Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

SIMULATION OF A MULTIPLE OPERATING ROOM SURGICAL SUITE

Brian T. Denton
Ahmed S. Rahman

Division of Health Care Policy & Research
Mayo Clinic
Rochester, MN 55905, U.S.A.

Heidi Nelson
Angela C. Bailey

Division of Colon and Rectal Surgery
Mayo Clinic
Rochester, MN 55905, U.S.A.

ABSTRACT

Outpatient surgery scheduling involves the coordination of several activities in an uncertain environment. Due to the very customized nature of surgical procedures there is significant uncertainty in the duration of activities related to the intake process, surgical procedure, and recovery process. Furthermore, there are multiple criteria which must be traded off when considering how to schedule surgical procedures including patient waiting, operating room (OR) team waiting, OR idling, and overtime for the surgical suite. Uncertainty combined with the need to tradeoff many criteria makes scheduling a complex task for OR managers. In this article we present a simulation model for a multiple OR surgical suite, describe some of the scheduling challenges, and illustrate how the model can be used as a decisions aid to improve strategic and operational decision making relating to the delivery of surgical services. All results presented are based on real data collected at Mayo Clinic in Rochester, MN.

1 INTRODUCTION

According to the National Coalition on Health Care (<www.nchc.org/facts/cost.shtml>) health care spending in the United States was $1.6 trillion in 2003 which represents 15.3 % of the Gross Domestic Product. Moreover, health care expenditures increased by 7.7 % in 2003, which was four times the rate of inflation. A recent joint study by the National Academy of Engineering and the Institute of Medicine (NAE 2005) states that "[t]he $1.6 trillion health care sector is now mired in deep crises related to safety, quality, cost, and access that pose serious threats to the health and welfare of many Americans." The report also indicates that the health care system is riddled with inefficiencies that significantly increase the cost of medical care and health insurance for employers and workers.

Operating rooms (ORs) have been estimated to account for more than 40% of a hospital's total revenues (HFMA 2005) and a similarly large proportion of their total expenses, which makes them a hospital's largest cost center as well as its greatest revenue source. Furthermore, recent studies indicate that OR efficiency metrics, such as utilization, overtime, and on-time start performance are well off of achievable targets at most hospitals (CAB 2001). Therefore, surgical suite management is an area with significant potential for realizing greater efficiencies within health care organizations. More efficient management will result in improvements to capacity utilization and therefore faster patient access to surgical services. Improvements to capacity utilization will also generate cost savings that can be redirected to basic health care research or to provide wider access to surgical services through new additional capacity investments.

Improving OR efficiency is a computationally challenging problem for several reasons. First, finding a schedule of patient arrival times that balances patient waiting with resource utilization (e.g., OR, surgeon, nurses, etc.) is a combinatorial problem, which includes decisions such as arrivals' sequencing, allocation of patients to ORs, and matching of patients with surgical teams. Second, ORs are not isolated resources, they are typically grouped as part of a surgical suite which houses multiple ORs that share common resources involved in the patient intake and recovery process. Therefore patient flow through the suite is driven by up and downstream resources. Third, there is significant uncertainty in several of the activities involved in the delivery of surgical care, including the surgical procedure itself, which makes advanced planning very difficult. This uncertainty leads to unpredictable waiting time for surgeons, nurses, anesthesiologists, patients, and critical auxiliary resources (e.g., specialized diagnostic equipment), as well as overtime staffing costs associated with late closure of the surgical suit. It also creates the need to balance physician demands (who are best served when an OR is dedicated for their exclusive use) against fiscal prudence (which demands high utilization of OR resources).

The challenge of balancing competing criteria to improve surgery scheduling is not new. (A review of pertinent literature can be found in Section 3). However, the extant literature deals largely with a single-OR and ignores constraints imposed by up and downstream resources. Figure 1 illustrates the flow of patients through a typical multi-OR surgical suite. It illustrates three key activities: intake, surgery, and recovery. All three activities have significant uncertainty associated with them (as we demonstrate in Section 7) which makes the multi-OR surgical suite much complex to model.

The simulation model we describe in this article is part of a long-term ongoing research study to simultaneously improve patient access and surgical suite efficiency. As part of this effort we are investigating answers to the following important questions:

Strategic Questions:

- What is the appropriate number of operating rooms and investment in mobile auxiliary equipment for a surgical suite?
- What is the best ratio of surgeons to ORs?
- What is the optimal number and mix of surgery types to be scheduled on a given day?

Operational Questions

- What is the optimal schedule of patient arrivals for a given number and mix of surgeries?
- How should up and downstream activities be staffed to balance patient flow and staff workload?
- What is the best scheduled time for an urgent add-on case to an existing schedule?

In this article we illustrate our progress towards answering these questions. We describe the design and operation of a general multi-OR surgical suite, discuss the details of our monte-carlo simulation model, and present numerical results based on real data to illustrate how the model can be used to answer some of the important strategic and operational questions described above to improve the delivery of surgical services. Our numerical results are based on real data collected for the outpatient endoscopy suite at Mayo Clinic, in Rochester MN.

2 SURGERY SCHEDULING PROCESS

Perioperative services encompass all stages of surgery including *preoperative*, *intraoperative*, and *postoperative* stages of patient care. Preoperative care begins with the patients decision to have surgery, and ends with the transfer of the patient to the OR bed. It can include a variety of activities such as patient education, a patient visit to an anesthesia outpatient clinic, preparation for the day of surgery, and arrival at the designated location for surgery. Intraoperative care is defined as the time between when the patient reaches the OR bed, and the time when they are admitted to the recovery area which may be a post-anesthesia care unit (PACU), Intensive Care Unit (ICU), or other post-procedure recovery area. Postoperative care is the time between arrival in the recovery area and the time that the surgeon terminates follow-up care with the patient. Each of these stages are critical to the successful delivery of surgical services to the patient.

