Published online ahead of print October 22, 2012

INFORMS Journal on Computing
Articles in Advance, pp. 1–11
ISSN 1091-9856 (print) | ISSN 1526-5528 (online)

http://dx.doi.org/10.1287/ijoc.1120.0530
© 2012 INFORMS

Stochastic Operating Room Scheduling for High-Volume Specialties Under Block Booking

Oleg V. Shylo, Oleg A. Prokopyev, Andrew J. Schaefer
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261
{olegio@gmail.com, prokopyev@engr.pitt.edu, schaefer@pitt.edu}

Scheduling elective procedures in an operating suite is a formidable task because of competing performance metrics and uncertain surgery durations. In this paper, we present an optimization framework for batch scheduling within a block booking system that maximizes the expected utilization of operating room resources subject to a set of probabilistic capacity constraints. The algorithm iteratively solves a series of mixed-integer programs that are based on a normal approximation of cumulative surgery durations. This approximation is suitable for high-volume medical specialities but might not be acceptable for the specialties that perform few procedures per block. We test our approach using the data from the ophthalmology department of the Veterans Affairs Pittsburgh Healthcare System. The performance of the schedules obtained by our approach is significantly better than schedules produced by simple heuristic scheduling rules.

Key words: surgical suite; operating room scheduling; block booking; chance-constrained programs
History: Accepted by Allen Holder, Area Editor for Applications in Biology, Medicine, and Healthcare; received June 2011; revised October 2011, March 2012, July 2012; accepted August 2012. Published online in *Articles in Advance*.

1. Introduction

Surgical suites account for the largest part of hospitals' budgets (Healthcare Financial Management Association 2003). Rising demand and increasing healthcare costs necessitate the efficient management of operating rooms (ORs). In this paper, we focus on scheduling of elective surgeries, which permit some flexibility regarding an actual surgery date, thereby improving the utilization of ORs' resources.

Currently, the scheduling of surgical suites follows either an *open booking* or *block booking* framework (Erdogan and Denton 2011). Under open booking, surgeons ask for OR time by submitting their cases to the scheduling team, who accommodates their request subject to the operating room capacity. Exact surgery dates and OR assignments are usually determined on a first-come-first-served basis. Under block booking, medical departments (or surgeons) that provide certain types of services (e.g., ophthalmology, orthopedics, cardiology) are assigned fixed blocks of time that are used to divide access to ORs among different specialties. Such assignments, or *block schedules*, are usually designed based on the current demand trends and historical utilization records. Despite potential inefficiencies because of unbalanced block schedules, this framework is widely accepted because of its convenience for both surgeons and managers (Erdogan and Denton 2011).

In our study we consider a single surgical suite that belongs to the Veterans Affairs Pittsburgh Healthcare System (VAPHS), a large Veterans Administration Hospital. This operating theater consists of 10 operating rooms shared among different medical departments. Scheduling is organized using the block booking system with annual updates of block schedules.

Within a block booking scheduling system, because the surgeries booked by different surgeons or medical departments are associated with different sets of blocks, the whole problem can be decomposed into a set of nonoverlapping scheduling subproblems. Each subproblem addresses the scheduling of surgeries for a specific medical department with a patient inflow that is independent from other specialties. This decomposition is valid when the availability of various resources, e.g., cleaning crews or recovery beds, is not a pressing issue, which is the case at the VAPHS.

Intuitively, the scheduling problem under the block booking paradigm consists of allocating surgery cases to the available time blocks in order to find a reasonable balance between overtime and utilization for each scheduling block. The processing sequence of surgical cases is usually defined by the surgeon based on his or her availability, other personal preferences, patients' status, and case complexity. The exact sequence has no (or limited) impact on the performance measures unless the cases depend on each other (e.g., two surgeries in different ORs are performed by the same surgeon or require the same unique piece of equipment), as discussed, for example, by Batun et al. (2011). In our model we assume that such issues can be addressed when making the sequencing decisions within each block.

Surgery scheduling under block booking systems represents an extension of a stochastic knapsack problem (Kleywegt et al. 2002, Goyal and Ravi 2010). Managing and planning of operating suites activities is an active research area (Blake and Carter 1997, Cardoen et al. 2010). The uncertainty of surgery durations, conflicting priorities, and limited capacity of the surgical suite all contribute to the difficulty of this problem. High utilization leads to larger surgery volumes, thus reducing the patients' waiting times, and minimizing the overtime prevents additional costs associated with extra work hours. We will briefly review some of the closely related recent research papers that address the assignment of surgeries to ORs without sequencing within each room. An extensive list of publications on operating room scheduling is maintained by Dexter (2011).

Denton et al. (2010) apply two-stage stochastic programming and robust optimization to minimize a weighted sum of the total cost of opening ORs and the total overtime. The sets of surgical blocks are used to group together the surgeries that should be performed sequentially in the same OR on a given day. The model incorporates two types of decisions: the number of ORs to open on a given day, and the assignment of surgery blocks to specific ORs.

Min and Yih (2010) apply sample average approximation (Kleywegt et al. 2002, Luedtke and Ahmed 2008) to minimize average overtime and patient cost, where patient cost is a predefined price associated with assignments of patients to blocks. Patient cost is strictly increasing in waiting time and is used to prioritize the patients. A surgical intensive care unit capacity constraint is imposed to avoid blocking, which occurs when a patient cannot leave the OR because of capacity limitations. All the surgery durations within the same surgical specialty are assumed to be identically and lognormally distributed.

Dexter et al. (1999) provide a simulation study of deterministic online bin-packing scheduling rules and the relationship between utilization and patients' waiting times. The authors emphasize the importance of moving the control of the surgical date from the surgeon and the patient to the OR suite in order to achieve the maximum utilization of OR resources.

Hans et al. (2008) develop constructive and local search heuristics for maximization of utilization and minimization of the overtime risk. In their model, all surgeries are assigned to operating rooms based on their expected durations. To cope with randomness of surgery processing times, a planned time slack (buffer) is reserved in each scheduling block, which is a function of total mean and variance of surgeries assigned to the corresponding scheduling block. When determining an appropriate size of the planned slacks, the authors assume that the sum of surgery durations follows a normal distribution.

Most of the existing approaches that model the stochastic surgery durations rely on scenario-based approximations (Denton et al. 2010, Min and Yih 2010, Batun et al. 2011). In this paper, we investigate a different approach based on normal approximation for the sum of surgery durations. This approximation can be justified when scheduling the high-volume specialties (e.g., four–seven surgeries in each scheduling block). The use of normal approximation is, for example, the current scheduling practice at the Erasmus Medical Center (Hans et al. 2008).

The contributions presented in this paper are as follows.

