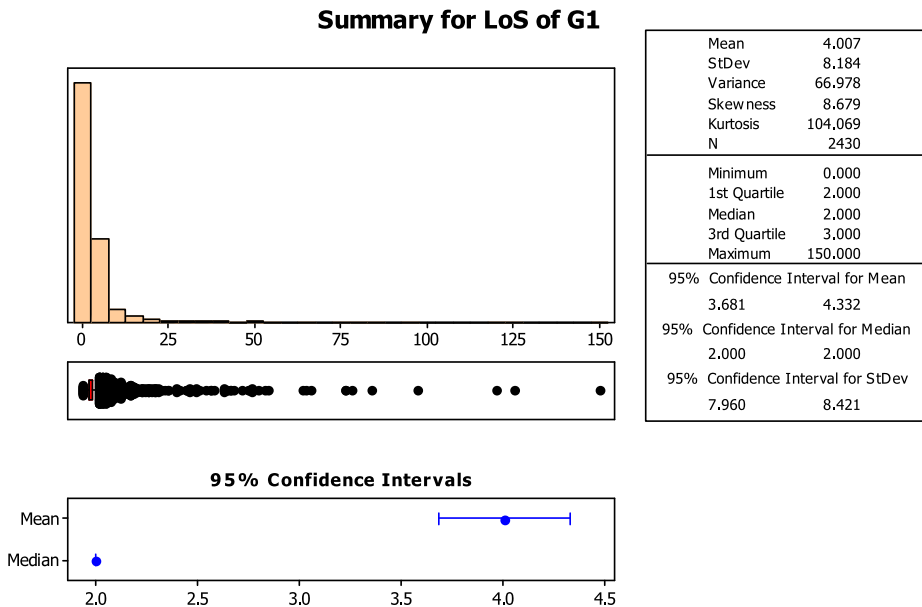
In our simulation model, we use Poisson distributions for groups G2 to G7 and empirical discrete probability distributions for groups G1 and G8, with values between 0 and 4. These distributions are displayed in Table 2.

### 2.3 Length of stay analysis

We also found that the stochastic behavior of LoS is different for each group of patients. All showed a high percentage of long stays, far from the mean value. As an example, Fig. 1 presents a summary of the G1 group LoS data. Note that the median and mean values are quite different and that the mean is even greater than the third quartile. The box plot also shows that many of the observations are outliers. These empirical findings suggest the suitability of modeling with weighted-tail distributions. Earlier studies have noticed that LoS distributions are skewed and contain outliers (Kim et al. 2000; Marazzi et al. 1998 and Rauner et al. 2003) Nevertheless, common distributions, including those with weighted tails, showed poor fit to the data, as shown in Fig. 2 for group G1, when we consider the lognormal and loglogistic distribution families.

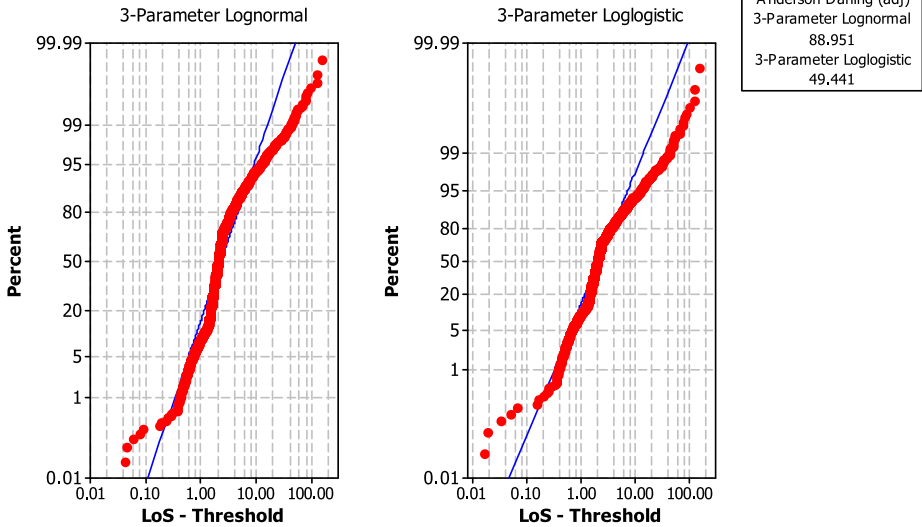
Various outlier trimming rules are commonly used in the literature in order to deal with this problem of lack of fit to the original data (Marazzi et al. 1998). This kind of procedure provides good results when the interest is in the robust estimation of parameters or in assessing the influence of a set of independent variables (Rauner et al. 2003). Nevertheless, outliers can provide important information for the study of various aspects of health service management, including ICU bed occupancy levels. Outliers should therefore not simply be eliminated, but rather carefully represented and included in the statistical model.

