Flex Serv Manuf J (2018) 30:198–223
https://doi.org/10.1007/s10696-016-9253-7

Performance evaluation of operating room schedules in orthopedic surgery

Zexian Zeng[1] · Xiaolei Xie[2] · Heidi Menaker[3] · Susan G. Sanford-Ring[3] · Jingshan Li[4]

Published online: 31 August 2016
© Springer Science+Business Media New York 2016

Abstract In orthopedic surgery department, multiple surgeries are carried out in the same operating room (OR) every day. Each surgery may have a random duration, which results in room idle time and patient waiting time. One of the major factors affecting the idle and waiting times is the schedule of surgeries in the OR. To better sequence multiple surgeries to reduce idle and waiting time, an effective performance evaluation method is needed. Although discrete event simulation can be used to evaluate the performance of surgery schedules, the long simulation time makes scheduling optimization more computationally intensive. In this paper, an analytical model to evaluate the performance of OR schedules in orthopedic surgery department is introduced. First, closed formulas to evaluate the case of two surgeries are developed and used as a building block in subsequent studies. Then, an iteration

✉ Jingshan Li
jingshan.li@wisc.edu

Zexian Zeng
zexian.zeng@northwestern.edu

Xiaolei Xie
xxie@tsinghua.edu.cn

Heidi Menaker
hmenaker@uwhealth.org

Susan G. Sanford-Ring
sanford-ring@uwhealth.org

1 Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

2 Department of Industrial Engineering, Tsinghua University, Beijing 100084, People's Republic of China

3 University of Wisconsin Hospital and Clinics, Madison, WI 53792, USA

4 Department of Industrial and Systems Engineering, University of Wisconsin, Madison, WI 53706, USA

procedure is presented by aggregating every two surgeries into one using the two-surgery formula, and continuing to approach the next one, until all surgeries are aggregated into one. Using such a model, the expected idle time and waiting time for a given surgery schedule can be calculated quickly and accurately. Such a work provides an efficient performance evaluation tool that can be used for optimization of OR schedules and investigation of the impacts of different schedules.

Keywords Operating room · Orthopedic surgery · Idle time · Waiting time · Scheduling · Aggregation

1 Introduction

The operating room (OR), or operating theater, is the largest cost center in a hospital, but it also generates about 42 % of a hospital's revenue (Health Care Financial Management Association 2003). A hospital's income and performance are significantly affected by the effectiveness of OR. However, managing the OR is difficult due to the complexity in surgeries, the limited resources, and various needs and expectations from patients, families, and support systems. For each surgery, multiple departments are required to work together to ensure processes such as tool preparation, anesthesia, operations, and post anesthesia being carried properly. Therefore, significant amount of attentions have been paid from the hospital's governing body and from researchers to organize surgical care with least cost. OR scheduling is one of the central issues.

Orthopedic surgery deals with conditions involving the musculoskeletal system. The surgery is practiced to restore the function of bones, joints, tendons, ligaments, nerves, or skin as a result of injury or disease. Compared to other surgeries in general, many orthopedic surgeries, for example, total knee replacement, have standardized treatment protocols. Thus, orthopedic surgeries typically have more standardized processes with fewer complications and less variations. Based on a study of 10 surgery types from 5269 events collected at University of Wisconsin Hospital and Clinics (UW Health), it is found that the variability of orthopedic surgery time is statistically significantly smaller than that of the non-orthopedic ones. This enables the possibility of accurate prediction of average surgery completion time.

As one knows, an optimal surgery schedule depends on accurate evaluation of the schedule performance, such as room idle time, patient waiting time (or tardiness), and average completion time. Typically, either deterministic models or discrete event simulations are used in performance evaluation of OR schedules. However, the variability and resulting idle and waiting times are ignored in deterministic models, and simulations suffer from substantial computation intensity. Moreover, in addition to mean time, variability also plays an important role. An effective method to evaluate the mean and variability performance of OR schedules quickly and accurately is necessary and important, and it also contributes significantly to helping solve the OR schedule optimization problem. Unfortunately, to our best knowledge, an analytical model addressing the idle and waiting times as well as surgery completion time and its variance is still not available. The main contribution of this paper is in developing such a method. Specifically, an aggregation method is presented to approximate the mean surgery completion time, the associated variance, the patient waiting time, and the room idle time for a given surgery sequence. To evaluate the effectiveness of the proposed method, a case study using the data collected on the hospital floor at UW Health is carried out. A software suite is developed to implement the algorithms and used in a pilot study at UW Health.

