A REVENUE MANAGEMENT APPROACH FOR
MANAGING OPERATING ROOM CAPACITY

Alia Stanciu

Bucknell University
School of Management
322 Taylor Hall
Lewisburg, PA, 17837, USA

Luis Vargas
Jerrold May

University of Pittsburgh
Joseph M. Katz Graduate School of Business
Mervis Hall
Pittsburgh, PA, 15260, USA

ABSTRACT

The advanced scheduling of patients for elective surgeries is challenging when the operating room capacity usage by these procedures is uncertain. We study the application of some revenue management concepts and techniques to operating rooms for several surgical procedures performed in a multi-tier reimbursement system. Our approach focuses on booking requests for elective procedures, under the assumption that each request uses a random amount of time. We create and use a modified version of Belobaba's well-known EMSRb algorithm (Belobaba 1989) to decide on near-optimal protection levels for various classes of patients. Under the random resource utilization assumption, we decide, for each planning horizon, how much time to reserve for satisfying the demand coming from each class of patients, based on the type of surgical procedure requested and the patient's reimbursement level.

1   INTRODUCTION

In order to cope with the increasing demand for healthcare, and in particular, for surgery, a surgical suite is required to efficiently and effectively balance a high utilization of the operating room (OR) with the need for maintaining or improving the quality of care. Two factors affect the efficient use of existing resources: the challenges imposed by the increasing demand for healthcare, and the variability in both demand and service time. In an OR setting, this translates in both an uncertain number of patients in need for surgery, as well as inherent variability in the surgery times across types of surgeries, surgeons and patients.

When managing demand for a relatively fixed and perishable capacity, revenue management (RM) techniques prove very successful. One of the pillars of revenue management objective is the optimal allocation of capacity to various demand classes, and two of the important questions that RM studies refer to how many requests to accept from discount customers and how much capacity to reserve for customers that are willing to pay full price in order to maximize revenues (profits) over a planning horizon. After its success in the airline industry in the 1970s, numerous efforts and research studies tried to adapt RM approaches to the needs of other industries, ranging from oil and gas pipelines to healthcare to made-to-order manufacturing (Phillips 2005).

In the OR setting, where available surgery time is relatively limited and perishable, one can apply capacity-based revenue management to better manage demand from the various patients. In this context, the specific managerial decision that we address is how much OR capacity should the scheduling department reserve or protect in advance, over the planning horizon, for various patient classes, where a class can be defined as a combination of patient reimbursement level and the type of surgery requested. Because of the inherent variability in the service rendered, each accepted and scheduled request for surgery uses up a va-

riable amount of time, which makes the patient scheduling problem not trivial. Thus, the main difference between the capacity allocation problem in airlines and healthcare is the resource utilization aspect: while a passenger requests only one seat, a patient scheduled for surgery would use a random amount of resource. The exact amount of time that is consumed is known with certainty only after the procedure is performed.

Stanciu (2009) and Vargas, May and Stanciu (2010) obtained optimal partitioned protection levels assuming both random demand for surgery services and random utilization of the resource. However, the optimization problem formulated in these papers becomes computationally intractable, due to the exponential growth in both the number of variables and in the number of constraints. Here, we provide a different methodological approach to this problem, by devising a heuristic algorithm to obtain near-optimal protection levels for various classes of patients. Through simulation-based optimization, based on several sets of surgical procedures, we compute the near-optimal expected revenues as well as approximate capacity allocations, and compare the results to the known corresponding optimal results (Stanciu 2009). Preliminary results show that the heuristic algorithm could perform very well in practice and be used as a viable alternative to the optimal mathematical model.

The results of the simulation model presented here incorporate the idea of accepting/postponing requests for service from several competing classes of patients that present fluctuating demands. The patient segmentation in various fare classes is based on the contractual revenue expected to be paid by each patient category. We can think of this segmentation in terms of patients' ability to pay, rather than their willingness to pay (Karaesmen and Nakshin 2007).

It is widely accepted that inefficient OR scheduling results in delays and cancellations of surgical or other procedures, with a negative impact on hospital and patients. The patients suffer because a delay or cancellation may be detrimental to their health condition with a negative effect on quality of life in general. From the hospital's perspective it may result in deferred or lost revenue (if the surgery is ultimately performed by another hospital or in another country altogether), loss of goodwill, over and/or underutilization costs, to mention just the most obvious ones. Traditionally, the primary focus of the healthcare operations literature, has been the reduction of the operating costs and the increase of the OR utilization. Since the survival and prosperity of the surgical suite in the long run also depends on the revenue it generates, we examine how to better manage the mix of patients that request elective surgeries, with the goal of increasing the expected revenue generated by the surgical department.

