Journal of the Operational Research Society (2013) 64, 912–924 0160-5682/13

www.palgrave-journals.com/jors/

Determination of number of dedicated OR's and supporting pricing mechanisms for emergent surgeries

JA Paul and L MacDonald*

Kennesaw State University, GA, USA

Inefficient management of emergent surgeries in hospitals can, in part, be attributed to a lack of rigorous analysis appropriate to capturing the underlying uncertainties inherent to this process and a pricing mechanism to ensure its financial viability. We develop a non-preemptive multi-priority queueing model that optimally manages emergent surgeries and supports the resource allocation decision-making process. Specifically, we utilize queueing and discrete event simulation to develop empirical models for determining the required number of emergent operating rooms for a hospital surgical department. We also present algorithms that estimate the appropriate pricing for patient surgeries differentiated by priority level given the patient demand and the resources reserved to meet this demand.

Journal of the Operational Research Society (2013) **64,** 912–924. doi:10.1057/jors.2012.92

Published online 3 October 2012

Keywords: hospitals; practice of operations research; emergent surgeries; simulation; queueing; decision support systems

Introduction

As operating costs continue to rise, hospitals continue to explore additional avenues for enhancing revenues. One such direction is strategic price modelling, which was more common in the past before fixed-fee payments became so prevalent (HFMA, 2004). One area in particular that could benefit from strategic pricing is the surgical department, given that it is estimated to account for more than 40% of a hospital's total revenues (HFMA, 2003). The surgical department handles three main types of surgeries: scheduled, add-on and emergent procedures. Currently, fees are typically comprised of a fixed charge, based on the type of surgery, and a variable charge based on the duration, that is, an hourly rate. However, the emergent and add-on surgeries may have additional overhead and costs associated with possible disruptions to the scheduled procedures resulting from their unpredictable arrival patterns as well as the severity, and hence the timing of the required surgical procedure. Yet, current practices involving surgical charges do not differentiate between surgery types at the operational level. As stated by Edward B. Carlson, Vice President and CFO at Munson Healthcare, 'You need some sense of what the relationship of cost-to-charge is at the procedure level, as opposed to just an overall ratio' (HFMA, 2004).

From an operational perspective, the emergent and add-on surgeries have been handled in one of the two primary ways. One, hospitals and operating room (OR) managers create slack in the existing schedule to accommodate these surgeries by cancelling (primarily elective) scheduled cases (Ozkarahan, 2000; Blake *et al*, 2002). The second approach involves managing emergent surgeries separately from scheduled surgeries by reserving rooms based on the emergent surgery demand. This approach has been shown to be effective in reducing the high variability in OR scheduling as well as improve patient outcomes (Denton *et al*, 2007). A recent study utilized simulation and optimization models to implement this approach to meet these objectives at the study participating hospital (Persson and Persson, 2010). Another study on emergent surgery planning, however, concluded that closing rooms reserved specifically for emergent surgeries and reserving capacity in elective ORs is a more cost effective strategy (Wullink *et al*, 2007). Thus, a pricing mechanism that is different from one used for scheduled surgeries is necessary to ensure the economic viability of a system wherein emergent surgeries are handled separately via reservation of ORs.

In this study, we evaluate and determine the number of dedicated ORs required, as well as develop a costing mechanism to ensure long-run economic viability of this approach, with an ultimate goal of maximizing patient outcomes. The remainder of the paper is organized as

*Correspondence: *L MacDonald, Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, Kennesaw, GA 30144, USA.*

E-mail: lmacdon4@kennesaw.edu

follows. We first present a discussion of existing literature relevant to this research effort and then describe our approach. Next, we present models that help reduce variability in OR operations and manage emergent surgeries. Specifically, we look at queueing and simulation approaches to estimate the optimal number of ORs and corresponding patient waiting time. We also propose pricing mechanisms that help apportion the total OR cost of operations among patients based on their priority class. We then demonstrate the applicability of the models via case studies. Finally, we present conclusions and directions for future research.

Background

A number of modelling and quantitative approaches have been brought to bear on various managerial issues within the health-care sector, including OR management. These include, but are not limited to staff scheduling, including size, type and timing issues, process and patient flows, as well as OR scheduling. The review paper by Guerriero and Guido (2011) provides a thorough overview of the application of operations research to the operating theatre. Additional details can be found in the review by Cardoen *et al* (2010). Thus, in our review below, we focus on OR management in which the direct or indirect impacts of urgent or emergent surgeries are considered.

Simulation has been a common approach for studying the complex OR environment. In one of the earliest papers, Goldman and Knappenberger (1968) use discrete event simulation (DES) to determine the optimal number of ORs subject to uncertain demand. In this paper, the authors consider direct OR costs as well as those attributed to patient wait times. Similarly, Pai *et al* (1997) use simulation to optimize the number of ORs in a hospital. van Oostrum *et al* (2008) also use DES to determine emergency OR staffing levels for the night shift, given the uncertainty in both arrivals and surgery duration, with consideration of patient safety due to the time-sensitive nature of the procedures. Sier *et al* (1997) study the surgical scheduling problem involving a number of hard and soft constraints, including those imposed by the uncertainty of emergent procedures. They develop a simulated annealing heuristic to aid in the decision-making process for this multi-criteria objective problem. Denton *et al* (2006) develop a simulation model for a multiple OR surgical suite to provide decision support for efficient handling of surgeries.

Other quantitative methods have also been applied to various problems related to the OR theatre. In an early paper using stochastic dynamic programming, Gerchak *et al* (1996) develop a model to schedule elective surgeries given uncertainties associated with both elective and emergent procedures. Denton *et al* (2007) develop a stochastic optimization model and heuristics for computing OR schedules when faced with uncertainty in surgery durations and sequencing, while Tucker *et al* (1999) use queueing theory to determine the optimal OR staff needed during the night shift at a Level II trauma centre. Zonderland *et al* (2010) study semi-urgent surgeries using queueing techniques to first determine the required number of ORs to meet an uncertain demand, then analyze the tradeoffs between accommodating these patients by cancelling elective surgeries and possible underutilization of OR capacity. In addition, Lamiri *et al* (2008) use a column generation approach to plan utilization of OR resources when faced with the problem of managing emergent and elective surgeries together. Stanciu *et al* (2010) use a unique approach in which resource allocation techniques typical of revenue management are used to allocate capacity for surgeries to maximize expected revenue, including costs associated with schedule interruptions and overtime. In addition, surgery classes are categorized by revenues, not by traditional priority. In Zhang *et al* (2009), the authors optimize allocation of OR capacity to specialties, considering priority emergency surgeries, using a mixed-integer programming model to reduce costs. Jebali *et al* (2006) use a mixed-integer programming approach in their two-step model for scheduling surgeries. The first step deals with the assignment of surgeries to ORs after consideration of the relevant constraints such as availability of an OR, surgeon, equipment and so on. The second step then determines the optimal surgery sequence while minimizing OR overtime.

One gap in the above research is a generic methodology that evaluates the key parameters for allocation of resources regardless of hospital size. This is the primary emphasis of this research endeavour. In particular, we focus on the emergent surgery management problem, and determine the required number of ORs to be reserved based on a designated service level. In addition, an efficient pricing mechanism that apportions the cost of the ORs among the different emergent patient categories to ensure the economic viability of this set-up is notably lacking. This forms the second motivation for this research effort.