The remainder of this paper is structured as follows: The related literature is reviewed in Sect. 2. Section 3 introduces the system and formulates the problem. The performance evaluation method is presented in Sect. 4 and validated in Sect. 5. In addition, Sect. 6 provides an illustration of the software implementing the results obtained in this paper. Finally, conclusions and future work are summarized in Sect. 7. All proofs are provided in the "Appendix".

2 Literature review

Substantial efforts have been devoted to scheduling in healthcare delivery systems (see, for instance, reviews by Cayirli et al. 2006; Wright et al. 2006; Gupta and Denton 2008; Cardoen et al. 2010; Erdogan and Denton 2010; Guerriero and Guido 2011; May et al. 2011 and papers by Robinson and Chen 2003; Muthuraman and Lawley 2008; Chakraborty et al. 2010; Liu et al. 2010; Mancilla and Storer 2012; Zacharias and Pinedo 2014). In particular, OR scheduling has been one of the central issues. The focus is mainly on improving the theater's efficiency, turnover rate, patient outcome, and surgical department capacity. For example, to clarify the different concepts from physicians, nursing researchers, administrators and management scientists, Blake and Carter (1997) describe a conceptual framework for surgical process scheduling. It shows that even if the operational aspects of advance and allocation scheduling are well understood, resolving scheduling issues at strategic and administrative levels is needed and the techniques to integrate OR scheduling with other hospital operations are required. Cardoen et al. (2010) provide a review of OR capacity planning and scheduling research and summarize the research trends and key areas to be addressed in future work. Gupta and Denton (2008) introduce appointment scheduling of selective surgeries and discuss the complicated factors, such as arrivals, services, and patient and provider preferences. Erdogan and Denton (2010) present an overview of the most important parts in surgery delivery system affecting surgery scheduling. A taxonomy of the literature based on the type of operations research methods used is provided and open challenges are discussed. Finally, a structured literature review is given in Guerriero and Guido (2011) on how operations research techniques can be applied to the surgical planning and scheduling processes, with a focus on mathematical (optimization and simulation) models and solution approaches. In addition, the literature of OR management decision making on daily surgeries is reviewed in Dexter et al. (2004). It also discusses decisions to reduce OR over-utilization and patient waiting time.

From the methodology point of view, mathematical programming is typically used to study OR scheduling. In most of the studies, deterministic surgical time is assumed. For instance, Blake and Donald (2002) introduce a case study of using an integer-programming model and a post-solution heuristic to allocate OR time to five surgical divisions at Toronto's Mount Sinai Hospital. Kuo et al. (2003) use linear programming technique to optimize the allocation of OR time among a group of surgeons to maximize revenue or minimize costs. Using the job shop scheduling method, referred to as multi-mode blocking job shop, a scheduling approach is proposed in Pham and Klinkert (2008) for elective and add-on surgeries by formulating it as a mixed integer linear programming problem.

To accommodate the randomness or uncertainty in surgical operations, stochastic optimization methods are often used. Denton et al. (2007) present a stochastic optimization model and practical heuristics to determine OR schedules that hedge against the uncertainties in surgery durations. Lamiri et al. (2008) describe a stochastic model for OR planning with both elective and emergency surgeries. The elective cases are assigned to different periods to minimize the sum of related costs and overtime cost, and a stochastic programming model is proposed to combine Monte Carlo simulation with mixed integer programming for optimization. Zhang et al. (2014) introduce a multistage stochastic programming model to dynamically assign a given set of surgeries to multiple identical operating rooms with planned surgeon arrival times. By considering patient priority, Min and Yih (2010) present a stochastic dynamic programming model to schedule elective surgery with a limited capacity. However, such methods usually are suitable for small problems with specific assumptions. Computation efforts increase substantially when the number of surgeries is increasing.