To illustrate the importance of including extreme values when fitting statistical models to data, we present a simple, hypothetical, numerical example. Let us suppose that patients with LoS of more than 15 days are considered as outliers and assume that they account for 5% of the total number of patients. Let 4 be the mean LoS for non-outlier patients and let 60 be the mean LoS for outlier patients (then mean LoS is  $4 \times 0.95 + 60 \times 0.05 = 6.8$ ). In this example, outlier patients account for over 44% of ICU occupancy. Some studies compensate for the elimination of outliers by reducing both the sample and the number of available beds



**Fig. 1** Statistical results for LoS of G1 patients

### Probability Plot for LoS of Group G1 ML Estimates-Complete Data



**Fig. 2** Statistical fitting of pathology type G1 LoS to Lognormal and Logistic families

by the same percentage. The weakness of this approach is obvious, however, since a 5% reduction in capacity leads to ICU service over-performance (95% of the available resources are retained to serve 56% of the workload). A 44% reduction in bed capacity would also be incorrect, since, even assuming average values, the new scenario will have a different degree of variance, and queuing theory holds that some important queue performance measures depend strongly on variances.

Observe that a patient with LoS of 100 days uses the same amount of resources as 100 patients who spend only 1 day in the ICU. Due to the significant impact on ICU resources of patients with high LoS values, an appropriate statistical model of LoS should accurately capture the behavior of the distribution tail. For each patient group, we analyze the dependence between LoS and other variables (age, sex, infections, APACHE II score and exitus). The data (see Table 3) show that LoS is positively correlated with the APACHE II score in the case of patients leaving the ICU following recovery (that is, exitus = 0) and negatively correlated with the APACHE II score in cases of mortality (exitus = 1). No significant correlation is found for LoS vs. age and LoS vs. sex pairs.

The variables found to be significantly related with LoS are included in the statistical model of each group by using regression models for the location parameters of their distribution function. Specifically, let us denote by  $F(\mu, \sigma)$  a distribution function for a family of distributions with location and scale parameters  $\mu$  and  $\sigma$ , respectively. It is assumed that the distribution function of the variable  $LoS_{G_k}$ , LoS for group  $G_k$ , belongs to one of these families with location and scale parameters.

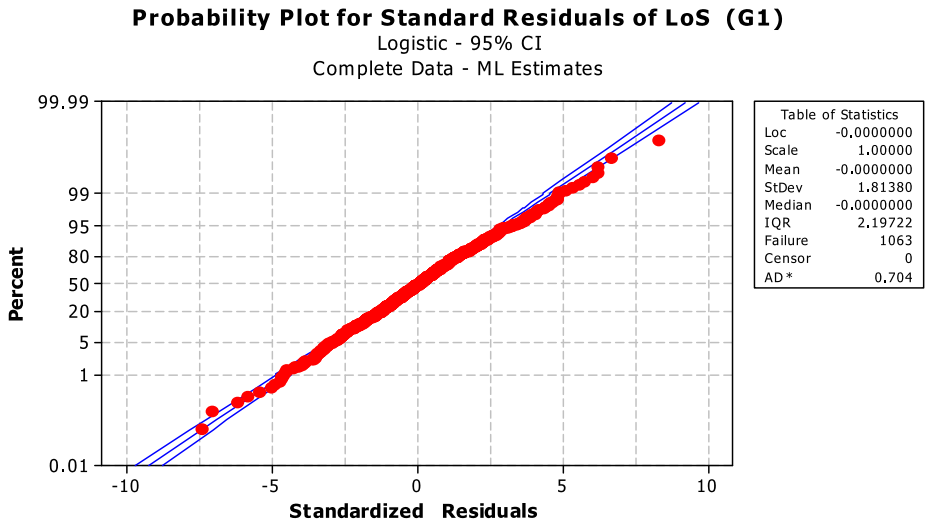
$$LoS_{G_k} \xrightarrow{d} F_k(\mu_k, \sigma_k) \tag{1}$$

The influence of explanatory variables such as *APACHE II* and *Exitus* is introduced into the distribution function through a regression model for the location parameter:

$$\mu_k = \beta_{0,k} + \beta_{1,k} \times APACHE + \beta_{2,k} \times EXITUS \quad \text{for } k = 2, 3, \dots, 8 \tag{2}$$

**Table 3** Correlations between LoS and APACHE II and *p*-values (in brackets)

	G1	G2	G3	G4	G5	G6	G7	G8
Exitus = 0.1	0.388 (0.000)	-0.009 (0.781)	0.031 (0.509)	-0.067 (0.066)	0.095 (0.013)	0.089 (0.156)	0.132 (0.019)	-0.253 (0.008)
Exitus = 0	0.420 (0.000)	0.155 (0.000)	0.141 (0.066)	-0.067 (0.000)	0.360 (0.011)	0.409 (0.000)	0.414 (0.000)	0.137 (0.328)
Exitus = 1	-0.011 (0.944)	-0.160 (0.127)	-0.177 (0.051)	-0.249 (0.000)	-0.266 (0.000)	-0.428 (0.002)	-0.154 (0.177)	-0.341 (0.009)