Methodology

This study was in part motivated by a regional hospital with a surgical department that handles about 7000 patient cases per year within their 11 OR suites. The patients in the OR were classified in three categories based on how they were incorporated in the OR schedule: scheduled, add-on and emergent cases. These three patient types could be further categorized by surgery type, for example, general, colon rectal, orthopedic and so on. Block scheduling was used to allocate time slots to surgeons based on historical caseloads, previous utilization rates and their preferences. However, the addition of emergent surgeries into the regular schedule were leading to long length of stays for patients, frequent and excessive wait times for surgery, overtime for personnel and so on. All of these issues were affecting patient satisfaction as well as hospital revenues. This led the hospital to conduct a study that looked at efficiency and throughput improvement alternatives, as there was a significant underutilization of ORs in addition to patient dissatisfaction issues. One key decision that resulted from this analysis, which is also well supported in the literature (see, eg, Litvak and Long, 2000 and Litvak *et al*, 2001 and Levtzion-Korach *et al*, 2010), was to reserve ORs for emergent surgeries. Our goal in this study is to develop a generic model that would be applicable to hospitals of varying size in determining the appropriate number of ORs to reserve for emergent procedures.

In general, the management of unscheduled or emergent surgeries in an OR environment is a classic non-preemptive multi-priority queueing problem. In addition, there is the added complexity of patients transitioning to a higher severity/priority class due to prolonged waiting. Thus, unlike standard priority queues, where the highest priority clients are served before lower priority clients, in a hospital the customers are patients whose severity or priority can change due to unreasonable wait times. This might warrant providing service to a previously lower priority patient ahead of a comparatively higher priority patient that has recently arrived. Given the complexity of the system under consideration, we opted to utilize simulations to study the properties of this system and its sensitivity to patient characteristics including volume, mix, average surgery time and so on.

We developed a DES model for this purpose as a DES model has the ability to represent activities in a complex system, such as an OR, as a network of interdependent and discrete events. Moreover, DES models can depict events using the priority-based rules and decision making that drive actual OR operations. Once built, the DES model can then be used to study the sensitivity of changes to inputs on various measures of performance. We obtained data from the previously described surgery department as initial inputs for the simulation model. The data provided by the study hospital included surgery times, OR downtime, OR setup time, population (arrivals) of patients belonging to each priority and so on from July 2008 to June 2009. This data were then used to estimate parameters and probability distributions to incorporate in the simulation model.

As the model was being developed to study a new operation, in which the emergent surgeries would be handled separately, there was no data available from the participant hospital to validate the simulation directly. However, as all the key inputs such as surgery time, patient mix, downtime, OR setup time and so on are based on real data and real-world inputs, and this setup is supported by successful real-time practices in other hospitals as noted from the extant literature (Denton *et al*, 2007; Levtzion-Korach *et al*, 2010), we have considerable confidence in the validity and reliability of the model outputs. In addition, we performed a power analysis to determine the number of replications for our simulation studies (Cohen, 1988). The significance level was set to 0.05, with the power at 0.80 and effect size set equal to 0.3. We assumed a relative error of 2%, and the coefficient of variation was obtained from pilot runs involving 50 replications. The power analysis indicated the appropriate number of replications to be approximately 100, and therefore we chose this value for all our subsequent simulation runs.

We initiated several simulation studies using ProModel (2010) to further our understanding of the emergency surgery process, in order to aid in the development of our models and heuristics. Given arrivals for 'r' priorities as $\lambda_1, \lambda_2, \ldots, \lambda_r$ with the highest severity assigned to Priority 1, then by definition the acceptable wait times (ω) can be ordered as $\omega_1 < \omega_2 < \cdots < \omega_r$. The patient priority is continuously updated based on the time they have already waited when compared with their initial acceptable wait time (ω_r). As an OR becomes available, the patient with the lowest remaining wait time is scheduled for surgery. The classification of patients into different severity levels or priorities is based on the maximum allowable time the patients could wait before receiving care, which we refer to as survivability time. In the study, participant hospital a five-level system for prioritizing patients based on wait times was used. Specifically, the classifications used and acceptable wait times were: emergent (within 1 h), emergent (within 2 h), urgent (within 4 h), semi-urgent (8 h) and non-urgent (within 24 h). This approach is supported in the extant literature as well. For example, a recent article dealing with emergent surgery planning showed that a five-level classification complements the policy of handling emergent surgeries separately (Levtzion-Korach *et al*, 2010). Given the above, we selected a five-level classification for our simulations and subsequent analysis.

Initial simulations were conducted using arrival and surgery time parameters based on the data provided by the participant hospital. We assumed the arrivals process followed a Poisson distribution, which is a common assumption for queueing systems (see Ross, 2003). In addition, the hospital data were consistent with the properties of the Poisson distribution, in that the mean and variance of the arrivals are approximately the same. The surgery times, using surgery data provided by the hospital, were fitted to an Erlang distribution, with a shape parameter of 3 and scale of 41.4 min, for a mean surgery time of 124.2 min. Figure 1 displays a histogram of the original data overlaid with the fitted distribution. Table 1 shows the patient priorities, arrival rates as well as the maximum time to treatment allowed for each priority type, with the

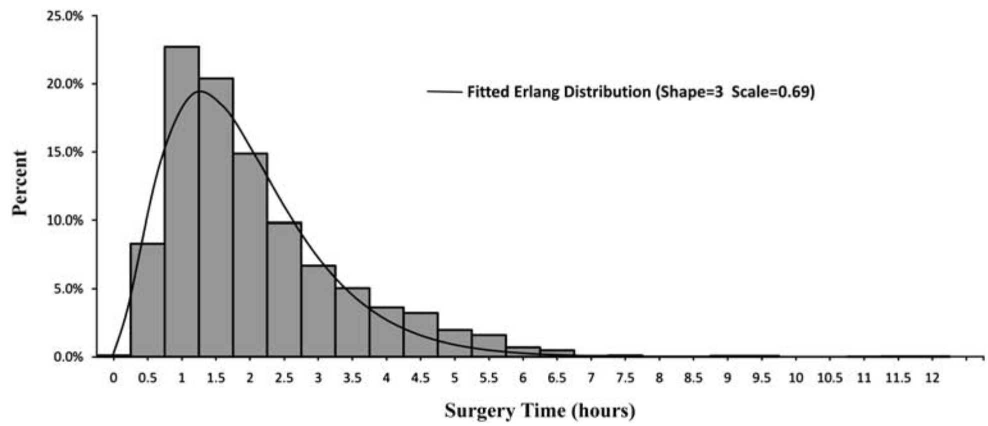

Figure 1 Fitted surgery time—Erlang distribution.

