

A Model for Understanding the Impacts of Demand and Capacity on Waiting Time to Enter a Congested Recovery Room

Tor Schoenmeyr, M.S.,* Peter F. Dunn, M.D.,† David Gamarnik, Ph.D.‡ Retsef Levi, Ph.D.,§ David L. Berger, M.D.,|| Bethany J. Daily, M.H.A.,# Wilton C. Levine, M.D.,** Warren S. Sandberg, M.D.,Ph.D.††

Background: When a recovery room is fully occupied, patients frequently wait in the operating room after emerging from anesthesia. The frequency and duration of such delays depend on operating room case volume, average recovery time, and recovery room capacity.

Methods: The authors developed a simple yet nontrivial queuing model to predict the dynamics among the operating and recovery rooms as a function of the number of recovery beds, surgery case volume, recovery time, and other parameters. They hypothesized that the model could predict the observed distribution of patients in recovery and on waitlists, and they used statistical goodness-of-fit methods to test this hypothesis against data from their hospital. Numerical simulations and a survey were used to better understand the applicability of the model assumptions in other hospitals.

Results: Statistical tests cannot reject the prediction, and the model assumptions and predictions are in agreement with data. The survey and simulations suggest that the model is likely to be applicable at other hospitals. Small changes in capacity, such as

addition of three beds (roughly 10% of capacity) are predicted to reduce waiting for recovery beds by approximately 60%. Conversely, even modest caseload increases could dramatically increase waiting.

Conclusions: A key managerial insight is that there is a sensitive relationship among caseload and number of recovery beds and the magnitude of recovery congestion. This is typical in highly utilized systems. The queuing approach is useful because it enables the investigation of future scenarios for which historical data are not directly applicable.

THE flow of patients through the perioperative process involves multiple resources as patients are transferred from preoperative locations to operating rooms (ORs) and then to recovery areas and beyond. Operating room suites cannot always smoothly meet all of their demand in part because of congestion in adjacent perioperative areas, “smoothly” being defined as the OR suite meeting demand on time and without requiring intensive intervention by managers. When workload exceeds a certain critical level, patients encounter delays because demand outstrips the available resources. In the case of a congested postanesthesia care unit (PACU), this means the patients recover in the OR after emerging from anesthesia. In addition, these delays have significant negative effects on resource utilization within the hospital (e.g., OR time) as well as the morale of staff. To confidently yet efficiently solve such a problem, one must be able to predict how much capacity to add to the congested unit in the workflow to reduce or eliminate delays. Alternatively, one must be able to predict the impact of increased OR volume in the face of constrained PACU capacity.

Similar issues arise in the context of future planning of physical capacity and personnel resource allocation. Our hospital is constructing a new OR building that will increase the total number of ORs by about 40%. Capacity planning of the various components in the new perioperative system is a key issue in ensuring a smooth transition to the new building and the elimination of bottlenecks that exist in the current system. For future planning purposes, one wishes to make robust decisions that will optimize against various possible future scenarios of surgery demand volume. In particular, in our hospital we have observed unpredictable changes in the volume of surgeries throughout the recent years, and the new perioperative system should be planned to accommodate this variability.

When making changes in a perioperative system, whether to optimize the current system or planning to accommodate multiple future scenarios, the financial

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org).

* Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts; † Assistant Professor of Anesthesia, Harvard Medical School, Anesthetist, Department of Anesthesia and Critical Care, and Executive Medical Director of the Operating Rooms, Massachusetts General Hospital, Boston, Massachusetts; ‡ Assistant Professor of Operations Research, § Assistant Professor of Operations Management, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts; || Assistant Professor of Surgery, Harvard Medical School, and Associate Visiting Surgeon, Massachusetts General Hospital, Boston, Massachusetts; # Administrative Director, Operating Room Business and Information Systems, Massachusetts General Hospital, Boston, Massachusetts; ** Instructor in Anesthesia, Harvard Medical School, and Clinical Director, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, Massachusetts; †† Associate Professor of Anesthesia, Harvard Medical School, and Assistant Anesthetist, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, Massachusetts.

Received from the Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, Massachusetts. Submitted for publication April 12, 2008. Accepted for publication January 26, 2009. Supported by a Center for Integration of Medicine and Innovative Technology (Cambridge, Massachusetts) Career Development Award to Dr. Sandberg. Additional departmental support was provided by the Department of Anesthesia and Critical Care, and by the Operating Room Administration Division of Massachusetts General Hospital (Boston, Massachusetts). Professor Levi was partially supported by grants DMS-0732175 and CMMI-0846554 (CAREER Award) awarded by the National Science Foundation, Washington DC, and by the Singapore-MIT Alliance grant and the Bauchsbau grant, awarded by MIT, Cambridge, Massachusetts. Presented in part at the Annual Meeting of the American Society of Anesthesiologists Annual Meeting, San Francisco, California, October 17, 2007.

Address correspondence to Dr. Sandberg: Department of Anesthesia and Critical Care, Massachusetts General Hospital, 55 Fruit Street, Clinics 3, Boston, Massachusetts 02114. wsandberg@partners.org. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

stakes are very high. Healthcare organizations function with very small net margins, so decisions about committing resources must be made with a high degree of confidence that the investment will lead to the desired result. Implementing congestion relief solutions to workflow problems may entail both monetary and intangible costs. Monetary costs include construction costs and the capital and personnel costs of implementing changes in patient flow. There is also an intangible cost, in terms of job satisfaction and potential personnel turnover, of modifying the roles and responsibilities of perioperative personnel. In addition, such changes entail a risk of disrupting the nominal function of the perioperative system without actually relieving congestion. Thus, it is useful to develop a model of the perioperative system within which to test proposed solutions to congestion problems before committing to a given course of action.

Effective modeling would save time and money over trial-and-error approaches, while providing some confidence that the solutions being implemented are likely to yield the desired result. On the other hand, perioperative workflow appears complex, which may limit the applicability of some modeling tools for developing successful models of perioperative system function. Nevertheless, simple approaches, such as, for example, queueing models are attractive because they are readily applied in other fields with similarly complex systems, thus providing a body of knowledge that may transfer to the perioperative system. They also provide managerial insights about fundamental trade-offs between resource allocation and procedural decisions and the performance of the system.

The major methodology used in this study is based on queueing tools. The main goal of this study was to develop and test a high-level queueing model of the PACU demand (that is generated by the ORs), and throughput to illuminate the relationship between key system parameters, such as caseload, average recovery time, and number of PACU beds on one hand, and the magnitude of the PACU waiting time on the other. By "high-level queueing model," we mean a tool that can be used to support operational-strategic decisions (*i.e.*, the number of beds to have in the PACU) rather than ongoing staffing decisions. In particular, we sought to understand the following issues: (1) whether proposed modest increases of PACU capacity would have a meaningful impact on the waiting time for PACU beds at our hospital; (2) how long-term changes in the overall case volume might affect waiting for PACU beds and perioperative congestion in our hospital and other similar medical centers. In the longer term, we hope to use this tool in the workflow planning of the new OR building that is currently under construction.

Queueing modeling and analysis have enjoyed a wide range of applications, beginning with the classic work in telephony by Erlang.¹ Currently, the application areas include telecommunications, wireless communication,

manufacturing, computer systems, networks, and a wide range of service areas such as call (contact) centers.² In a compact but insightful way, the theory captures fundamental trade-offs among the limited resources, implied waiting times, and inherent randomness of congested systems. However, effective application of queueing theory tools to capacity management in the healthcare industry has thus far been relatively limited. Relevant references in this area include work on inpatient bed occupancy,³ access times for outpatient departments,⁴ and access to critical care beds.^{5,6}

Other authors have considered the possibility of reducing PACU waitlists by modifying the OR schedule, thus "shaping" the profile of demand for PACU resources throughout the OR day.⁷ Our results are mostly applicable in scenarios where case scheduling is, by design or by accident, mostly random, leading to a steady release of work into the PACU. In this paper, we will test whether this is the case in our hospital, but we were sensitive to the fact that our model may not be applicable in contexts with highly structured case scheduling. To get a rough sense of how commonly random (as opposed to well-structured) scheduling policies are encountered (and by extension, how broadly our model is applicable) we conducted an informal survey of scheduling policies in medium and large OR suites.

