SURGERY PLANNING AND SCHEDULING

S. Ayca Erdogan
Brian T. Denton
Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina

In 2007, total health-care spending in the United States reached US $2.3 trillion, and continues to rise at the fastest rate in US history. The design and implementation of better planning and scheduling systems is an important area to study in order to reduce high costs and improve access to services in the health-care system. Surgery scheduling, in particular, is an area with significant potential for realizing greater efficiencies. Poor scheduling prevents health-care providers from matching patient demand with available capacity, causing inefficient use of resources, decreased return on investment, and long waiting lists for patients. It has been estimated that surgery accounts for more than 40% of a hospital's total revenues and expenses. Recent studies indicate that resource utilization, overtime, and on-time start performance within surgical suites could be improved at most hospitals. These important performance measures are influenced in part by the surgery scheduling systems and policies that are used in practice.

Surgery scheduling systems impact a variety of expensive resources including operating rooms (ORs), the postanesthesia care unit (PACU), intensive care unit (ICU), hospital beds, equipment resources such as mobile diagnostic imaging devices, and human resources including surgeons, nurses, anesthesiologists, and other staff. The unpredictable nature of surgery results in uncertainty in the duration of surgery and patient recovery. This can be caused by many factors including the varying experience of surgeons and OR teams, the presence of residents or surgical fellows in academic medical centers, or patient characteristics like age, weight, or unobservable physiological factors that only become known at the time of surgery. Unexpected *add-on cases* (urgent surgeries that arise on short notice), patient cancellations, and unanticipated staff or resource unavailability are additional sources of uncertainty that can create problems on the day of surgery.

Surgery scheduling is also difficult because there are a number of competing performance criteria. The efficiency and effectiveness of a surgery schedule may be judged differently depending on the individual's perspective. For instance, OR managers are often concerned with utilization of resources such as ORs, expensive equipment, and staff. Patients, on the other hand, are more concerned with waiting time.

In this article, we provide a general overview of the most important parts of the surgery delivery system that affects surgery scheduling. We focus on the surgical suite itself since it comprises the most expensive resources in the delivery of surgical care. We describe the most significant factors that complicate surgery scheduling, and characterize the types of uncertainty that affect scheduling. Our review focuses on the literature related to surgery scheduling that has appeared in both the operations research and the health-care literature. We particularly emphasize recent studies that consider new and important aspects of surgery scheduling systems. We provide a taxonomy of the literature based on the type of operations research methodology used. We also provide a small example to illustrate how OR methods can be used in surgery scheduling. Finally, we briefly discuss some open challenges and opportunities that exist for future research.

SURGERY SCHEDULING

Surgery may be performed on an *inpatient* or an *outpatient* basis. In the hospital outpatient setting, patients are admitted to the

Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran

hospital and leave the hospital on the same day of the surgery. In such cases the patient recovers at home. In the inpatient setting, patients either arrive on the day of surgery or they are assigned to a hospital bed ahead of the day of surgery. Inpatients stay in the hospital to recover for one or more days following surgery. Owing to increasing volume, many outpatient surgeries are now performed in ambulatory service centers (ASCs), which are distinct from hospitals (though often associated with a hospital). ASCs include an intake area, often several ORs, and a recovery area. Unlike hospitals, ASCs are not equipped for emergencies or overnight stays and most of them are open 8–10 h a day.

The overall surgery process involves several activities before, during, and after the actual surgical procedure. These include *preoperative*, *intraoperative*, and *postoperative* stages. Figure 1 illustrates these stages as well as the related activities. Preoperative care starts with the patient's and surgeon's decision to have surgery. It may involve a visit to a preoperative clinic where an examination and follow-up tests may be completed. It also typically involves some preparation prior to the day of surgery. On the day of surgery, the patient arrives at the hospital or ASC at a preassigned time. A number of administrative activities may be required, and the patient may require a brief physical examination or medical tests (e.g., blood test) before surgery. Intraoperative care comprises the activities performed in the OR. After surgery, the postoperative stage begins and patients are admitted to a recovery area such as a PACU or ICU depending on the nature of the surgery. In the case of inpatient surgery, the patients may return directly to their hospital bed. The complete surgery process ends when a patient has fully recovered and no additional follow-up is needed by the surgeon. This may take weeks, months, or even years. In this review, we focus on activities performed in the surgical suite on the day of surgery. Owing to the high cost of surgical suite resources, this is often a bottleneck in the overall process.

Surgical suites at hospitals must be equipped to handle the regular daily surgery schedule as well as urgent add-on cases. ASCs, on the other hand, are designed to perform nonurgent outpatient surgeries with minimal resources and therefore at a lower cost. Many characteristics of the OR environment are the same whether the surgery is performed in a hospital or in an ASC. A typical surgical suite is composed of several individual ORs sharing resources such as an equipment storage area, sterilization resources, preoperative