In this paper we are not concerned with the OR daily schedule or procedure sequencing, but rather with the booking decisions that lead to accepting or rejecting a particular request and assigning it OR time during the open booking period, if accepted. The surgical procedures use a variable amount of time and usually cannot be well confined within the allocated time slot, resulting in over- and under-utilization of both the OR's and doctors' time. This is also why we are looking at the time resource as a continuous variable, rather than discretizing it into the classical slots.

We label our simulation-based heuristic algorithm Expected Marginal Capacity Revenue for Operating Rooms (EMCR-OR). EMCR-OR can be used to implement an advanced dynamic booking policy for the OR scheduling personnel. The simulation-based protection levels provide an average optimality gap of only about 0.7%, thus making this approach quite attractive in practice.

The paper is organized as follows. After the literature review in section 2, we present in section 3 the analytical model for computing the protection levels under random resource utilization. Section 4 presents the description of the algorithm, followed by preliminary numerical results in sections 5, an extension of the heuristic in section 6, and final conclusions and remarks in section 7.

2   LITERATURE REVIEW

The research stream that opened the road for revenue management implementations in healthcare can be traced back to the 1990s. Chapman and Carmel (1992) used threshold curves to determine whether and when to apply discounts in order to increase the capacity utilization and revenue yield within Duke University's diet and fitness center. Gerchak, Gupta and Henig (1996) develop an advanced reservation plan-

ning policy for elective surgery patients when the operating room capacity is common for both elective and emergency surgeries. In a more recent article, Green et al. (2006) analyze the patient scheduling problem faced by an MRI diagnostic facility and identify threshold policies to manage patient demand and the capacity allocation (appointment scheduling and dynamic priority) by using a finite-horizon dynamic program (in contrast, our approach incorporates the stochasticity of the service time).

Olivares, Terwiesch and Cassorla (2008) analyzed the situation of OR time allocation to a single surgical procedure with random service time. Their paper provides a general structural model to estimate the overage and underage costs in a newsvendor setting, with an application in reserving OR time. Specifically, the decision on OR time allocation to a specific surgical case (emergency or elective) is analyzed from the perspective of the factors that influence demand, while providing insight into what the cost parameters are for the hospital under study. From past observations one can derive the overage/underage cost ratio which becomes the input for the decision of how much OR time to reserve for a particular surgical case. Gupta and Wang (2008) propose several heuristics to help clinics decide how to ration the available slots between walk-ins (same-day appointments) and regular patients (advance booking) who may have a preference for both the slot time (assumed to take up the same amount of time) and their primary care physician.

The classical approach when dealing with multiple-class patient scheduling takes the form of priority queues. Solution approaches are simulation and stochastic linear and multi-objective mathematical programming. Various decision support models for tactical decisions in the day-to-day hospital admission and scheduling for surgery have been proposed, like the studies made by Everett (2002), Lowery (1996), and Ivaldi, Tanfani and Testi (2003), among others. The simulation models are regularly used as an operational tool to balance hospital availability and patients' needs while comparing the effectiveness of different alternative policies in this multi-criteria decision setting. A first-come-first-served rule within a class of urgency is adopted and usually no considerations are given to various classes' financial characteristics.

Reducing the operating costs and increasing the utilization using a more efficient OR and patient scheduling were the focus of the largest OR literature. Most models in the related literature deal with computing performance measures for the proposed control policies, like average utilization, patients and resources waiting time, waiting queue length, etc. (see, for example, Taylor and Templeton (1980), Rege and Sengupta (1996)). Rather than computing performance measures for various policies, we are deriving near-optimal protection levels that lead to increased expected revenue for the surgical department. We consider that focusing on increasing the revenue from better surgical scheduling may become as important as containing costs, because in the end, the survival and good functioning of any unit, be it in manufacturing or healthcare, would depend on the financial soundness of that unit, and how well it balances related costs and revenues.

3   ANALYTICAL MODEL

Unlike previous research where priority classes usually correspond to the degree of urgency for the procedure, in the present study the priority classes correspond to the expected revenue per unit of time expected to be obtained by performing a procedure on a patient from that class. This choice of patient segmentation, based on the reimbursement category (level of the insurance coverage) for the type of procedure requested lends itself naturally to analyzing the problem as one of expected revenue maximization. Traditionally, maximizing the expected revenue is the objective of choice when there is a limitation of resources and when there are several customer classes that compete over a relatively fixed amount of resource over a specified time period. Additionally, EMCR-OR can also be used to track the expected revenue distribution and answer further questions of interest for the decision maker, for example, the probability of incurring a certain revenue.