Materials and Methods

The main methodology in this study is based on queueing tools. We start with a general but necessary basic background on queueing, then we will explain how queueing models can be used to model our problem.

Discussion of Queueing Models

Basic queueing models consist of the following components. There is a stream of jobs arriving to the system in a random fashion, one after the other. The randomness in the job arrivals is modeled through a probability distribution of the time that elapses between the arrivals of two consecutive jobs. This creates an arrival process that is *a priori* unpredictable. Each of the jobs arriving to the system needs to be served by one of multiple servers with identical performance. However, the time it takes to serve a given job is, in general, also random and follows its own probability distribution. This creates another source of unpredictability in the system. Jobs are served according to some queueing discipline, in our case on a first-in-first-out basis. Upon arrival, a job may have to wait in the queue until there is an idle server, and all the jobs that arrived before that job are already served or in service. The arrival and service times are unpredictable; therefore, it is never *a priori* certain exactly how long a given job will have to wait until it will be served. Many queueing models have been developed to

characterize the unpredictable behavior of the system based on certain performance measures (e.g., the average waiting time of a job).

In this study, we adapt a queuing model in a nontrivial way to model the flow of patients through a congested PACU. In our model, jobs correspond to patients to be recovered from anesthesia after surgery, and the servers correspond to the staffed beds in the PACU. The jobs (i.e., patients) arrive to the servers and need to be served (i.e., recover from anesthesia). For each patient (job), the service time is simply the time they are in the PACU system. However, unlike most queuing models, the jobs in our model do not wait in the queue until they can be served. In particular, patients start their recovery from anesthesia after emergence, regardless of whether there is an available bed in the PACU. To capture this fact, we adapt the standard model and assume that there are infinitely many servers, out of which some are “real” servers (i.e., actual beds in the PACU) and some are “dummy” servers (i.e., waitlist slots for patients recovering in the ORs). This represents a nontrivial adaptation of standard queuing models because it accounts for a reality in medical care; the service times of jobs start upon their arrival into the queue and not upon the time they are delivered to the server (as would traditionally be the case in a classic queueing model). This is a fundamental difference between the jobs in our study (patients recovering from anesthesia) and the jobs in a traditional queueing system; therefore, we believe that the adaptation is novel and not a trivial departure from established modeling techniques. In the model we assume that patients arrive to the PACU with exponentially distributed interarrival times (between consecutive patients), but we impose no specific assumptions about the distribution of time in recovery and on the waitlist, and we allow maximum modeling flexibility of this aspect. The assumption of exponential interarrival times is justified based on known theory. We then want to analyze the fraction of time in which the dummy servers are used, which is exactly the time in which the PACU is completely full and patients have to recover in the ORs. The results from the theoretical model are then compared to and verified with actual data from our OR statistics database.

Next, we describe the model in more detail. In queueing theoretic terminology, we have modeled the waitlist plus the PACU occupancy as a collection of identical servers (beds), fed by an exogenous process (the ORs). More specifically, the model can be denoted as M/G/∞; M because the arrival process is assumed to be Markovian (exponentially distributed interarrival times), G because the “processing” time (the time a patient spends in the PACU and on the waitlist) is assumed to have a general distribution (that is, we make no specific assumptions about this distribution), and ∞ because we place no limit on the number of patients in the joint

PACU+waitlist system. Because there is no limit on the number of patients, our model actually does not have a queue *per se*; all patients in the system are assumed to be in a state of processing (i.e., time spent on the waitlist “counts” towards total time required to recover from anesthesia). This reflects the fact that patients start the recovery phase immediately after surgery, regardless of whether there is an available bed in the PACU. The way in which we distinguish patients in waitlists *versus* those in PACU is a novel aspect of our modeling, and it will be detailed below.

We denote the rate of completed operations Λ (e.g., PACU admissions per hour in our setting as λ), and the (arithmetic) mean length of stay in the joint PACU+waitlist system as $1/\mu$. It is a well-known result from queueing theory⁸ that, under these assumptions, the steady-state distribution of the number of patients in the system follows a Poisson distribution, and the probability of having i patients in the system can be expressed as follows:

$$P(i) = \frac{(\lambda / \mu)^i e^{-\lambda / \mu}}{i!} \quad (1)$$

Mathematically, this model does not yet make any distinction between patients in PACU and patients on the waitlist. If the capacity of the PACU is N patients (i.e., N beds), then the number of patients i_{PACU} in the PACU is:

$$i_{\text{PACU}} = \min(i, N). \quad (2)$$

Similarly, the length of the waitlist i_{Waitlist} is

$$i_{\text{Waitlist}} = \max(0, i - N). \quad (3)$$

That is, if there are N or fewer patients in the system, then all of them are in the PACU, but patients in excess of N will be on the waitlist.

By using equations (1) and (3), we can calculate the average number of waitlist cases i_{Waitlist} as follows:

$$\begin{aligned} E[i_{\text{Waitlist}}] &= \sum_{i=0}^{\infty} \max(0, i - N) P(i) \\ &= \sum_{i=N+1}^{\infty} (i - N) P(i) \quad (4) \end{aligned}$$

To verify the accuracy of the queueing model in this context, we performed several tests, which are discussed in detail below and in the Results section. (1) We used real OR process time data to test statistically whether the underlying assumptions of our model are appropriate for our hospital. (2) We conducted simulation experiments to further validate these assumptions. (3) We conducted a survey to check whether the OR scheduling policies of other hospitals are likely to satisfy the underlying assumptions of the model.

Statistical Tests of Appropriateness of Underlying Model Assumptions

First, we considered the different mathematical assumptions that are necessary for equation (1) to hold. The model makes no assumptions about the statistical distribution for recovery time; therefore, the main concern was about the realism of exponentially distributed arrivals with an average arrival rate that was constant throughout the day. We have attempted to validate this assumption in several ways. We used goodness of fit and R^2 statistics to test the hypothesis that interarrival times of patients to the PACU follow an exponential distribution and whether, more generally, the queueing model gives good predictions about the system performance.

To estimate the parameters for the model and to perform the empirical tests of the model assumptions, we used de-identified OR throughput data from calendar year 2006. We also used data from calendar year 2005 and the first three quarters of calendar year 2007 in tests of the modeling assumptions. The data comprised time stamps of requests for PACU beds and time stamps for patients entering and leaving the facilities of interest. We verified that the average arrival rate into the PACU was constant during the middle part of the day, when we planned to use the queueing model, using timestamps for patient arrivals and departures from the PACU. Next, we estimated λ for the model by using the average arrival rate over the window of constant arrivals, which was $\lambda = 6.26$ patients per hour for the year 2006 (this is the only measured parameter used in the exponential distribution). We investigated whether this arrival rate was constant over the year before 2006 and the 9 months after 2006. We further investigated whether the arrival rate in 2006 was constant by calendar quarter and by day of the week.