Table 1 Volume and survivability time by priority class-base arrival rate scenarios

Priority	*Survivability (min)*	*Default (Patients/ day)*	*Scenario 1 (Patients/ day)*	*Scenario 2 (Patients/ day)*	*Scenario 3 (Patients/ day)*	*Scenario 4 (Patients/ day)*	*Scenario 5 (Patients/ day)*	*Scenario 6 (Patients/ day)*
1	60	0.224	0.551	1.142	0.403	0.403	0.403	0.403
2	120	0.443	0.551	0.403	1.142	0.403	0.403	0.403
3	240	1.142	0.551	0.403	0.403	1.142	0.403	0.403
4	480	0.641	0.551	0.403	0.403	0.403	1.142	0.403
5	1440	0.324	0.551	0.403	0.403	0.403	0.403	1.142

Table 2 Average surgery time scenarios by priority class

Priority	*Arrivals (Patients/ day)*	*Default surgery time (min)*	*Surgery time P1 high (min)*	*Surgery time P2 high (min)*	*Surgery time P3 high (min)*	*Surgery time P4 high (min)*	*Surgery time P5 high (min)*
1	0.551	124.2	186.3	108.675	108.675	108.675	108.675
2	0.551	124.2	108.675	186.3	108.675	108.675	108.675
3	0.551	124.2	108.675	108.675	186.3	108.675	108.675
4	0.551	124.2	108.675	108.675	108.675	186.3	108.675
5	0.551	124.2	108.675	108.675	108.675	108.675	186.3

parameters for the default scenario based on the original hospital data.

To determine the impact of patient mix, that is, the volume or arrival rates of the individual priorities, as well as the overall patient arrivals, various scenarios were run with modified individual and total arrivals. Total patient volume or arrival rate, denoted as λ_T, was varied from one-quarter to triple the default level of 2.77 patients per day at levels of one-quarter, one-half, double and triple. As well, different patient volumes by priority within these total patient volumes were simulated. These included equal arrival rates for each priority as well as an extreme volume for an individual priority, with the arrivals of the remaining priorities taken as equivalent. For this second approach, each priority in sequence was assigned the highest arrival rate from the data (eg, the highest default arrival rate was for Priority 3 at 1.142 patients per day), with the remaining volume equally distributed over the other priorities. These scenarios are labelled as Scenarios 1 through 6 in Table 1.

To evaluate the effects of surgery duration on survivability, we ran two sets of simulations. First, we varied the average surgery time for each priority individually to determine the impact of varying the surgery times across priorities. Specifically, the average surgery time was increased by 50% for one of the priorities in sequence, with all other priorities set at an equal though reduced level to maintain the same overall average surgery time as the default scenario. All arrivals were taken as equal, with total patient volume identical to the default level. The parameter values for these simulations are given in Table 2. In the second set of simulations related to surgery time, which were motivated by the results obtained from the first round (discussed in detail in the next section), the average surgery time was reduced by a total of 60 min in 10-min increments, as well as increased by increments of 25% up to three times the current average. In this second set of simulations, the surgery time was not varied by priority, that is, the same average was applied to all priorities. Each of these surgery times was then simulated across all arrival rates listed above, as well as the default patient volume. To measure the impact of the patient arrivals and surgery times, we determined the percentage of patients of each priority that exceeded their stated survivability in minutes. For the initial analysis a single server or OR was considered available. Subsequent analysis then increased the number of OR's to determine the required OR's to meet specific operating goals.

Simulation results

The first observation arising from the simulation studies was the lack of a significant impact of the patient volume mix by priority. Figure 2 shows the percentage of patients for each priority that exceeded their survivability times for each of the scenarios presented in Table 1 using the original surgery time distribution. On the basis of the figure, we observed that the general trend was relatively flat across the different scenarios, thus indicating that priority volume mix was not significant in evaluating the emergency OR requirements. ANOVA results comparing the mean percentage exceeding survivability across the different scenarios for each priority is presented in Table 3. With the exception of Priority 5 patients, which could not be analyzed, all p-values exceeded 0.1, thus we were unable to conclude there was a difference in the means, indicating that patient mix was not a critical variable in determining the percentage exceeding survivability time for a given priority.

The second key observation arising from the studies involving surgery time variations was that as long as the overall weighted average of the surgery times for all patient priorities was the same, the survivability results across scenarios did not differ significantly.[1] Figure 3 shows the percentage of patients for each priority that exceeded their survivability times for each of the scenarios presented in Table 2 using the original total patient volume. The general trend was relatively flat across the different scenarios, thus indicating that average surgery time by priority was not significant in evaluating the emergency OR system or requirements.

[1]The findings might not hold if the surgery distribution did not fit a γ-type distribution (of which the Erlang is a special case), in which the distribution averages are additive.

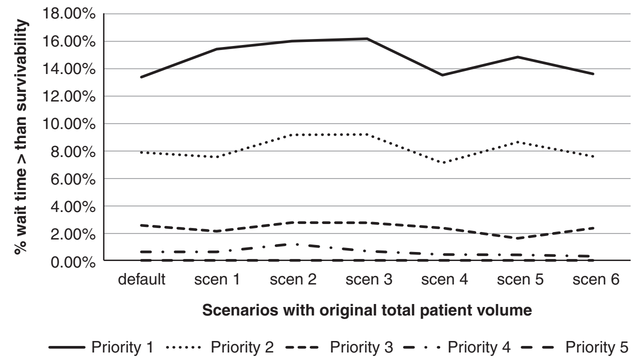

Figure 2 Percentage over survivability across total patient volumes.

Table 3 Welch (ANOVA) test of equality of mean percentage over survivability

	Welch test statistic*	*df1*	*df2*	*Significance*
Priority 1	1.648	6	306.657	0.134
Priority 2	1.634	6	306.829	0.137
Priority 3	1.043	6	306.185	0.397
Priority 4	1.515	6	303.879	0.173
Priority 5†	—	—	—	—

*Welch test used as p-values for Levene tests of equal variances were all <0.000

†Could not be tested as one or more of the groups had 0 variance.

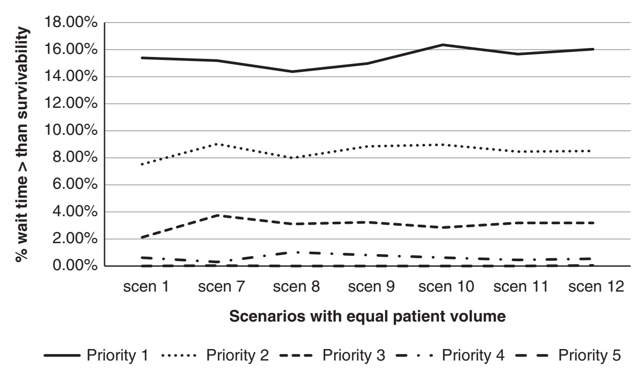

Figure 3 Percentage over survivability with surgery times varying by priority.

Given the above results, the problem becomes somewhat simplified, though still complex, in that it is not necessary to distinguish different patient mix volumes or average surgery times by priority, and can determine surgical OR requirements based on total patient volume and an overall average surgery time distribution where all priority classes are taken as the same. It should be noted that once the required number of ORs have been determined, we assume that the corresponding number of OR teams are available to perform the procedures. Before proceeding, we first needed to establish a threshold for the percentage of patients for all priorities that exceeded their stated survivability times. For the simulation scenarios studied, we found that using 5% as a threshold, we were able to achieve a more than 90% compliance rate for all surgeries together. We used this as a metric for all our analysis presented in the subsequent sections as this was found to be appropriate based on prior literature (Litvak *et al*, 2001).