- We provide theoretical properties of overtime and undertime functions under the assumption that the sum of surgery durations is normally distributed. This assumption is common in the literature and in practice, and it fits high-volume medical specialties, such as the ophthalmology department at the VAPHS. The theoretical justification for the normal approximation is given by the central limit theorem.
- We are the first to use a chance-constrained model of overtime for the OR scheduling.
- Using the properties of overtime and undertime functions, we formulate a mixed-integer program of OR scheduling that provides lower and upper bounds for the optimal solution of the original stochastic scheduling problem and allows us to obtain near-optimal solutions for realistic instances.
- We empirically explore the quality of the normal assumption using the historical data of the ophthalmology department provided by the VAPHS (see the data description in §4) and compare it with the scenario-based approaches.
- We compare the performance of the proposed algorithm to the first-fit scheduling heuristic that is commonly used in OR scheduling practice.

The remainder of this paper is organized as follows. We give a general optimization model of batch scheduling in §2. Based on this model, we provide an approximate solution approach in §3. We describe the simulation model that is used to verify the validity and potential of the proposed approach in §4. Section 5 provides a conclusion.

2. General Optimization Model

Let $S = \{s_1, \ldots, s_n\}$ be a set of surgeries to be allocated into an ordered set of time blocks $B = \{b_1, \ldots, b_m\}$, where b_m is the block corresponding to the latest date. Each block $b \in B$ has a fixed duration denoted by $l(b)$. The duration d_s of each surgery $s \in S$ is a random variable. Here, we assume that the time required to clean an operating room after performing the surgery s is factored into the case duration d_s.

Let $x_{s,b}$ be a binary variable that equals one iff surgery s is allocated to block b. The random variables $\max\{0, \sum_{s\in S} d_s x_{s,b} - l(b)\}$ and $\max\{0, \sum_{s\in S} l(b) - d_s x_{s,b}\}$ are referred to as the *overtime* and *undertime* of the block b, respectively.

Currently, the scheduler at the VAPHS uses a point estimate of the expected duration of each case based on historical data and the estimate provided by the surgeon to determine the feasibility of the schedules. The block assignments are considered feasible if the sum of the point estimates of surgery durations assigned to the same scheduling block and the intervening cleaning times do not exceed the block length. The goal of this policy is to make sure that all of the scheduled cases can be finished within the allocated block length, thereby avoiding overtime. Let est_s denote a point estimate of the expected duration of surgery $s \in S$. Then the current feasibility conditions can be modeled by the following set of constraints:

$$\sum_{s\in S} \mathrm{est}_s \cdot x_{s,b} \le l(b), \quad \text{for all } b \in B. \tag{1}$$

Unfortunately, the deterministic constraints (1) ignore the variability of surgery times and cannot provide any guarantee on the expected overtime levels. Such a guarantee can be rendered by imposing the probabilistic constraints to capture the stochastic nature of the underlying processes (Charnes et al. 1958, Prékopa 1995).

The probabilistic counterpart for the deterministic constraint (1) can be introduced by requiring that the probability of an overtime exceeding a threshold L be no more than a scalar $0 \le \alpha \le 1$,

$$\Pr\left\{\sum_{s\in S} d_s x_{s,b} - l(b) > L\right\} \le \alpha, \quad \text{for all } b \in B. \tag{2}$$

The chance-constrained scheduling problem can be formulated as the following nonlinear integer program:

$$\min_{x_{s,b}} \left\{ \sum_{b\in B\backslash b_m} \mathbb{E}\left[\left(l(b) - \sum_{s\in S} d_s x_{s,b}\right)^+\right]\right\} \tag{3a}$$

$$\Pr\left\{\sum_{s\in S} d_s x_{s,b} - l(b) > L\right\} \le \alpha, \quad \text{for all } b \in B, \tag{3b}$$

$$\sum_{b\in B} x_{s,b} = 1, \quad \text{for all } s \in S, \tag{3c}$$

$$x_{s,b} \in \{0, 1\}. \tag{3d}$$

The surgeries from S are assigned to the blocks from B so as to minimize the average undertime for the blocks in the set $B\backslash\{b_m\} = \{b_1, \ldots, b_{m-1}\}$. The last available block $b_m \in B$ does not have to be fully utilized, because it can be filled later with newly arrived surgeries. The chance constraints guarantee the acceptable levels of the overtime, whereas minimizing the undertime is equivalent to maximizing room utilization.

The choice of L and α can be used to manage the performance metrics. To reduce the variance and provide stability with respect to the average waiting times, one can apply a scheduling policy that dynamically adjusts its scheduling settings based on the preferences of the OR suite management (Shylo et al. 2011). The management of the OR suite can use a control chart to track the current average waiting time for all patients on the waiting list. The upper (lower) control limits of the control chart can be used to indicate a need for a more (less) aggressive scheduling policy that can be realized by a proper change of the parameters L and α.

Chance-constrained problems and their applications have been extensively studied in the past, but remain computationally intractable in general (Prékopa 1995). The complexity of chance-constrained problems may come from the difficulty of evaluating the probabilistic feasibility, which usually involves multivariate integration. Furthermore, even if the feasibility validation is not difficult, the feasible region itself is generally nonconvex (Luedtke et al. 2010). If none of these complications are present, the chance-constrained models can usually be reformulated as deterministic programs. Otherwise, one can discretize the random distribution and formulate a deterministic combinatorial optimization problem that approximates the original one. The sample average approximation (SAA) method is an example of this approach (Pagnoncelli et al. 2009), where the original problem is approximated by a chance-constrained problem with a discrete distribution based on a Monte Carlo sample. Normally, each sample point is modeled by a binary variable, thus even moderate sample sizes can lead to computationally intractable approximations (Luedtke and Ahmed 2008). Nemirovski and Shapiro (2006) introduce convex approximations of chance constraints that yield solutions feasible subject to original nonconvex chance constraints.

Durations of common surgical cases can be modeled by a lognormal distribution, which has been validated in a variety of hospital settings (Strum et al. 2000, Stepaniak et al. 2010). In the literature, it is also indicated that the models based on a gamma distribution and the models that assume lognormal distribution are often interchangeable (Wiens 1999). The known analytical expression for the convolution of gamma or lognormal distributions is complicated and does not lend itself easily to exact optimization models (Moschopoulos 1985). Because the overtime and undertime of each block depend on a sum of surgery durations, the distribution of the overall processing time must be approximated. Monte Carlo simulation or scenario-based approximations are the most popular approaches to such situations (Hans et al. 2008, Denton et al. 2010, Min and Yih 2010).