Simulation Experiments to Test Model Assumptions

We first investigated the pattern of completed operations in numerical simulations, using our hospital's OR time data as the basis for simulations created in a general programming language (MATLAB, Mathworks, Natick, MA). Procedure times (patient in to patient out) are best assumed to come from a lognormal distribution, which is consistent with past work,⁹ as well as data from our own hospital. For the simulation, we sampled the distribution of actual procedure times. This distribution of procedure times can be described by their geometric mean (the antilog of the mean of the log-transformed data) and a measure of dispersion from the mean, the SD, here given as the antilogs of (mean of the log-transformed data minus SD of the log-transformed data) and (mean of the log-transformed data PLUS SD of the log-transformed data). Using this convention, the mean procedure time was 1.95 h (SD 0.97 h, 3.94 h), as expected in a right-skewed distribution. In the simulation, additional cases were added to each OR until the last case had a sched-

uled finish time later than 2:30 PM, at which point no more cases were added. Cases were added without consideration of their duration; in effect, random scheduling was simulated. The simulated OR days began at 8:00 AM, and 30 min of scheduled turnover time (our hospital's average target) was added between each case. However, the actual time between surgeries using the convention above was described by a lognormal distribution with mean of 1.25 h (SD 0.68 h, 2.29 h). This longer and more variable time between cases corresponds to the actual time between cases measured at our hospital, and it is driven both by turnover time and longer scheduled and unscheduled gaps between cases. The simulation was repeated for 5, 10, . . . , 50 ORs, although for some questions the number of ORs did not matter. For each number of ORs, the simulation was run for 1,000 days. The output of the simulation was the interarrival time between completed cases, and this was compared to the exponential distribution. We also used the simulations to generate the number of completed operations (*i.e.*, PACU arrivals) per hour, again varying the number of hypothetical ORs feeding the PACU.

Next, we simulated smaller OR suites using our OR time data. To do this, we segmented the data from calendar year 2006 to create simulated 5-OR suites. In other words, ORs 1 through 5 were considered to be a small, independent OR suite, ORs 6 through 10 were considered to be another such suite, and so on until 8 unique mock suites of 5 ORs were created, representing the OR suites of typical small hospitals. We then calculated the number of completed operations (*i.e.*, PACU arrivals) for each half-hour of the day using the actual OR time data from each of the different suites.

Survey of Scheduling Policies at Other Hospitals

After these tests of the modeling assumptions, there was still the concern that our approach might not be generalizable if a hospital, through its OR scheduling practice, had succeeded in creating an OR schedule that did not produce exponentially distributed arrivals with a constant average interarrival time. For example, a hospital that consistently scheduled all of its short cases first would release a burst of completed cases into the PACU in a relatively short time frame, violating the constant average interarrival time assumption. To get some idea of how likely this might be, we conducted an informal survey of knowledgeable anesthesia clinicians at a convenience sample of 32 medical centers, asking them about what, if any, case sequencing strategies were employed in their hospitals.

We contacted the institution's Clinical Director of Anesthesia, or an equivalent person, *i.e.*, a person in the anesthesia department with personal, direct knowledge of how the daily OR schedule was constructed. We then asked how many beds were in the unit considered to be the main PACU and how many ORs were in the largest

contiguous group of ORs feeding that PACU. Next, we asked “Who decides how the cases on the OR schedule in the largest block of ORs feeding the main PACU are sequenced?” We then asked all respondents a follow-up question: “Are any strategies such as shortest-cases-first or ambulatory-cases-first pursued at your institution?” Finally, we asked directly: “Does your institution pursue any strategies to shape the flow of work entering the PACU such that it would be different from the patient stream occurring from surgeons simply adding cases to the schedule until their block was filled?” To analyze the results, we condensed the results of “Are any strategies such as shortest-cases-first or ambulatory-cases-first pursued at your institution?” and “Does your institution pursue any strategies to shape the flow of work entering the PACU such that it would be different from the patient stream occurring from surgeons simply adding cases to the schedule until their block was filled?” into a yes/no categorical response for whether case-sequencing was applied.

The assumptions about arrival distribution are important to the model, and they were subject to substantial verification efforts and discussion. On the other hand, we are fortunate that the queueing model and equation (1) are true regardless of the distribution of processing times.⁸ That is, one can use the average time $1/\mu$ regardless of what the actual distribution is. In our case, we calculated the average processing time by taking the difference between the time a PACU bed was requested and the time of PACU departure. Sometimes patients in the PACU have to wait for beds to become available elsewhere in the hospital. This “waiting to leave PACU” congestion effect was not dynamically modeled; however, the average time spent waiting for floor beds was included in the processing time $1/\mu$. Even though equation (1) does not rely on any distributional assumptions of processing time, it does assume that the average time is the same throughout the day. However, our data suggested that the average time in PACU tended to increase somewhat over the day, primarily because departing patients more often had to wait for floor beds later in the day. We will return to this factor when we compare the predicted and actual patient distributions in the Results section.

Next we will briefly discuss the identification of patients as being in PACU, equation (2), or waitlisted, equation (3), based only on the joint number of such patients and the PACU capacity N . The physical capacity of the PACU at our hospital is 28 beds. However, even when the system is busy it is rare that 28 patients are in the PACU simultaneously. Occasionally, the beds are not fully staffed. More importantly, when the system is busy, many patients are transitioning in and out of the recovery area; as a result, there are usually a few PACU beds that have just released one patient but not yet received the next waitlisted one from the ORs. We found that

when a patient gets waitlisted during the steady-state period from 3:00 to 5:00 PM, there were 24.1 ± 2.1 (mean \pm SD) patients in the PACU. We will use the (constant) value $N = 24$ of “effective capacity” to represent the current situation.

Finally, the queueing model was used to predict how many additional beds are required to eliminate most of the waitlist. The model was robust and flexible enough to give predictions to future scenarios in which some of the system parameters may change (*e.g.*, surgery volume, length of recovery, *etc.*).

Results

Evaluation of Modeling Assumptions

The first modeling assumption is that of a steady, “memory-less” arrival process to the PACU. In other words, the intervals between arrivals should be exponentially distributed, leading to a constant arrival rate over time when the queueing model is expected to work. Detailed measurement of the distribution of inter-arrival times from the regular work-day time window showed excellent agreement with the exponential curve, as shown in figure 1. Moreover, the release pattern of work into the PACU, as plotted in figure 2, showed that the arrival rate was approximately constant from about 10:00 AM to 4:30 PM. Specifically, the average arrival rate over the window of constant arrivals was $\lambda = 6.26$ patients per hour for the year 2006. This arrival rate, λ , is the only measured parameter required for the model, but the utility of the model can be affected if λ changes (by day of week, for example). We investigated the period over which the arrival rate was stable. For the calendar year 2005 and for the first three quarters of 2007 (the end of our dataset), the arrival rate to the PACU during the steady-state portion of the day was the

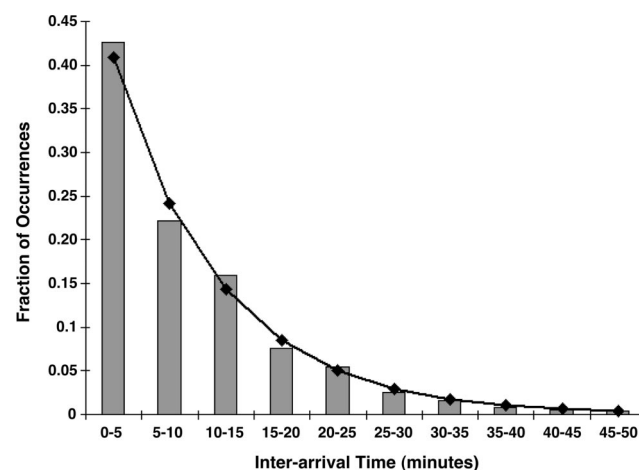


Fig. 1. Actual (bars) versus predicted (line) distribution of inter-arrival times to the postanesthesia care unit (PACU) in minutes. For the line, we assumed exponential distribution with constant parameter (6.26 patients per hour, the empirical average from 10:00 AM to 4:30 PM (see fig. 2)).