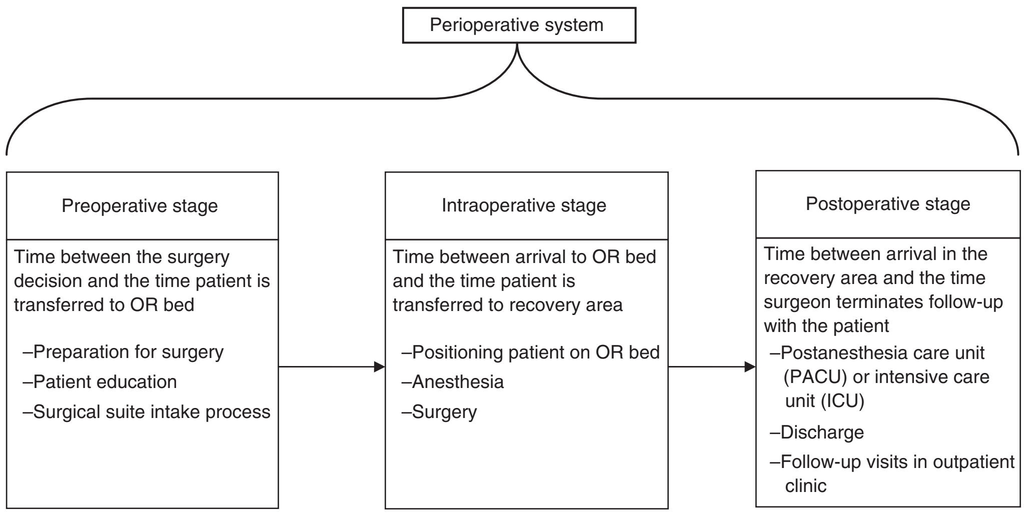

Figure 1. Stages of a perioperative system.

intake, and postoperative recovery rooms. Depending on the type of surgery, some ORs might have different technological resources. For instance, some types of cardiac surgery require cardiopulmonary bypass equipment. In some cases, these special resources may be dedicated to the ORs, while in others they may be mobile and transferred between ORs as needed.

The design and layout of a surgical suite can influence surgery scheduling. Traditional surgical suites have separate intake, surgery, and recovery areas. The patients enter the intake area, progress to surgery, and then eventually transition to a separate recovery area. Figure 2 illustrates the patient flow through this type of a surgical suite. More modern designs have experimented with a common intake and recovery area. The patient is allocated to a room, referred to as a *pre/post room*, where the intake process is carried out. The patient is subsequently taken to surgery, and then returned to their pre/post room after surgery [1].

The process for scheduling surgeries may differ from one hospital or ASC to another. One common system used by many hospitals is known as *block-booking*. In a block-booking system, blocks of uninterrupted OR time are reserved for individual surgeons or surgical specialties in a periodic (often weekly or monthly) schedule. Surgeons book cases into their allotted block time only if the surgery can be completed within the assigned time (or if an administrative exception is granted). The amount of block time is based on the surgeon's or surgical specialty's historical OR usage or some other performance criteria. Other constraints such as surgeon preferences and resource availability are also taken into account when assigning block times.

Another scheduling approach used by some hospitals is *nonblocked* or *open-booking*. In these systems, surgeons submit requests for OR time. Surgeries are scheduled on a first-come-first-serve basis until either a maximum number of cases is reached or a predetermined OR capacity (planned available time) is reached [2]. Surgeons can submit their cases up until the day before surgery at which point a schedule is generated for the surgical listing.

In both block-booking and open-booking systems, a detailed schedule of surgeries is generated which includes surgery-to-OR assignments and estimated start times. On the day of surgery, the schedule (whether based on block-booking or open-booking) is revised by the OR manager (often an anesthesiologist or the head nurse) throughout the day to accommodate additional add-on cases, and to compensate for cases that run longer or shorter than expected. Both of these systems have advantages and disadvantages. In open-booking systems, surgeries are scheduled on short notice and uncertain demand gives rise to variable OR utilization rates [2]. Block-booking often overcomes this disadvantage by allocating blocks of OR time to surgical specialties based on historical utilization rates. However, it does not allow blocked OR time to be used for other specialties or surgeons, which can result in unused OR time.

Many hospitals use systems that are combinations of open-booking and block-booking systems. For instance, surgeons or surgical

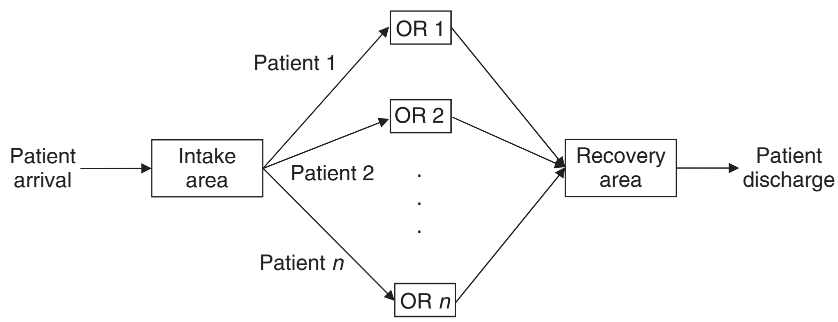

Figure 2. Patient flow through a surgical suite.

specialties may share their assigned block time if it is unutilized within a certain lead time prior to the day of surgery. Furthermore, within the same surgical suite, some ORs may strictly follow a block-booking system, whereas others may follow an open-booking system.