In the next section, we follow a different approach. Specifically, we provide a scheduling framework, based on the empirical observation that a normal approximation to the convolution of gamma distributions representing surgery durations of high-volume specialties (approximately four–seven surgeries in each scheduling block) leads to accurate estimates of average overtime/undertime in our problem setting (see §4 for further details). This allows us to formulate the problem as a mixed-integer *linear* programming problem and develop an algorithm for batch scheduling that can be used for real-time scheduling. The proposed approach is appropriate for the medical specialties that perform more than four surgeries per scheduling block, e.g., ophthalmology. On the other hand, the low-volume specialties that perform one–two surgeries per block, e.g., cardiac surgery, are not suitable for the proposed scheduling model. However, the low number of cases that are usually scheduled by these specialties may imply limited scope for the optimization.

3. Approximation Model

We assume that d_s has a normal distribution for each $s \in S$ with mean μ_s and variance σ_s^2. Therefore, the overtime value of $-l(b) + \sum_{s\in S_b} d_s$ is itself a normally distributed random variable with mean $\mu_b = \sum_{s\in S_b} \mu_s - l(b)$ and variance $\sigma_b^2 = \sum_{s\in S_b} \sigma_s^2$. Then the resource capacity chance constraint (2) can be rewritten as follows:

$$L \geq \mu_b + \phi^{-1}(1-\alpha)\sigma_b, \quad \text{for all } b \in B, \tag{4}$$

where $\phi(\cdot)$ denotes the cumulative distribution function of the standard normal variate. Furthermore, the average overtime of block b, denoted by $O(\mu_b, \sigma_b)$, is a function of μ_b and σ_b:

$$O(\mu_b, \sigma_b) = \int_0^\infty t \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(t-\mu_b)^2}{2\sigma_b^2}\right) dt. \tag{5}$$

Similarly, the average undertime of block b is defined as

$$\begin{aligned} U(\mu_b, \sigma_b) &= O(\mu_b, \sigma_b) - \mu_b \\ &= \int_{-\infty}^0 (-t) \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(t-\mu_b)^2}{2\sigma_b^2}\right) dt. \end{aligned} \tag{6}$$

Proposition 1. *The overtime, $O(\mu_b, \sigma_b)$, is a monotonically increasing convex function of its arguments.*

Proof. The partial derivatives of (5) are nonnegative:

$$\frac{\partial O(\mu_b, \sigma_b)}{\partial \mu_b} = \int_{-(\mu_b/\sigma_b)}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \geq 0, \quad \text{and} \tag{7}$$

$$\frac{\partial O(\mu_b, \sigma_b)}{\partial \sigma_b} = \int_{-(\mu_b/\sigma_b)}^\infty t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \geq 0. \tag{8}$$

Furthermore,

$$\begin{aligned} \frac{\partial^2 O(\mu_b, \sigma_b)}{\partial \mu_b^2} &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_b} \exp\left(-\frac{1}{2}\frac{\mu_b^2}{\sigma_b^2}\right); \\ \frac{\partial^2 O(\mu_b, \sigma_b)}{\partial \sigma_b^2} &= \frac{1}{\sqrt{2\pi}} \frac{\mu_b^2}{\sigma_b^3} \exp\left(-\frac{1}{2}\frac{\mu_b^2}{\sigma_b^2}\right); \\ \frac{\partial^2 O(\mu_b, \sigma_b)}{\partial \mu_b \partial \sigma_b} &= -\frac{1}{\sqrt{2\pi}} \frac{\mu_b}{\sigma_b^2} \exp\left(-\frac{1}{2}\frac{\mu_b^2}{\sigma_b^2}\right). \end{aligned}$$

The necessary result directly follows from the fact that the Hessian matrix of $O(\mu_b, \sigma_b)$ is positive semidefinite. □

Remark. If we define O as a function of mean and variance, then it will remain monotonic, but not necessarily convex.

The following propositions follow trivially from Proposition 1.

Proposition 2. *The undertime, $U(\mu, \sigma)$, is a convex function of its arguments.*

Proposition 3. *The undertime, $U(\mu, \sigma)$, is a monotonously increasing function of σ.*

Proposition 4. *For any $m \in \mathbb{R}^1$ and $v \in \mathbb{R}^1_+$, the following holds* (*because of convexity*):

$$\begin{aligned} U(\mu, \sigma) \geq U(m, v) &- \frac{\partial U(m, v)}{\partial \mu} \cdot (m - \mu) \\ &- \frac{\partial U(m, v)}{\partial \sigma} \cdot (v - \sigma). \end{aligned} \tag{9}$$

Because OR scheduling is a continuous process, the scheduling blocks might be partially filled with the surgeries that arrived earlier. Therefore, for each block $b \in B$ we introduce two input parameters: initial mean and variance values denoted by μ_b^{init} and $(\sigma_b^{\text{init}})^2$, respectively.

Based on the previous results, we can formulate the chance-constrained batch scheduling problem as the following mixed-integer nonlinear problem:

$$z^{\text{opt}} = \min \sum_{b\in B\setminus b_m} z_b \tag{10a}$$

$$z_b \geq U(\mu_b, \sigma_b), \quad \forall\, b \in B\setminus b_m, \tag{10b}$$

$$\sum_{b\in B} x_{s,b} = 1, \quad \forall\, s \in S, \tag{10c}$$

$$\mu_b = \mu_b^{\text{init}} + \sum_{s\in S} \mu_s x_{s,b} - l(b), \quad \forall\, b \in B, \tag{10d}$$

$$\sigma_b^2 = (\sigma_b^{\text{init}})^2 + \sum_{s\in S} \sigma_s^2 x_{s,b}, \quad \forall\, b \in B, \tag{10e}$$

$$L \geq \mu_b + \phi^{-1}(1-\alpha)\sigma_b, \quad \forall\, b \in B, \tag{10f}$$

$$x_{s,b} \in \{0,1\}, \quad \forall\, s \in S,\ \forall\, b \in B, \tag{10g}$$

$$z_b \geq 0, \quad \forall\, b \in B \backslash b_m. \tag{10h}$$

Constraints (10b) assure that in the optimal solution cost z_b is equal to the undertime of the block $b \in B$. Each surgery must be assigned to a single block, enforced by (10c). The total mean and variance for each block are modeled by constraints (10d) and (10e), respectively. The overtime of each block is constrained by (10f).

If there is a feasible solution such that none of the surgeries are assigned to the last block b_m, then there is no need to use block b_m in the problem formulation and the set of blocks can be reduced to $B\backslash\{b_m\}$. Here, we assume that for every feasible solution there is at least one surgery assigned to block b_m. To achieve this, for example, one can solve a sequence of feasibility problems (10c)–(10h) to identify the minimum number of blocks that provide at least one feasible solution, or remove the last block b_m from the set of available blocks B, whenever a feasible solution that does not assign any surgery to b_m is found.