**Fig. 3** Standard residuals for group G1

The estimation of the parameters  $\beta_{0,k}, \beta_{1,k}$  and  $\sigma_k$  gives distribution functions that closely fit the LoS data for all groups except G1. The latter requires the use of explanatory variables to capture four different types of infections. We consider binary variables ( $INFECTION_i, i = 1, 2, 3, 4$ ), that takes a value of 1 if the patient is infected by infection type  $i$  and 0 otherwise.

$$\mu_1 = \beta_{0,1} + \sum_{i=1}^4 \beta_{i,1} \times INFECTION_i + \beta_{5,1} \times APACHE + \beta_{6,1} \times EXITUS \quad (3)$$

Table 4 shows the estimated distributions obtained to model LoS for each patient group. The location and scale distribution functions belong in all cases to the lognormal ( $k = 2, 3, 5, 6, 7$ ) and the loglogistic ( $k = 1, 4, 8$ ) distributions. The parameters are estimated by the maximum likelihood method and the models validated by analyzing the standard and Cox-Snell residuals. By way of illustration, Fig. 3 plots the standard residuals for group G1.

Observe that, to simulate the LoS from these distributions, we also need to simulate values for the explanatory variables included in the regression models. The distribution fits of these variables are straightforward and they also are included in Table 4. The software used for the statistical analysis was the Minitab® 15.1.30.0 Package.



**Table 4** Exitus, Apache II and LoS distributions

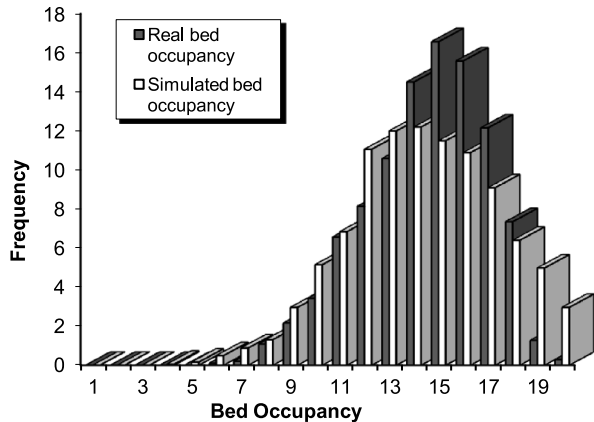
Group	Exitus	(EX) %	APACHE II	Length of stay
G1	EX = 0	95.7%	WEIB(12.9, 2.22)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.230098 + infection_1 \times 1.57507 + infection_2 \times 1.10923 + infection_3 \times 1.05665 + infection_4 \times 0.903656 + exitus \times 1.34681 + apache \times 0.0542346 - exitus \times apache \times 0.06969$ $\sigma: 0.302509$
	EX = 1	4.3%	LOGLOG(0.173, 2.95)	
G2	EX = 0	68.8%	WEIB (20.48, 2.277)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.859410 + apache \times 0.0338953 + exitus \times 1.85895 - exitus \times apache \times 0.094334$ $\sigma: 1.17510$
	EX = 1	31.2%	NORM (29.27, 8.955)	
G3	EX = 0	72.7%	GAMMA(5.91, 2.819)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.992703 + apache \times 0.0420803 + exitus \times 1.62775 - exitus \times apache \times 0.085255$ $\sigma: 1.13368$
	EX = 1	27.3%	WEIB (28.49, 3.373)	
G4	EX = 0	63.2%	LOGIST(4.465, 18.79)	Loglogistic ( $\mu, \sigma$ ): $\mu: 1.22707 + apache \times 0.0221330 + exitus \times 2.32357 - exitus \times apache \times 0.103393$ $\sigma: 0.661459$
	EX = 1	36.8%	NORM(28.23, 8.936)	
G5	EX = 0	68.3%	NORM(15.94, 6.742)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.83441 + apache \times 0.0664619 + exitus \times 2.60450 - exitus \times apache \times 0.157931$ $\sigma: 1.09284$
	EX = 1	31.7%	LOGLOG(0.180, 3.15)	
G6	EX = 0	79.9%	WEIB(15.52, 2.102)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.988974 + apache \times 0.0750502 + exitus \times 2.08852 - exitus \times apache \times 0.159748$ $\sigma: 1.04561$
	EX = 1	20.1%	LOGIST(4.347, 27.88)	
G7	EX = 0	75.2%	WEIB(14.78, 1.964)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.844700 + apache \times 0.0831031 + exitus \times 2.08434 - exitus \times apache \times 0.156644$ $\sigma: 1.07846$
	EX = 1	24.8%	NORM(27.04, 6.568)	
G8	EX = 0	65.7%	WEIB(3.91, 9.801)	Loglogistic ( $\mu, \sigma$ ): $\mu: 0.769804 + apache \times 0.0355266 + exitus \times 0.263105 - exitus \times pache \times 0.06273$ $\sigma: 0.542519$
	EX = 1	34.3%	WEIB(29.27, 3.322)	