Managing a surgical suite is complicated because it involves many levels of decision making that affect or are affected by surgery scheduling systems [3]. Strategic decisions take place over long time periods (e.g., one year) and include decisions such as whether to use an open-booking or block-booking system, major capacity decisions such as the number of ORs to utilize, investment in new technology, and the number and type of staff to employ. Tactical level decision making involves reserving OR time and capacity for surgical specialties, finding an efficient surgery mix, and determining how much time to plan for each type of surgery. These decisions involve balancing the cost of allocating too much time which causes idle time both for the OR and the staff with the cost of allocating too little time which may result in overtime. Operational decisions involve sequencing surgeries in a particular OR, estimating start times for individual surgeries [4], and setting appointment times for patients to arrive for surgery so as to minimize on-site waiting and achieve high patient satisfaction. In addition, decisions about how to manage add-on cases, and whether to cancel scheduled cases must be made on the day of surgery [3].

COMPLICATING FACTORS IN SURGERY SCHEDULING

Each activity in the perioperative period is critical to the successful delivery of surgical services to patients. The types of activities and their durations may differ from inpatient to outpatient settings and from one hospital to another. One commonality is that most of these activities have significant uncertainty in duration. This is well known for surgical procedures themselves and has also been reported for preoperative and postoperative time in ASCs [1]. Preoperative time, for instance, can depend on many factors including whether the patient requires some prescreening (e.g., a patient with diabetes might require a blood test prior to surgery) or if the patient requires a translator. The surgery itself can be highly uncertain in duration. Factors affecting surgery include the type of procedure, physical characteristics of the patient such as body-mass-index, age, gender, and the surgical environment (e.g., academic vs nonacademic medical center or hospital vs outpatient procedure center) [5]. Uncertainty in postoperative time can result from the outcome of surgery and the patient's response to anesthesia. Previous research indicates that surgery duration and anesthesia duration often fit a lognormal distribution [4]. Although perfect prediction of surgery durations is impossible, improved estimates can have a positive impact on costs associated with underutilization and overutilization of the surgery suite.

Another complicating factor in surgery scheduling is the uncertainty in the number of patients to be scheduled on a particular day. Surgery at hospitals can be divided into two major categories such as *elective* and *nonelective* cases [6]. For elective cases, surgery may be planned well in advance (e.g., months) to be performed on a future date. For nonelective cases, on the other hand, the surgery is unanticipated. These cases must be worked into the existing schedule, either by using intentionally reserved or otherwise available space in the schedule, or by creating space by canceling elective cases. Some hospitals allocate separate ORs for emergencies and add-ons, whereas others allow slack time in the schedule [7,8].

Other causes of uncertainty in the number of surgeries on a particular day include patient punctuality, no-shows, or cancellations [9]. Outpatient settings are more affected by patient punctuality and no-shows. In inpatient settings, due to the greater complexity and risk of surgery, cancellations are a significant source of uncertainty. For instance, a cancellation may occur due to unexpected changes in the patient's medical status.

Variation and uncertainty in daily surgery mix is another complicating factor due to the varying resource requirements for different types of surgery. Hospitals often control the mix by allocating OR capacity to specific surgical groups (e.g., thoracic, orthopedic, and colorectal). Such allocation, often called *block-booking*, is done based on historical utilization by surgical groups.

Surgical suites have high fixed costs, the large proportion of which are associated with the labor cost of the OR team. Surgical suites also have variable costs related to scheduling. For instance, ORs typically have a planned utilization time (e.g., 8 h) 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 consideration. Efficient surgery scheduling also affects the amount of waiting by the OR team, and other critical resources. The recovery area can also be the bottleneck for the surgical suite at certain times of the day. Therefore, resources such as PACU, ICU, and other hospital beds required after surgery and staff should be considered during the scheduling process [10].

High fixed costs encourage scheduling a sufficient number of cases to fully utilize capacity on the day of surgery. This pressure combined with complex interdependency between activities, and the need to schedule a large number of activities with uncertain duration within a fixed time during the day makes for a very challenging scheduling problem. Not surprisingly, much of the literature has focused on stochastic models (e.g., discrete-event simulation, queuing, stochastic programming, and Markov decision processes). In the next section, we provide a detailed review of the relevant literature on surgery scheduling.

LITERATURE REVIEW ON SURGERY SCHEDULING

Surgery scheduling has been a widely studied topic in the medical and operations research fields. There are a number of previous literature reviews related to surgery scheduling. An early review [11] focuses on classifying the previous research based on various stages of decision making, performance measures such as OR utilization and costs, and the scheduling process. A more recent review article provides detailed classifications based on surgery durations (deterministic or stochastic), patient arrivals (elective and nonelective), and operations research methodology [6]. Previous reviews of literature on general health care appointment scheduling (e.g., primary care clinics [12]) are also relevant to surgery scheduling. In our review, we emphasize recent studies that consider new and important aspects of surgery scheduling systems. We categorize the literature based on the type of OR methodology used, and use our literature review to identify opportunities for future research which we summarize in the section titled "Open Challenges and Future Research Areas."

Queuing Models

There have been numerous queuing-based studies presented over the past several decades on the problem of scheduling patients for surgeries and outpatient clinic appointments [13–15]. Queuing research has focused on single server problems for which appointment decisions are economically significant. Although not directly applicable, scheduling problems from other contexts (e.g., manufacturing lines and other service systems) are also relevant to surgery scheduling. Much of this research is relevant to single-OR scheduling with open-booking systems. The aim of the single-OR scheduling problem is to assign start times for a certain number of surgeries to be scheduled in a particular OR on a given day. The scheduler must consider the variability in surgery durations. Clearly, the amount of waiting and idling between surgeries will be impacted by the choice of start times. Also, depending on the planned OR time, some overtime may occur at the end of the day. Figure 3 demonstrates the situations when these conditions occur in an OR. The section titled "Example" provides a specific example in this context.