Let W be a set of k points $\{(m_1, v_1), (m_2, v_2), \ldots, (m_k, v_k)\}$, $k \geq 2$. We use the set W and valid inequality (9) to approximate the undertime function $U(\mu, \sigma)$. Additionally, we use a piecewise linear approximation to obtain a lower bound on the value σ using σ^2. The points $v_1, v_2, \ldots, v_k$ define a grid for such approximation. We omit the modeling details here because piecewise linear (PWL) approximations are well known in the literature (see the detailed discussion by Vielma and Nemhauser 2011), and denote a set of constraints and variables that model the piecewise linear approximation using the following notation:

$$\tilde{\sigma} = \text{PWL}(\sigma^2, W).$$

The piecewise linear approximation should satisfy the following relationship: $\tilde{\sigma}_b \leq \sqrt{\sigma_b^2}$, which can be easily achieved because of concavity of the square root function. Using the set points W, the problem (10) can be approximated using the following mixed-integer linear program, which is further referred to as $P(W)$:

$$z^{\text{opt}}(W) = \min \sum_{b\in B\backslash b_m} z_b \tag{11a}$$

$$z_b \geq U(m,v) - \frac{\partial U(m,v)}{\partial \mu}\cdot(m-\mu_b) - \frac{\partial U(m,v)}{\partial \sigma} \cdot (v - \tilde{\sigma}_b), \quad \forall\, b \in B\backslash b_m,\ \forall\,(m,v)\in W, \tag{11b}$$

$$\sum_{b\in B} x_{s,b} = 1, \quad \forall\, s \in S, \tag{11c}$$

$$\mu_b = \mu_b^{\text{init}} + \sum_{s\in S} \mu_s x_{s,b} - l(b), \quad \forall\, b \in B, \tag{11d}$$

$$\sigma_b^2 = (\sigma_b^{\text{init}})^2 + \sum_{s\in S} \sigma_s^2 x_{s,b}, \quad \forall\, b \in B, \tag{11e}$$

$$\tilde{\sigma}_b = \text{PWL}(\sigma_b^2, W), \quad \forall\, b \in B, \tag{11f}$$

$$L \geq \mu_b + \phi^{-1}(1-\alpha)\tilde{\sigma}_b, \quad \forall\, b \in B, \tag{11g}$$

$$x_{s,b} \in \{0,1\}, \quad \forall\, s \in S,\ \forall\, b \in B, \tag{11h}$$

$$z_b \geq 0, \quad \forall\, b \in B\backslash b_m. \tag{11i}$$

Note that the partial derivatives $(\partial U(m,v))/\partial\mu$ and $(\partial U(m,v))/\partial\sigma$ as well as the function value $U(m,v)$ at point (m,v) in (11b) can be calculated using simple numerical integration applied to (5), (7), and (8).

The following proposition formally describes the relation between z^{opt} and $z^{\text{opt}}(W)$.

Proposition 5. $z^{\text{opt}}(W) \leq z^{\text{opt}}$.

Proof. The necessary result directly follows from Propositions 3 and 4 using the relationship $\tilde{\sigma}_b \leq \sqrt{\sigma_b^2}$. □

Note that a feasible solution to problem (11) might be infeasible for the initial problem (10) because of violation of block capacity constraints (10f). Because the quality of the bound provided by $z^{\text{opt}}(W)$ can be improved by refinement of the approximation points in W, we next present an iterative algorithm that provides a converging sequence of lower bounds for z^{opt}.

The pseudocode of our approach is presented in Figure 1. The algorithm repeatedly solves the problem (11) and updates set W in order to improve

```
Initialize a set of approximation points:
    $W = \{(-\max_{b\in B} l(b), 0), (\sum_{s\in S}\mu_s - \min_{b\in B} l(b), \sum_{s\in S}\sigma_s)\}$
$\text{LB}^{best} = 0$; $\text{UB}^{best} = \infty$
while $(\text{UB}^{best} - \text{LB}^{best})/\text{UB}^{best} \geq \epsilon$ do
    Solve the approximation problem $P(W)$. Let $x^*_{s,b}$,
        $b \in B$, $s \in S$, be a set of optimal values for the
        surgery assignments (11c), and $z^*_b$, $b \in B$, be a set
        of optimal values for the cost function
        approximations (11a).
    $\mu^*_b = \sum_{s\in S}\mu_s x^*_{s,b}$, $\forall\, b \in B$
    $\sigma^*_b = \sqrt{\sum_{s\in S}\sigma_s^2 x^*_{s,b}}$, $\forall\, b \in B$
    $\beta \in \arg\max_{b\in B}[U(\mu^*_b, \sigma^*_b) - z^*_b]$, (find the block with
        the worst approximation quality).
    $V = \{(\mu^*_b, \sigma^*_b)\colon L \leq \mu^*_b + \phi^{-1}(1-\alpha)\sigma^*_b\}$, (find all
        violations of capacity constraint).
    $W = W \cup V \cup \{(\mu_\beta, \sigma_\beta)\}$
    if $V = \varnothing$ then
    Update the upper bound:
        $\text{UB}^{best} = \min\{\text{UB}^{best}, \sum_{b\in B, b\neq b_m} U(\mu^*_b, \sigma^*_b)\}$
    Update the lower bound:
        $\text{LB}^{best} = \max\{\text{LB}^{best}, \sum_{b\in B, b\neq b_m} z^*_b\}$
return approximate solution to $P(W)$
```

Figure 1 Pseudocode of the Stochastic Batch Scheduling Algorithm

the lower bound of the optimal solution (lines 3–12). Initially, the set W is initialized with two points (line 1) that represent extreme cases: (i) there are no surgeries assigned to the block (zero mean, zero variance); (ii) all surgeries are assigned to the block with the smallest length.

Given an optimal assignment of surgeries for problem $P(W)$, we evaluate the corresponding exact value for the mean μ_b^*, the standard deviation σ_b^*, and the exact value of the function $U(\mu_b^*, \sigma_b^*)$ for each block. The difference $U(\mu_b^*, \sigma_b^*) - z_b^*$ represents the quality of approximation for block b. At each iteration, the means and variances of the blocks that violate capacity constraints or represent the worst approximation quality are inserted into W. The algorithm terminates whenever the gap between the upper and lower bounds is sufficiently small.

4. Computational Experiments

To take into account the intrinsic uncertainty associated with surgery durations, we develop a prediction system based on generalized linear models (Firth 1991) that represents the durations as gamma variates. Dexter et al. (2010), Eijkemans et al. (2010), Strum et al. (2000), and Stepaniak et al. (2010) provide a thorough discussion of the prediction models of surgery durations and analysis of relevant predicting factors. To predict the mean and variance of the new surgical cases, our model uses the specific procedure type as identified by a common procedure terminology (CPT) code (Strum et al. 2000), the surgeon's experience given by the total number of surgeries performed within the OR suite, and an estimated duration provided by a surgeon.