There are different types of surgery delivery systems. Hospitals provide many services, and are typically equipped with a broad range of capabilities including an emergency department for handling cases resulting from unpredictable adverse events. More recently, a new delivery system called Ambulatory Service Centers (ASCs) have emerged (Bowers and Mould 2005). ASCs service elective (equivalently *deferrable* and *scheduled*) surgeries that can be performed safely in an outpatient setting with minimal supporting resource. At hospitals there are different degrees of urgency associated with patient care. Often surgery can be performed on an elective basis on an agreed upon future date. This is true of many types of surgery in which there is not an immediate need for intervention. Urgent or emergent cases, on the other hand, are cases in which the timing is critical. Urgent cases arise on short notice and speed of intervention directly affects the patient's safety and potential for recovery. Emergent cases typically arrive at an emergency department via ambulance or airlift in need of immediate attention. Urgent and emergent cases are simultaneously high priority and unpredictable, and therefore present difficult challenges from a planning perspective.

Whether surgery is performed on an inpatient or outpatient basis, at an ASC or hospital, or on an elective or emergent basis, many aspects of the OR environment are the same. From a facilities perspective ORs tend to be housed in a suite, in which several individual ORs are located that share central resources such as an equipment storage area, sterilization resources, preoperative and recovery rooms. From a staffing perspective the OR team is composed of a variety of uniquely skilled individuals including the surgeon, one or more surgical assistants, an anesthesiologist, nurse anesthetist, and a scrub person.

ORs have very high fixed costs, the large proportion of which is associated with the labor cost of the OR team. Typically ORs have a planned utilization time (e.g., 8 hours) beyond which overtime costs for some members of the OR Team begin to accrue. Therefore on-time surgery start performance, to the extent it affects overtime, is an important metric. Efficient surgery scheduling also affects the amount of waiting for surgeons, anesthesiologists, OR Teams, and other critical resources.

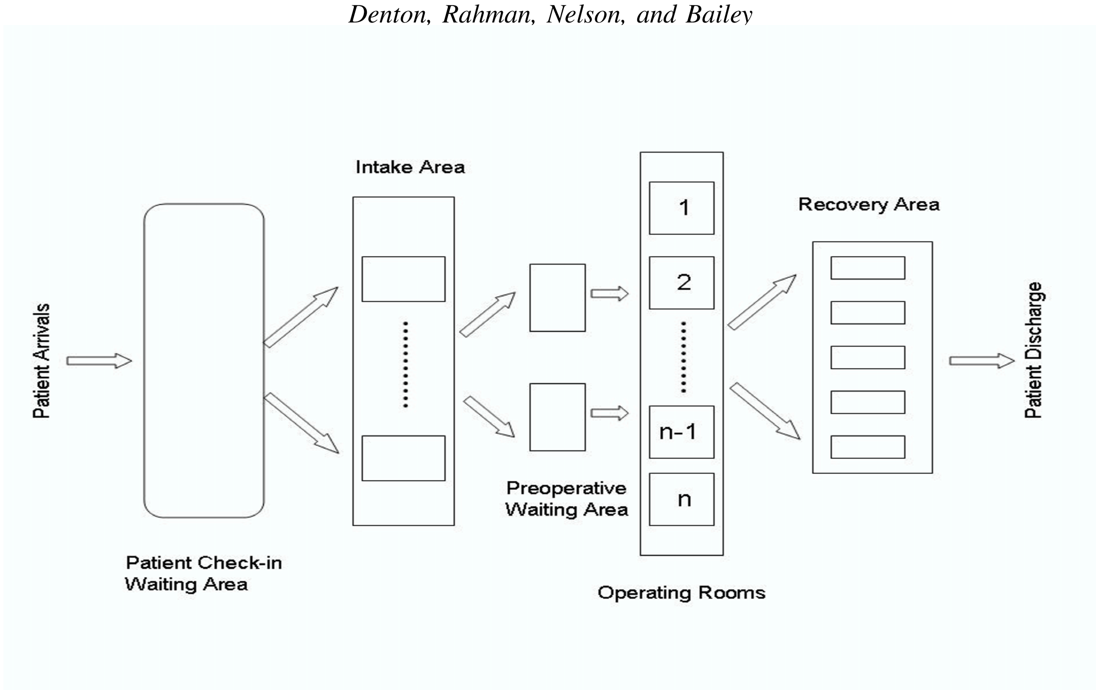

Figure 1: Illustration of Inpatient Patient Flow Through a Multi-OR Surgical Suite from Arrival to Recovery.

There are two well known processes for advance scheduling of surgeries known as *block-booking* and *open-booking*. Under a block-booking system individual surgeons or surgical groups are assigned times in a particular OR in a periodic (e.g., weekly or monthly) schedule. Each period surgeons book cases into their assigned block time. On the other hand, in open-booking systems surgeons submit cases up until the day of surgery and by-and-large all accepted cases are scheduled. Individual surgeries are then allocated to ORs to create a schedule immediately prior to the day of surgery.

The simulation model we describe in this article is motivated by the endoscopy suite at Mayo Clinic. The suite is dedicated for colorectal screenings. It is located in an outpatient setting and all procedures are scheduled in advance of the day of surgery. Patients are scheduled by allocating them to a *slot* which is associated with a predefined arrival time on a particular day. As opposed to the more complete perioperative process described above, our focus in this article is on the delivery of the surgical procedure itself, from patient check-in to discharge, since this activity is simultaneously the most costly, the most affected by uncertainty, and the most difficult to plan.