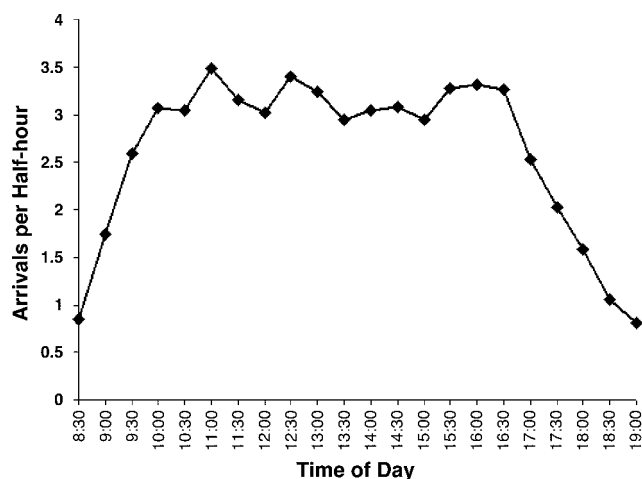


Fig. 2. The rate of new patient arrivals into the postanesthesia care unit (PACU) as a function of clock time (in 0.5-h increments). PACU arrival rate is approximately constant from 10 AM to 4:30 PM.

same as found for 2006. We also segmented the data by calendar quarter or by day of the week (*i.e.*, Mondays *vs.* Tuesdays *vs.* Wednesdays, *etc.*), and we still found the same arrival rate during the steady-state period. We concluded that the work released into the PACU was constant during the study. The impact of λ that changes throughout the day and how to deal with such changes will be addressed in the Discussion section.

The result above indicates that the fundamental modeling assumption (randomly distributed, steady stream of arrivals) is reasonable for our setting, but it is silent about whether the simple queueing model is useful in other hospitals. We attempted to illuminate this question using numerical simulations, overlaying our hospital's case time performance on hypothetical OR suites between 5 and 50 ORs in size. The simulations suggest that exponentially distributed interarrival times and a steady release of work into the PACU is not unique to our setting. Specifically, figure 3 shows that the exponential distribution is descriptive even for as few as five ORs. The steady average arrival rate (fig. 4) also seems to hold rather well under the simulated conditions. Specifically, an exponential process with constant parameter explains more than 99% of the variation in the interarrival times ($R^2 > 0.99$). The results of the simulation have also been listed in table 1. Note that, as can be seen in figure 4, the temporal distribution of arrivals is independent of the number of ORs. That is, although changing the number of ORs will change the absolute number of PACU arrivals, figure 4 and our empirical data demonstrate that the average proportion of arrivals in different hours of the day only depends on the durations of operations and turnover times and the scheduling policy.

Numerical simulations based on averages from our OR time data indicate that the model applies to ORs of various sizes, so we next simulated smaller OR suites using our actual OR time data. We segmented the data

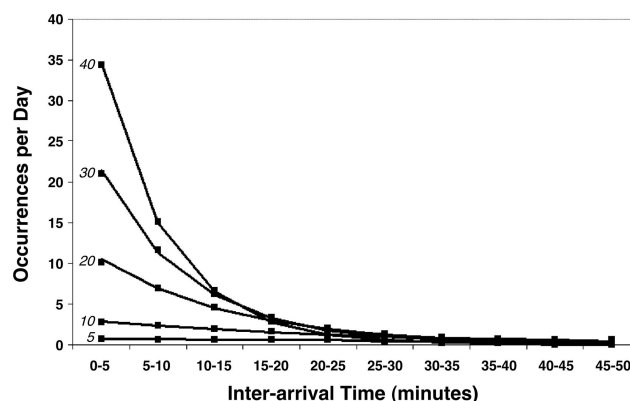


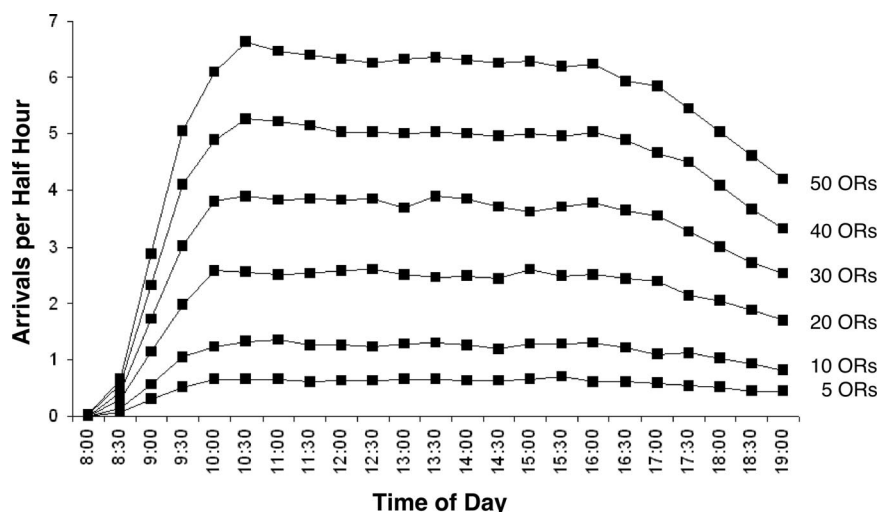
Fig. 3. Results of numerical simulations. Theoretical (lines) and simulated (boxes) interarrival times between completed surgeries (*i.e.*, postanesthesia care unit [PACU] arrivals) for hospitals with different numbers of operating rooms (ORs). The number of ORs in each hypothetical suite (5, 10, 20, 30, or 40) is shown next to the corresponding curve.

from calendar year 2006 to create eight unique, nonoverlapping simulated 5-OR suites typical of smaller hospitals. Then, we calculated the number of completed operations (*i.e.*, PACU arrivals) for each 0.5 h of the day using the actual OR time data from each of the different suites. Figure 5A shows eight different grayscale lines plotting the completed operations from each of the 5-OR suites. The basic shape of each curve is identical to figure 2 (the actual rate of completed operations from the complete OR suite) and figure 4 (the numerical simulations). Specifically, the release of work from each 5-OR suite is constant from 10:00 AM to 4:30 PM. Figure 5B highlights the arrival rate to the PACU from one of the mock 5-OR suites and shows the 95% confidence interval of an constant arrival rate model of 0.36 patients per hour. The actual curve compares quite well to this model.

Comparison of Actual and Modeled Patient Distributions

From the data set of arrivals and departures, it was also possible to infer the number of patients in PACU or waitlisted (*i.e.*, emerged from anesthesia and waiting to enter the PACU) at a given time. This brings us to the next part of our verification effort: checking directly if the distribution of equation (1) fits with empirical data. A central question here is when and if the system, which starts the day empty, reaches the steady-state conditions assumed in equation (1). We must also consider the previously noted issue that the average processing time tended to increase somewhat over the day. Analyzing the distribution of patients at different times of the day revealed the following pattern. In the morning, the OR suite-PACU system was filling, and the number of patients was much smaller than the steady-state equation suggested. However, in the afternoon, the theoretical distribution¹ predicted the actual distribution of patients in the combined (PACU + waitlist) queue, assuming that

Fig. 4. Simulated number of postanesthesia care unit (PACU) arrivals per 30 min, given randomly scheduled cases with log-normal durations. The shape of the curve does not depend on the number operating rooms (ORs). The number of ORs in each hypothetical suite (5, 10, 20, 30, 40, or 50) is shown next to the corresponding curve.



we used a PACU processing time relevant for the corresponding time window (figs. 6 and 7). That is, we measured the average time spent in PACU around 1:00 PM and around 4:00 PM (the values were $1/\mu_{1PM} = 3.21$ h/patient and $1/\mu_{4PM} = 3.64$ h/patient), and we used these values in equation (1). Specifically, we measured the average time spent in PACU+waitlist for patients leaving between 12:00 PM and 2:00 PM and 3:00 PM and 5:00 PM, respectively. Stays longer than 5 h typically represented cases from the previous day, and were truncated at 5 h. We then compared the theoretical curve with the empirical distribution of patients around the same times, and we specifically performed a χ^2 hypothesis test on the 4: PM data, with the bins (≤ 12 , 13–14, 15, 16, . . . 31, 32–33, ≥ 34). The hypothesis that the model could generate the distribution we observed was not rejected ($P = 0.217$). In figures 6 and 7, we have also drawn the ranges within which the occupancy levels would be expected to be 95% of the time.