In the above referenced queuing literature, the authors assume an infinite horizon and stationary service and arrival processes which are not typical in surgery scheduling environments. Some researchers propose more realistic queuing models with a finite horizon. For instance, Brahimi and

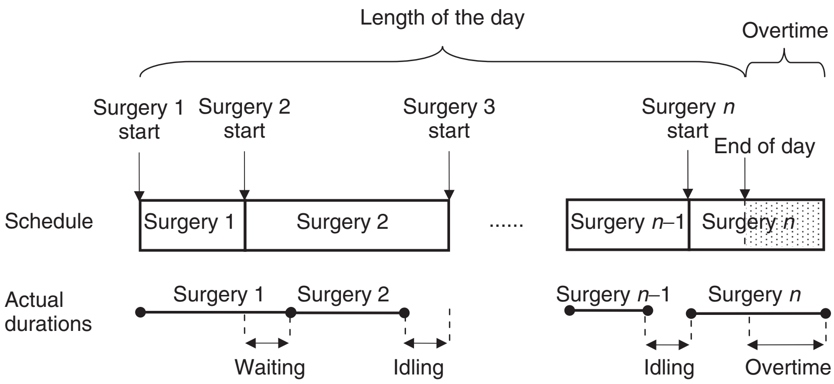

Figure 3. Single OR scheduling problem. The scheduler selects surgery start times. The solid line depicts one possible scenario which includes some waiting, idling, and overtime.

Worthington [16] develop a queuing system with a time-dependent Markovian arrival rate and discrete service time distributions (M(t)/G/s queue) where the number of customers at any time is assumed to be finite. Wang [17] studies the scheduling problem both as a static scheduling problem in which a finite number of customers are scheduled at once, and as a dynamic scheduling problem in which an additional number of customers are scheduled one at a time after an initial batch of customers have been scheduled. He uses phase-type distributions to investigate the transient solution of a Markovian server with general arrival distribution (S(n)/M/1 queue) to determine optimal start times for each customer.

Vanden Bosch and Dietz [18] present a queuing model with phase-type service durations and deterministic arrivals where patient no-shows are also considered. In their study, patients are classified into different groups depending on their service durations. Analyzing the special structure of the model, they propose an algorithm which is guaranteed to find the optimal schedule along with the optimal sequence efficiently.

Although not directly applicable to surgery scheduling, it is worth mentioning the recent study by Zeng *et al*. [19] on overbooking in clinical appointment schedules to reduce the negative impact of no-shows. The authors develop a queuing model with patients having different no-show probabilities. They propose that overbooking is beneficial for open-booking systems, and they demonstrate that clustering patients according to their no-show probabilities will produce better schedules.

Queuing models offer valuable general insights about the effect of uncertainty on scheduling decisions. They also have several shortcomings in the context of surgery scheduling. Some of the queuing articles referred above assume that the system reaches steady state. However, surgical suites typically operate for some specific period of time during the day (e.g., 8–12 h) after which a small number of ORs are available for emergencies. Thus, the system is terminating, and the steady-state assumption is a significant approximation that has not been carefully validated. Queuing models also often require strong assumptions such as exponentially distributed surgery durations, which is not reasonable for most types of surgery that tend to better fit a lognormal distribution [4].

Simulation

Simulation models have found considerable application to surgery planning and scheduling. Although they do not provide closed form solutions like the queuing models, simulation models are more flexible in terms of assumptions about the probability distributions of surgery or other activity durations. However, they are also more computationally intensive than queuing models.

Ho and Lau [20] used simulation in their studies to compare the performance of simple scheduling rules in a single server system. They propose that the performance of the scheduling rules are affected by the *environmental conditions* of the operating environment such as the probability of no-shows, number of patients to schedule, and the service distribution. They consider simple rules such as scheduling two, three, or more patients simultaneously at the beginning of the session and scheduling the later patients with intervals equal to the mean of the service distribution. They also modified these rules by changing the interval length between patients. In total, they evaluated more than 50 different rules with several different combinations of the three environmental conditions. They present the Pareto set of scheduling rules with respect to expected idle time of the server and expected patient waiting time.

Many researchers have used simulation methods to analyze more complex multi-OR surgical environments with recovery areas. Schmitz and Kwak [21], for example, used simulation to analyze a multi-OR surgical suite with recovery rooms to determine the number of ORs to open on a day given that the number of surgeries to be performed is known. They also study the impact of an increase in the number of ORs on recovery room usage and determine the need for recovery room capacity given that some of the surgeries are done in the outpatient setting. Lowery [22] simulates patient flow through a hospital's critical care units including ORs, ICU, and PACU beds to determine the bed requirements for these recovery units.

Simulation methods have been widely used for testing the performance of scheduling heuristics. In [1], a discrete event simulation model is developed for a newly designed outpatient surgical suite to study the impact of scheduling and start time heuristics and the daily surgical mix on the expected patient waiting time and overtime. The authors tested the performance of seven combinations of OR allocation heuristics and sequencing rules such as scheduling surgeries according to longest processing time first (LPT), shortest processing time first (SPT), increasing variance (VAR), and increasing covariance (COV), and found that in general LPT and combination of LPT and VAR rules (LPT−VAR) perform better than other rules. The authors also found that arrival time schedules and daily surgery mix have significant effects on the performance measures while scheduling rules have smaller effects on the performance measures. The authors also compare the performances of the heuristics with a bi-criteria genetic algorithm which finds near optimal solutions.