Considering the class definition described above, it is not always true that higher patient classes would correspond to patients who have full health insurance/coverage, while lower priority classes correspond to classes of patient who have low or partial health insurance. The expected revenue per unit of

time dictates the rankings across classes, where both the revenue per surgery and the surgery durations are influencing factors. For this reason, two hospitals practicing the same prices, but where surgeons performing those surgeries differ in their effectiveness, may have different rankings for those patient classes, resulting in different protection levels. This means that a surgeon's efficiency and effectiveness are direct factors influencing the optimal protection levels and ultimately the revenue expected to be incurred by the surgical department.

We assume that the hospital has evidence and records of all or most past requests for elective surgeries, including those that were not finally honored (patients either gave up, died, or went to a different hospital). This assumption avoids demand censoring situations that are common in airline and hotel industry (van Ryzin and McGill 2000), where it is difficult to keep track of every unfilled request that was made online, for example. In the hospital environment, where requests for surgery are taken by the scheduling personnel, it is not unrealistic to assume that there is a complete or large enough set of records of all past requests.

The EMCR-OR simulation algorithm is a generalization of the EMSRb heuristic (Belobaba 1989). The method makes use of forecasted demand for the next planning period, which is based on historical demand distribution for each type of surgery. The distribution of the reimbursement category would follow the national distribution of the main insurance categories, but each institution is free to use the distribution that better fits its market demand. These demand distributions would be updated as the time horizon progresses, at intervals that would make sense to suspect that there are relevant changes in the demand distributions that warrant changes in the protection levels.

New requests for bookings for elective surgeries arrive each day from patients, based on their doctors' recommendations. Under unconstrained capacity conditions, these procedures would be performed as soon as possible, but in reality OR capacity is limited in a given day, considering the number of available doctors, nurses, equipment etc., so accepting too many requests for a certain day or time period will, in most cases, result in excessive overtime, or even turning away previously scheduled patients or emergency cases. An additional difficulty arises from the fact that the exact time it takes to perform various elective surgical procedures is not known with certainty at the time of booking Shukla, Ketcham and Ozcan 1990). The problem the hospital faces in general, and the surgical scheduling department in particular, at the beginning of each planning period or booking window, is to decide how many of the additional requests for elective surgery to accept for that time interval for which the booking process is open, with the goal of maximizing the expected revenue, while also considering regular and overtime utilization, as well as emergency treatments.

EMCR-OR does not make any assumptions about the order of patients' arrival. That is, we assume random customer arrivals, with stochastic total demand per customer class. This mimics what happens in practice, where customers from different classes arrive based on needs, concurrently rather than sequentially, and not necessarily in some order determined by their reimbursement category. All patients' requests for surgery type j possessing coverage level i contribute to class ij's demand. Hence, a class' demand is function of the surgery j's duration and the fraction of patients holding type i reimbursement level. The information can be obtained by analyzing historical data on past surgeries of that type performed in that surgical unit or medical facility, and by analyzing historical data about the population structure on surgical needs and insurance coverage.

The data we use in this research come from a large teaching hospital, collected over six years. The information on the actual cost incurred and revenue generated by the surgeries performed is not available, because it was not recorded when the surgeries took place. To compensate for this, we make some assumptions on the relative values between the revenues r_{ij} generated by a surgery of type j when requested by a patient holding a reimbursement plan type i, that are reimbursed to the hospital for the services performed. We also assume that we can establish an ordering between the expected revenues per unit of time across classes of patients. As each surgery takes a variable amount of time, t_j, the revenue generated per unit of time will vary even across same type surgeries.