We thus estimate the number of ORs required to meet a 5% survivability threshold with total patient volume and average surgery time as predictor variables. To determine the optimal number of ORs, for each combination of arrival rate and surgery times described previously, the number of OR's assumed available was increased until the proportion for all patient priorities met the 5% threshold at a 95% confidence interval. The rationale for including the confidence level in the determination of the required number of ORs results from the understanding that planning strictly on the basis of average is not always reliable (Albright *et al*, 2009). This is illustrated in Figure 4, where the mean as well as the upper and lower confidence levels of the percentage of patients exceeding their respective survivability limits for each priority are given for a particular scenario. As can be observed, though the mean for Priority 1 is less than the 5% threshold, the upper 95% confidence level exceeds the required limit. An ordinal probit regression was used to determine the functional relationship between the dependent variable, the number of ORs required, and the independent variables total patient volume and average surgery time. The parameter values for the ordered probit model were estimated using STATA 11.0, with all coefficient p-values <0.05. The resulting model is given below.

$$P(OR = 1) = \Phi\,(13.181 - [3.195x_1 + 0.0809x_2]) \qquad (1)$$

$$P(OR = 2) = \Phi\,(33.878 - [3.195x_1 + 0.0809x_2]) - \Phi\,(13.181 - [3.195x_1 + 0.0809x_2]) \qquad (2)$$

$$P(OR = 3) = \Phi\,(42.251 - [3.195x_1 + 0.0809x_2]) - \Phi\,(33.878 - [3.195x_1 + 0.0809x_2]) \qquad (3)$$

$$P(OR = 4) = \Phi\,(49.797 - [3.195x_1 + 0.0809x_2]) - \Phi\,(42.251 - [3.195x_1 + 0.0809x_2]) \qquad (4)$$

$$P(OR = 5) = \Phi\,(\infty - [3.195x_1 + 0.0809x_2]) - \Phi\,(49.797 - [3.195x_1 + 0.0809x_2]) \qquad (5)$$

where x_1 = total daily patient volume and x_2 = average surgery time (min) and $\phi(\cdot)$ is the standard normal distribution.

Equations (1)–(5) estimate the cumulative probability that the required number of ORs is equal to a specific value, in our Case 1 through 5. The probability of each

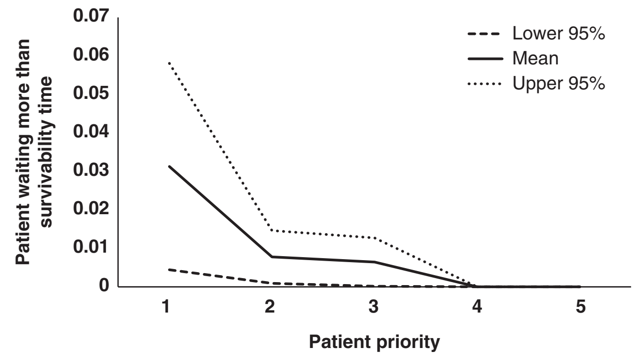

Figure 4 The 95% confidence interval for percentage over survivability ($\lambda_T = 0.69$; $E(B) = 124.2$ min).

discrete outcome (number of ORs) is calculated based on the standard normal probability distribution using the value of the derived expression for the independent variables total patient volume and average surgery time. One interpretation is that the number of required ORs is equal to the standard normal probability of the magnitude of the derived expression based on the total patient volume and average surgery time. For example, applying the equations using the original hospital data, where $x_1 = 2.77$ patients per day and $x_2 = 124.2$ min, we find the number of required ORs is two with a probability approaching 1, while the probabilities for all the other possibilities are approximately 0. As expected, the lower the numerical value of this expression, as a result of either low average surgery times or patient arrivals or both, the higher the likelihood that the required number of ORs is low.

Once the required number of ORs was determined based on the equations above, we then estimated the relationship between the proportion of patients in each priority that waited more than their survivability time and a set of independent variables. We used the GLM option with link (logit) and the binomial distribution for this analysis as the dependent variable was a proportion. We found that for the optimal number of ORs, average surgery time and total patient volume were significant in influencing the probability of a patient exceeding their survivability time. The results are provided below for all the five patient priorities, with all coefficients found to be statistically significant at the 0.05 level.

$$p_1 = \frac{e^{-4.091+0.285x_1+0.007x_2-1.046x_3}}{1+e^{-4.091+0.285x_1+0.007x_2-1.046x_3}} \qquad (6)$$

$$p_2 = \frac{e^{-5.256+0.226x_1+0.009x_2-1.0328x_3}}{1+e^{-5.256+0.226x_1+0.009x_2-1.0328x_3}} \qquad (7)$$

$$p_3 = \frac{e^{-7.692+0.256x_1+0.0140x_2-1.120x_3}}{1+e^{-7.692+0.256x_1+0.0140x_2-1.120x_3}} \qquad (8)$$

Table 4 Comparison of percentage of patients exceeding survivability time (default scenario)

		Parameters		*Number of ORs = 2 = Optimal*	
		λ_T (Patients/ day)	*E(B) (min)*	*Parametric model (%)*	*Simulation results (%)*
Patient priority	1	2.77	124.2	1.07	0.99
	2	2.77	124.2	0.38	0.25
	3	2.77	124.2	0.06	0.00
	4	2.77	124.2	0.00	0.00
	5	2.77	124.2	0.00	0.00

$$p_4 = \frac{e^{-14.036+0.933x_1+0.0316x_2-2.284x_3}}{1+e^{-14.036+0.933x_1+0.0316x_2-2.284x_3}} \quad (9)$$

$$p_5 = 0$$

where p_i = proportion of patients belonging to priority 'i' that waited more than their survivability time when the number of OR is optimal, with x_3 = optimal number of ORs, and x_1 = total daily patient volume and x_2 = average surgery time (min) as given previously.

To demonstrate the validity of the parametric models, in Table 4 we present a comparison of the results on survivability data at the hospital considered in this study (default problem scenario) when the number of operational ORs was equal to two, the optimal value established using the parametric models given in Equations (1)–(5). These results show that the developed parametric models are quite good at predicting the proportion of patients for each priority exceeding their respective survivability limits as compared with the simulation results, with the parametric results all within the 95% confidence interval of the simulated percentages. Similar equations can be estimated for other threshold values as determined to be appropriate to the hospital by the OR manager. For example, the hospital may be willing to accept a 10% threshold as opposed to the 5% used above. Regardless of the threshold used, the equations should provide a robust estimate of the number of required OR's as well as determine the proportion of patients exceeding the OR department's service goals.

Transform development and application

An alternative to the models developed above is the use of transforms and the resulting mathematical moments (eg, mean and variance) to evaluate the emergent surgery queueing system. An obvious advantage of this approach is that analytical results are easier to generalize than those from simulation models, including being readily extended to any number of priority systems. However, it is limited to settings where a single OR is optimal or is otherwise dedicated to emergent surgeries due to the complexity of the equations.