Figure 2 illustrates the typical process by which a patient navigates the endoscopy suite on the day of surgery. Upon arrival patients check-in and wait for a nurse to initiate the intake process. Intake involves several activities including pre-surgery consultation with a nurse, transfer to the change room, change of dress, and transfer of the patient to an OR, or a pre-operative waiting area. For our purposes we define the start of the surgical procedure as the point at which the OR Team arrives to begin preparation (e.g., administration of IV, monitoring). When the procedure is complete the patient waits to be transferred by a nurse to the recovery area. Transfer is contingent on a recovery bed being available, and adequate nurse coverage (nurse to patient ratio) in the recovery area. From the patient's perspective recovery from the affects of anesthesia begins immediately after surgery is complete, whether they wait in the OR or are transferred directly to the recovery area. However, if the patient waits in the OR they must be monitored by the OR Team nurse, which may delay the start of the next surgery. Thus the smooth flow of patients to and from the OR may be affected by uncertainty in the duration of the intake process, as well as uncertainty in the duration of recovery which affects the number of beds available.

3 LITERATURE REVIEW

More extensive reviews than the following can be found in (Blake and Carter 1997; Goldman, Knappenberger, and Shearson 1970; Magerlein and Martin 1978; Przasnyski 1986). In our brief review we separate the literature into two areas that are relevant to the simulation model we present, long range systems design decisions, and short range advance scheduling decisions.

3.1 System Design

System design is concerned with long-term strategic decisions such as the number of ORs to be located at a facility, investment in equipment resources (e.g., diagnostic tools), and decisions about the specialization of ORs for certain types of surgery. The literature in this area considers policies

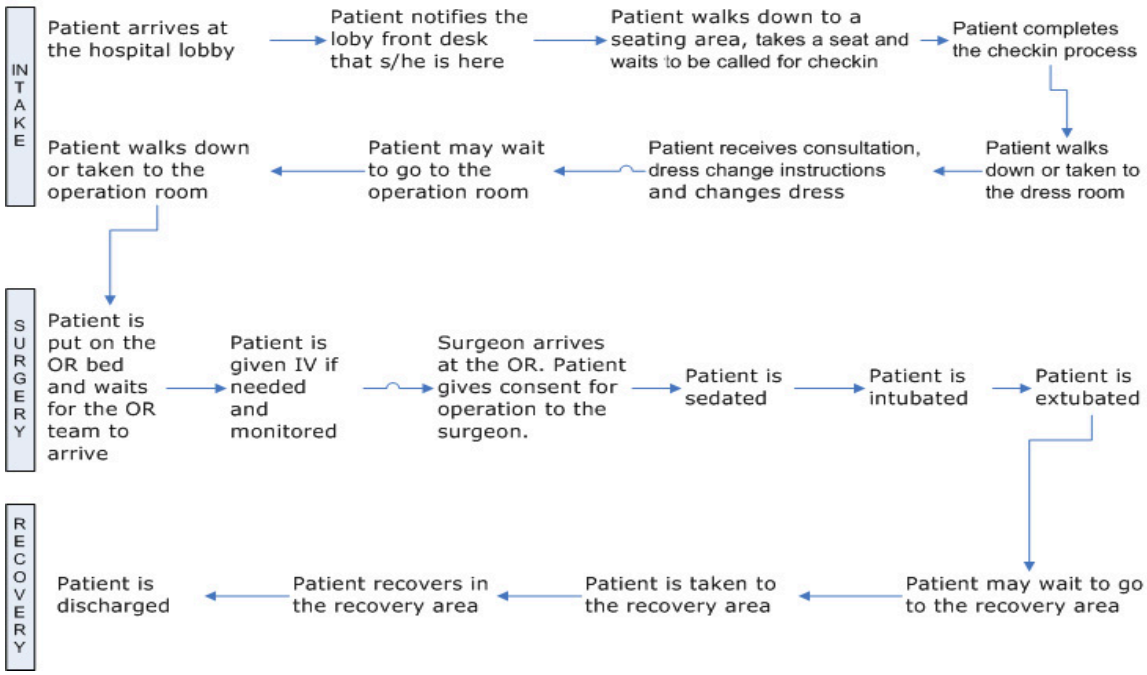

Figure 2: Illustration of the Detailed Process for Patient Flow Through the Endoscopy Suite from Check-In to Discharge.

for how to organize surgical schedules (e.g., block-booking or open-booking), and which types of surgeries to accommodate at a particular facility. For instance, in (Bowers and Mould 2005) the authors consider the effects of introducing ASCs by reassigning elective cases from hospitals to ASCs. They evaluate the combined effect on delivery system efficiency resulting from the reallocation of specific high volume procedure types from hospitals to ASCs. In (Lovejoy and Li 2002) the authors present a multi-criteria stochastic model for OR capacity expansion decisions. They describe optimization models that reflect the goals of several constituents involved in the decision making process. In (Dexter, Ledolter, and Wachtel 2005) the authors discuss models for evaluating OR expansion decisions and they provide examples based on real data. Their model considers tactical decisions for how to allocate time among surgical specialty areas based on demand estimates and financial criteria.

3.2 Advance Scheduling

Advance scheduling involves allocating OR time among surgical groups in advance of the day of surgery. Of the two areas we describe this is the most developed. The single-OR scheduling problem is the simplest version of the advance scheduling problem. It concerns the setting of start times in the presence of uncertainty in surgery durations. The objective is to balance relevant metrics including surgeon and OR Team waiting, patient waiting, idling of the OR, and overtime costs for running later than the scheduled closing time. The evaluation of the expectations of these quantities requires the evaluation of multidimensional integrals that typically have no closed-form solutions. Accordingly, a number of previous studies have used simulation to study the performance of heuristic rules for setting start times. In (Ho and Lau 1992) monte-carlo simulation is used to compare the performance of many simple scheduling heuristics.

The single-OR scheduling problem also arises in many other contexts in which appointment decisions are economically significant. In Sabria and Daganzo (1989) the authors consider scheduling the arrival of cargo ships at a seaport. In their treatment of the problem the costs of underutilization of a seaport are traded off against the cost of cargo ship waiting. On the other hand, Wang (1993) discusses the problem in a manufacturing setting where the objective is to schedule the arrival of parts on the shop floor such that work-in-process inventory and machine idling are minimized. There have been numerous other simulation and queuing based studies presented in operations research, statistics, and health care journals over the past three decades on the problem of assigning start time for surgeries and outpatient clinic appointments (for example, Bailey 1952, Charnetski 1984, Dexter et al. 1999, Ho and Lau 1992, Jansson 1966, Mercer 1973, Rohleder and Klassen 2002, Soriano 1966, Welch 1964, and references therein).