### 3 Simulation model building and validation

The ICU was mathematically modeled as a queuing system, where the clients are the patients, the servers are the 20 beds and there is no waiting room. Any patient arriving when the ICU is full is transferred to an alternative ICU (in a local private facility or a neighboring region). The simulation model is therefore structurally quite simple and can be run using the statistical modeling techniques presented in Sect. 2. Patient arrival times are simulated using Poisson processes and empirical distributions. Their service times (LoS) are simulated from statistical distributions. The location parameters of these also need to be calculated from simulated values of the explanatory variables, as described in the previous section. This queuing model was implemented in ARENA 11 software. After verifying the model, 20-year simulation experiments were run, with a 3-year warm-up period. These were sufficient for the ICU system to reach a stationary state as verified by bed occupancy data.

**Table 5** Two-sample test on bed occupancy real data vs simulated data

	<i>N</i>	Mean	St. dev	SE mean
<i>t</i> -Test: $H_0: \mu(\text{real}) = \mu(\text{simulated})$				
Real bed occupancy	20	14.44	0.87	0.19
Simulated bed occupancy	36	14.19	1.46	0.24
<i>T</i> -test of diff. = 0 (vs. <> 0): <i>t</i> -value = 0.79, <i>p</i> -value = 0.434				

**Fig. 4** Real vs. simulated bed occupancy frequencies

### 3.1 Model validation

The most definitive test of a simulation model’s validity is whether its output data closely resemble the output data of the actual system (Law 2006). This validation approach is applicable to our simulation model because we are using real output data from an existing ICU. That is, we use the bed occupancy variable to compare the real ICU with the simulated ICU. If the two sets of data resemble each other *closely* then our simulation model of the existing ICU can be considered *valid*.

We use the ICU daily bed occupancy records to validate the simulation model. A test of equality of means does not reject the null hypothesis of equality of mean bed occupancy for the recorded data and the simulation output data (see Table 5). To perform this *t*-test, we take the two streams of data in 50-day batches and calculate the averages. This gives samples with independent observations.

This equality of means cannot be extended to the whole distribution, however, as suggested by the frequency graphs of both data sets (see Fig. 4 and Table 6). In fact, the Kolmogorov-Smirnov test clearly rejects the hypothesis of equal distributions (*p*-value < 0.0001).

Therefore, these results do not completely validate our simulation model. A visual inspection of Fig. 4 and Table 6 suggests that the length of stay of some patients might be influenced by the ICU bed occupancy level. The real data show a higher frequency of values 14, 15, 16 and 17 and the simulated data show higher frequencies at both ends of the distribution, especially for values of 19 and 20.

The only explanation for these discrepancies is that the ICU system does not work as “automatically” as its mathematical simulation. This suggests the influence of management

**Table 6** Frequency values of real and simulated bed occupation data

Daily bed occupancy	Real data frequencies	Simulated data frequencies
1	0	0
2	0	0
3	0	0
4	0	0.05
5	0	0.16
6	0.1	0.49
7	0.2	0.88
8	1.08	1.31
9	2.16	2.96
10	3.43	5.15
11	6.57	6.84
12	8.14	11.06
13	10.59	12.37
14	14.51	12.70
15	16.57	11.71
16	15.59	10.89
17	12.16	9.09
18	7.35	6.40
19	1.27	4.98
20	0.29	2.96

policies for patient admission and discharge on the bed occupancy distribution. The ICU staff confirmed that some discharge decisions are made in order to keep the number of occupied beds between 12 and 17. These figures correspond with the general recommendation that ICUs should have 60–70% average bed occupancy (The Intensive Care Society 1997). Clearly, therefore, these human decisions are based on the ICU bed occupancy level, the ICU bed requirement forecast and the medical condition of the ICU patients. Medical staff may therefore slightly delay discharge or authorize slightly earlier discharge if the patient's condition allows.

### 3.2 ICU management decision modeling

Since these decisions appear to have a crucial impact on ICU performance, they must be introduced into the simulation model. There is no written protocol for managers to determine patient admission and discharge automatically; these decisions are subject to the judgment of the intensive care consultant. Our approach to modeling this decision-making process requires us to define a set of rules to determine the conditions for the slightly earlier discharge of certain patients or the slightly delayed discharge of others. These conditions should depend on the bed occupancy level. Specifically, we mathematically model early and delayed discharge by considering the following rules:

1. If the bed occupancy level is *high* ( $n > 15$ ) and some patients' length of stay conditions are satisfied (*if the patient's prognosis is good and the remaining LoS is less than  $PR_n\%$  and less than  $DR_n$  days*), then some patients may be discharged early.