Examining a surgery planning from a strategic perspective, Everett [23] developed a decision support system based on a discrete event simulation model to analyze the waiting times of surgery patients in a public health-care system with multiple hospitals in Australia. His model includes a single list for all surgical patients categorized by urgency and surgery type and can be used to match availability of the hospitals' and patients' needs. This simulation model considers each hospital as an entity in a large health-care system which may reflect the publicly funded health systems, such as in Canada, United Kingdom, and Australia, where there is more centralized control over the assignment of patients to health providers. The article provides evidence that the health-care policies used in different countries may influence the process as well as the efficiency in surgery scheduling.

Simulation models have generally found wide usage in the area of surgery planning and scheduling. Their flexibility allows them to consider many important factors relevant to complex surgical suites. Furthermore, they are descriptive models, meaning that they can be used to compare alternative policies or rules for planning and scheduling. However, there are also some shortcomings. First, they are computationally intensive. Secondly, generating results can be time-consuming compared to queuing models for which analytic solutions are available. A third important shortcoming of simulation models is that they do not provide optimal solutions.

Optimization

Optimization techniques have been widely applied to surgery scheduling. Many researchers have used deterministic models for scheduling surgeries. Some have also used stochastic optimization models in order to incorporate uncertainties related to surgery durations and patient arrivals. We begin by describing deterministic models and then move on to stochastic optimization models.

Ozkarahan [2] proposed a goal programming model for scheduling surgeries in a surgical suite where block-booking is used for scheduling. Her model includes multiple objectives such as OR utilization, OR and surgeon preferences, and intensive care capacity. For each of these objectives, a goal is specified and the objective is to minimize the deviations from these goals.

Pham and Klinkert [24] extended the job shop scheduling problem to a *multimode blocking job shop* in order to satisfy resource constraints (modes) specific to the OR environment. They use a mixed integer programming (MIP) model formulation with the objective of minimizing the weighted sum of makespan and the sum of starting times of all procedures. They propose that add-on and emergency surgeries can be scheduled by adding new constraints using *job insertion*. Job insertion inserts new surgeries into the established schedule and bumps some elective surgeries to the following day, if necessary, to perform emergent surgeries. Fei *et al*. 25 have developed an integer programming (IP) model to assign elective surgeries with deterministic durations into multiple ORs. The objective is to minimize the weighted sum of overutilization and underutilization of ORs. They solve instances optimally with a branch and price algorithm.

There are also a number of stochastic optimization models for surgery scheduling. The majority assume a single OR, and determine optimal start times for surgeries (see Fig 3). Weiss [26] provides the optimal start times for a simple two-surgery problem with uncertain surgery durations in a single OR. Balancing surgeon's idle time and waiting time, the author shows that the problem is equivalent to the well-known *news vendor* problem from inventory theory, for which the optimal solution is known in closed form. Weiss also examines the optimal sequence of two cases by examining the expected cost resulting from the sequences and shows that the sequencing decision is related to the dispersion of the density function of surgery durations. He proposes that scheduling surgeries that have distributions with "fatter" tails first results in higher total expected costs of waiting and idling due to the impact on the following surgery.

Denton and Gupta [27] study a general two-stage stochastic linear programming formulation of the single server appointment scheduling problem. They determine the optimal start times for surgeries. The criteria in their model are expected patient waiting time, expected OR idle time, and expected overtime with respect to a defined length of day. They develop an iterative approximate method based on upper and lower bounds on the optimal solution to the problem with continuously distributed surgery durations. More recently, Begen and Queyranne [28] studied the same problem with random service durations given by a joint discrete probability distribution. They prove that there exits an optimal appointment schedule which is integer and can be found in polynomial time.

Gerchak *et al*. [7] use a stochastic dynamic programming model for a multiperiod surgery scheduling problem. In particular, they use a Markov decision process to investigate the optimality of using *cut-off policies* in an open-booking system. Cut-off policies suggest rejecting future elective cases in order to be able to accept possible emergent surgeries if a certain number of ORs are already occupied by elective cases. The authors show that these cut-off policies are not necessarily optimal; however, they are typically close to optimal. Lamiri *et al*. [29] propose a stochastic model for scheduling elective surgeries over a planning horizon, however, they consider OR capacity for both elective and add-on patients. They use simulation to generate samples for emergency surgeries. The samples are used to estimate deterministic capacity requirements for an MIP model that minimizes total overtime costs and other costs associated with hospitalization, and penalties for waiting time of the patients, violating patient's and surgeon's preferences and deadlines. Their model is solved via branch-and-bound algorithm and then compared to the solution of a simulation optimization approach. The authors conclude that simulation optimization offers improvements even for small sample sizes.

Denton *et al*. [30] proposed a two-stage stochastic MIP formulation to find optimal allocation of surgery blocks to ORs. They also develop a *robust formulation* which aims to minimize the maximum cost (worst-case outcome) that may result from the set of uncertain surgery durations. The robust formulation is advantageous when limited data is available for surgery durations. It is also much less computationally intensive than their stochastic programming model. Batun *et al*. [31] model a multi-OR scheduling problem as a stochastic MIP to determine the number of ORs to open on a given day, allocation of surgeries to ORs, sequence of surgeries within each OR, and start times for each surgeon. The focus of their study is on parallel scheduling of surgeries in which a surgeon may be allocated two or more ORs on the day of surgery.