Another avenue of research for single-OR scheduling is the study of optimization models. In Weiss (1990) and Robinson, Gerchak, and Gupta (1996) the authors solve two and three surgery problems, respectively, which can be solved relatively easily owing to the low dimensionality.

Special cases for $n > 3$ are considered in Wang (1993) in which job durations are exponentially distributed and computational advantages of phase-type distributions can be exploited. In Vanden Bosch and Dietz (2000) the authors present an algorithm for a similar problem for the case of phase-type distributions in which appointment slots are integer multiples of a discrete slot parameter. Denton and Gupta (2003) study a general two-stage stochastic linear programming formulation of the OR scheduling problem and provide efficient methods for solving larger instances of the problem.

The literature on advance planning in the context of multiple ORs is very sparse. Notable references include the following. Blake and Donald (2002) present a deterministic integer programming formulation of a model for setting block-booking schedules for multiple ORs. Dexter, Epstein, and Marcon (2005) consider policies under a block-booking schedule in which unutilized OR time is released prior to the day of surgery. They consider the trade-offs regarding the timing of release, where such problems are analogous to those found in the revenue management literature.

In contrast to the above referenced literature, our goal is to use a simulation model to study the impact of uncertainty on strategic design and operational scheduling decisions in the multi-OR context. In addition to considering multiple ORs we also explicitly model the dependency of performance metrics on up and downstream resource availability. Therefore our model incorporates many of the criteria described in the referenced articles, while allowing a deeper analysis of the broader delivery system.

4 SINGLE-OR OPTIMIZATION MODEL

The single-OR scheduling problem is similar to the S(n)/G/1 queuing model in which n customers arrive at a server according to a deterministic schedule of arrival times, $\{a_i, i = 1, ..., n\}$), to receive service of uncertain duration with probability distribution $G(\cdot)$. To define the model for this problem we let Z_i denote the random surgery duration for surgery i, and let W_i and S_i denote the waiting and idling times associated with surgery i respectively. To simplify the articulation of our model we let x_i denote the time allocated for surgical procedure i (note that specifying x_i is equivalent to specifying a_i). In our model expected waiting and idling times represent performance metrics. The waiting and idling times can be written as the following recursive equations.

$$\begin{aligned} W_i(\omega) &= \max(W_{i-1}(\omega) + Z_{i-1}(\omega) - x_{i-1}, 0), \forall i, \\ S_i(\omega) &= \max(-W_{i-1}(\omega) - Z_{i-1}(\omega) + x_{i-1}, 0), \forall i. \end{aligned}$$

It is assumed that waiting and idling associated with the first surgery are zero, $W_1(\omega) = S_1(\omega) = 0$. Uncertainty is denoted by a scenario ω that defines the collective outcomes of the random surgery durations, $\mathbf{Z}$, having support $\Xi \subseteq \Re^n$ and probability distribution P on Ξ. In addition to waiting and idling another important performance metric is the *overtime* with respect to an established *length-of-day*, which we denote by L and d respectively. Overtime can be written as

$$L = \max(W_n(\omega) + Z_n(\omega) + \sum_{i=1}^{n-1} x_i - d, 0).$$

The above equations illustrate the dependency of the performance metrics on the selection of arrival times. For instance, if arrivals are spaced out (i.e., x_i's are increased) patients tend to wait less but at the expense of greater idle times and overtime. On the other hand, compressing the schedule results in lower idling and overtime at the expense of greater patient waiting. (See Denton and Gupta 2003 for a thorough treatment of the single-OR model.)

Assuming linear costs for waiting, idling and tardiness, this tradeoff can be modeled explicitly as the following optimization problem:

$$Z = \min_{\mathbf{x}} \left\{ \sum_{i=1}^{n} c_i^w E[W_i] + \sum_{i=1}^{n} c_i^s E[S_i] + c^\ell E[L] \right\}.$$

This unconstrained non-linear optimization problem can be re-formulated as a two-stage stochastic linear program. The structure of this problem was leveraged to develop fast solution methods in Denton and Gupta (2003). In Denton and Vogl (2006) we have investigated potential improvements from the application of the single-OR model using real data, and comparing optimal schedules to actual schedules for an OR at a large urban hospital. The significant improvements we observed for the single operating room problem encouraged our further investigation of the more complex multi-OR surgical suite model to optimize the strategic design and operational scheduling decisions.

5 MULTI-OR SIMULATION MODEL

The full multi-OR surgical suite is much more complex than the single-OR model described above, which makes an optimization model is more difficult. Therefore our initial investigation has been a discrete event simulation model. In Section 6 we describe how this simulation can be used as the basis for a simulated annealing algorithm to compute improved schedules. Figure 2 of Section 1 illustrates the complete set of activities for patients, and Figure 1 represents the flow for our simulation model. On a given day patients arrive according to their assigned appointment times. Patients nearly always arrive on time for their surgery, and therefore the arrival process is reasonably treated as deterministic. Appointments are assigned in advance of the day of surgery according to a predefined set of slots, and the same schedule is used each day of the week. The process begins with patients waiting for a nurse to initiate the intake process. The intake process can be viewed as a set of parallel servers which represent a combination of resources necessary for intake (e.g., nurse, consultation room, change room).

After the intake process patients are taken to the first available OR if an OR is available. Otherwise, they wait in a pre-operative area for an OR to become available. After being taken to the first available OR the patient waits for the OR Team to arrive to begin the procedure. Each OR Team has one or more ORs in the suite pre-assigned for their use on the day of surgery. An OR Team's first surgery of the day commences in any of the available ORs (all ORs are clean and ready for use at the beginning of the day). Once surgery is complete in an OR there is a setup time to prepare the OR for the next patient (e.g., cleaning). If the OR Team has been allocated multiple ORs then the team may move to the next available OR and start surgery immediately, provided the OR has been cleaned and the next patient has been transferred to the room. By and large patients in the endoscopy suite are not pre-assigned a specific OR Team, therefore the first available OR Team treats the first available patient (our model easily accommodates the more general case of pre-assignment of patients to OR Teams as well).