Survey of Case Sequencing Strategy Implementation

To test the possibility that hospitals might use case sequencing strategies that could invalidate the assumption of constant average arrival rates into the PACU, we

contacted clinicians with knowledge of scheduling practices at 32 other large hospitals and asked about how cases were scheduled. The average number of ORs in the largest suite feeding into the largest PACU was 22, with a range of 6 to 50 and an SD of 10. Thirty institutions reported that the OR schedule was constructed in the order that the cases were booked by the surgeons. Two institutions reported that they had policies mandating ambulatory cases first. Upon closer discussion, it became clear that these policies were not consistently obeyed, that not all services followed them, and that the duration of the ambulatory cases was indistinguishable from the average duration of nonambulatory cases. In other words, there were no scheduling policies in place that would clearly render the arrival rate into the PACU anything but random. This is consistent with the findings of Marcon and Dexter regarding the impact of surgeons' noncoordinated case-sequencing decisions on the pattern of work released into the PACU.¹⁰

Discussion

We developed a queueing model of the joint OR-PACU congestion problem. We then tested the model, first by

Table 1. Number of Observed Occurrences of Different Arrival Times during 1,000 Days

Operating Rooms	Interarrival Time (minutes)											
	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55+
5	12,384	11,676	11,140	10,116	9,494	8,911	8,119	7,582	7,003	6,423	5,941	47,688
10	52,810	44,800	37,442	31,102	25,978	21,423	17,778	14,564	11,974	9,490	7,614	27,283
15	117,623	88,636	66,642	50,236	36,933	27,012	19,704	14,376	10,401	7,491	5,193	11,252
20	202,629	137,845	92,201	61,417	40,753	26,405	17,331	10,931	7,008	4,348	2,806	4,194
25	304,995	185,244	112,134	66,729	39,206	22,645	12,975	7,453	4,127	2,297	1,272	1,408
30	421,010	230,311	123,891	66,505	34,585	17,701	9,006	4,562	2,313	1,114	563	460
35	549,074	269,717	130,576	62,155	28,765	13,379	6,039	2,676	1,106	481	214	141
40	685,789	302,708	131,013	56,274	23,321	9,378	3,867	1,558	560	218	89	55
45	833,007	328,807	128,506	48,665	17,967	6,684	2,337	839	274	92	28	9
50	986,271	350,979	121,702	41,314	13,840	4,388	1,442	473	124	45	11	5

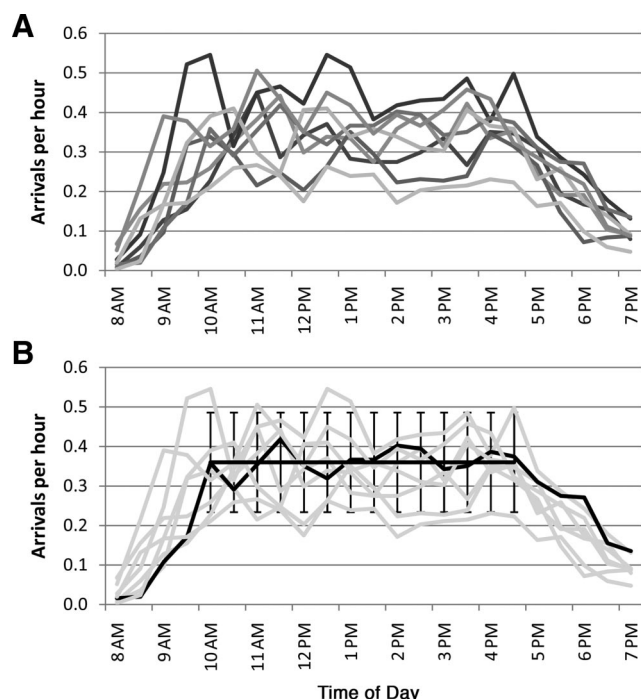


Fig. 5. The rate of new patient arrivals into mock 5-operating room (OR) postanesthesia care units (PACUs) as a function of clock time (in 0.5-h increments). (A) Each grayscale curve shows the arrival rate in 0.5-h intervals into the PACU from 5-OR suites created from unique groups of five ORs from our OR process time database. For each curve, the arrival rate is constant from 10:00 AM to 4:30 PM. (B) One of the arrival rate curves from panel A is highlighted and compared with the 95% confidence interval of a constant arrival rate model with an arrival rate of 0.36 patients per hour.

verifying that the necessary assumptions were applicable and then directly by checking that the predicted occupancy and waitlist levels matched the observed ones during the afternoon (when the PACU has the highest occupancy and also when the model is the most applicable). Naturally, the model omits many of the real-life complexities of perioperative workflow, yet it showed good agreement with current system performance. A

survey and a simulation suggest that our model is applicable for many medium and large hospitals (five or more ORs).

Our study is largely complementary to previously performed work on how to best align nurse staffing schedules with PACU demand,^{11,12} but it addresses a different problem. We assume that nurses are indeed available when needed and that the key system constraint being considered is physical bed availability. Our model might also be useful when staff is the scarce resource; however, we do assume that the scarce resource is constant. Thus, our work may not be applicable in situations when staffing resources are both variable and drive waitlist problems. At least in our hospital, nurse staffing does not seem to be a first-order bottleneck in the PACU environment. In terms of methodology, we depart from previous work by introducing a queueing model, rather than seeking answers only directly from data on past events. While our approach necessitates testable assumptions, it is conducive to managerial insight, and it enabled us to analyze future and hypothetical situations for which data-driven analyses based on past data are not applicable. The latter point is particularly important in the context of long-term planning.

The use of a queueing model requires assumptions that are discussed below. First, we discuss the assumption that interarrival times of patients into the PACU are exponentially distributed. This might at first seem like an unrealistic assumption, because the actual process of generating “completed” operations, which “feeds” the PACU facility, is quite complex. Operations are scheduled at the discretion of the individual surgeons (with variable central coordination); the times for the individual operations have different distributions (typically not exponential); not all patients are sent to the PACU, and numerous sporadic phenomena further complicate the picture (e.g., cancellations, delays, urgent and emergent cases, to name prominent examples).

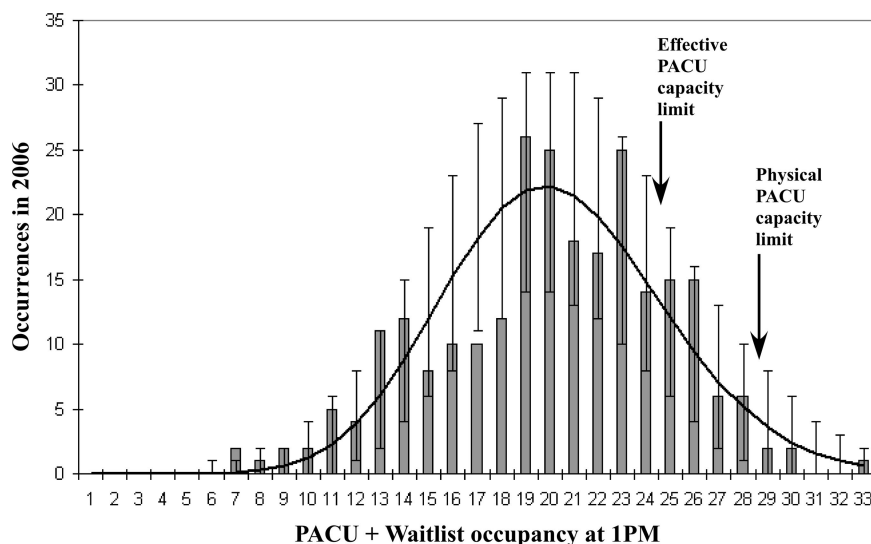
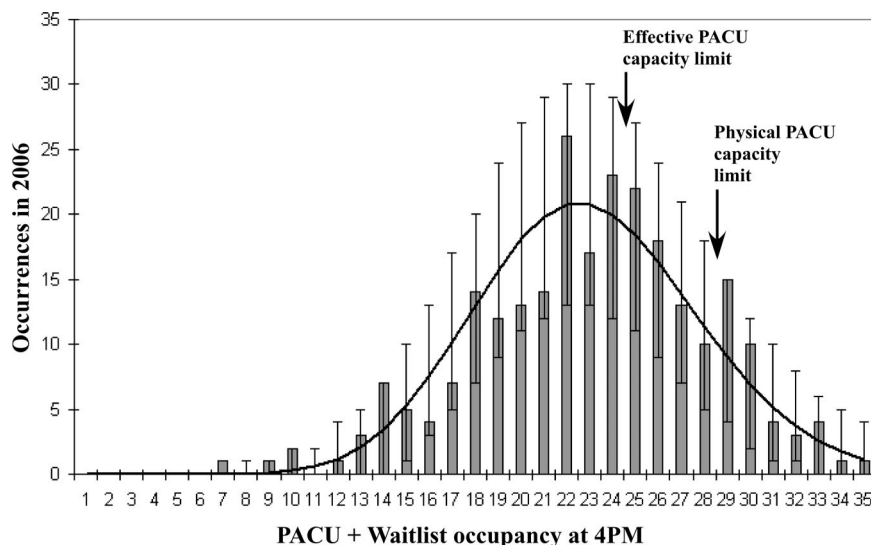