Given N surgeries (identified by their Current Procedural Terminology or CPT) and M reimbursement categories, a patient's class will be determined based on the type of surgery requested ($1,...,j,...N$) and the reimbursement level ($1,...,i,...M$). Let k be the class of the request (the combination of patient's reimbursement level and surgery requested), with $k = 1,...,K$ and $K = M \times N$. Each class k's demand is a random variable with a given probability density function, f_k, and cumulative distribution, F_k. Let D_k be a realization of such a demand, c_i be the reimbursement level for insurance type i, p_j be the price for surgery type j, and $r_k \equiv r_{ij} = c_i p_j$. With this formulation, the implied assumption is that the price of a surgery is independent of the reimbursement levels, which is a reasonable (and ethical) assumption to make in the context of managed care. The ranking across these k classes of patients is based on the order statistics of $\rho_k = r_k / t_j$, the revenue generated per unit of time. Let $\mu_k = E(\rho_k)$, the expected revenue per unit of time collected per class k. Suppose that all μ_k are ordered decreasingly, $\mu_{(1)} > ... > \mu_{(k)} > ... > \mu_{(K)}$. The question then becomes how many time units, $x_k - x_{k-1}$, should be protected in a partitioned fashion for each individual class k (reserved strictly for that class), or in a nested fashion, for classes k and higher, x_k, while observing the total system capacity C, in order to maximize the total expected revenue. We are looking at the non-trivial case for which the total demand does exceed the available system capacity. Under this assumption, the total expected revenue is of the form:

$$ER(x_1, x_2, ..., x_{K-1} \mid C) = \sum_{k=1}^{K} \mu_k \left[\int_0^{x_k - x_{k-1}} y f_k(y) dy + \int_{x_k - x_{k-1}}^{\infty} (x_k - x_{k-1}) f_k(y) dy \right]$$

Let $X = (x_1, x_2, ..., x_{K-1})$ be the vector of nested protection levels for the $K - 1$ classes, and X^* be the vector of optimal protection levels (defined in a corresponding fashion). Then, the problem can be formulated as

$$X^* = \arg\max \left\{ ER(X) \mid \sum_{k=1}^{K} x_k \leq C \right\}$$

With an increased number of classes, the above mathematical formulation suffers from the curse of dimensionality, and poses serious computational issues in practice. The problem reduces to a stochastic knapsack problem with random item (surgery) sizes, known exactly only after the selection decision has been made, and known item values at the time of the decision. The general knapsack problem is known to be NP-hard (Garey and Johnson 1979). In practice, heuristic approaches are preferred to the optimal calculations due to the intuition behind them, faster computation time, and small optimality gaps (observed 0.5% gap in the case of EMSRb (Talluri and van Ryzin 2004)). This is why in this paper we use simulation to compute the protection levels. The EMCR-OR output takes the form of near-optimal protection levels (time allocations) that would give the scheduler a very good insight into how many time units to protect for each reimbursement class within a surgery type or surgical subspecialty. This information would then become the basis for deciding, during each planning period, on the number of ORs to open and the number of surgeries to be performed within each class.

4 DESCRIPTION OF THE EMCR-OR ALGORITHM

The goal of the EMCR-OR approach is to obtain a distribution of protection levels (for each class of patients), the mean of which is reported as the near-optimal protection level for that class. The new idea we introduce here is that of identifying and analyzing a distribution of protection levels, rather than just a set of fixed protection levels, as is the case in the RM literature. The algorithm consists of the following steps:

Step 1 (Initialization): Pre-compute all revenues r_k and averages μ_k, with $k = 1,...,K$. All μ_k are ordered decreasingly, $\mu_{(1)} > ... > \mu_{(k)} > ... > \mu_{(K)}$.

Step 2: Let $A_k = \sum_{k=1}^{K} D_k$ be the aggregated future demand for classes $k, k-1, ..., 1$. Compute the weighted

average revenue for all k classes, $\bar{\mu}_k = \dfrac{\sum_{k=1}^{K} \mu_k E[D_k]}{\sum_{k=1}^{K} E[D_k]}$.

Step 3: The nested protection level for class k and higher, x_k, is chosen (Littlewood 2005) so that:

$$P(A_k \geq x_k) = \frac{\mu_{k+1}}{\bar{\mu}_k}.$$

Step 4: Update $X_N = (x_1, x_2, ..., x_{K-1})$, the vector of nested protection levels, with the values obtained in Step 3. Let $x_K = \max\{0, C - x_{K-1}\}$, where C = system's capacity.

Step 5: Partitioned (i.e., disaggregated) protection levels are obtained so that $x_k - x_{k-1}$ time units are allocated to class *k*.