Surgery for a patient that has completed the intake process begins at the later of the following three times (a) their arrival time at the OR (b) the time that an OR becomes available and (c) the time that an OR Team becomes available. Figure 3 illustrates the rotation of OR Teams among ORs based on two different scenarios for the case of a four-OR suite. In Scenario 1 a single OR Team is allocated each of the four ORs, and utilizes each sequentially as the day of surgery progresses. In Scenario 2 two OR Teams share the suite, each having access to two ORs which they circulate through independently.

Upon completion of the surgical procedure the patients recovery process begins. The time for recovery is a random variable which depends on each patients response to the anaesthetic. The process begins upon completion of the surgical procedure independent of whether they are located in the OR or the recovery area. However, until the patient is transferred from the OR to the recovery area preparation of the OR for the next patient can not begin. Furthermore, a nurse from the OR Team must monitor the patient until they are admitted to the recovery area. Therefore, dependence on the availability of a recovery bed downstream may result patients waiting upstream in the preopertive waiting area for a room and/or OR Team to become available.

5.1 Scenario Generation

Each day a predefined number of patients are assigned arrival times to the endoscopy suite. Our model assumes each patient involves three main activities with uncertain durations: intake, surgery, and recovery. Detailed collection of patient flow data through the endoscopy suite began in 2004 and a large database of event timings is now available. The timings include a larger number of activities (see Figure 2) which we have isolated and aggregated into intake, surgery, and recovery. Our simulation model generates a *scenario* by randomly selecting elements from a list of historical samples for each of these three activities, and for each patient scheduled on a particular day. Scenarios are generated by sampling with replacement, with a sample size of 2376 corresponding to each of the three activity durations for endoscopy patients seen in 2005. For the simulations that we report on in Section 7 we use a sample size of $10,000$ scenarios. Our numerical experimentation indicates that this is typically sufficient to achieve a 95% confidence interval that is less than 1% of the mean.

5.2 Performance Measures

Performance measures for the operation of the endoscopy suite fall into two main categories: patient waiting time and overtime of the endoscopy suite. Patient waiting is associated with the first two activities, i.e., patients wait for intake, and wait for surgery. There is no waiting time in the recovery process since recovery begins immediately after surgery, whether the patient is in a recovery bed, or in the OR waiting for a recovery bed. Waiting for intake and surgery is viewed negatively since it increases the total time the patient waits in anticipation of the surgical procedure, which increases total flow time through the suite and is a source of stress for patients. Overtime in the endoscopy suite occurs when the discharge of the last patient occurs after the planned closure time of the suite. Late closure results in overtime costs for nurses and other staff members involved in the operation of the suite. Output from our simulation model includes expected waiting time for each patient for intake and surgery as well as expected overtime. We let W_i^I and W_i^S denote waiting for intake and surgery respectively, and (similar to the single-OR model) we use L to denote overtime. In the numerical examples of the next section we report the following aggregate estimators

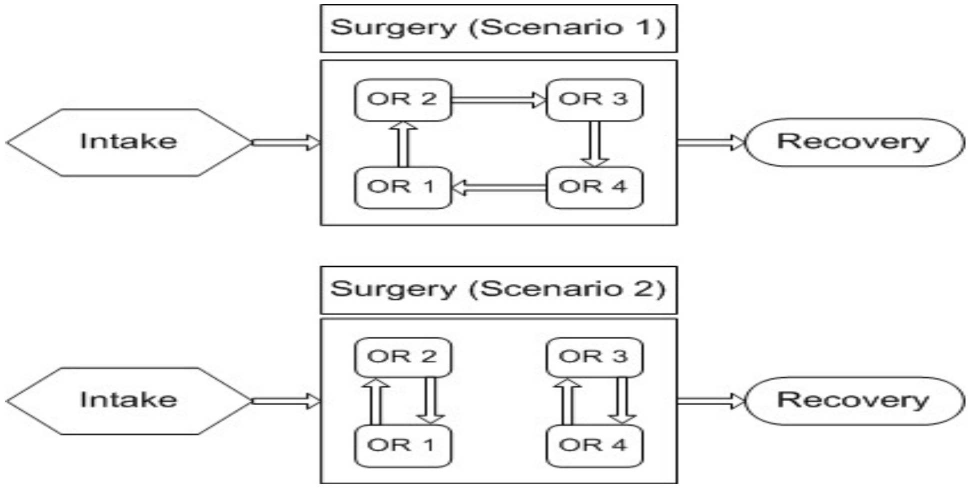

Figure 3: Illustration of the Rotation of OR Teams Among Allocated ORs, where in Scenario 1 (Top) One OR Team Uses 4 ORs, and in Scenario 2 (Bottom) Two OR Teams Use Two ORs Each.

for total waiting time, and overtime.

$$E[W^I] = \sum_{i=1}^{n}\sum_{k=1}^{K}\frac{W_i^I(\omega_k)}{K} \quad (1)$$

$$E[W^S] = \sum_{i=1}^{n}\sum_{k=1}^{K}\frac{W_i^S(\omega_k)}{K} \quad (2)$$

$$E[L] = \sum_{i=1}^{n}\sum_{k=1}^{K}\frac{L(\omega_k)}{K} \quad (3)$$

where ω_k denotes the discrete set of K sampled scenarios ($K = 10000$ in the result presented in Section 7).