Fig. 6. Total number of patients in the postanesthesia care unit (PACU) and waitlist at 1 PM. Data (bars) versus model (curve with error bars). The horizontal axis shows the total number of patients in the PACU queue at 1:00 PM (sum of patients in the PACU plus those waiting in the operating room (OR) for a PACU bed). The physical capacity of the PACU is 28 patients, but the unit operates at its functional capacity of 24 patients. The vertical axis is the number of occurrences of a given occupancy in 2006. The error bars indicate the 95% confidence interval for getting a given observation if the model is true. The figures show performance from regular workdays only (251 in 2006).

Fig. 7. Total number of patients in the postanesthesia care unit (PACU) and waitlist at 4:00 PM. Data (bars) versus model (curve with error bars). The horizontal axis shows the total number of patients in the PACU queue at 4:00 PM (sum of patients in the PACU plus those waiting in the operating room [OR] for a PACU bed). The functional and physical capacities of the PACU are 24 and 28 patients, respectively. The vertical axis is the number of occurrences of a given occupancy in 2006. The error bars indicate the 95% confidence interval for getting a given observation if the model is true. The figures show performance from regular workdays only (251 in 2006).



However, as already mentioned, these findings are not surprising in light of the Palm-Khintchine theorem, which states that, under certain assumptions, a superposition of many independent arrival processes converges to a Poisson process with the rate equal to the sum of the rates of the individual processes.¹³ This observation was predicted to be true in our numeric simulations (fig. 4) and verified for eight different 5-OR mock suites created by segmenting our actual OR process time data (fig. 5). In figure 5, each mock suite has a different average arrival rate; for each, however, the rate is constant from 10:00 AM to 4:30 PM. This indicates that even for as few as five ORs, the superposition of arrival processes indeed converges to a Poisson process, with the rate equal to the sum of the rates of the individual processes. We observed this for eight mock OR suites created from our own actual data, so it is likely that the key assumption of steady-state arrivals is borne out in many OR suites.

On the other hand, if case durations and turnaround times can be accurately predicted, then it would be possible to engineer other arrival patterns by enforcing particular scheduling policies (e.g., shortest case first). Indeed, such policies have been proposed as a means to reduce PACU congestion.⁷ We think such benefits are difficult to realize in practice, especially at large hospitals, partly because of organizational resistance (patients and surgeons have other considerations than PACU congestion when scheduling cases) and partly because of substantial divergences between scheduled and actual events in real-world hospital environments. Finally, Marcon and Dexter found that the scheduling policies used in actual practice had little impact on peak PACU patient load.¹⁰

Additional Modeling Considerations

Our results above appear to leave the major assumption of the queueing model intact, but there are a few other factors to consider. First, we note that because we

model PACU/waitlist arrivals as an exogenous process, we are not capturing the feedback effect that waitlist cases can delay subsequent operations. We think this is reasonable in the present context because, as we shall see, our model is mostly used in the afternoon, and cases started (and potentially delayed) during this period usually end later in the evening, when the congestion has usually cleared up. Moreover, at our hospital, administrators preferentially allot the available PACU beds to those ORs with further cases to perform.

The data analysis involved in this verification effort also gave us a clearer picture of the waitlist problem in general. As is apparent from figures 6 and 7, waiting for slots in the PACU was common. This is particularly true at 4:00 PM (fig. 7). Between 3:00 PM and 5 PM on any given day, 5 to 10% of all ORs were waiting for a PACU bed. This observation is confirmed by a query of our anesthesia information management system, where anesthesia personnel report delays waiting for PACU beds (not shown). Another illuminating analysis was to directly plot (fig. 8) the number of waitlisted (patients waiting in the OR) cases as a function of the hour of the day, for three representative weeks. The pattern in figure 8 is that of a full hospital. Waitlists peak in the middle of the week, when hospital beds are occupied by patients from earlier in the week. Then, the congestion eases later in the week as patients begin to be discharged.

Digressing slightly, we point out that it is possible to use figure 8 (or rather, the underlying data set) to give a data-driven answer to the question "How many additional PACU beds are needed to reduce the waitlist?" Each new bed added to the system has the approximate impact of removing one waitlisted patient, which can be intuitively visualized as lowering the curve in figure 8 one step (but never below zero). This way of calculating the waitlist reduction is quite unlike the mathematical queueing analysis that is the focus of this paper, and fewer assumptions are necessary. However, only the

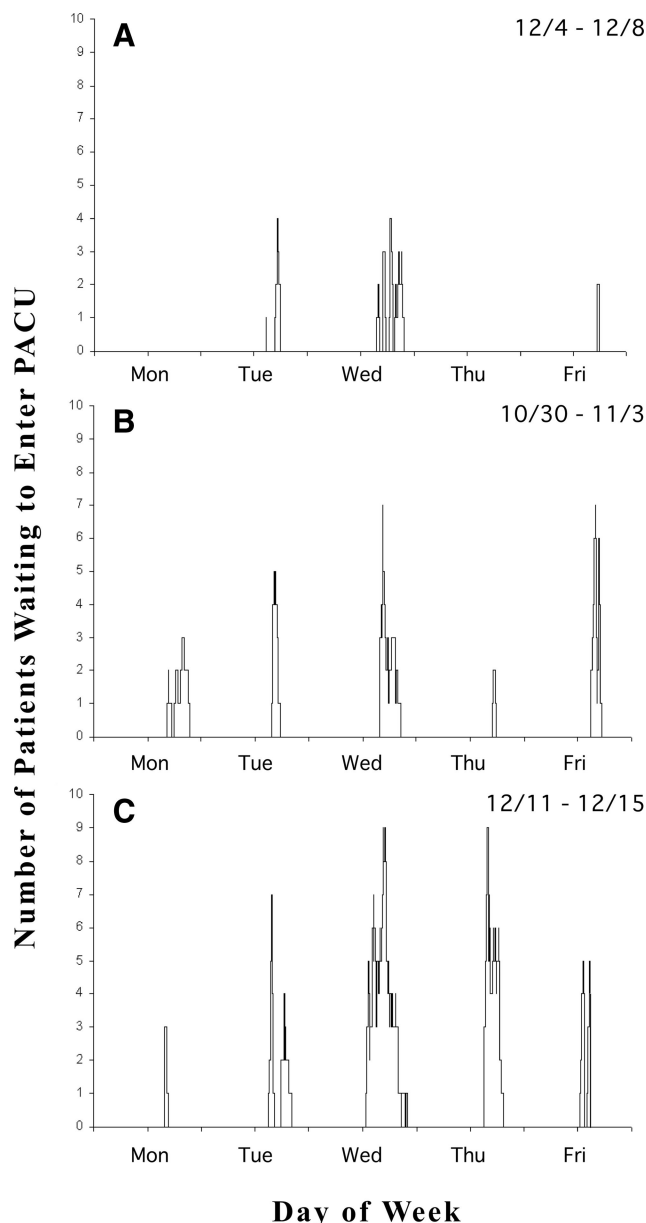


Fig. 8. Timeline of five successive weekdays from three representative workweeks, showing the number of patients on a waitlist to enter the postanesthesia care unit (PACU) as a function of time of day. Panels *A*, *B* and *C* illustrate progressively larger waitlists to enter the PACU. The time scale is continuous, with *ticks* at midnight and noon for each day. The day of week labels are below the noon ticks.

impact of additional PACU beds can be calculated in this way, not the impact of changed caseload or length of stay.