Step 6: Simulate the distributions for each disaggregated protection level. The means of these distributions of protection levels become the near-optimal protection levels reported for each class. Report the vector of partitioned protection levels, $X_P = (x_1^*, x_2^* - x_1^*, ..., x_{K-1}^* - x_{K-2}^*)$. Report total revenue,

$$\sum_{k=1}^{K} (x_k^* - x_{k-1}^*) \mu_k.$$

If the surgical department is interested in pooling the total available time for all *N* surgical types provided, and then compute protection levels for all *N*M* possible class combinations, then it needs to generate *N*M* - 1 protection levels. If it is interested in deciding how much OR time to allocate for patients in one reimbursement class within one particular surgery, then it needs to compute only *M* - 1 protection levels for that surgical procedure. If the department is interested in deciding how much time to allocate for the first *k* procedures, then it only needs to use the nested protection levels found at steps 3 and 4.

5   PRELIMINARY NUMERICAL RESULTS

We present several examples to show the algorithm's implementation and performance, and compare its results with the optimal ones obtained by Stanciu (2009) using the same set of surgical procedures. In the general case with *n* classes, the *n* - 1 optimal protection levels are found by solving a system of *n* - 1 equations of first order conditions. They are set to balance the expected revenue losses between any two adjacent classes (when all are ordered function of their expected revenue per unit of resource), by not protecting enough capacity for these classes.

First, consider two surgeries, defined by their CPTs: Surgery 1, CPT 36489 (Placement of central venous catheter) with a duration following a 2-parameter lognormal distribution with $\mu = 1.15$ hours and $\sigma = 0.6$ hours, and Surgery 2, CPT 52000 (Cystourethroscopy of bladder) with a duration following a 2-parameter lognormal distribution with $\mu = 0.68$ hours and $\sigma = 0.32$ hours. In the first example, we assume that the capacity for the next planning horizon is $C = 40$ hours, $c_1 = 100\%$ and $c_2 = 75\%$. In the second example, $C = 50$ hours, $c_1 = 100\%$, $c_2 = 75\%$ and $c_3 = 50\%$. Based on past history, we were able to estimate the demand distribution for each of these four and six classes, respectively, along with the relative procedure prices. Since the individual surgical durations follow the lognormal distribution, the aggregated demand for a particular class, as a summation over multiple lognormal distributions, lends itself to an approximation by the normal distribution (by the central limit theorem). Empirically, the normal distribution was also found to be a very good fit for the demand across each class, with the extremes of the distribution more than six standard deviations away from the mean. We chose to use it, for simplicity, over the

Weibull distribution (another good fit), and, as a precautionary measure, we truncated from the left the normal distribution in the simulation runs, so that no negative surgery times are generated.

Applying the EMCR-OR heuristic described above, and using the Crystal Ball add-in for MS Excel (11th edition, 10,000 iterations) we get the partitioned protection levels (x) that are displayed in Figure 1 along with the optimal ones (x^*) obtained in our previous work (Stanciu 2009). Note that if $x_k = 0$ for some class k, then there is no time protected for class k over the planning period under analysis, and those patients would be postponed until later, when it becomes economically feasible to serve that class.
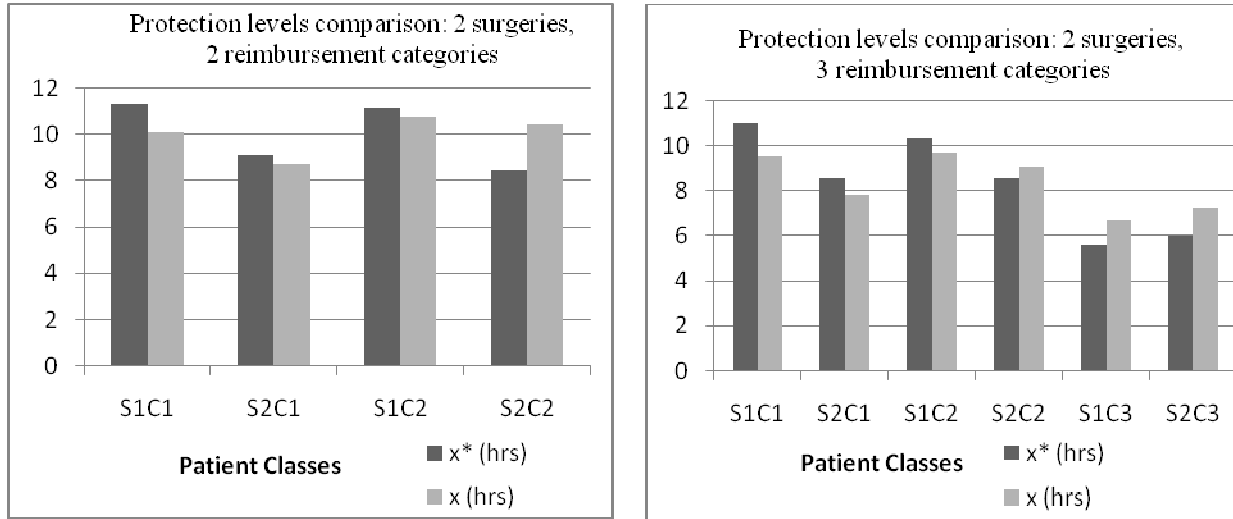