6 MULTI-OR OPTIMIZATION MODEL

In this section we briefly describe an implementation of an optimization model that utilizes the monte-carlo simulation model to evaluate schedules. The goal of the model is to minimize an objective function that considers the dual criteria of waiting time and overtime. The optimization problem can be written as

$$Z = \min_{\mathbf{x}} \left\{ c^I E[W^I] + c^S E[W^S] + c^L E[L] \right\} \quad (4)$$

where $\mathbf{x}$ represents the schedule, and c^I, c^S, and c^L, denote cost coefficients associated with expected waiting and intake, waiting at surgery, and overtime.

We use a simple simulated annealing (SA) algorithm to search for improved patient arrival schedules (the reader is referred to a detailed description of SA algorithms in Glover and Kochenberger 2003). Our implementation leverages the intuition of decision makers by starting with the actual schedule used in practice. We consider an initial population of multiple schedules (e.g., 5000) which are generated by progressively perturbing the initial solution by randomly moving arrival times earlier or later in the schedule, with a random step size of 5, 10, or 15 minutes. At each iteration all initial schedules in the population are evaluated using the monte-carlo simulation model. Schedules are retained in the population at each iteration of the algorithm with an *annealing probability* of $exp(-\delta/k)$. The parameter k (referred to as the *temperature*) is selected a priori as an input into the algorithm, and parameter δ is the difference between the current objective function value and the candidate objective function. The algorithm proceeds iteratively with the parameter k becoming progressively lower at rate α per iteration. Accepted schedules at each iteration are further perturbed until k drops to some predetermined value, ε, and the best solution in each iteration is retained.

7 NUMERICAL RESULTS

In this section we provide some summary statistics about the uncertain activities involved in the surgery delivery process and two specific examples to illustrate the types of questions that can be investigated using the simulation model we have developed. The first (strategic planning) example involves the evaluation of alternative staffing models, and the second (operational scheduling) example illustrates the use of the model to improve the assignment of patient arrival times to balance patient waiting with overtime.

7.1 Model Implementation

Our simulation model was implemented in C/C++ on a SUN Unix platform. An open source pseudo-random number generator was used to generate random deviates between 0.0 and 1.0 which were subsequently converted to integers corresponding to elements of an array containing the historical sample time for intake, surgery, and recovery activities. Input data for the simulator includes (a) three files containing sample points for the intake, surgery, and recovery activities (b) a file containing a user defined schedule of patient arrival times (e.g., 7:30am, 8:00am, 8:40am, etc.) and (c) a *model configuration* file that defines several elements for a particular instance of the simulation including: sample size, length of day for the endoscopy suite, turn-over time to clean an OR after surgery, number of servers for each activity, number of ORs per OR Team, and the pseudo-random number generator seed. Output data includes a list of expected waiting times for patients at intake, surgery, and recovery, as well as expected overtime based on the difference between the discharge time of the last patient and the planned length of day (if it is nonzero). Computation time for typical instances of the model are a few seconds with a sample size of $10,000$ on a SUN V440 with CPU speed of 1GHz and 8GB of RAM.

7.2 Summary Statistics

Figure 4 depicts the probability distribution associated with intake, surgery, and recovery. The probability distributions are based on the frequency of observations in 5 minute increments. All activities have a fairly large probability mass in a relatively small interval with a long tail. Such a tail has been noted for many types of surgeries, and results from unexpected or unanticipated events during the surgical procedure which occur with low but finite probability. In the intake process a long tail denotes additional waiting that may be necessary for patients that have not followed pre-surgery directions (e.g., fasting), and long recovery times occur for patients that recover more slowly because of natural variation in patients ability to process the anesthetic after surgery. Summary statistics for the three activities are illustrated in Table 1.

Table 1: Summary Statistics for Intake, Surgery, and Recover.

Activity	*Mean*	*Std. Dev.*	*Min.*	*Max.*
Intake	14.2	7.8	0	115
Surgery	28.4	14.0	3	114
Recovery	50.9	15.3	1	131

7.3 Staffing Example

The allocation ratio of ORs to OR Teams is an important strategic decision that affects long term staffing and capacity investment. Allocating a single OR causes OR Teams to wait between surgeries for their OR to be cleaned, while allocating multiple ORs means the OR Team can move on to the next available OR immediately. However, allocating many ORs to one OR Team may result in some ORs sitting idle between surgeries. Due to uncertainty in the intake, surgery, and recovery processes it is difficult to evaluate the best allocation. We illustrate the use of our simulation model to consider two scenarios (a) a single OR Team utilizing all four ORs and (b) two OR Teams using two ORs each (Figure 2 is an illustration of the two different scenarios). In both cases we assume two intake servers and four recovery beds, which is typical of the actual staffing of a four-OR endoscopy suite. Also, consistent with actual operating procedure, we assume that the planned length of day for the suite is five hours. The schedule for Scenario 2 is based on the actual schedule in use at the endoscopy suite, and that of Scenario 1 is based on the best judgment of the endoscopy suite manager (without the aid of our simulation model).

Since detailed data collection for OR setup times has not been collected, in this example we have used an estimate of 15 minutes for OR setup time based on the informed judgment of the endoscopy suite director. Increasing setup time tends to improve the results for Scenario 1 (since the OR Team is less likely to be delayed by setup time when additional ORs are available) whereas decreasing setup time tends to favor Scenario 2. Based on the results in Table 2 the waiting times for intake and surgery are quite different, with Scenario 2 exhibiting much higher waiting at the surgery stage and lower expected waiting at the intake stage. Total waiting for Scenarios 1 and 2 is 725.5 and 222.8 respectively. Expected overtime for Scenario 1 and 2 is 40.79 and 110.4 respectively. From an OR Team utilization perspective Scenarios 1 and 2 are quite different since Scenario 1 has 16 patients scheduled and Scenario 2 has only 12. Therefore the number of patients seen per OR Team is $1/3$ higher under Scenario 1.

Table 2: Numerical Example Contrasting the Performance Measures Under Two Different OR Allocation Scenarios.