The most relevant closed form expressions available in the queueing literature for determining system properties involve Laplace–Steiltjes (L–S) transforms. However, even for an M/G/1 queue, that is, single server with Poisson arrivals and general service distribution, only expressions for expected wait time are available for multi-priority queues. We extend the existing results for M/G/1 queues to develop approximations to L–S transforms to aid in the determination of general properties of multi-priority queues as studied in this paper. The resulting transform for the wait time for priority patient 'i' is:

$$\tilde{W}_i(s) = \frac{\left(1-\sum_{j=1}^{r}\rho_j\right)^2}{(1-(\rho_1+\rho_2+\cdots\rho_i))(1-(\rho_1+\rho_2+\cdots\rho_{i-1}))\left(1-\sum_{j=1}^{r}\rho_j\tilde{R}(s)\right)} \quad (10)$$

where

- i patient priority class, $i = 1$ to r
- ρ_i server utilization due to Priority i patient $= \lambda_i E(B)$
- λ_i patient arrival rate for patients belonging to Priority i
- $E(B_i)$ expected service (surgery) time for Priority i
- $\tilde{R}_s(s)$ residual service time transform $= (1-\tilde{B}_i(s))/sE(B_i)$ for Priority i
- $\tilde{B}_i(s)$ service time (surgery time) transform for a Priority i patient $= (\mu/(\mu+s))^k$ for the Erlang distribution with scale parameter k
- $E(L_i^q)$ the expected number of Priority i patients in the queue
- $\widetilde{L_i^q}(s)$ transform for number waiting in queue that belong to priority class $i = \lambda_i\tilde{W}_i(s)$

Using the general principles of L–S transforms, the individual moments for wait-time can be derived. Thus, using the wait-time transform in (10), we can find the nth moment, $E(W_i^n)$, as: $\tilde{W}_i^n(0) = (-1)^n \cdot E(W_i^n)$, where $\tilde{W}_i^n$ is the nth derivative with respect to 's' for the above wait-time transform. For example, the mean or expectation of wait-time, $E(W_i)$, is derived by taking the first derivative of $\tilde{W}_i(s)$, then setting $s = 0$, resulting in:

$$E(W_i) = \frac{\sum_{j=1}^{r}\rho_j \cdot E(R_j)}{(1-(\rho_1+\rho_2+\cdots\rho_i))(1-(\rho_1+\rho_2+\cdots\rho_{i-1}))} \quad (11)$$

with $E(R_j) = (E(B_j^2)/2 \cdot E(B_j))$ and the first and second moments of the surgery time distribution based on the Erlang distribution with a scale parameter of three as described previously. Health care managers can use these transforms and resulting expectations in OR planning for all situations in which a single OR is being utilized.

Table 5 Comparison of mean wait times in minutes from simulations and transforms by priority

Parameters		Priority 1		Priority 2		Priority 3		Priority 4		Priority 5	
λ_T* (patients/ day)	$E(B)$ (min)	Simulation Average	Trans $E(W_1)$	Simulation Average	Trans $E(W_2)$	Simulation Average	Trans $E(W_3)$	Simulation Average	Trans $E(W_4)$	Simulation Average	Trans $E(W_5)$
0.69	124.2	5.61	4.98	4.46	5.05	7.77	5.23	5.76	5.44	6.36	5.56
0.69	124.2	6.15	4.98	6.07	5.10	5.63	5.22	5.36	5.35	5.60	5.49
0.69	124.2	5.27	5.04	7.09	5.22	4.26	5.31	3.02	5.41	3.96	5.51
0.69	124.2	4.85	4.96	5.74	5.13	4.46	5.31	6.80	5.41	9.10	5.51
0.69	124.2	5.02	4.96	4.38	5.05	5.08	5.22	6.00	5.41	6.94	5.51
0.69	124.2	5.75	4.96	5.22	5.05	8.66	5.14	5.58	5.32	5.46	5.51
0.69	124.2	5.74	4.96	6.32	5.05	6.50	5.14	5.56	5.23	7.16	5.42
0.69	64.2	1.42	1.33	2.89	1.34	2.14	1.36	3.52	1.39	1.74	1.4
0.69	94.2	2.88	2.86	2.39	2.89	4.13	2.97	4.73	3.06	4.87	3.11
0.69	114.2	3.35	4.2	7.82	4.26	4.66	4.4	5.69	4.56	5.68	4.66
1.39	64.2	4.62	2.66	3.26	2.7	4.36	2.8	3.94	2.92	4.46	2.98
2.77	64.2	6.75	5.35	7.61	5.51	8.50	5.93	8.77	6.46	8.61	6.78

*Total patient arrival rate. The proportion for priority was taken as identical to the default arrival rates (see Table 1).

Table 6 Comparison of standard deviations of wait times (min) from simulations and transforms by priority

Parameters		Priority 1		Priority 2		Priority 3		Priority 4		Priority 5	
λ_T* (Patients/ day)	$E(B)$ (min)	Simulation std dev	Trans $\sigma(W_1)$	Simulation std dev	Trans $\sigma(W_2)$	Simulation std dev	Trans $\sigma(W_3)$	Simulation std dev	Trans $\sigma(W_4)$	Simulation std dev	Trans $\sigma(W_5)$
0.69	124.2	27.6	26.7	30.4	26.9	38.0	27.4	27.8	27.9	33.4	28.2
0.69	124.2	28.0	26.7	26.1	27.0	25.2	27.4	28.1	27.7	25.3	28.0
0.69	124.2	24.0	26.9	32.7	27.3	18.9	27.6	28.0	27.8	24.7	28.1
0.69	124.2	24.0	26.7	25.6	27.1	28.2	27.6	28.2	27.8	40.4	28.1
0.69	124.2	29.4	26.7	23.9	26.9	24.4	27.4	31.1	27.8	32.0	28.1
0.69	124.2	22.5	26.7	24.3	26.9	28.1	27.1	24.7	27.6	26.0	28.1
0.69	124.2	29.2	26.7	27.9	26.9	31.0	27.1	28.0	27.4	35.0	27.8
0.69	64.2	4.9	9.8	13.6	9.9	9.0	9.9	13.4	10.0	10.4	10.1
0.69	94.2	18.0	17.6	13.9	17.6	17.5	17.9	20.2	18.1	16.8	18.3
0.69	114.2	18.9	23.5	26.9	23.7	21.9	24.0	28.0	24.5	26.4	24.7
1.39	64.2	16.6	14.1	12.9	14.2	16.2	14.4	17.0	14.7	17.9	14.9
2.77	64.2	18.9	20.4	22.0	20.7	24.8	21.4	27.4	22.3	26.9	22.8

*Total patient arrival rate. The proportion for priority was taken as identical to the default arrival rates (see Table 1).

We confirm the validity of these transforms by comparing the average or expected wait time and standard deviation in minutes using the transforms above with those of the simulation results for various arrival and surgery time conditions. These results are presented in Tables 5 and 6, respectively, with both the average and standard deviations generally similar to those generated by the simulations, with all analytic expected values within the 95% confidence interval of the simulation results.