2. If the bed occupancy level is *low* ( $n < 14$ ) and some patients' length of stay conditions are satisfied (if the patient's initial LoS is less than  $DI_n$  days), then some patients' stay can be slightly extended ( $PI_n$ ).

The definition of these rules depends on the following parameters:

- $DR_n$ : upper limit for the reduction in the number of LoS days when bed occupancy level is  $n$ ,  $n = 16, \dots, 20$ .
- $PR_n$ : percentage reduction in LoS when bed occupancy level is  $n$ ,  $n = 16, \dots, 20$ .
- $Num_n$ : upper limit for the number of patients where LoS can be reduced when bed occupancy level is  $n$ ,  $n = 16, \dots, 20$ .
- $PI_n$ : percentage increase in LoS when bed occupancy level is  $n$ ,  $n = 0, \dots, 13$ .
- $DI_n$ : upper limit for the number of LoS days for which an increase is allowed, when bed occupancy level is  $n$ ,  $n = 0, \dots, 13$ .

### 3.3 Finding the best set of rules to describe the management decisions

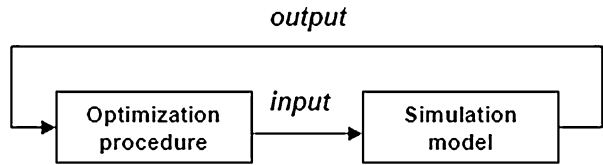
Observe that the set of rules defines infinite management policies for the ICU: one for each set of parameter values. The simulation model only allows us to assess the performance of one set of parameters in each simulation run. By running the simulation model for some reasonable values of the rule parameters, we confirm their influence on the distribution of the bed occupancy frequencies and the approximation of the simulated to the real frequency distribution. This brings us to the question of how to find the best set of parameter values. We address this by defining an optimization problem aimed at matching the simulated bed occupancy output with the real one, denoted by  $simulfrequ_i$  and  $realfrequ_i$ , respectively.

Specifically, we consider a mathematical programming problem where the decision variables are the parameters of the rules described above ( $DR_n, PR_n, Num_n, PI_n, DI_n$ ). The upper and lower bounds for these variables ( $upperDR, lowerDR, upperPR, lowerPR, upperNum, lowerNum, upperPI, lowerPI, upperDI$  and  $lowerDI$ ) are parameters of the optimization problem. The sum of squared differences of the simulated and real frequencies of bed occupancy is minimized (4). The constraints represent the monotonicity of the values for the five groups of decision variables in accordance with the decision rules described above and their upper and lower bounds (5)–(9). As an example, constraints (5) establish that reductions in length of stay when 20 ICU beds are occupied ( $DR_{20}$ ) must be no less than when 19 beds are occupied ( $DR_{19}$ ). The following nonlinear integer programming problem is considered:

$$\text{Min } \sum_{i=0}^{20} (realfrequ_i - simulfrequ_i)^2 \tag{4}$$

subject to

$$\left\{ \begin{array}{l} upperDR \geq DR_{20} \geq DR_{19} \geq \dots \geq DR_{16} \geq lowerDR \tag{5} \\ upperPR \geq PR_{20} \geq PR_{19} \geq \dots \geq PR_{16} \geq lowerPR \tag{6} \\ upperNum \geq Num_{20} \geq Num_{19} \geq \dots \geq Num_{16} \geq lowerNum \tag{7} \\ upperPI \geq PI_1 \geq PI_2 \geq \dots \geq PI_{13} \geq lowerPI \tag{8} \\ upperDI \geq DI_1 \geq DI_2 \geq \dots \geq DI_{13} \geq lowerDI \tag{9} \\ DR_i, Num_i, DI_j \text{ integer, } \quad i = 16, \dots, 20, \quad j = 1, \dots, 13 \end{array} \right.$$

**Fig. 5** Combination of simulation with optimization techniques**Table 7** Two-sample test on real bed occupancy data vs simulated data using decision rules

	<i>N</i>	Mean	St. dev	SE mean
<i>t</i> -Test: Ho: $\mu(\text{real}) = \mu(\text{simulated})$				
Real bed occupancy	20	14.44	0.87	0.19
Rules-simulated bed occupancy	36	14.46	1.40	0.12
<i>t</i> -Test of diff. = 0 (vs. <0): <i>t</i> -value = $-0.11$ , <i>p</i> -value = 0.911				

Observe that the objective function can only be evaluated through simulation. Therefore, an optimization with simulation procedure has to be used to solve the proposed optimization problem. Simulation and optimization are combined (see April et al. 2003) as follows. The optimization procedure determines a set of parameter values or solution, that is, a value for the decision variables of the problem. This system configuration is simulated. The output of this simulation is used in the optimization procedure to evaluate the random objective. Using this information and its own search method, the optimization procedure decides the next solution to be evaluated (Fig. 5). This process continues until the stopping conditions of the optimization method are met.