To evaluate the performance of surgical schedules, discrete event simulations are often utilized. Wullink et al. (2007) use discrete event simulation to study reserved surgical capacity for emergency department. It is found that the room utilization and overtime can be significantly improved if there is one dedicated operating room in all elective operating rooms for emergencies. Zhang et al. (2009) develop a mixed integer programming model to determine a weekly OR allocation template to minimize inpatient's cost measured by length of stay. The solution of the analytical model is used as an input to a simulation model that captures the randomness of the process and non-linearities. However, as one can expect, long simulation time is a significant limitation in such an approach.

From the performance improvement point of view, different strategies for OR scheduling have been investigated. Dexter and Traub (2002) study elective surgery scheduling to adjust anesthesia and nursing staffing to maximize the efficiency of OR usage with two patient-scheduling rules: earliest or latest start time. Guinet et al. (2003) consider the operation theatre planning in two steps. First, patients are assigned to operating rooms. Second, each room is scheduled individually. A similar two-step approach for OR scheduling is introduced in Jebali et al. (2006). In addition, Oostrum et al. (2008) present a cyclic operation room schedule that includes a master surgical schedule, a mathematical program with stochastic constraint and a column generation approach to maximize OR utilization and level requirements for subsequent hospital beds. Aiming at open scheduling strategy, Fei et al. (2009) provide a mathematical model to assign surgical cases to operating rooms and also use a column-generation-based heuristic procedure to find a solution with the best performance. Again, when applying the models, the above mentioned limitations still exist. Moreover, in all the studies, variance is not investigated.

In spite of the efforts, developing efficient algorithms to achieve optimal utilization of operating rooms and to reduce patient waiting times is still needed. Using such algorithms to replace simulations, the computation intensity in optimization of OR schedules can be reduced. To achieve this, an accurate estimation of the mean surgery completion time and its variance is necessary. Therefore, this paper introduces an analytical method to efficiently evaluate patient waiting time and room idle time for a given surgery schedule.

3 System description and problem formulation

Due to the complexity in surgeries, when multiple surgeries are scheduled in one room, it is not uncommon to observe that one surgery finishes earlier than next surgery's appointment time. The theater and other resources remain idle and unutilized until the next surgery starts as scheduled. It is also common to see one surgery finishing later than next surgery's scheduled time. Therefore, patients and resources for later appointments have to wait. If additional time is needed at the end of the day, then overtime occurs.

To reduce room idle time and patient waiting time (or tardiness), the surgery sequence plays a key role. To study such an issue, we focus on daily surgery scheduling in one operating room. Each surgery type is modeled with a random duration described by a probability density function. Then the goal is to develop a method to evaluate completion time, as well as room idle and patient waiting times for a given surgery sequence efficiently.

3.1 Assumptions

The following assumptions address the orthopedic surgery schedules studied in this paper:

1. There are N surgeries, $S_1, S_2, \ldots, S_N$, to be scheduled in one orthopedic operating room per day.
2. All patients arrive in time. No patients will arrive later than the scheduled arrival time, which is typically 90 min before the scheduled surgery time.
3. All pre-surgery operations will be finished by the scheduled surgery starting time. In other words, a surgery will not be delayed due to incomplete preparation.
4. The surgeries are scheduled based on the mean surgery time. The turnover time between surgeries is included in the surgery duration.
5. There are K types of surgeries. The duration of surgery type $j, j = 1, \ldots, K$, is described by a random variable with unimodal probability density function $f_j(t)$. The mean time and standard deviation of surgery type j are defined as τ_j and σ_j, respectively.

6. The first surgery always starts on time. Surgery cancellation and postponement are not considered when scheduling is made.
7. If a surgery finishes earlier than the scheduled finishing time, the next surgery will not begin until the scheduled starting time. The gap between two surgeries is referred to as the room *idle time*.
8. If a surgery finishes later than the scheduled finishing time, the next surgery will start immediately. The overtime of the previous surgery contributes to the patient *waiting time*.

Remark 1 Among 5269 surgeries carried out in UW Health in 2012, only 328 patients, i.e., 6.2 %, arrived later than the scheduled arrival time. The average late time is 13 $\pm$ 1.9 min. Therefore, the assumption of in-time patient arrival is reasonable.

Remark 2 The goal of this paper is to present a performance evaluation method rather than an optimization algorithm of surgery schedules. In addition, developing the optimal and robust schedules relies on efficient and accurate schedule evaluation. Thus, the method introduced in this paper provides a foundation for optimization of surgery schedules with possible cancellations and delay. In the case study we carried out, cancellation or delay is not a severe concern when schedules are made. Therefore, at current stage, cancellation or delay is not considered in performance evaluation. In future work, these issues will be studied.