In our study, we use data provided by the VAPHS. The data set includes historical surgery durations and turnover times for all surgical cases performed by the ophthalmology department of the VAPHS between 2006 and 2009. Information about the CPT code, scheduled duration, and surgeon's identification number is available for each surgical case. The CPT codes that are used in our study correspond to the historical records, i.e., these codes are entered after completion of each surgery. Therefore, they might differ from the scheduled CPT codes, which are not available in our data set (Dexter et al. 2010). However, the data set does include approximately 60 different CPT codes. Around 80% of all surgeries in our data set are cataract surgeries, and the remaining 20% are relatively rare procedures (20 or fewer occurrences for each code). The prediction model for ophthalmology is fitted using data from 2006–2008 (1,673 cases) and tested on historical data from 2009 (619 cases). The model demonstrates a reasonable predictive potential: out of 619 cases in the validation set, 53% fall under the predicted 50th percentile, 79% fall under the predicted 80th percentile, and 88% fall under the predicted 90th percentile (the percentiles differ across different cases).

To test the proposed approach, we implemented a discrete event simulation model of scheduling using the OMNeT++ package (http://www.omnetpp.org), which is a C++ library for building simulation models. It is used to simulate scheduling decisions for the ophthalmology department at the VAPHS. The length of each simulation run is set to one year. Patients seeking surgery are assumed to arrive according to a Poisson distribution with a mean of 14 per week. This value is based on the average historical rate for the ophthalmology department during 2009. Each set of simulation parameters is used to generate 100 replications (each replicate simulates schedules for one year). The warm-up period is set to 20 days, providing an initial backlog of 20 days. The block schedule used in the simulation is identical to the one that was used by the ophthalmology department at the VAPHS in 2009,

(i) one block from 8 A.M. until 5 P.M. every Tuesday;

(ii) one block from 8 A.M. until 3 P.M. every Thursday; and

(iii) one block from 8 A.M. until 12 noon every other Friday.

The simulation algorithm of the scheduling process that we implemented can be described as follows.

Step 1. *Initialize the current simulation date and the set B of blocks available for scheduling on the current date.*

The current date is incremented during the simulation and allows us to extract the day of the week and week numbers, used to identify clinic days and the blocks available for scheduling. Each block has a certain realization date and duration determined by the block schedule.

Step 2. *If the current date is a clinic date, generate a set of new surgeries S, otherwise proceed to Step* 4.

The number of arriving surgeries n is generated according to a Poisson distribution with a fixed arrival rate. A random sample of n surgery records is selected from the historical data. The CPT code of each surgery, the surgeon's experience given by the total number of surgeries performed within the OR suite, and the scheduled duration are used as inputs for the predictive model of surgery durations. This model provides the estimates of distribution parameters for each surgery in S.

Step 3. *Sequentially assign each surgery in S to a block from B using a scheduling rule* Π. *If there is no feasible assignment for some surgery, add a new block to B according to the block schedule and assign the surgery to the new block.*

Two scheduling rules were implemented for the purposes of the current study.

(i) Π_{FFD} (*first-fit deterministic*). *First-fit scheduling rule using the mean values of surgery durations*. As modeled by the set of constraints (1), the surgery is assigned to the first available block for which the sum of mean durations is less than its length after the assignment.

(ii) Π_{FFP} (*first-fit probabilistic*). *First-fit scheduling rule under the probabilistic constraints*.

This scheduling rule has two parameters: the threshold value L and the probability of overtime α. As modeled by the set of constraints (2), the surgery is assigned to the first available block for which the probabilistic capacity constraint is satisfied after the assignment.

Step 4. *Increment the current date* (*next day*).

Step 5. *Process all the blocks from B that are scheduled for the current date and remove them from B*.

Different performance metrics (overtime, utilization, waiting time) are calculated for each removed block.

Step 6. *Stop the simulation if the current date exceeds the end date of the simulation, otherwise proceed to Step* 2.

Surgery cancellations, changes in surgery dates, and the addition of emergency, add-on, urgent, and emergent cases lead to alterations of the actual schedule on a daily basis. The decisions about such changes are almost impossible to predict, because they are made in real time based on expert opinion (both the surgeon and the scheduling team are involved), the current state of the schedule, the availability of add-on procedures, and the willingness of the surgeon to accept such changes. In our model these situations are omitted from consideration, because they can be managed to a large extent by adjusting the scheduling parameters based on the actual cancellation rates and the arrival rates for add-on, emergent, and urgent cases (e.g., overbooking for cancellations or underbooking for add-on cases). These rates differ substantially across the set of medical specialties, some of which (particularly ophthalmology) have negligible cancellation rates at the VAPHS.

4.1. Quality of Normal Approximation

In the first set of experiments, we compare the scheduling decisions by the probabilistic rule Π_{FFP} that uses Monte Carlo estimates of the probabilistic feasibility modeled by (2), and the decisions based on a normal approximation that uses a simplified chance constraint given by (4). As before, 100 one-year replications are simulated for the ophthalmology department. Decisions about whether to include a new arrival to a certain block based on a sample of 10,000 realizations of random gamma distributed durations are considered to be "true" decisions in our tests. These decisions are cross-checked with recommendations provided by the normal approximation scheme and the Monte Carlo method with varying sample sizes. The percentages of false positives and false negatives as compared to the true recommendations are presented in Table 1. In addition, we report the average number of performed surgeries (throughput), average overtime values, and average utilization percentages. These metrics are estimated using an additional 100,000 replications after the actual scheduling decisions are made.

The results of these experiments indicate that the normal approximation provides extremely accurate decisions for the mix of surgeries and duration distributions that are typical for the ophthalmology medical department at the VAPHS. The implications of the observed accuracy of the normal approximation are twofold. First, expensive Monte Carlo calculations within a simulation model can be substituted with a much simpler evaluation of the formula (4) without any significant loss of accuracy. Second, unlike other approaches, the use of the normal approximation allows us to provide an optimization model that does not rely on a generation of possible scenarios. To obtain performance comparable to the normal approximation one needs a large number of samples (see Table 1). The average performance metrics for the scheduling decisions that are made under normality assumption, such as average overtime and utilization, are close to those calculated using the Monte Carlo method with 10,000 samples. On the other hand, there is an evident difference between these metrics when

Table 1 Number of False Scheduling Recommendations for Normal Approximation and the Monte Carlo Simulation Based on Different Number of Scenarios

Approximation	Throughput	Overtime	Utilization (%)	False positives (%)	False negatives (%)
MC 50 scenarios	652.17	1,063.02	72.82	1.1	2.5
MC 100 scenarios	648.35	952.25	72.57	0.7	1.6
MC 300 scenarios	641.71	822.11	72.19	0.3	1.2
MC 500 scenarios	640.77	794.10	72.02	0.2	1.0
Normal	638.67	748.74	71.86	0.0	0.4
MC 10,000 scenarios	636.25	753.85	71.88	0.0	0.0

Note. Monte Carlo simulation with 10,000 sample points was used to calculate the "true" recommendations.

the feasibility of the chance constraint (2) is estimated using a small number of samples, which is due to chance constraint violations.