In addition to estimating key operational parameters such as mean wait time, the transforms can be used to estimate the probability or proportion of patients who wait longer than their respective survivability times $P(W_i > \omega_i)$. The most direct approach would involve inverting the wait-time transform to determine the associated probability distribution. Unfortunately, the above transform cannot be inverted directly. However, we may approximate the wait-time distribution by decomposing the wait-time components. We specifically focus on Priority 1 patients, as from the parametric Equations (6)–(9), if these patients do not exceed their survivability time, the other patients are also within their respective limits. For Priority 1 patients, the wait time is composed of three components; the proportion of time the system is free, the residual service time of any patient currently in service, and the service time for any Priority 1 patients in the queue. The first component is straightforward, and is given by $1-\sum_{j=1}^{r}\rho_j$. The residual service distribution, along with the probability that a patient is in service, $\sum_{j=1}^{r}\rho_j$, gives the second component. For an Erlang distribution with shape parameter 'k', the probability distribution resulting from inverting the transform is:

$$f_R(t) = \frac{1}{k} e^{-\mu t} \sum_{n=0}^{k-1} \frac{\mu^{k-n} t^{k-n-1}}{(k-n-1)!} \equiv \frac{1}{k} \sum_{K=1}^{k} E_k \quad (12)$$

Thus, the residual distribution is the average of the sum of Erlang distributions with shape parameters from 1 to k. Therefore, the resulting cumulative distribution function for the residual, denoted as $F_R(t)$, is:

$$F_R(t) = P(R \leqslant t) = \frac{1}{k} \left(\sum_{K=1}^{k} F_K(t) \right) \quad (13)$$

where $F_K(t)$ represents the cumulative Erlang distribution with shape 'k'. Using just the first two components as an approximation results in:

$$\begin{aligned} P(W_1 > \omega_1) &= 1 - P(W_1 \leqslant \omega_1) \\ &= 1 - \left[\left(1 - \sum_{j=1}^{r} \rho_j \right) + \sum_{j=1}^{r} \rho_j F_R(\omega_1) \right] \end{aligned} \quad (14)$$

Direct determination of the third component is not possible, but under low arrival/utilization rates, would likely be modest in that most of the time the server (OR) is not facing lengthy queues in order to meet reasonable thresholds for patients exceeding their respective survivability times. Thus, there is a high probability that there would be no Priority 1 patients in the queue, with the corresponding 0 contribution to the wait time for an arriving patient. Table 7 shows a comparison of the estimated probability of exceeding the wait time obtained from the simulations and the approximate cumulative distribution. As would be expected, the calculations based on the transforms perform better with lower overall arrival and utilization rates, as the third component is likely to be less significant under those scenarios. Overall, the results compare quite well with those obtained from the simulation, and for higher utilizations indicate where excessive risk or the probability of exceeding survivability limits occurs.

Table 7 Comparison of simulations and transforms for proportion exceeding survivability

Parameters			$P(W_1 > \omega_1)$	
λ_1 *(Patients/day)*	λ_T *(Patients/day)*	$E(B)$ *(min)*	*Simulation*	*Trans*
0.056	0.69	124.2	0.031	0.032
0.224	2.77	64.2	0.038	0.031
0.112	1.39	124.2	0.074	0.065
0.224*	2.77	124.2	0.134	0.13
0.448	5.55	124.2	0.341	0.260
0.112	1.39	186.3	0.125	0.123
0.224	2.77	186.3	0.294	0.246
0.112	1.39	248.4	0.180	0.182
0.224	2.77	248.4	0.403	0.36

*Default scenario.

It should be noted that the above transforms and resulting expected wait times are only an approximation as they do not account for the priority transition that can occur in an OR setting. For the highest priority customers, the expected wait time will always be greater with than without transitions in the queue, while for the lowest priority patients 'r', wait times will be strictly lower with transition. However, the impact on the wait times for intermediate priorities is dependent on the physical characteristics of the system, such as the individual arrival rates and transition probabilities. Details of this analysis are presented in Appendix A. Numerical estimation of these transition impacts have indicated that these effects are relatively small (typically $<1\%$) and therefore, the wait time approximation is representative of the system of interest.

Techniques for apportioning cost

Having developed the relationship between the required number of ORs and the significant variables influencing this decision, we present techniques for equitably apportioning the costs to ensure the economic sustainability of dedicated ORs for emergent surgery. We distinguish two cases; the first is where a single OR is sufficient to meet the required threshold for all patient priorities, and the second deals with situations requiring multiple ORs.

The need for a pricing mechanism is largely driven by the requirement for an equitable as well as defensible method of allocating costs across the different patients/priorities. As discussed in the introduction, patient outcomes and satisfaction can be improved by separating emergent surgeries by dedicating OR facilities. This is likely in part due to minimizing the disruptions in surgery schedules caused by emergent surgeries. However, it has also been found that dedicated facilities may not be the most cost-effective. Therefore, to justify maintaining such facilities, and the patient benefits that arise, we need to effectively and fairly allocate the costs. This in turn requires a change from the standard fixed/variable cost structure currently applied to these surgeries. When considering the allocation of costs, it would seem obvious that, to a large extent, higher priorities drive the need for dedicated facilities. That is, lower priorities, as defined by their survivability or acceptable wait-times, would cause less disruption in the overall surgery schedule given the ability to delay these surgeries. This would in turn provide an argument or justification for allocating a larger proportion of the costs to the higher priority patients.

Pricing for a single OR

For a single OR, an approach based on the transforms and expected wait times discussed above can be used to apportion the costs among the different priorities. The rationale is based on the fact that those that wait the least derive a larger benefit from the dedicated OR facilities, and this can therefore be used to apportion costs. We first define P_i as the price charged for surgery for a patient belonging to priority 'i' and C_i the proportion of the total surgery costs allocated to these patients. Then, the expectations of wait time can be used as follows:

$$C_i = \frac{C_T \cdot R_i \cdot \lambda_i}{\sum_{i=1}^{r} R_i \cdot \lambda_i} \quad \text{and} \quad P_i = \frac{C_T \cdot R_i}{\sum_{i=1}^{r} R_i \cdot \lambda_i} = \frac{C_i}{\lambda_i} \qquad (15)$$

where

P_i price charged for surgery for a patient belonging to Priority i,
C_i proportion of total costs charged to Priority i,
C_T $C_f + C_V =$ total cost of operating the OR, with
C_f fixed cost and $C_v =$ variable cost,

$$R_i = 1 - \frac{E(W_i)}{\sum_{i=1}^{r} E(W_i)} \qquad (16)$$

λ_I average daily OR patient arrivals for Priority i.

Here, R_i expresses the ratio of the expected wait times for each priority, with $R_1 > R_2 > \cdots > R_r$. The ratio of expected wait time is then applied in Equation (15) to allocate the total fixed and variable costs (C_T) with maintaining the emergent surgery OR for each priority (C_i). Then, based on arrival rates, the price per patient in each priority (P_i) is determined. This ensures that the total costs are recovered and are equitably distributed based on priority. The rationale for this pricing policy is based on the argument that patients that wait longer on average (ie, lower priorities) should get charged less, as the primary reason for the longer wait times is due to the higher priority patients. Further, as the discrepancy in wait times diminish, given sufficient available capacity, the price charged per patient across priorities converges to the same rate. Thus, any difference in pricing across priorities decreases, going to 0, as the advantage to higher priorities based on expected wait time diminishes.