Different types of surgeries may be fitted by various probabilistic distributions. To start with, we assume Gamma distribution for surgery time, as Gamma distribution has been widely used to model the time length (Coit and Jin 2000), and it has two parameters, α_j and β_j, which provide the freedom to fit mean and variance. Specifically, they can be obtained by solving the equations:

$$\tau_j = \frac{\alpha_j}{\beta_j}, \quad \sigma_j = \sqrt{\frac{\alpha_j}{\beta_j^2}}, \quad j = 1, \ldots, N, \tag{1}$$

which leads to

$$\alpha_j = \frac{\tau_j^2}{\sigma_j^2}, \quad \beta_j = \frac{\tau_j}{\sigma_j^2}, \quad j = 1, \ldots, N. \tag{2}$$

Remark 3 Clearly, the surgery durations may not exactly follow Gamma distributions. In Sect. 5, based on the information extracted from 5269 surgeries at UW Health in 2012, we justify that the system performance is mainly dependent on the mean and coefficient of variation of the surgical time, rather than the complete distribution. In other words, it does not depend on the distribution type but primarily depend on the mean and coefficient of variation. Then using Gamma distribution (characterized by mean and variance), accurate estimation of waiting time and idle time can be obtained. Such properties also exist in other healthcare and manufacturing applications.

3.2 Problem formulation

To evaluate the performance of OR schedules, patient waiting time and room idle time are important measurements. Introduce T_{w_i} as the patient waiting time for the i-th surgery, $i = 2, \ldots, N$, i.e., from the scheduled starting time of surgery i to its actual starting time, in case of late finishing of surgery $i-1$. Let T_{e_i} denote the room idle time of surgery $i, i = 1, \ldots, N-1$, i.e., from the time surgery S_i finishes to the scheduled starting time of surgery S_{i+1}, during which the room is empty of patient. Then the total waiting time and idle time are the summation of the corresponding time of each surgery.

$$T_w = \sum_{i=2}^{N} T_{w_i}, \quad T_e = \sum_{i=1}^{N-1} T_{e_i}. \tag{3}$$

In addition to mean idle time and waiting time, the mean and the variance of the completion time of all surgeries are also of interest. Let C_i represent the average time to complete surgeries 1 to $i, i = 2, \ldots, N$, and V_i as its variance. Evaluation of C_i and V_i will also be pursued.

Thus, the problem is formulated as follows: *Under assumptions (1)–(8), develop a method to evaluate the room idle time T_e and patient waiting time T_w, as well as surgery completion time C_i and its variance V_i.* The solutions to the formulated problem are introduced next.

4 Performance evaluation method

The goal of this paper is to develop a method to calculate the mean completion time, variance, the patient waiting time and room idle time when multiple surgeries are scheduled in one operating room. Under assumption 5), the duration of surgery $S_i, i = 1, \ldots, N$, has parameters α_i and β_i following Gamma distribution,

$$g(x; \alpha_i, \beta_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i - 1} e^{-\beta_i x}, \tag{4}$$

where α_i and β_i are defined in (2) and $\Gamma(s)$ is the gamma function,

$$\Gamma(s) = \int_0^{\infty} t^{s-1} e^{-t} dt. \tag{5}$$

To evaluate the system performance, we start with a two-surgery scenario, and then extend to N surgeries.

4.1 Two surgeries

First we consider the case that surgery S_1 finishes before the scheduled time (i.e., mean time τ_1). Such a probability, p_{e_1}, can be calculated as

$$p_{e_1} = \int_0^{\tau_1} g(x)dx = \frac{\gamma(\alpha_1, \alpha_1)}{\Gamma(\alpha_1)}, \tag{6}$$

where $\gamma(s,x)$ is the lower incomplete gamma function,

$$\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt. \tag{7}$$

In this case, room idle time will occur. Since the first surgery S_1 finishes before its scheduled time τ_1, and the second surgery S_2 will start at the scheduled time, it is equivalent to view that S_1 still "completes at time τ_1". Therefore, the mean completion time of two surgeries (S_1 and S_2) equals the scheduled surgical time. In addition, S_1 can be viewed as a "constant" without variance, and the two surgeries are "consecutive" so that the variance of two surgeries equals to the variance of the second one. Therefore, the mean completion time of two surgeries and its variance can be calculated as:

Lemma 1 *Under assumptions (1)–(8) with $N = 2$, if surgery S_1 finishes before the scheduled time τ_1, then the mean completion time $C_{2|early}$ and the variance $V_{2|early}$ can be calculated as*

$$C_{2|early} = \frac{\alpha_1}{\beta_1} + \frac{\alpha_2}{\beta_2} = \tau_1 + \tau_2, \tag{8}$$

$$V_{2|early} = \frac{\alpha_2}{\beta_2^2} = \sigma_2^2. \tag{9}$$

Next we consider the scenario that surgery S_1 goes over the scheduled finishing time τ_1. The probability of such an event is

$$p_{w_1} = \int_{\tau_1}^{\infty} g(x)dx = \frac{\zeta(\alpha_1, \alpha_1)}{\Gamma(\alpha_1)}, \tag{10}$$

where $\zeta(s,x)$ is the upper incomplete gamma function,

$$\zeta(s,x) = \int_x^{\infty} t^{s-1}e^{-t}dt. \tag{11}$$

Under the condition that surgery S_1 finishes after its scheduled time τ_1, the mean time of surgery S_1 will be equal to

$$\int_{\tau_1}^{\infty} xg(x|x > \tau_1)dx = \frac{\zeta(1 + \alpha_1, \alpha_1)}{\beta_1 \zeta(\alpha_1, \alpha_1)}. \tag{12}$$

Remark 4 Note that the above integral and much of the subsequent derivation are obtained using *Mathematica*.

As surgery S_2 will start immediately after S_1, summing up the surgical time of S_2 will give the mean completion time of both surgeries. Similarly, the variance of delayed surgery S_1 can be calculated as

$$\int_{\tau_1}^{\infty} x^2 g(x|x > \tau_1)dx - \frac{\zeta^2(1+\alpha_1,\alpha_1)}{\beta_1^2\zeta^2(\alpha_1,\alpha_1)} = \frac{\zeta(\alpha_1,\alpha_1)\zeta(2+\alpha_1,\alpha_1) - \zeta^2(1+\alpha_1,\alpha_1)}{\beta_1^2\zeta^2(\alpha_1,\alpha_1)}. \tag{13}$$

Then the variance of both surgeries will be equal to the sum of their individual ones. Therefore, the mean completion time and the variance of both surgeries can be calculated as:

Lemma 2 *Under assumptions (1)–(8) with* $N = 2$, *if surgery* S_1 *finishes after the scheduled time* τ_1, *then the mean completion time* $C_{2|late}$ *and the variance* $V_{2|late}$ *can be calculated as*

$$C_{2|late} = \frac{\zeta\left(1+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1}{\sigma_1^2}\zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)} + \tau_2, \tag{14}$$

$$V_{2|late} = \frac{-\zeta^2\left(1+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1^2}{\sigma_1^4}\zeta^2\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)} + \frac{\zeta\left(2+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1^2}{\sigma_1^4}\zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)} + \sigma_2^2. \tag{15}$$

By considering both scenarios, the mean surgery completion time C_2 and variance V_2 for both surgeries can be evaluated. In addition, the room idle time T_{e_1} (when S_1 finishes early) and patient waiting time T_{w_2} (when S_1 finishes late) are important measures of the efficacy of surgical schedules. These measures can be calculated as follows:

Proposition 1 *Under assumptions (1)–(8) with* $N = 2$, *the mean surgery completion time* C_2 *and the variance* V_2 *as well as the average room idle time* T_{e_1} *and average patient waiting time* T_{w_2} *can be calculated as*

$$C_2 = \tau_1 + \tau_2 + \frac{\zeta\left(1+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right) - \frac{\tau_1^2}{\sigma_1^2}\zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1}{\sigma_1^2}\Gamma\left(\frac{\tau_1^2}{\sigma_1^2}\right)}, \tag{16}$$

$$V_2 = \frac{-\zeta^2\left(1+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right) + \zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)\zeta\left(2+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1^2}{\sigma_1^4}\Gamma\left(\frac{\tau_1^2}{\sigma_1^2}\right)\zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)} + \sigma_2^2, \tag{17}$$