Lamiri *et al*. [32] used a column generation approach to find the assignments of the elective surgeries into ORs on a given day in the planning horizon. The authors formulate the problem as a stochastic IP model and reserve a random capacity for emergency surgeries. Column generation is used as a solution technique in which columns represent possible surgery-to-OR assignments in a given planning horizon.

Despite the wide literature on optimization methods in OR scheduling problems, there is still a lack of literature in optimization of multi-OR surgical suite scheduling. The current literature suggests that considering the stochastic aspects of the surgical environment is essential for solving realistic problems. However, optimization methods have disadvantages as the problem size gets larger. Finding the optimal solution becomes computationally harder and sometimes impossible for large-scale realistic scheduling problems. Developing efficient stochastic optimization methods to solve larger and more realistic problems, which consider multiple ORs and other critical resources in the surgical suite, remains as an open challenge for researchers.

Heuristic Methods

As described in the previous section, the stochastic and combinatorial nature of surgery scheduling problems gives rise to computationally challenging optimization models. Many researchers have addressed this problem by proposing fast heuristics. This enables consideration of more complex and more realistic systems such as a multi-OR surgical suite with recovery rooms.

Denton *et al*. [33] consider a two-stage stochastic MIP model of the single OR scheduling problem with the addition of surgery sequencing. Decisions in their model are the optimal start times for surgeries and the optimal sequence of surgeries. A simple interchange heuristic is used to find near-optimal sequences, and start time decisions are computed to optimality for each sequence generated by the interchange heuristic. They also prove for a two-surgery example, conditions under which a stochastic ordering defines the optimal sequence, and use numerical experiments to demonstrate that sequencing in order of increasing variance of surgery duration is typically optimal or near-optimal. Intuitively, this is because delaying highly uncertain surgeries to the end of the day limits their impact on other surgeries in the schedule. Vanden Bosch and Dietz [34] propose a local heuristic method to sequence the customers in a first-come-first-serve appointment system when customers differ by waiting cost, no-show probability, and service time distributions. They also propose a solution method to find the optimal schedule given a fixed sequence.

Beliën and Demeulemeester [35] propose a simulated annealing approach as well as several different MIP heuristics to create a master surgery schedule for a block-booking system with the objective of minimizing the expected shortage of beds. Their MIP heuristics involve solving a number of MIPs, which aim to minimize the maximum expected duration, maximum expected bed occupancy, and maximum weighted sum of mean and variance of the daily bed occupancy respectively. Their simulated annealing approach, on the other hand, starts with a random surgery block and does pairwise exchanges of blocks to find an improved schedule with respect to bed shortage.

Bin-packing algorithms have been widely used to find the mix of surgical cases for multi-OR surgical environments [36]. In this context, the ORs are available for a fixed time during the day, and represent the bins. The surgeries, which have an estimated duration, are the items to be packed in the bins. There are two types of bin-packing problems that have been studied: *on-line* and *off-line*. In the on-line bin-packing problem, urgent surgeries are scheduled sequentially one at a time. Many heuristics have been proposed for this problem. One well-known heuristic, called *best fit*, schedules surgeries into the OR which has the least amount of remaining time that is sufficient to fit the case. Another heuristic, called *worst fit*, schedules the surgery into the emptiest OR which has enough time to fit the case.

In the off-line version of the bin-packing problem, surgeries are batched and simultaneously assigned to ORs. LPT is the best-known heuristic for off-line bin packing. To apply LPT, surgeries are sorted based on their mean durations and then assigned sequentially to the ORs. Many variants of LPT exist including *best fit descending* in which surgeries are assigned sequentially to the fullest OR after being sorted by LPT rule and *worst fit descending* in which they are assigned to the emptiest OR. Dexter *et al*. [36] compare several of these heuristics for scheduling add-on surgeries into remaining open OR times. They found that off-line algorithms are able to fit more add-on surgeries and in particular *best fit descending with special constraints* is more likely to maximize OR utilization.

Hans *et al*. [37] study the bin-packing heuristics with the *portfolio effect* to consider the uncertainty in surgery durations. Portfolio management is used to reduce the risk (minimize variance) or increase the profitability (maximize expected return) by distributing the investments into various different projects instead of investing in a single project. They propose both constructive and local search heuristics to minimize the total planned slack time between surgeries depending on the variances of different surgery durations. They found that, as a result of the portfolio effect, surgeries with similar duration variability are often scheduled on the same day for each specialty.

In general, heuristics provide a powerful trade-off between descriptive models, such as simulation and queuing models, and prescriptive (optimization) models for problems that are very computationally challenging. The main shortcoming of heuristics is that they do not necessarily provide an optimal solution and it is often difficult to develop good bounds on their performance.