	$E[W^I]$	$E[W^S]$	$E[L]$
Scenario 1	36.4	689.1	40.8
Scenario 2	18.0	204.8	110.4

7.4 Schedule Optimization Example

Selecting arrival times for patients is an important and difficult operational decision. Uncertainty in service durations makes scheduling difficult even in the single-OR case described in Section 4, let alone a multi-OR suite. In this example we use simulated annealing (SA) to find a schedule that simultaneously improves waiting and overtime for Scenario 1. Both criteria were weighted equally in our objective function and the results are presented in Figure 5.

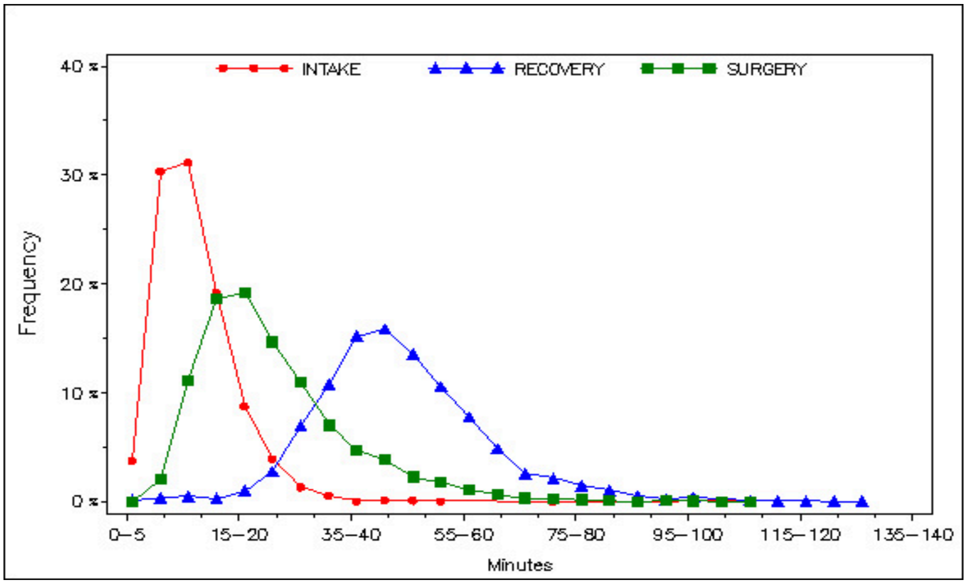

Figure 4: Illustration of the Empirical Probability Distribution for Intake, Surgery, and Recovery Based on Historical Data for 2005

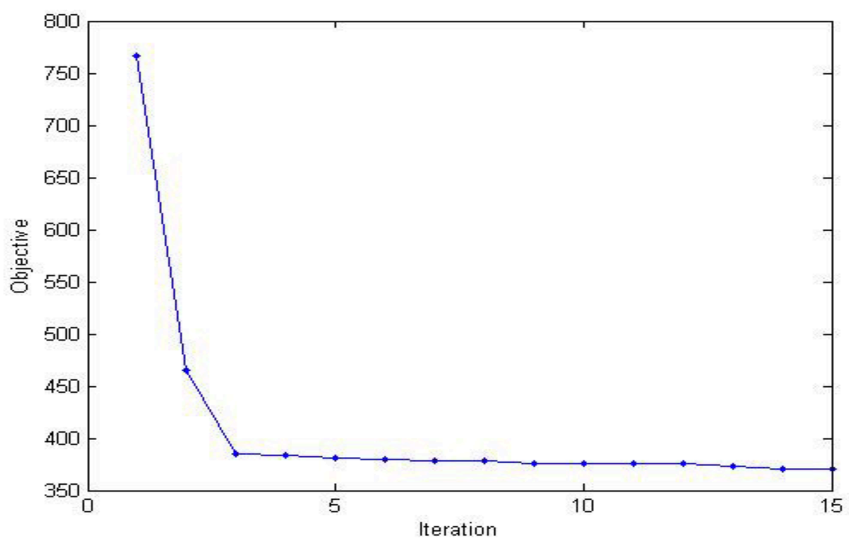

Figure 5: The Total Waiting and Overtime with Respect to Iterations of the Simulated Annealing Algorithm.

The SA algorithm starts with the actual schedule used in practice, and populates an initial solution set by progressively perturbing this schedule. Each generated schedule is used as the seed to generate the next perturbed schedule. A total of $5,000$ first generation schedules are created for the first iteration of the algorithm. Each schedule is evaluated via the monte-carlo simulation model using a sample size of $10,000$. The initial *temperature* is $k = 100$ which is reduced iteratively at a cooling rate of $\alpha = 1$ per iteration to $k = 85$ for a total of 15 iterations. Based on Figure 5 the SA algorithm achieves substantial improvements in early iterations, followed by much slower convergence. Such slow convergence in later iterations is a common characteristic of SA algorithms. In spite of the slow convergence the total improvement with respect to the schedule used in practice is approximately 50%. Therefore the Scenario 1 schedule can be significantly improved using a simple SA algorithm. These promising results encourage future investigation of other more advanced metaheuristics such as genetic algorithms.

8 CONCLUSIONS

This article illustrates the application of a monte-carlo simulation model and simulated annealing to multi-OR surgical suite scheduling based on real data from an outpatient surgical suite. We demonstrate how the model can be used to evaluate multiple competing criteria for different staffing scenarios. Our analysis indicates that even a simple scheduling heuristic based on scheduling of the bottleneck (surgery) activity can lead to simultaneous improvements in expected patient waiting time and overtime. This encourages further future investigation of scheduling decisions via simulation based optimization.

The use of simulation and optimization models for manufacturing systems and other types of service systems are very well established. However, research into the use of such models for surgery scheduling and other service systems is less mature. Using a specific example we have illustrated the use of simulation models to answer open questions about the design and operation of surgical suites. Our initial optimization results encourage future investigation of methods to improve surgical suite schedules. Furthermore, surgical suites are typically part of an even more complex system in hospitals that has dependency on admissions via emergency rooms, outpatient clinic scheduling, and hospital bed resources. Therefore, significant opportunities for future research exist, including expansion of the scope of our model to the broader hospital delivery system.