Figure 1: Protection levels comparison (x and x*) for 4 and 6 classes

The classes are shown in decreasing order of their expected revenue per time unit, as presented in the heuristic algorithm above. Note that this order can be easily changed (by surgery or by reimbursement category), but a different order would not always preserve the information on relative expected revenues for two adjacent classes.

Figure 1 shows a slight difference between the optimal and simulated protection levels. What is more relevant, though, is the expected revenue comparison. We can report an average revenue gap, in both examples, of only 0.7%, with a maximum gap of 1.5%. Similar simulation runs were performed using other CPTs, and the expected revenue gaps were consistently around 0.7%.

We now extend the two-surgery, three-class example mentioned above, by adding one more cardiovascular surgery (Surgery 3, CPT 33512, Three vein coronary artery bypass graft) with a duration following a 2-parameter lognormal, with $\mu = 6.25$ hours and $\sigma = 1.2$ hours. We now have nine final patient classes. Assuming a capacity of $C = 120$ hours, we present in Figure 2 the simulation-based protection levels (x) along with optimal ones (x^*). As in Figure 1, the classes are shown in decreasing order of their expected revenue per time unit.
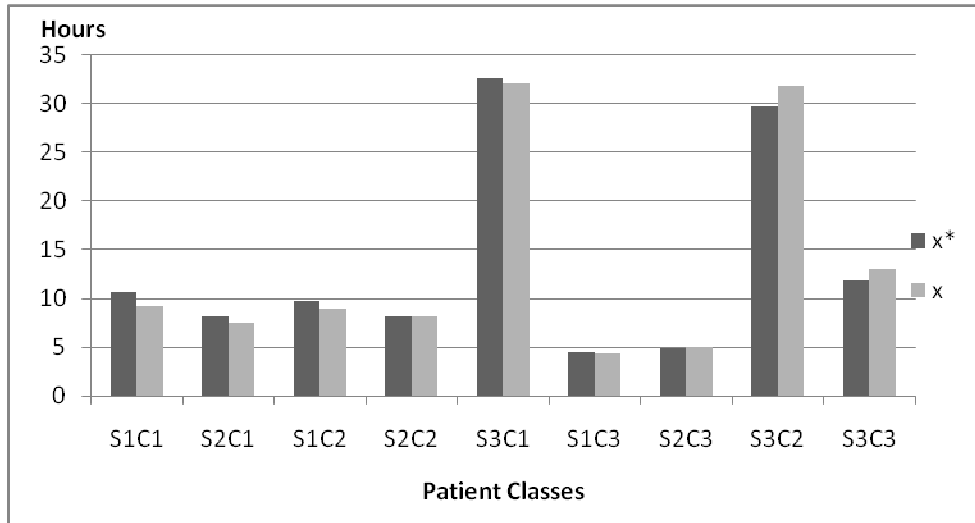
Figure 2: Protection levels comparison for 9 classes

The expected revenue gap, computed using 10,000 iterations and random patient arrivals is about only 0.7%, with a maximum gap of about 1.4%. The heuristic is attractive in practice, with just slight loss in expected revenue, but with the advantage of less computational time.

6   EMCR-OR EXTENSION: INCORPORATING OVERTIME

Overtime is necessary in the situations where a surgery goes beyond the end of the budgeted OR time for that day, since we assume that no surgery continues on the next day. Subsequently, once a decision is made to start a surgery, in the event that it goes beyond the budgeted time, it incurs overtime cost, which usually is higher than the cost of regular time.

Strum and Vargas (1997) developed a minimal cost analysis (MCA) approach to decide on the time allocations across subspecialties in order to minimize under and overutilization costs, based on the ratio of overtime to regular time cost. When overtime is unavoidable but allowed, we can incorporate this idea when making the decision of capacity allocation across classes of demand. The idea we propose is that after obtaining the distribution of protection levels for each class, instead of reporting the mean as the near-optimal protection levels, we can decide on a threshold, or cut-off value based on this MCA analysis, that would represent the time the surgical department should allocate for that surgery-reimbursement class combination, in an attempt to increase revenue and decrease overtime costs. For example, if the ratio between overtime and regular costs per time unit is 2, then we can use the 66.7th percentile of each probability distribution for the number of units of time to be assigned for that patient class.