Therefore, the optimization models based on the normal approximation scheme have the potential to provide better solutions when compared to scenario-based models, in particular, for scheduling surgical specialties with the mix of surgeries and duration distributions similar to the ones observed in our data set.

In the second set of experiments, we investigate the quality of the normal approximation for the estimation of an average overtime. We randomly select a random subset of surgeries from historical data together with the distribution parameters of their durations provided by the predictive model and assume that all of them are assigned to the same scheduling block. Then, the average overtime $\mathbb{E}[(\sum_{s\in S} d_s - l(b))^+]$ is estimated using the Monte Carlo method, where d_s is a gamma random variable with shape parameter k_s and scale parameter θ_s, S is a sample set of surgeries assigned to block b of length $l(b)$. Three sets of possible block lengths are tested: $l(b) \in \{270, 450, 540\}$ minutes corresponding to block lengths used by the ophthalmology department at the VAPHS. The normal approximation for the average overtime is calculated as $\int_0^\infty (x - l(b)) f(x)\,dx$, where $f(x)$ is a probability density function of the normally distributed random variable with mean $\sum_{s\in S} k_s\theta_s$ and variance $\sum_{s\in S} k_s\theta_s^2$, which is an approximation for the convolution of gamma variables from S. For each sample size (two to 10 surgeries in the same block), we test 1,000 random combinations of surgery sets and use 100,000 samples in Monte Carlo estimation. The maximum and average values of the absolute difference between the Monte Carlo and normal approximation (in minutes) for the average overtime and corresponding 95% confidence intervals for each sample size are presented in Table 2. The worst approximation is achieved for five or six surgeries in the block, which roughly corresponds to the tight packing. Again, the normal approximation provides great accuracy in our problem domain with the maximum absolute error less than 90 seconds.

Table 2 Quality of Normal Approximation for Average Overtime

Number of surgeries	Maximum absolute difference (minutes)	Average absolute difference (minutes)	Maximum length of 95% confidence interval (minutes)
2	0.15	0.00	0.03
3	0.95	0.02	0.19
4	0.94	0.17	0.75
5	1.32	0.37	0.78
6	1.08	0.44	0.96
7	0.61	0.17	0.97
8	0.65	0.14	1.14
9	0.57	0.15	1.13
10	0.87	0.16	1.12

Table 3 Percentile Comparison of the Empirical Distributions (Sum of Surgery Durations) and the Corresponding Normal Distributions

Number of surgeries	$\eta_{0.05}$ (%)	$\eta_{0.15}$ (%)	$\eta_{0.25}$ (%)	$\eta_{0.40}$ (%)	$\eta_{0.50}$ (%)	$\eta_{0.60}$ (%)	$\eta_{0.75}$ (%)	$\eta_{0.85}$ (%)	$\eta_{0.95}$ (%)
1	2.9	14.4	26.5	43.4	53.6	63.3	76.7	85.1	93.8
2	3.3	14.5	26.2	42.8	53.1	62.8	76.4	85.1	93.9
3	3.5	14.6	26	42.4	52.7	62.5	76.3	85.1	94.1
4	3.7	14.6	25.9	42.2	52.5	62.2	76.2	85.1	94.1
5	3.8	14.7	25.9	42	52.3	62.1	76.1	85.1	94.2
6	3.9	14.7	25.8	41.9	52.1	61.9	76	85.1	94.2
7	4.0	14.7	25.8	41.8	52.1	61.9	76	85.1	94.2
8	4.0	14.7	25.7	41.7	51.9	61.8	75.9	85.1	94.3
9	4.1	14.8	25.7	41.6	51.8	61.7	75.9	85.0	94.3
10	4.1	14.8	25.7	41.6	51.8	61.6	75.8	85.0	94.3

Table 3 shows the quality of the normal approximation on the duration of the set of surgeries. We randomly select random subsets of one to 10 surgeries from historical data together with the distribution parameters of their durations provided by the predictive model and assume that all of them are assigned to the same scheduling block. For each subset, we calculate the percentiles of the corresponding normal distribution, where η_p stands for $100 \cdot p$th percentile. For each sample size (one to 10 surgeries in the same block), we test 1,000 random combinations of surgery subsets and use 100,000 samples to determine the average empirical distribution function values for each of the percentiles. From Table 3, it is clear that the normal percentiles closely follow the empirical data.

The results of the previous experiments described suggest the validity of the algorithms based on the normal approximation and show that the proposed approach provides a competitive framework compared to the scenario-based approximation when applied to high-volume specialties. Even though our computational study is focused on the ophthalmology department at the VAPHS, the normal approximation for OR scheduling has been effectively used by other large healthcare providers (Hans et al. 2008). Thus, we can reasonably assume that similar results might hold for high-volume specialties across other hospitals, requiring further verification.

4.2. Comparing Batch Scheduling to Sequential Scheduling

The batch scheduling algorithm described in §3 provides a rescheduling tool that can improve the utilization of the OR compared to the sequential scheduling using Π_{FFP}. Such rescheduling delays an assignment of surgeries to scheduling blocks. Because it is always desirable to provide a certain surgery date as soon as possible, the adjustment of each surgery assignment should be done within a certain time from the original request date for surgery. Multiple changes in the schedule should also be avoided. Taking into account the above consideration, we investigate a weekly rescheduling strategy. For example, the new ophthalmology surgeries arrive on Mondays and Wednesdays and the rescheduling can be performed on Thursday morning. After rescheduling, all surgery dates are fixed and cannot be altered afterward. The delay of any surgery assignment is no more than three days. Thus, this minor change to a sequential scheduling currently used by the VAPHS can easily be incorporated into the current practice.