We use ARENA and OptQuest software to solve this optimization problem and to obtain the optimal value of the rule parameters for the simulation model.

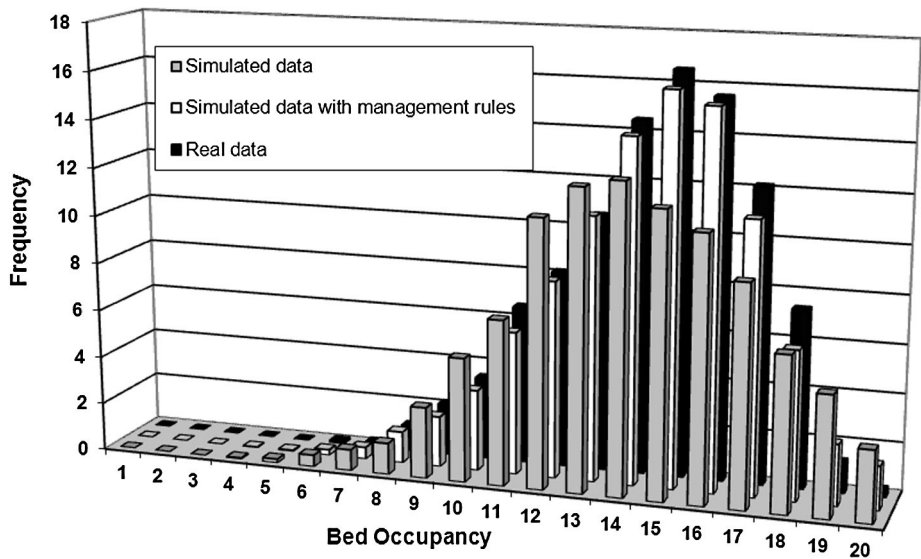
### 3.4 Validation of the simulation model including management decisions

To validate the new simulation model, we check that the daily bed occupancy generated by the simulation model matches the real data, both in mean values and in distributions. The equality of means cannot be rejected (see Table 7) and can be extended to the whole distribution. Graphical and numerical experimental results of real and simulated bed occupancy frequencies are presented in Fig. 6 and Table 8, respectively. The Kolmogorov-Smirnov test does not reject the null hypothesis of equal distributions ( $p$ -value = 0.531).

Figure 6 shows the daily bed occupancy frequencies for the real data (dark gray bars), the frequencies for the simulation model including the optimal rule parameters (white bars), and the frequencies for the simulation model without management rules (pale gray bars). This figure and the numerical values in Table 8 confirm the need to introduce management rules into the simulation model. We also checked that the application of these management rules does not modify the distribution of the data. These results allow us to validate the simulation model.

## 4 Conclusions and discussion

This paper describes a simulation model that may be viewed as a reliable tool for ICU sizing and capacity analysis and the testing of different ICU management policies. We illustrate the proposed methodology with a simulation of the ICU at the Hospital of Navarre, Spain.



**Fig. 6** Real vs. simulated (with and without rules) bed occupancy frequencies

We focus particularly on the statistical modeling of stochastic elements, thus reaching beyond the usual scope of simulation studies. We model the arrival process for each patient pathology group. All patient groups except groups G1 and G8 fit Poisson processes. Group G8 is a residual group, but G1 has a major impact on the arrival process, since it accounts for 40% of ICU admissions. More research is required to find a better statistical model for this group and obtain more information about surgery scheduling and operating theater performance. We have initiated the search for this information. We illustrate the importance of constructing an accurate LoS model. Furthermore, since our data show weighted tails, and therefore do not fit standard statistical models, we include explanatory variables to improve the fit of the LoS data by fitting them to non-normal regression models.

Using descriptive statistical analysis, we check for trends and seasonality components in both arrivals and LoS averages. We find a seasonality component in the arrivals for programmed surgery due to physicians' vacations, which we take into account by fitting different statistical models. We also find a change in the rate of arrivals of group G6 patients in 2005 due to the decrease of traffic accidents (resulting from new traffic regulations including harder penalties for lawbreakers). For this group we only fit the data for the period 2005 to 2008.

One of the main issues addressed in this study is the validation of the model. We use a bed occupancy variable to compare the real ICU against our first ICU simulation model. Further analysis reveals that, despite having equal means, the simulated distribution does not match the real one. The explanation for this was that we initially included the common assumption of independency between LoS and workload in the ICU. The validation process showed this assumption to be incorrect, however, since patient length of stay is sometimes influenced by the bed occupancy level. Our results are consistent with accounts given in interviews by ICU managers and medical staff. They reported their personal experience to be that some commonly-applied bed-management decisions imply early discharge for some patients and slightly delayed discharge for others.