To illustrate this extension, we employ the example with nine patient classes presented above. Table 1 shows the simulation results of the mean and the 66.7th percentile values for these (empirical) distributions of protection levels for each of the nine classes. The values for the 66.7th percentiles should be used as fixed protection levels when overtime is allowed. The last column displays the probability of not allocating any time to that particular class during the planning horizon (probability that $x = 0$, i.e., the class is closed because it is not financially feasible). The classes are presented in decreasing order of their expected revenue per time unit.

Table 1: Protection levels for 9 classes.

Rank	Class	Mean (hrs.)	66.7th percentile (hrs.)	Prob($x = 0$)
1	S1C1	9.23	10.18	0.00
2	S2C1	7.51	8.52	0.04
3	S1C2	8.91	10.23	0.08
4	S2C2	8.11	9.45	0.11
5	S3C1	32.03	33.86	0.00
6	S1C3	4.38	5.90	0.30
7	S2C3	4.95	6.55	0.28
8	S3C2	31.79	35.22	0.00
9	S3C3	13.03	17.82	0.10

We further used simulation to compute the increase in expected revenue when using as protection levels the 66.7th percentile of the protection level distributions, rather than the mean of those distributions. The results obtained show that the new protection levels would result in an 11.5% increase in expected revenue, which comes with an average capacity usage of about 123 hours; this represents an average overtime of only 3 hours, 2.5% of the initial capacity of 120 hours. We consider this to be a remarkable expected revenue increase, which is due to recognizing the fact that procedure variability plays a large role in the financial situation of the surgical unit.

While overtime is not desirable, it may become necessary in situations where unexpected complications arise, or when longer surgeries are scheduled during a day. While the former situation cannot be predicted, the latter should take into account the probability of going over the allocated time, due to the potential costs involved. The models presented here should be coupled with rigorous forecasting and prediction methods, helping the decision maker determine and implement more accurate protection levels.

7   CONCLUSIONS

This work focuses on determining the reservation of a relatively fixed capacity across multiple customer classes, in the case where the accepted customers' requests use a variable amount of the resource under consideration. We look especially at situations faced by a surgical unit that is in the process of scheduling elective surgeries over a certain period. We propose a simulation algorithm that computes the protection levels (both partitioned and nested) for the classes of patients considered over this planning period. We analyze a very realistic situation in which the actual service duration is known with certainty only after the procedure is performed, but scheduling customers' requests should be performed days before this happens. Customers arrive in a random order, and we assume that customers can be segmented based on two criteria: their need for a specific surgery, and their reimbursement level. The latter criterion could be dropped in the case of cosmetic surgeries, for example, which are not usually covered by health insurance. The use of protection levels would determine which patients to be accepted and which to be postponed during the planning period under study, resulting in maximizing the expected revenue incurred by the surgical unit. Patients are accepted given the protection level for that class, and the postponed ones are scheduled for a later date when the protection level for that class can accommodate that patient. Preliminary results on a selected sample of surgeries provided very encouraging results, with small revenue gaps, which renders our heuristics to be very attractive in practice. Further testing should provide more insight into the performance of the heuristics over longer planning periods and more patient classes.

While our examples are from the healthcare area, the simulation model has applicability in many other sectors characterized by stochastic service times. While there are many tradeoffs and criteria that decision makers need to consider when budgeting time and booking customers' request for service, the model we present here can be used as a decision and planning tool to improve the operational decision making in many service settings, with operating room being one of them. Efficient time allocation across

classes of patient in the form of protection levels is of great importance for improving the revenue of the health care unit. Incremental revenue could be used for capacity expansion, so that, when coupled with adequate personnel, could help the surgical unit accommodate more surgeries in the long run, thus improving the quality of care.

This study could offer a viable alternative to the queuing approach when analyzing a service environment's complex behavior, characterized by uncertain service duration and random customer arrivals. We recognize that the model can be extended to account for other various situations encountered in practice, additions that would make the problem we present here of an even greater importance. As a possible future extension, the decision maker can incorporate a class related deadline, after which a penalty is incurred for postponed requests. At the same time, the model could be enhanced by including penalties for surgical errors, which affect the quality and duration of care. Penalties could take different forms, both monetary and non-monetary, from deterioration in the health condition of a patient, to some monetary penalty imposed by the hospital for postponed, or even lost, surgical requests, or, in extreme cases, malpractice expenses, all affecting the efficiency of medical care at various degrees.