The rescheduling step is implemented in conjunction with the probabilistic scheduling rule (Π_{FFP}). The rescheduling algorithm described in §3 is applied every Thursday to reschedule the assignment of surgeries that arrive on clinic days after the last rescheduling step. The block assignments provided by Π_{FFP} are used to provide an input for the rescheduling problem: a set of available blocks B, a set of surgeries S. The initial means and variances, μ_b^{init} and $(\sigma_b^{\text{init}})^2$, for each block $b \in B$ account for the surgeries that cannot be rescheduled (surgery assignments are fixed after every rescheduling step). The threshold value L and overtime probability α used by the rescheduling algorithm are identical to the parameters of Π_{FFP}, allowing the use of the block assignments provided by Π_{FFP} for calculating an initial upper bound on the optimal solution of the rescheduling problem.

The performance of the Π_{FFP} scheduling rule and the batch scheduling algorithm (maximum three-day delay) are evaluated using the different settings for the threshold parameter L and overtime probability α. Table 4 presents the average total number of processed patients (throughput) for the Π_{FFP} scheduling rule and the batch scheduling algorithm. The batch scheduling increases the total annual number of surgeries by 2.95%–3.75%. Thus, even a three-day delay in surgery assignments provides a noticeable improvement in service rates. The average throughput for the Π_{FFD} is 660 patients and the 95% confidence interval (CI) is [659.31, 661.13]. Similar service rates are achieved by the batch scheduling algorithm with the following sets of parameters: $\{L = 30,\ \alpha = 0.15\}$ and $\{L = 60,\ \alpha = 0.05\}$.

Table 5 presents the average final number of backlog days for the Π_{FFP} scheduling rule and the batch scheduling algorithm. As noted earlier, the warm-up period (initial backlog) is set to 20 days. It is clear that for every set of parameters the service rate is less than the incoming rate (arrival rate 7.0) with the exception of $\{L = 60, \alpha = 0.15\}$ for the batch scheduling. The batch scheduling decreases the backlog by 15%–36%, which is roughly equivalent to nearly

Table 4 **Total Number of Performed Surgeries (Throughput)**

	Π_{FFP}		Batch scheduling		
Parameters	Mean	95% CI	Mean	95% CI	Improvement (%)
$L = 0,\ \alpha = 0.05$	563.92	[562.8, 565.04]	585.06	[584.08, 586.04]	3.75
$L = 0,\ \alpha = 0.10$	584.49	[583.42, 585.56]	605.95	[604.85, 607.05]	3.67
$L = 0,\ \alpha = 0.15$	599.05	[598.06, 600.04]	620.39	[619.37, 621.41]	3.56
$L = 30,\ \alpha = 0.05$	601.89	[600.74, 603.04]	623.69	[622.52, 624.86]	3.62
$L = 30,\ \alpha = 0.10$	622.77	[621.56, 623.98]	645.28	[644.05, 646.51]	3.61
$L = 30,\ \alpha = 0.15$	638.04	[636.86, 639.22]	660.84	[659.61, 662.07]	3.57
$L = 60,\ \alpha = 0.05$	640.99	[639.79, 642.19]	662.37	[660.78, 663.96]	3.34
$L = 60,\ \alpha = 0.10$	662.90	[661.79, 664.01]	684.85	[683.42, 686.28]	3.31
$L = 60,\ \alpha = 0.15$	676.94	[675.21, 678.67]	696.67	[694.57, 698.77]	2.91

Table 5 **Final Number of Backlog Days (Maximum Waiting Time)**

	Π_{FFP}		Batch scheduling		
Parameters	Mean	95% CI	Mean	95% CI	Improvement (%)
$L = 0,\ \alpha = 0.05$	106.87	[103.33, 110.41]	90.04	[86.63, 93.45]	15.75
$L = 0,\ \alpha = 0.10$	90.71	[87.47, 93.95]	74.96	[71.81, 78.11]	17.36
$L = 0,\ \alpha = 0.15$	77.60	[74.45, 80.75]	62.53	[59.55, 65.51]	19.42
$L = 30,\ \alpha = 0.05$	75.86	[72.43, 79.29]	61.15	[57.93, 64.37]	19.39
$L = 30,\ \alpha = 0.10$	60.66	[57.65, 63.67]	46.21	[43.49, 48.93]	23.82
$L = 30,\ \alpha = 0.15$	52.44	[49.65, 55.23]	38.64	[35.94, 41.34]	26.32
$L = 60,\ \alpha = 0.05$	48.75	[46.11, 51.39]	36.01	[33.53, 38.49]	26.13
$L = 60,\ \alpha = 0.10$	38.52	[35.65, 41.39]	25.94	[23.28, 28.6]	32.66
$L = 60,\ \alpha = 0.15$	28.68	[26.26, 31.1]	18.16	[16.05, 20.27]	36.68

Table 6 **Total Overtime in Minutes (Gamma Distribution)**

	Π_{FFP}		Batch scheduling	
Parameters	Mean	95% CI	Mean	95% CI
$L=0$, $\alpha=0.05$	78.77	[77.75, 79.79]	122.65	[121.56, 123.73]
$L=0$, $\alpha=0.10$	162.09	[160, 164.18]	250.65	[248.71, 252.6]
$L=0$, $\alpha=0.15$	253.33	[250.27, 256.39]	386.17	[383.09, 389.25]
$L=30$, $\alpha=0.05$	283.93	[280.69, 287.18]	456.95	[453.6, 460.31]
$L=30$, $\alpha=0.10$	514.52	[509.12, 519.93]	812.38	[807.28, 817.48]
$L=30$, $\alpha=0.15$	747.86	[740.13, 755.58]	1155.22	[1148.18, 1162.27]
$L=60$, $\alpha=0.05$	833.52	[823.71, 843.33]	1291.00	[1281.77, 1300.23]
$L=60$, $\alpha=0.10$	1346.17	[1333.61, 1358.73]	2006.82	[1990.91, 2022.72]
$L=60$, $\alpha=0.15$	1837.20	[1821.14, 1853.25]	2575.39	[2545.66, 2605.12]

Table 7 **Total Overtime in Minutes (Normal Distribution)**

	Π_{FFP}		Batch scheduling	
Parameters	Mean	95% CI	Mean	95% CI
$L=0$, $\alpha=0.05$	44.42	[43.74, 45.1]	79.62	[78.86, 80.38]
$L=0$, $\alpha=0.10$	112.13	[110.54, 113.72]	193.25	[191.65, 194.85]
$L=0$, $\alpha=0.15$	194.95	[192.32, 197.58]	323.84	[321.08, 326.6]
$L=30$, $\alpha=0.05$	224.19	[221.24, 227.15]	395.21	[391.95, 398.47]
$L=30$, $\alpha=0.10$	450.14	[444.93, 455.36]	756.16	[751.06, 761.27]
$L=30$, $\alpha=0.15$	688.56	[680.86, 696.26]	1112.03	[1104.86, 1119.21]
$L=60$, $\alpha=0.05$	776.21	[766.16, 786.26]	1250.28	[1240.76, 1259.81]
$L=60$, $\alpha=0.10$	1307.53	[1294.45, 1320.61]	1994.45	[1978.06, 2010.84]
$L=60$, $\alpha=0.15$	1818.33	[1801.7, 1834.97]	2584.31	[2553.71, 2614.92]

16 days of backlog reduction (per year). The average final number of backlog days for the Π_{FFD} is 38.98 days, 95% CI is [36.43, 41.53]. Again, similar results are achieved by the batch scheduling algorithm with the parameters: $\{L=30, \alpha=0.15\}$ and $\{L=60, \alpha=0.05\}$.