**Table 8** Bed occupancy frequency values for the new simulation model

Daily bed occupancy	Real data frequencies	Rule-simulated frequencies
1	0	0.00
2	0	0.00
3	0	0.00
4	0	0.03
5	0	0.05
6	0.1	0.23
7	0.2	0.49
8	1.08	1.31
9	2.16	2.11
10	3.43	3.37
11	6.57	5.93
12	8.14	8.16
13	10.59	10.91
14	14.51	14.15
15	16.57	16.04
16	15.59	15.48
17	12.16	11.19
18	7.35	6.16
19	1.27	2.53
20	0.29	1.82

The selection of patients for admission to and discharge from the ICU is a triage problem that physicians face on a daily basis. This problem of triage is particularly difficult when the ICU is full or almost full. Despite the impact of these triage decisions on the everyday running of the ICU, to the best of our knowledge they have never been included in a simulation model in previous research. Thus, our attempt to represent them by using a mathematical model that can be included in the simulation model is one of the main contributions of this paper. Furthermore, we show that a simulation model omitting these management decisions would not be valid.

We model these patient discharge decisions using a set of rules expressing in mathematical language physicians' declarations regarding earlier discharge for some patients without compromising their health when the ICU is fully, or almost fully, occupied. The rule for shortening LoS is then described by two variables that establish, in absolute and relative terms, respectively, the conditions under which early discharge can take place by limiting the reduction time. We need a pair of these variables for each level of bed occupancy. We model extended length of stay in a similar way.

The ICU staff, not trained in mathematical models, understood the purpose of the mathematical rules and accepted them as a good way to represent what they do. In fact, this study has made them think for the first time about how they were making the decisions. This issue has until now never been addressed from a global perspective in a structured way (working in an ICU is usually stressful and physicians are focused on healthcare decisions concerning the patients at hand).

From a technical point of view, these rules depend on a set of parameters whose "best" values are determined by the solution of an optimization problem requiring the combination

of simulation and optimization. Comparison of this new simulation model with the real ICU reveals close similarity between the two, which not only validates the simulation model; it also demonstrates its credibility for use as a decision-support tool.

Nevertheless, since the first version of this paper they (the physicians) and we (mathematicians) have continued studying this modeling problem. Currently we are trying to improve the rules by incorporating other medical criteria, by distinguishing between the triage decisions for different patient groups, for example. The most challenging improvement, however, stems from the consideration that interventions to modify a patient's normal length of stay (from entry to recovery, or death) should be kept to a minimum. The consideration of this new element leads to the formulation of bicriteria optimization problems in which historical data are no longer the only criterion used to determine parameter values. This is our work in progress.

Another issue has to do with the meaning of management decisions at the ICU. The fact that a single physician makes all the decisions required for each patient raises a question: are there significant differences in the decisions made by different physicians? ICU staff think not, but they do not know for sure. There has been a stable team of physicians over the period and they usually discuss the patients' health, but these circumstances alone do not guarantee identical decisions. Moreover, the information in the database is aggregated for all physicians, so comparison using historical data is not possible. Given the interest that this issue has for physicians, and also for us, we plan as future work to develop an interactive simulation model for the ICU in which the decisions are made by the user. Its purpose is to record information about how different physicians reached their decisions, in order to test the equality of their management decisions.

We therefore achieved our initial aim of building a credible simulation model for ICU managers. It serves as a starting point from which to develop capacity-planning studies. As an example of its use, consider the problem of determining the number of beds needed to meet an increase of patients at the ICU due to the increase of patients coming from programmed surgery. One of the main problems of the public health care system is the length of waiting lists for some specialist surgical procedures. To reduce waiting lists and improve service quality, the hospital is considering extending some operating theater hours, which would increase the number of patients from elective surgery (group G1) by 50%. Simulation results show that two more ICU beds are required to keep rejected admissions to a maximum of 5%, and six more to keep them below 1%.

Although the results obtained in this study are directly related to the input data and the characteristics of this ICU, it should be noted that the proposed methodology could be used to model other ICUs.

**Acknowledgements** The authors thank the managers and staff of the ICU of the Hospital of Navarre for providing the data and for their help in increasing our understanding of ICU performance. Part of this work was done during the stay of the Author Fermín Mallor in the Missouri University of Science and Technology, who is grateful to the Ministerio de Educación Español for the grant PR2010-0430 of the "Salvador de Madariaga" program.

## References

- Ahmed, M. A., & Alkhamis, T. M. (2009). Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research*, 198, 936–942.
- de Angelis, V., Felici, G., & Impelluso, P. (2003). Integrating simulation and optimisation in health care center management. *European Journal of Operational Research*, 50, 101–114.