REFERENCES

Belobaba, P. P. 1989. Application of a probabilistic decision model to airline seat inventory control. *Operations Research* 37: 183-197.

Chapman, S. N. and J. I. Carmel. 1992. Demand/capacity management in health care: an application of yield management. *Health Care Management Review* 17: 45-55.

Everett, J. E. 2002. A decision support simulation model for the management of an elective surgery waiting system. *Health Care Management Science* 5: 89-95.

Garey, M., and D. Johnson. 1979. *Computers and intractability: a guide to the theory of NP-Completeness*. W. H. Freeman Publishing.

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

Green, L. V., S. Savin, and B. Wang. 2006. Managing patient service in a diagnostic medical facility. *Operations Research* 54: 11-25.

Gupta, D., and L. Wang. 2008. Revenue management for a primary-care clinic in the presence of patient choice. *Operations Research* 56: 576-592.

Ivaldi, E., E. Tanfani, and A. Testi. 2003. Simulation supporting the management of surgical waiting lists. Discussion Paper della Sezione di Economica Politica e Studi Economici Internazionali.

Karaesmen, I., and I. Nakshin. 2007. Applying pricing and revenue management in US hospitals - New perspectives. *Journal of Revenue and Pricing Management* 6: 256-259.

Littlewood, K. 2005. Forecasting and control of passenger bookings. *Journal of Revenue and Pricing Management* 4: 111-123.

Lowery, J. C. 1996. Design of hospital admissions scheduling system using simulation. In *Proceedings of the 1996 Winter Simulation Conference,* ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 1199-1204. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Olivares, M., C. Terwiesch, and L. Cassorla. 2008. Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science* 54: 41-55.

Phillips, R. 2005. *Pricing and revenue optimization.* Stanford University Press.

Rege, K. M., and B. Sengupta. 1996. Queue-length distribution for the discriminatory processor-sharing queue. *Operations Research* 44: 653-657.

Shukla, R. K., J. S. Ketcham, and Y. A. Ozcan. 1990. Comparison of subjective versus data base approaches for improving efficiency of operating room scheduling. *Health Services Management Resources* 3: 74-81.

Stanciu, A. 2009. Applications of revenue management in healthcare. Ph.D. thesis, Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania. Available via <

http://etd.library.pitt.edu/ETD/available/etd-07292009-144217/unrestricted/Stanciu_etd2009Pitt.pdf>
[accessed April 6, 2010].

Strum, D. P., L. G. Vargas, J. H. May, and G. Bashein. 1997. Surgical suite utilization and capacity planning: a minimal cost analysis model. *Journal of Medical Systems* 21: 309-322.

Talluri, K. T., and G. J. van Ryzin. 2004. *The theory and practice of revenue management*. Kluwer Academic Publishers.

Taylor, I. D. S., and J. G. C. Templeton. 1980. Waiting time in a multi-server cutoff-priority queue, and its application to an urban ambulance Service. *Operations Research* 28: 1168-1188.

van Ryzin, G. J., and J. McGill. 2000. Revenue management without forecasting or optimization: an adaptive algorithm for determining airline seat protection levels. *Management Science* 46: 760-775.

Vargas, L. G., J. H. May, and A. Stanciu. 2010. Revenue management with random resource utilization. Working paper.

AUTHOR BIOGRAPHIES

ALIA C. STANCIU is a Visiting Assistant Professor in the School of Management at Bucknell University. She got her Ph.D. in Operations Research from University of Pittsburgh in 2009. Her research interest is in the areas optimization and simulation applied in healthcare, including operating room capacity management and scheduling, and patient appointment scheduling. Her research also focuses on the capacity aspect of revenue management, and she is interested in incorporating revenue management in the healthcare area.

LUIS G. VARGAS has been a faculty member of Katz Graduate School of Business at University of Pittsburgh since 1979. Luis has focused his research on decision theory, practical applications of the Analytic Hierarchy Process (AHP), artificial intelligence in manufacturing, the use of artificial intelligence techniques for scheduling, measurement of resource utilization, group decision making, Bayesian networks, and forecasting. For more information, see <http://www.business.pitt.edu/faculty/vargas.html>.

JERROLD H. MAY has been a faculty member of Katz Graduate School of Business at University of Pittsburgh since 1974. He is currently interested in data mining, and is a member of INFORMS and AAAI. For more information, see <http://www.business.pitt.edu/faculty/may.html>.