Tables 6 and 7 present the average total annual overtime for schedules provided by the Π_{FFP} scheduling rule and the batchscheduling algorithm. The maximum absolute total error provided by the normal approximation is around 60 minutes, which is the sum of approximation errors over approximately 130 blocks. The average total annual overtime for the Π_{FFD} is 1,284.88, 95% CI is [1271.59, 1298.16] (for the gamma distribution). As seen earlier, the service rates and backlog statistics of the Π_{FFD} are similar to the batch scheduling with the following parameters: $\{L=30, \alpha=0.15\}$ and $\{L=60, \alpha=0.05\}$. However, the resulting overtime for the batch scheduling algorithm with the first set of parameters is improved by 15% (the second set of parameters produces similar statistics).

The results in Tables 4–7 highlight the effect of competing priorities that are common to OR scheduling problems; the patient backlog and room utilization can be improved only at the cost of increased overtime. The set of scheduling policies provides an easy way to control the output performance metrics based on management preferences. This is difficult to achieve using only point estimates for scheduling.

5. Conclusions

In this paper, we present a chance-constrained optimization model of batch scheduling for high-volume specialties within the OR suite that uses a block booking system. We develop an algorithm based on a normal approximation for the sum of surgery durations to provide near-optimal solutions to the stochastic scheduling problem. This approximation is particularly suitable for high-volume medical specialities. We test our approach using the historical data from the ophthalmology department provided by the Veterans Affairs Pittsburgh Healthcare System. A set of computational experiments with the discrete simulation model of the scheduling process reveals the high accuracy of our method. We compare the sequential scheduling heuristic (first-fit) to the optimal batch scheduling and show the superiority of the latter approach.

Acknowledgments

The authors research benefited from the supercomputing resources provided by the University of Pittsburgh Center for Simulation and Modeling. The authors thank Mark Wilson, William Stevens, Robert Monte, Scott Moore, Matthew Jenkins, Ian Joos, Deborah Owens of the VAPHS, and Jayant Rajgopal, Winston Yang, Louis Luangkesorn of the University of Pittsburgh for their assistance and feedback. The authors also thank Area Editor Allen Holder, an anonymous associate editor, and two anonymous referees for their insightful and constructive comments. This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, and is also partially supported by National Science Foundation [Grant CMM1-0620780] and Air Force Office of Scientific Research Award FA9550-11-1-0037.

References

Batun S, Denton BT, Huschka TR, Schaefer AJ (2011) Operating room pooling and parallel surgery processing under uncertainty. *INFORMS J. Comput.* 23:220–237.

Blake JT, Carter MW (1997) Surgical process scheduling: A structured review. *J. Soc. Health Systems* 5:17–30.

Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. *Eur. J. Oper. Res.* 201:921–932.

Charnes A, Cooper WW, Symonds GH (1958) Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Sci.* 4:235–263.

Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. *Oper. Res.* 58:802–816.

Dexter F (2011) Surgical services management. Accessed September 9, 2011, http://www.FranklinDexter.net/bibliography_SurgicalServices.htm.

Dexter F, Dexter EU, Ledolter J (2010) Influence of procedure classification on process variability and parameter uncertainty of surgical case durations. *Anesthesia and Analgesia* 110: 1155–1163.

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

Eijkemans MJC, van Houdenhoven M, Nguyen T, Boersma E, Steyerberg EW, Kazemier G (2010) Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology* 112:41–49.

Erdogan SA, Denton BT (2011) Surgery planning and scheduling: A literature review. Cochran J, Cox T, Keskinocak P, Kharoufeh JP, Smith JC, eds. *Wiley Encyclopedia of Operations Research and Management Science* (John Wiley & Sons, New York), 5414–5427.

Firth D (1991) Generalized linear models. Hinkley DV, Reid N, Snell EJ, eds. *Statistical Theory and Modelling. In Honour of Sir David Cox*, Chap. 3. (Chapman & Hall, London), 55–82.

Goyal V, Ravi R (2010) A PTAS for the chance-constrained knapsack problem with random item sizes. *Oper. Res. Lett.* 38:161–164.

Hans E, Wullink G, Houdenhoven M, Kazemier G (2008) Robust surgery loading. *Eur. J. Oper. Res.* 185:1038–1050.

Healthcare Financial Management Association (2003) Achieving operating room efficiency through process integration. *Healthcare Financial Management* 57(Suppl.):1–7.

Kleywegt AJ, Shapiro A, Homem-de-Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* 12:479–502.

Luedtke J, Ahmed S (2008) A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.* 19:674–699.

Luedtke J, Ahmed S, Nemhauser G (2010) An integer programming approach for linear programs with probabilistic constraints. *Math. Programming* 122:247–272.

Min D, Yih Y (2010) Scheduling elective surgery under uncertainty and downstream capacity constraints. *Eur. J. Oper. Res.* 206: 642–652.

Moschopoulos P (1985) The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.* 37:541–544.

Nemirovski A, Shapiro A (2006) Convex approximations of chance constrained programs. *SIAM J. Optim.* 17:969–996.

Pagnoncelli BK, Ahmed S, Shapiro A (2009) Sample average approximation method for chance constrained programming: Theory and applications. *J. Optim. Theory Appl.* 142:399–416.

Prékopa A (1995) *Stochastic Programming* (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Shylo O, Luangkesorn L, Prokopyev O, Rajgopal J, Schaefer A (2011) Managing patient backlog in a surgical suite that uses a block-booking scheduling system. Jain S, Creasey RR, Himmelspach J, White KP, Fu M, eds. *Winter Simulation Conf. Proc.* (IEEE, New York), 1314–1324.

Stepaniak PS, Heij C, de Vries G (2010) Modeling and prediction of surgical procedure times. *Statistica Neerlandica* 64:1–18.

Strum DP, May JH, Vargas LG (2000) Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology* 92:1160–1167.

Vielma JP, Nemhauser GL (2011) Modeling disjunctive constraints with a logarithmic number of binary variables and constraints. *Math. Programming* 128:49–72.

Wiens BL (1999) When log-normal and gamma models give different results: A case study. *Amer. Statistician* 53:89–93.