A potential limitation of our approach, as discussed in detail in earlier sections, is that our mathematical model relies on specific assumptions about the various aspects of hospital operations. Although these assumptions seemed to hold quite well in our setting, they may of course be less applicable elsewhere. For example, our model assumes a constant rate of arrivals into the PACU. Figure 2 indicates that this assumption is accurate in our

environment, and our numerical simulations indicate that it is robust to the overall size of the OR suite (figs. 3 and 4). Furthermore, others have also observed that the net result of multiple ORs sending work to the PACU is a steady workload.¹⁰ However, this generalization must be treated cautiously. In one instance, the observed arrival rates are best described by a triangular distribution—one in which the arrival rate is never constant.¹⁴ Investigators wishing to use our model in a different hospital would have to check that these assumptions are applicable in that setting. Specifically, if a hospital employed case sequencing strategies that could make the release of work into the PACU nonrandom, then the simple queueing model might not work. In that instance, it would be necessary to compare the actual and predicted distributions of patients in the combined PACU + waitlist queues by a goodness of fit test.

Numerical simulations and theoretical considerations suggest that the simple queueing model is most likely to apply in hospitals that are midsized or large, where case scheduling occurs without coordinating policies. We note that if caseload varies consistently by day of week or by season, then one can evaluate the model (*i.e.*, measure the arrival rate and recovery time for each scenario) for each day of the week and plan capacity accordingly. For example, in a hypothetical 5-OR suite in which one of the ORs releases one 8-h case into the PACU on Mondays and 15 short cases per day into the PACU on Tuesdays, it is likely that separate model parameters will be needed for the 2 days.

Differences between hospitals that are nevertheless consistently present within each hospital are well handled by our approach. For example, in a physician-only practice with no one assigned to the PACU, patients may wait to leave the PACU (and hence, other patients may wait to enter the PACU) because no physician is available to sign patients out of the PACU. However, this source of waiting will be subsumed into the institution's measurement of the PACU recovery time.

One could envision modified and expanded queueing models for systems with other idiosyncrasies; although there is generally no guarantee that a given system can be described with a compact formula such as equation (1). An alternative is to use numerical discrete-event simulation, which can typically handle more complex situations. For example, the simulation study by Marcon *et al.*¹⁵ captures several aspects (such as transfer time and porter availability) not explicitly encompassed by our model. On the other hand, analytical formulae such as equation (1) also have benefits; they reveal key relationships in a compact and instructive way.

Because there was good agreement between data and model in our setting, we are in a position to make predictions about how the waitlist problem would be affected by changes in the time patients spend in the PACU, the number of PACU beds, and the OR caseload.

In particular, this might become useful in the workflow planning of the new OR building that is currently under construction at our hospital. We thus hope that the results presented here will complement past research on PACU congestion, which has primarily focused on the impact of various managerial policies.⁷

We will first consider changes in the average waitlist-PACU length of stay $1/\mu$. It should be clear, both intuitively and from equation (1), that if $1/\mu$ could be decreased, for example by actions that speed recovery, the waitlist will be reduced as well. As noted, the PACU length of stay parameter is calculated by including time spent waiting for beds to become available elsewhere in the hospital (when relevant). Therefore, if the waiting time to leave the PACU could be reduced, this would reduce $1/\mu$, and then the waitlist will decrease in turn. In fact, we have already implemented “decompressive” strategies by selectively routing patients so that they are released directly to home from PACU,^{16,17} rather than first going to (and potentially waiting for) other hospital beds. According to the model, if such efforts could have a meaningful impact on the average length of stay, this would translate to a reduced waitlist problem as well. However, our previous work in this setting suggests that, at least for one patient population, the practical impact of this strategy was quite limited.¹⁶ On the other hand, we also had data (not shown) on what portion of time in PACU was spent waiting for beds elsewhere. If this portion could be taken out of $1/\mu$, then calculating waitlists from equations (1) to (3) suggests that the OR-to-PACU waitlist would be all but eliminated. In this sense, the PACU is not the “true” capacity constraint at our hospital.

The exit condition from the PACU (full hospital *vs.* free access) is important because it affects the impact of interventions to reduce $1/\mu$ by speeding ‘wakeup’ in the PACU. If much of $1/\mu$ is attributable to waiting for hospital beds then it is not beneficial to expend resources to speed recovery.

A potentially stronger way to model the leaving-the-PACU waitlist problem would be to expand our queueing model to dynamically model congestion in other parts of the hospital and how this interacts with the PACU congestion. Our hospital is physically constrained from adding additional inpatient beds in the next 5 yr, so we leave such improvements in modeling technique for future research. However, adding PACU beds directly to the OR suite was a feasible option, which we will discuss shortly.

We will now turn our attention to predicting the impact of changed N and λ . As we have seen, because of staff availability and especially turnover times for PACU beds, the effective PACU capacity is less than the actual physical capacity at our hospital; it is the former that should be used as N in our model. If one could bring the effective capacity closer to physical capacity, for exam-

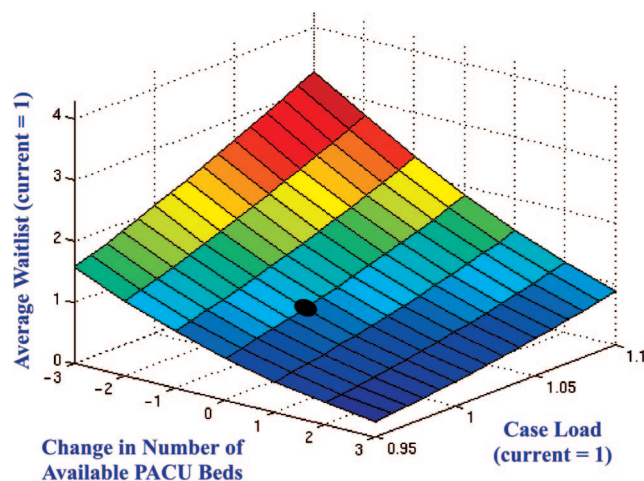


Fig. 9. Three-dimensional surface of cumulative postanesthesia care unit (PACU) waitlist time as a function of PACU capacity and operating room (OR) caseload. Current state is indicated by the black dot. Functional PACU capacity is limited to 24 beds. This gives an average weekly cumulative PACU waitlist of about 21 patient-hours in the current state.

ple, by using improved nurse scheduling procedures⁷ and/or by reducing PACU turnover times, then the model would predict the resulting reduction of waitlists. We primarily wanted to understand the impact on waitlists of changing demand or changing PACU capacity; for a small increase it seems reasonable to assume that the effective number of beds will increase by approximately the same amount (*e.g.*, that if the physical capacity goes from 28 to 31, then the effective capacity will go from 24 to 27).

Figure 9 shows a three-dimensional surface predicting the interaction among caseload, effective PACU capacity, and waiting time for PACU beds based on the M/G/ ∞ queueing model. The current state, shown as “1” on the left axis represents 21 cumulative hours per week of waiting for PACU beds. Using the model, we can predict with our current PACU capacity that raising surgical volume by 5% or reducing staffed PACU beds by 2 (the impact of a sick call by a nurse) would effectively double the amount of waiting for PACU slots. On the other hand, adding three staffed PACU beds, bringing the total to 31 (and the effective capacity to 27), would reduce waiting from approximately 21 h per week to 8 bed-hours per week. Stated differently, a capacity expansion of about 10% would reduce the congestion problem by approximately 60%. This very sensitive relationship may seem surprising, but it is in fact rather typical for congested queueing systems (see figure 3.4 in reference 18). It is also in good agreement with the result of the data-driven analysis outlined at the end of the Results section (*i.e.*, lowering the curve in fig. 8). It should also be acknowledged, with reference to figure 9, that the marked sensitivity to small changes in the input parameters makes the model results sensitive to estimation errors.