EXAMPLE

In this section, we provide a small stochastic programming example on surgery scheduling. In this example, we consider a problem similar to the one illustrated in Fig. 3 with five surgeries to be scheduled on a given day. Surgery durations are distributed uniformly between 60 and 120 min. The length of the day is assumed to be 450 min. The optimal start times for the surgeries can be found by solving the following minimization problem:

$$\min_x \left\{ E\left[c^w \sum_{i=1}^{5} w_i + c^\ell \ell \right] \right\} \qquad (1)$$

where $x_i, i = 1, \ldots, 4$, are the time allowances between start times of each of the surgeries, $w_i, i = 1, \ldots, 5$, are the waiting times, and ℓ is the overtime. Overtime cost and waiting cost are assumed to be equal, that is, $c^w = c^\ell = 1$ in this example. The model in Equation (1) can be formulated as a two-stage stochastic linear program. The x_i are the first stage decision variables, and w_i and ℓ are second stage decisions which are made after the first stage decisions [27]. It is assumed that the first surgery is scheduled at time zero. Thus the waiting time for this surgery is zero.

The optimal solution to this two-stage stochastic programming problem is $x_1 = 95.11, x_2 = 99.73, x_3 = 99.40, x_4 = 96.01$ which means that after first surgery is scheduled at time zero, second one will be scheduled at time 95.11, third one at time 194.84, fourth one at time 294.24, and the last one at time 390.25. The total expected waiting time plus overtime cost of this solution is 65.39.

A common practice is to set x_i equal to the mean of the duration distribution estimated based on historical data. Stochastic programming takes this uncertainty into account by sampling surgery duration scenarios to find a near-optimal solution which works well on average across all the possible scenarios. For this particular example, 10, 000 different duration scenarios were used to find the optimal solution. If the mean value solution was used for this example, the total expected waiting time plus overtime cost would be 79.93, which is 22% higher than the total cost of the stochastic programming solution. Note that the optimal solution trades off expected waiting time cost and idle time cost, which are competing performance measures. For instance, short allowances tend to result in longer waiting time but shorter idle time. Alternatively, long allowances tend to result in shorter waiting times and longer idle times.

OPEN CHALLENGES AND FUTURE RESEARCH AREAS

In this article, we have so far provided an overview of important aspects related to surgery scheduling and a discussion of some of the most significant complicating factors. Surgery scheduling has been a topic of investigation for several decades. However, there are many open research areas and opportunities still to be explored. We describe some of these below.

Most of the above referenced literature focuses on the design of daily schedules prior to the day of surgery. Little effort has been spent on developing rules for real time scheduling on the day of surgery. This is a common problem that OR managers have to face every day to manage the deviations caused by the surgeries that run longer or shorter than anticipated. In addition to the daily operational level scheduling problem, there are also aspects of long range surgery planning and scheduling that need more attention. For instance, the vast majority of the current literature focuses on patient waiting time on the day of surgery, when, in fact, the time spent in waiting lists may be more important from the health outcomes perspective.

Despite the rich literature considering the uncertainty in surgery durations, demand uncertainty is still largely unexplored. For instance, uncertainty in patient no-shows, case cancellations, and the addition of add-on cases to schedules on short notice can all have significant impact on surgery schedules. Most research has assumed that a fixed number of elective cases are to be scheduled. Uncertainty in demand from day to day can have a significant impact on surgery schedules resulting in substantial delays or cancellations on the day of surgery. Other causes of uncertainty such as unpredictable internal demand on recovery room resources on the day of surgery (e.g., ICU beds) are also in need of study.

Many of the studies referenced above focus on ORs; however, other parts of the surgical suite can also influence performance. For instance, the PACU can become a bottleneck, leading to patients backing up in ORs, and causing longer than anticipated turnover times between surgeries. This could ultimately result in overtime or surgery cancellations. Uncertainty in other critical activities like recovery time have been considered in the literature and demonstrated to be important considerations in scheduling. Some of the above referenced literature on simulation and deterministic mathematical programming consider this issue [10]; however, few attempts have been made to incorporate these aspects into stochastic optimization models.

Another research area that needs further attention is the design of surgical suites and processes within and around the OR. This has emerged as a way to increase OR utilization by reducing nonoperative time and increasing throughput. For example, Sandberg *et al*. [38] proposed a new OR suite design called *OR of the future* (*ORF*) for technologically intensive surgeries with an intake room and early recovery room attached to the OR. They are motivated by the goal of performing some of the preoperative and postoperative processes in parallel to increase throughput and decrease overtime. With this new structure, the goal is to perform anesthesia induction in the induction room while the OR is prepared for surgery. Early recovery is provided by an additional perioperative nurse, while the anesthesia personnel can move to the intake of the next patient. Further study is needed to analyze the impact of new designs, which will likely lead to opportunities to improve the efficiency of surgery scheduling.

In this article, we have discussed the basic concepts related to surgery planning and scheduling as well as the surgical environment and the resources. We have also presented a broad range of operations research and management science methodologies used in surgery scheduling. Though this is a very well-studied subject, there are still significant opportunities to improve.

Acknowledgments

This project was funded in part by grants CMMI-0620573 and CMMI-0844511 (Denton) from the National Science Foundation.