REFERENCES

Bailey, N. 1952. A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society* A14:185–189.

Blake, J. T., and M. Carter. 1997. Surgical process scheduling: A structured review. *Journal of the Society of Health Systems* 5 (3): 17–30.

Blake, J. T., and J. Donald. 2002. Mount Sinai hospital uses integer programming to allocate operating room time. *Interfaces* 32 (2): 63–73.

Bowers, J., and G. Mould. 2005. Ambulatory care and orthopaedic capacity planning. *Health Care Management Science* 8:41–47.

CAB 2001. Surgical services reform: Executive briefing for clinical leaders. Technical report, Washington D.C.: Clinical Advisory Board.

Charnetski, J. 1984. Scheduling operating room surgical procedure with early and late completion penalty costs. *Journal of Operations Management* 5:91–102.

Denton, B. T., and D. Gupta. 2003. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions* 35:1003–1016.

Denton, B. T., and A. Vogl. 2006. Stochastic optimization of surgery sequencing and start time scheduling decisions. Working Paper.

Dexter, F., R. H. Epstein, and E. Marcon. 2005. Strategies to reduce delays in admission into a post anesthesia care unit from operating rooms. *Journal of PeriAnesthesia Nursing* 20 (2): 92–102.

Dexter, F., J. Ledolter, and R. E. Wachtel. 2005. Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialtys future workload. *Anesthesia & Analgesia* 100:1425–1432.

Dexter, F., A. Macario, R. D. Traub, M. Hopwood, and D. A. Lubarsky. 1999. An operating room scheduling strategy to maximize the use of operating room block time: Computer simulation of patient scheduling and survey of patient preferences for surgical waiting time. *Anesthesia & Analgesia* 89:7–20.

Glover, F., and G. Kochenberger. 2003. *Handbook of heuristics*. Kluwer.

Goldman, J., H. A. Knappenberger, and W. J. Shearson. 1970. A study of the variability of surgical estimates. *Hospital Management* 110:46–46D.

HFMA 2005. Achieving operating room efficiency through process integration. Technical report, Health Care Financial Management Association Report.

Ho, C.-J., and H.-S. Lau. 1992. Minimizing total cost in scheduling outpatient appointments. *Management Science* 38:750–764.

Jansson, B. 1966. Choosing a good appointment system - a study of queues of the type (d,m,1). *Operations Research* 14:292–312.

Lovejoy, W. S., and Y. Li. 2002. Hospital operating room capacity expansion. *Management Science* 48 (11): 1369–1387.

Magerlein, J. M., and J. B. Martin. 1978. Surgical demand scheduling: A review. *Health Services Research* 13 (4): 418–433.

Mercer, A. 1973. Queues with scheduled arrivals: A correction simplification and extension. *Journal of the Royal Statistical Society, Series B* 35:104–116.

NAE 2005. Building a better delivery system: A new engineering/health care partnership. Technical report, National Academy of Engineering and Institute of Medicine.

Przasnyski, Z. 1986. Operating room scheduling: A literature review. *AORN Journal* 44 (1): 67–79.

Robinson, L. W., Y. Gerchak, and D. Gupta. 1996. Appointment times which minimize waiting and facility idleness. Technical report, McMaster University. Working Paper.

Rohleder, T. R., and K. J. Klassen. 2002. Rolling horizon appointment scheduling: A simulation study. *Health Care Management Science* 5:201–209.

Sabria, F., and C. F. Daganzo. 1989. Approximate expressions for queuing systems with scheduling arrivals and established service order. *Transportation Science* 23:159–165.

Soriano, A. 1966. Comparison of two scheduling systems. *Operations Research* 14:388–397.

Vanden Bosch, P. M., and D. C. Dietz. 2000. Minimizing expected waiting time in a medical appointment system. *IIE Transactions* 32:841–848.

Wang, P. P. 1993. Static and dynamic scheduling of customer arrivals to a single-server system. *Naval Research Logistics* 40:345–360.

Weiss, E. N. 1990. Models for determining the estimated start times and case orderings. *IIE Transactions* 22 (2): 143–150.

Welch, J. 1964. Appointment systems in hospital outpatient departments. *Operations Research Quarterly* 15:224–237.

AUTHOR BIOGRAPHIES

BRIAN T. DENTON is an Assistant Professor at Mayo Clinic in the College of Medicine. His primary research interests are in the development and application of OR methods to health care delivery, and his work relating to surgical scheduling won the IIE Transactions outstanding publication award in 2005. He completed his Ph.D. in Management Science at McMaster University in Hamilton, Ontario, Canada. His email address is <denton.brian@mayo.edu>.

AHMED S. RAHMAN is a Senior Information Research Analyst in the Division of Health Care Policy & Research, Mayo Clinic. He holds a bachelors degree in Computer Science from Winona State University, Winona, Minnesota. He has a wide range of experience in database design and development of analytical systems to support outcomes research. He has presented and published several technical papers in conference proceedings. His email address is <rahman.ahmed@mayo.edu>.

HEIDI NELSON chairs the Mayo Clinic Division of Colon and Rectal Surgery and Co-Chairs the GI Program in the Mayo Cancer Center. She has published more than 100 refereed articles. Dr. Nelson attended medical school at the University of Washington and completed surgical residency training at Oregon Health Sciences University. Dr. Nelson was appointed the Fred C. Anderson Professor in May 2005. Her email address is <nelsonh@mayo.edu>.

ANGELA C. BAILEY is the Director of Operations for the Colon and Rectal Surgery and GI Endoscopy Units at Mayo Clinic, Rochester. She did her undergraduate work at the University of Michigan. Angela has her Master of Healthcare Administration from the University of North Carolina with concentrations in both Finance and Quality. Her e-mail address is <bailey.angela1@mayo.edu>.