We have constructed a downloadable spreadsheet based on our simple model (see model, Supplemental Digital Content 1, an Excel [Microsoft, Redmond, WA] spreadsheet that readers can use to model their own OR suites; given input on PACU load and characteristics, the model predicts distribution of occupancy and waitlists during peak hours, <http://links.lww.com/A1197>). Readers can apply the model to their own setting (with due attention to the limitations and assumptions described above) by inputting values for PACU arrival rate, length of stay, and effective number of staffed PACU beds. The spreadsheet returns "average number of waitlist cases," "probability of at least one waitlist case," and the probability distribution for having a given number of patients in the combined PACU-OR system under steady-state conditions. The value for "average number of waitlist cases" actually means "average number at any time during steady-state conditions." If an OR suite's configuration returns a value of 1.37 for this field and the OR is in steady-state for 3 h per day then the model predicts $3 \times 1.37 = 4.11$ patient-hours per day of waitlists. However the model does not provide information about whether this is distributed over single or multiple cases. Of course, to make a decision, managers must consider not only waitlists, but numerous other monetary and nonmonetary factors, many of which will no doubt vary from institution to institution. Our model does not compute an optimal number of beds; rather, it is intended as a decision-support tool. The spreadsheet allows simple modeling of scenarios such as the one described above, wherein one OR of a small suite releases one case per day to the PACU on Mondays but 15 cases per day on Tuesdays. Under this extreme example situation, the arrival rate will definitely change between days, in turn changing the probability of PACU waitlists if capacity is kept constant.

Further reference to figure 9 indicates that adding three PACU beds would allow case volume to grow by 10% relative to current demand while still encountering fewer hours of waiting than in the current configuration. What would be the cost or benefit of making such a change? The construction costs are unique to each hospital, but they are quantifiable. Similarly, the cost of additional PACU staff can be estimated. Our hospital pays overtime to anesthesiologists and nurses to recover patients in the OR after regular OR hours or to finish cases that were delayed because of PACU waitlists earlier in the day. There is a tangible cost (in terms decreased margin for the cases using overtime) and an intangible cost (in terms of staff frustration, demoralization and turnover) of not reducing the waitlist. If a hospital had waitlists with none of these costs, then there is little reason to use resources to reduce the waitlist.

In separate but complementary work, we have been developing a pod of ORs designed to implement parallel processing to reduce the total process time for cases and raise throughput in these ORs.¹⁹⁻²⁴ Preliminary analysis of this strategy indicates that there would be sufficient time

saved and OR capacity realized to close a small OR and convert it to PACU space,^{25,26} with the impact of adding a net three PACU beds to the perioperative system. Thus, the empirical initiative to expand parallel processing in our ORs and the queueing model work described in this paper provide convenient opportunities to test the validity of both approaches and provide methods by which other congested academic medical centers may increase their functional capacity for growth in OR patient volume.

References

1. Erlang AK: The theory of probabilities and telephone conversations. *Nyt Tidsskrift Mat B* 1909; 20:33-9
2. Gans N, Koole G, Mandelbaum A: Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* 2003; 5:79-141
3. Gorunescu F, McClean SI, Millard PH: Using a queueing model to help plan bed allocation in a department of geriatric medicine. *Health Care Manag Sci* 2002; 5:307-12
4. Elkhuizen SG, Das SF, Bakker PJ, Hontelez JA: Using computer simulation to reduce access time for outpatient departments. *Qual Saf Health Care* 2007; 16:382-6
5. McManus ML, Long MC, Cooper A, Litvak E: Queueing theory accurately models the need for critical care resources. *ANESTHESIOLOGY* 2004; 100:1271-6
6. McManus ML, Long MC, Cooper A, Mandell J, Berwick DM, Pagano M, Litvak E: Variability in surgical caseload and access to intensive care services. *ANESTHESIOLOGY* 2003; 98:1491-6
7. Dexter F, Epstein RH, Marcon E, de Matta R: Strategies to reduce delays in admission into a postanesthesia care unit from operating rooms. *J Perianesth Nurs* 2005; 20:92-102
8. Kleinrock L: *Queueing Systems, Volume 1: Theory*. New York, Wiley-Interscience, 1975, pp 234
9. Strum DP, Sampson AR, May JH, Vargas LG: Surgeon and type of anesthesia predict variability in surgical procedure times. *ANESTHESIOLOGY* 2000; 92:1454-66
10. Marcon E, Dexter F: An observational study of surgeons' sequencing of cases and its impact on postanesthesia care unit and holding area staffing requirements at hospitals. *Anesth Analg* 2007; 105:119-26
11. Dexter F, Epstein RH, Penning DH: Statistical analysis of postanesthesia care unit staffing at a surgical suite with frequent delays in admission from the operating room—a case study. *Anesth Analg* 2001; 92:947-9
12. Dexter F, Wachtel RE, Epstein RH: Impact of average patient acuity on staffing of the phase I PACU. *J Perianesth Nurs* 2006; 21:303-10
13. Heyman DP, Sobel MJ: *Stochastic Models in Operations Research*. New York, New York, McGraw-Hill, 1982, pp 158
14. Dexter F, Tinker JH: Analysis of strategies to decrease postanesthesia care unit costs. *ANESTHESIOLOGY* 1995; 82:94-101
15. Marcon E, Khararaja S, Smolski N, Luquet B, Viale JP: Determining the number of beds in the postanesthesia care unit: A computer simulation flow approach. *Anesth Analg* 2003; 96:1415-23
16. Sandberg WS, Canty T, Sokal SM, Daily B, Berger DL: Financial and operational impact of a direct-from-PACU discharge pathway for laparoscopic cholecystectomy patients. *Surgery* 2006; 140:372-8
17. Seim AR, Andersen B, Berger DL, Sokal SM, Sandberg WS: The effect of direct-from-recovery room discharge of laparoscopic cholecystectomy patients on recovery room workload. *Surg Innov* 2006; 13:257-64
18. Kleinrock L: *Queueing Systems Volume 1: Theory*. New York, Wiley-Interscience, 1975, pp 98
19. Friedman DM, Sokal SM, Chang Y, Berger DL: Increasing operating room efficiency through parallel processing. *Ann Surg* 2006; 243:10-4
20. Hanss R, Buttgerit B, Tonner PH, Bein B, Schleppers A, Steinfath M, Scholz J, Bauer M: Overlapping induction of anesthesia: An analysis of benefits and costs. *ANESTHESIOLOGY* 2005; 103:391-400
21. Harders M, Malangoni MA, Weight S, Sidhu T: Improving operating room efficiency through process redesign. *Surgery* 2006; 140:509-16
22. Krupka DC, Sandberg WS: Operating room design and its impact on operating room economics. *Curr Opin Anaesthesiol* 2006; 19:185-91
23. Sandberg WS, Daily B, Egan M, Stahl JE, Goldman JM, Wiklund RA, Rattner D: Deliberate perioperative systems design improves operating room throughput. *ANESTHESIOLOGY* 2005; 103:406-18
24. Torkki PM, Marjamaa RA, Torkki MI, Kallio PE, Kirvela OA: Use of anesthesia induction rooms can increase the number of urgent orthopedic cases completed within 7 hours. *ANESTHESIOLOGY* 2005; 103:401-5
25. Sokal SM, Chang Y, Craft DL, Sandberg WS, Dunn PF, Berger DL: Surgeon profiling: A key to optimum operating room use. *Arch Surg* 2007; 142:365-70
26. Sokal SM, Craft DL, Chang Y, Sandberg WS, Berger DL: Maximizing operating room and recovery room capacity in an era of constrained resources. *Arch Surg* 2006; 141:389-93; discussion 393-5