Pricing for multiple ORs

When the optimal number of ORs is higher than one, the approach presented above could lead to an unfair allocation of costs among patients of different severity. For instance, consider a scenario in which the lowest severity patients are already below the 5% threshold even when the number of ORs is less than optimal. For example, at the study participating hospital, we found that optimal number of ORs was two, but the percentage of patients belonging to Priority 5 was below the 5% threshold even with a single OR. Therefore, the cost of the second OR should be allocated to those driving the need for it, in this case, patients belonging to Priorities 1 through 4. One method of accomplishing a fair allocation of costs in this situation is to compare the proportion of patients for each priority whose wait time exceeded their survivability with a single OR, or when the system has a stable utilization (discussed below), with that when the number of ORs is optimal. For example, Table 7 shows the percentage of patients for each priority that waited more than their respective survivability time at the arrival rate and surgery time of the participating hospital with a single OR as well as when the number of ORs was optimal, that is, where all percentages were below the 5% threshold. The net change is then used to allocate costs as described in detail below.

Before implementation, a second consideration must be taken into account, that of system stability. When the utilization exceeds or is near saturation (ie, 100%), the system performance metrics are not reliable. Figure 5 shows the percentage of exceeding survivability for each priority for different ORs available at a total patient volume of 8.32 arrivals per day and an average surgery time of 376.2 min. For utilization exceeding saturation, all priorities approach 100% exceeding their survivability time. Therefore, for apportioning costs, we compare improvement achieved when OR = optimal with OR = 1 or closest stable system for the patient mix and total volume under consideration.

The functional expressions obtained for the percentage of patients waiting beyond their survivability time when the number of ORs = 1 (based on simulation scenarios that had OR = 1 that resulted in a stable system) are as follows, with all factors were significant at the 0.05 level. In addition, we also present a com-

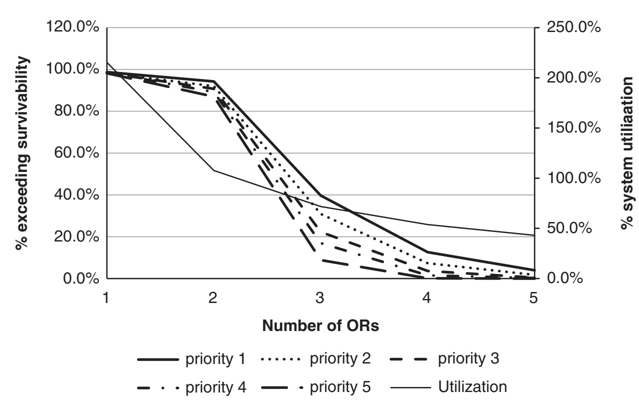

Figure 5 OR utilization and percentage exceeding survivability as f(#ORs).

Table 8 Comparison of percentage of patients exceeding survivability time (default scenario and single OR)

		Parameters		OR=1	
		λ_T (patients/day)	E(B) (min)	Parametric model (%)	Simulation results (%)
Patient priority	1	2.77	124.2	11.35	13.35
	2	2.77	124.2	5.96	7.86
	3	2.77	124.2	2.31	2.55
	4	2.77	124.2	0.59	0.62
	5	2.77	124.2	0.04	0.00

parison of the predicted values to those generated by the simulations in Table 8.

$$p_1^1 = \frac{e^{-4.378+0.435x_1+0.009x_2}}{1+e^{-4.378+0.435x_1+0.009x_2}} \quad (17)$$

$$p_2^1 = \frac{e^{-5.514+0.457x_1+0.012x_2}}{1+e^{-5.514+0.457x_1+0.012x_2}} \quad (18)$$

$$p_3^1 = \frac{e^{-7.431+0.577x_1+0.0168x_2}}{1+e^{-7.431+0.577x_1+0.0168x_2}} \quad (19)$$

$$p_4^1 = \frac{e^{-9.696+0.752x_1+0.020x_2}}{1+e^{-9.696+0.752x_1+0.020x_2}} \quad (20)$$

$$p_5^1 = \frac{e^{-14.059+1.082x_1+0.027x_2}}{1+e^{-14.059+1.082x_1+0.027x_2}} \quad (21)$$

where

p_i^1 proportion of patients of Priority 'i' that weird more than their survivability time when $OR=1$.

Taking all the above factors into account, we propose the algorithm provided in Appendix B. It is important to note, as above for a single OR, that the prices determined by the cost apportioning algorithm would result in a scenario where the cost and revenue break even. The hospital can modify this algorithm easily to accommodate their financial goals.

We applied the algorithm to determine the price to be charged for each priority class at the study participating hospital. We used the data on costs provided to us by the hospital and simulation results for optimal number of ORs of two and percentage over survivability as needed in the algorithm for this purpose. On the basis of data provided by the hospital, the total daily fixed cost for operating two ORs was US\$52 854.99 and total daily variable cost was \$24 614.22. This fixed cost also includes the personnel cost when reserving a certain number of ORs. Thus, the total cost of operations was \$77 469.21. After application of the algorithm, the prices to be charged per priority were determined, with the results presented in Table 9. As can be observed from the table, the highest cost is apportioned to the highest priority, which is intuitively correct as they are most directly responsible for driving the requirement for additional ORs.

Table 9 Costs for surgery by priority

Priority	Cost per patient (P_i)	Cost per priority (C_i)
1	\$34 933	\$7825
2	\$32 010	\$14 181
3	\$29 023	\$33 144
4	\$27 888	\$17 876
5	\$13 762	\$4459

Conclusions

In this study, we developed models for determining the optimal number of ORs that need to be reserved for emergent surgeries to meet specified threshold service levels. Specifically, we developed generic expressions that can be applied at hospitals regardless of size. We also developed transforms that could be used for problem scenarios wherein the number of ORs is equal to one. In addition, we developed pricing rules and algorithms required to maintain dedicated emergent ORs and to fairly apportion costs among the priority classes for resources reserved based on these models. We demonstrated the applicability of the algorithm via a case study on the data provided by the study participating hospital. The pricing mechanism was important for ensuring the financial viability of results proposed by our simulation and queueing models.

In future extensions of this research effort, we plan to study the sensitivity of various apportioning techniques to different problem settings. In the current study, due to data limitations, we assumed that all surgeries followed the same distribution. This might not hold true for all hospitals. This forms one of the motivations for future studies. We also aim to look at the sensitivity of the OR planning policies to different classification systems discussed in the healthcare literature in addition the five-level system used in this research endeavour.

References

Albright SC, Winston WL and Zappe CJ (2009). *Data Analysis and Decision Making, 4/e*. South-Western Cengage Learning: Mason, OH.

Blake JT, Dexter F and Donald J (2002). Operating rooms managers' use of integer programming for assigning block time to surgical groups: A case study. *Anesthesia & Analgesia* **94**(1): 143–148.

Cardoen B, Demeulemeester E and Beliën J (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research* **201**(3): 921–932.

Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates: Hillsdale, NJ.

Denton B, Rahman AS, Nelson H and Bailey AC (2006). Simulation of a multiple operating room surgical suite. In: Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM and Fujimoto RM (eds). *Proceedings of the 2006 Winter Simulation Conference*, Monterey, pp 414–424.

Denton B, Viapiano J and Vogl A (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science* **10**(1): 13–24.

Gerchak Y, Gupta D and Henig M (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science* **42**(3): 321–334.

Goldman J and Knappenberger H (1968). How to determine the optimum number of operating rooms. *Modern Hospital* **111**(3): 114–116.

Guerriero F and Guido R (2011). Operational research in the management of the operating theatre: A survey. *Health Care Management Science* **14**(1): 89–114.