- April, J., Glover, F., Kelly, J., & Laguna, M. (2003). Practical introduction to simulation optimization. In S. Chick, P. J. Sánchez, D. Ferrin, & D. J. Morrice (Eds.), *Proceedings of the 2003 winter simulations conference*.
- Azcárate, C., Mallor, F., & Gafaro, A. (2008). Multiobjective optimization in health care management. A metaheuristic and simulation approach, *Algorithmic Operations Research*, 3, 186–202.
- Brailsford, S. C., Gutjahr, W., Rauner, M. S., & Zeppelzauer, W. (2007). Optimal screen policies for diabetic retinopathy using a new combined discrete event simulation and ant colony optimization approach. *Computational Management Science*, 4, 59–83.
- Brailsford, S. C., Harper, P. R., Patel, B., & Pitt, M. (2009). An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3, 130–140.
- Eldabi, T., Paul, R. J., & Young, T. (2007). Simulation modelling in healthcare: reviewing legacies and investigating futures. *The Journal of the Operational Research Society*, 58, 262–270.
- Fetter, R. B., & Thompson, J. D. (1965). The simulation of hospital systems. *Operations Research*, 13, 689–711.
- Fone, D., Hollinghurst, S., Temple, M., Round, A., Weightman, A., Roberts, K., Coyle, E., & Bevan, G. (2003). Systematic review of the use and value of computer simulation modelling in population health and healthcare delivery. *Journal of Public Health Medicine*, 25, 325–335.
- Griffiths, J. D., Price-Lloyd, N., Smithies, M., & Williams, J. E. (2005). Modelling the requirement for supplementary nurses in an intensive care unit. *The Journal of the Operational Research Society*, 56, 126–133.
- Günal, M. M., & Pidd, M. (2010). Discrete event simulation for the performance modelling in health care: a review of the literature. *Journal of Simulation*, 4, 42–51.
- Jun, J. B., Jacobson, S., & Swisher, J. (1999). Application of discrete-event simulation in health care clinics: a survey. *The Journal of the Operational Research Society*, 50, 109–123.
- Katsaliaki, K., & Mustafee, N. (2010). Applications of simulation within the healthcare context. *The Journal of the Operational Research Society*. doi:10.1057/jors.2010.20.
- Kc, D., & Terwiesch, C. (2009). Impact of workload on service time and patient safety: and econometric analysis of hospital operations. *Management Science*, 55, 1486–1498.
- Kim, S. C., Horowitz, I., Young, K., & Buckley, T. A. (1999). Analysis of capacity management of the intensive care unit in a hospital. *European Journal of Operational Research*, 115, 36–46.
- Kim, S. C., Horowitz, I., Young, K., & Buckley, T. A. (2000). Flexible bed allocation and performance in the intensive care unit. *Journal of Operations Management*, 18, 427–443.
- Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical Care Medicine*, 13, 818–829.
- Kolker, A. (2009). Process modeling of ICU patient flow: effect of daily load levelling of elective surgeries on ICU diversion. *Journal of Medical Systems*, 33, 27–40.
- Law, A. M. (2006). How to build valid and credible simulation models. In L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, & R. M. Fujimoto (Eds.), *Proceedings of the 2006 winter simulation conference* (pp. 58–66).
- Litvack, N., van Rijsbergen, M., Boucherie, R. J., & van Houdenhoven, M. (2008). Managing the overflow of intensive care patients. *European Journal of Operational Research*, 185, 998–1010.
- Marazzi, A., Paccaud, F., Ruffieux, C., & Beguin, C. (1998). Fitting the distributions of length of stay by parametric models. *Medical Care*, 36, 915–927.
- Marshall, A., Vasilakis, C., & El-Zardi, E. (2005). Length of stay-based patient flow models: recent developments and future directions. *Health Care Management Science*, 8, 213–220.
- Oddoye, J. P., Jones, D. F., Tamiz, M., & Schmidt, P. (2009). Combining simulation and goal programming for healthcare planning in a medical assessment unit. *European Journal of Operational Research*, 193, 250–261.
- Rauner, M. S., Zeiles, A., Schaffhauser-Linzattti, M.-M., & Hornik, K. (2003). Modelling the effects of the Austrian inpatient reimbursement system on length-of-stay distributions. *OR-Spektrum*, 25, 183–206.
- Ridge, J. C., Jones, S. K., Nielsen, M. S., & Shahani, A. K. (1998). Capacity planning for intensive care units. *European Journal of Operational Research*, 105, 346–355.
- Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli*, 11, 1115–1128.
- The Intensive Care Society. (1997). Standards for intensive care units. <http://www.ics.ac.uk/>.
- Vasilakis, C., & Marshall, A. H. (2005). Modelling nationwide hospital length of stay: opening the black box. *The Journal of the Operational Research Society*, 56, 862–869.