REFERENCES

1. Gul S, Denton BT, Fowler J, *et al*. Bi-criteria scheduling of surgical services for an outpatient procedure center. Working Paper.
2. Ozkarahan I. Allocation of surgeries to operating rooms by goal programming. J Med Syst 2000;24(6):339–378.
3. Dexter F, Epstein RH, Traub RD, *et al*. Making management decisions on the day of surgery based on operating room efficiency and patient waiting time. Anesthesiology 2004;101(6):1444–1452.
4. Strum DP, May JH, Vargas LE. Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. Anesthesiology 2000;92(4):1160–1167.
5. Strum DP, Sampson AR, May JH, *et al*. Surgeon and type of anesthesia predict variability in surgical prodecure times. Anesthesiology 2000;92:1454–1466.
6. Cardoen B, Demeulemeester E, Beliën J. Operating room planning and scheduling: A literature review. FED Research Report KBI 0807 Katholike Universiteit Leuven; 2008.
7. Gerchak Y, Gupta D, Henig M. Reservation planning for elective surgery under uncertain demand for emergency surgery. Manage Sci 1996;42(3):321–334.
8. Klassen KJ, Rohleder TR. Outpatient appointment scheduling with urgent clients in a dynamic, multi period environment. Int J Serv Ind Manage 2003;15(2):167–186.
9. Basson MD, Butler TW, Verma H. Predicting patient nonappearance for surgery as a scheduling to optimize operating room utilization in a veterans' administration hospital. Anesthesiology 2006;104(4):826–834.
10. Marcon E, Dexter F. Impact of surgical sequencing on post anesthesia care unit staffing. Health Care Manage Sci 2006;9: 87–98.
11. Blake JT, Carter MW. Surgical process scheduling: a structured review. J Health Syst 1997;5(3):17–30.
12. Cayirli T, Veral E. Outpatient scheduling in health care: a review of literature. Prod Oper Manage 2003;12(4):519–549.
13. Welch J. Appointment systems in hospital outpatient departments. Oper Res Q 1964;15:224–237.
14. Mercer A. A queuing problem in which the arrival times of the customers are scheduled. J R Stat Soc Ser B 1960;22(1):108–113.
15. Soriano A. Comparison of two scheduling systems. Oper Res 1966;14(3):388–397.
16. Brahimi M, Worthington DJ. Queuing models for outpatient appointment systems: a case study. J Oper Res Soc 1991;42(9):733–746.
17. Wang PP. Static and dynamic scheduling of customer arrivals to a single-server system. Nav Res Logist 1993;40:345–360.
18. Vanden Bosch PM, Dietz DC. Minimizing expected waiting time in a medical appointment system. IIE Trans 2000;32:841–848.
19. Zeng B, Turkcan A, Lin J, *et al*. Clinical scheduling models with overbooking for patients with heterogeneous no-show probabilities. Ann Oper Res. DOI: 10.1007/s10479-009-0569-5 (published 12 June 2009).
20. Ho C, Lau H. Minimizing total cost in scheduling outpatient appointments. Manage Sci 1992;38(12):1750–1764.
21. Schmitz HH, Kwak NK. Monte Carlo simulation of operating-room and recovery-room usage. Oper Res 1997;20(6):1171–1180.
22. Lowery JC. Simulation of a hospital's surgical suite and critical care area. In: Proceedings

of the 1992 Winter Simulation Conference; Arlington (VA). 1992. pp. 1071–1078.

23. Everett JE. A decision support simulation model for the management of an elective surgery waiting system. Health Care Manage Sci 2002;5:89–95.
24. Pham DN, Klinkert A. Surgical case scheduling as a generalized job shop scheduling problem. Eur J Oper Res 2008;185(3): 1011–1025.
25. Fei H, Chu C, Meskens N, *et al*. Solving surgical cases assignment problem by a branch and price approach. Int J Prod Econ 2008;112:96–108.
26. Weiss EN. Models for determining estimated start times and case orderings in hospital operating rooms. IIE Trans 1990;22(2): 143–150.
27. Denton B, Gupta D. A sequential bounding approach for optimal appointment scheduling. IIE Trans 2003;35:1003–1016.
28. Begen MA, Queyranne M. Appointment scheduling with discrete random durations. In: Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms; New York. 2009. pp. 845–854.
29. Lamiri M, Xie X, Dolgui A, *et al*. A stochastic model for operating room planning with elective and emergency demand for surgery. Eur J Oper Res 2008;185(3):1026–1037.
30. Denton BT, Miller AJ, Balasubramanian HJ, *et al*. Optimal allocation of surgery blocks to operating rooms under uncertainty. Oper Res. In press.
31. Batun S, Denton B, Schaefer A, *et al*. Operating room pooling and parallel surgery processing under uncertainty. INFORMS J Comput 2010. In press.
32. Lamiri M, Xie X, Zhang S. Column generation approach to operating theater planning with elective and emergency patients. IIE Trans 2008;40:838–852.
33. Denton B, Viapiano J, Vogl A. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manage Sci 2007;10(1):13–24.
34. Vanden Bosch PM, Dietz DC. Scheduling and sequencing arrivals to an appointment system. J Serv Res 2001;4(1):15–25.
35. Beliën J, Demeulemeester E. Building cyclic master sugery schedules with leveled resulting bed occupancy. Eur J Oper Res 2007; 176(2):1185–1204.
36. Dexter F, Macario A, Traub RD. Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Anesthesiology 1999;91:1491–1500.
37. Hans EW, Wullink G, Van Houdenhoven M, *et al*. Robust surgery loading. Eur J Oper Res 2008;185(3):1038–1050.
38. Sandberg WS, Daily B, Egan M, *et al*. Deliberate perioperative systems design improves operating room throughput. Anesthesiology 2005;103(2):406–418.