HFMA (2003). Achieving operating room efficiency through process integration. *Health Care Financial Management* **57**(3): S1–S7.

HFMA (2004). Strategic price setting: ensuring your financial viability through price modeling. *Health Care Financial Management* **58**(7): 1–7.

Jebali A, Hadj Alouane AB and Ladet P (2006). Operating rooms scheduling. *International Journal of Production Economics* **99**(1–2): 52–62.

Lamiri ML, Xie D and Zhang S (2008). Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions* **40**(9): 838–852.

Levtzion-Korach O, Murphy KG, Madden S and Dempsey C (2010). For urgent and emergent cases, which one goes to the OR first? *OR Manager* **26**(7): 1,11–13.

Litvak E and Long MC (2000). Cost and quality under managed care: Irreconcilable differences? *American Journal of Managed Care* **6**(3): 305–312.

Litvak E, Long MC, Cooper AB and McManus ML (2001). Emergency department diversion: causes and solutions. *Academic Emergency Medicine* **8**(11): 1108–1110.

Ozkarahan I (2000). Allocation of surgeries to operating rooms by goal programming. *Journal of Medical Systems* **24**(6): 339–378.

Pai J-Y, Petry F and Fos P (1997). Employing simulation to optimize the number of operating rooms in hospitals. Technical Report. In: *Proceedings of the Medical Sciences Simulation Conference*. The Society for Computer Simulation International (SCS): Phoenix, pp 36–42.

Persson MJ and Persson JA (2010). Analysing management policies for operating room planning using simulation. *Health Care Management Science* **13**(2): 182–191.

ProModel (2010). Software website, http://www.promodel.com, accessed May 2010.

Sier D, Tobin P and McGurk C (1997). Scheduling surgical procedures. *Journal of the Operational Research Society* **48**(9): 884–891.

Stanciu A, Vargas L and May J (2010). A revenue management approach for managing operating room capacity. In: Johansson B, Jain S, Montoya-Torres J, Hugan J, and Yücesan E (eds). *Proceedings of the 2010 Winter Simulation Conference*. Baltimore, pp 2444–2454.

Ross SM (2003). *Introduction to Probability Models, 8/e*. Academic Press: San Diego, CA.

Tucker JB, Barone JE, Cecere J, Blabey RG and Rha CK (1999). Using queueing theory to determine operating room staffing needs. *Journal of Trauma* **46**(1): 71–79.

van Oostrum JM, *et al* (2008). A simulation model for determining the optimal size of emergency teams on call in the operating room at night. *Anesthesia & Analgesia* **107**(5): 1655–1662.

Wullink G, Houdenhoven MV, Hans EW, Oostrum JM, van der Lans M and Kazemier G (2007). Closing emergency operating rooms improves efficiency. *Journal of Medical Systems* **31**(6): 543–546.

Zhang B, Murali P, Dessouky MM and Belson D (2009). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society* **60**(5): 663–673.

Zonderland ME, Boucherie RJ, Litvak N and Vleggeert-Lankamp C (2010). Planning and scheduling of semi-urgent surgeries. *Health Care Management Science* **13**(3): 256–267.

Appendix A

Effect of priority transition

For highest priority customers, the expected wait time without transitions is given as:

$$E(W_1) = E\left(L_1^q\right) \cdot E(B) + \sum_{j=1}^{r} \rho_j \cdot E(R)$$
$$= \lambda_1 \cdot E(B) \cdot E(W_1) + \sum_{j=1}^{r} \rho_j \cdot E(R)$$

that is, the number of Priority 1 patients in the queue and the residual time for patients in service. With transition, Priority 1 would also have to wait for all the lower priority classes that have already waited longer than the difference in survivability time between the priorities, given as $\sum_{j=2}^{r} \lambda_j \cdot E(W_1) \cdot E(B) \cdot P_{j,1}$, where $P_{m,n} = P(wait\ time\ priority > \omega_m \omega_n)$, that is, the probability that priority '*m*' has waited longer than the difference in wait time of priority '*n*', where $m > n$ which is strictly positive, thus the expected wait time for Priority 1 will be higher.

For priorities >1 without transition, we can decompose the expected wait time into two components: the amount of residual service time of a patient currently in service as well as all patients with the same or higher priority in the queue upon arrival. The second is all the higher priority patients arriving during the patients waiting time:

$$E(W_i) = \sum_{j=1}^{i} E\left(L_j^q\right) \cdot E(B) + \sum_{j=1}^{r} \rho_j \cdot E(R) + \sum_{j=1}^{i-1} \lambda_j \cdot E(B) \cdot E(W_i), \quad i > 1$$

Two additional terms account for the possibility of transition of priority resulting from excessive wait time. The first concerns the presence of lower priority patients that have waited longer than the difference between the survivability times. The second addresses the fact that arriving higher priority patients will only take priority for those priority '*i*' patients that have not waited longer than the difference in survivability times. Adding these to the above equation results in

$$E(W_i) = \sum_{j=1}^{i} E\left(L_j^q\right) \cdot E(B) + \sum_{j=i+1}^{r} \lambda_j \cdot E(W_i) \cdot E(B) \cdot P_{j,i} + \sum_{j=1}^{r} \rho_j \cdot E(R) + \sum_{j=1}^{i-1} \lambda_j \cdot E(B) \cdot E(W_i) \cdot (1 - P_{i,j})$$

The second term increases the expected wait time, while the addition of $(1-P_{i,j})$ to the last term decreases the expected wait time, with the net effect dependent upon system dynamics. The only exception is for the lowest priority, as the second term is zero given that there are no lower priority patients. Thus, the lowest priority will always see a decrease in their expected wait time.

Appendix B

Table B1 Pseudocode for cost apportioning algorithm

Cost Apportioning Algorithm
Input:
p_i^{opt} – {set storing values of proportion of patients of priority '*i*' waiting more than their survivability time when number of ORs is optimal}
p_i^{base} – {set storing values of proportion of patients of priority '*i*' waiting more than their survivability time when number of ORs = 1 or for the closest stable system}
C_T – {total cost of ORs, including fixed and variable costs}
λ_T – {total patient arrivals or volume per day}
OR_{opt} = {optimal number of ORs}

Output:
$P(i)$ – {Price charged for patient surgery for priority '*i*' }

Process:
if $OR_{opt} = 1$ then
  $P(i)$ = Apply rules provided in (10)
else
  For patient severity $i = 1$ to r
    $Z(i) = p_i^{base} - p_i^{opt}$
  count = 0
  For patient severity $i = 1$ to r
    if $Z(i) = 0$ then
      $C(i) = C_T/(\lambda_T * OR_{opt})$
      count = count + 1
    else
      $C(i) = \frac{C_T}{\lambda_T \cdot OR_{opt}}\left[1 + \left(OR_{opt} - 1\right)\left(1 + count * \frac{Z(i)}{\sum_i Z(i)}\right)\right]$
    end if
  $C_i' = 0$
  For patient severity $i = 1$ to r
    $C_i' = C_i' + C_i \cdot \lambda_i$
  For patient severity $i = 1$ to r
    $P(i) = ((C(i) * C_T)/C_i')$
end if
return P

Received July 2011;
accepted June 2012 after two revisions