$$T_{e_1} = \frac{-\frac{\tau_1^2}{\sigma_1^2}\zeta\left(\frac{\tau_1^2}{\sigma_1^2}, \frac{\tau_1^2}{\sigma_1^2}\right) + \zeta\left(1 + \frac{\tau_1^2}{\sigma_1^2}, \frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1}{\sigma_1}\Gamma\left(\frac{\tau_1^2}{\sigma_1^2}\right)}, \tag{18}$$

$$T_{w_2} = \frac{-\frac{\tau_1^2}{\sigma_1^2}\zeta\left(\frac{\tau_1^2}{\sigma_1^2}, \frac{\tau_1^2}{\sigma_1^2}\right) + \zeta\left(1 + \frac{\tau_1^2}{\sigma_1^2}, \frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1^2}{\sigma_1^4}\Gamma\left(\frac{\tau_1^2}{\sigma_1^2}\right)}. \tag{19}$$

Proof See the "Appendix". □

Clearly, C_2 is monotonically increasing with respect to the mean time of the second surgery τ_2 and V_2 is monotonically increasing with respect to the standard deviation of the second surgery σ_2. However, C_2 is independent of σ_2 and V_2 is independent of τ_2. In other words, variance in the second surgery does not impact the average completion time, and the mean time of the second surgery does not affect the variance.

To study the monotonicity with respect to the mean time of the first surgery τ_1, we increase τ_1 under different values of standard deviation σ_1. An illustration is shown in Fig. 1, where τ_1 is increased from 30 to 300 by 1 for different σ_1 values starting from 30 to 270 with increments 60.

As one can see, C_2 is monotonically increasing with respect to τ_1 almost linearly. In addition, larger σ_1 will lead to longer C_2, i.e., the completion time is monotonically increasing with respect to the variability of the first surgery. In other words, reducing the variation of the first surgery could decrease the total completion time. Thus, standardized processes could help reduce the variance of surgery time, which reduces the total completion time of two consecutive surgeries.

Similarly, V_2 is monotonically increasing with respect to σ_2 and τ_1. In addition, it is also monotonically increasing with respect to σ_1. As shown in Fig. 2, when σ_1 is

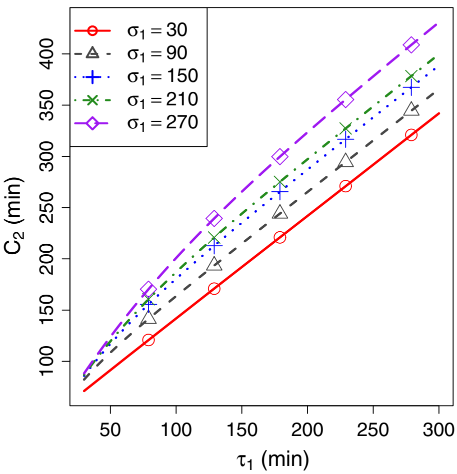

Fig. 1 Monotonicity of C_2 with respect to τ_1

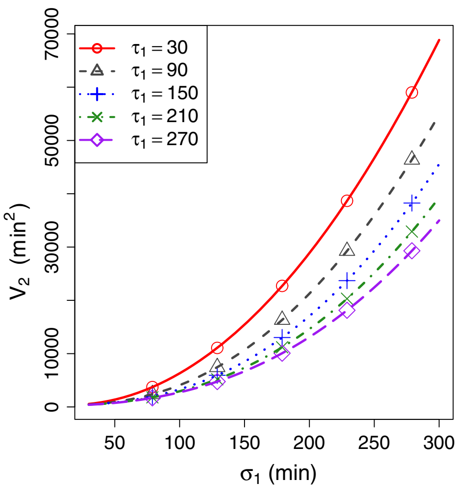

Fig. 2 Monotonicity of V_2 with respect to σ_1

increased from 30 to 300 by 1 for different values of τ_1 starting from 30 to 270 with increments 60, V_2 is monotonically increasing, and the growth rate is also increasing. Thus, larger τ_1 always results in larger V_2.

4.2 Three surgeries

When there are more than two surgeries, direct integral to derive the idle time and waiting time is difficult, since the number of possible scenarios will increase substantially. Therefore, an approximation method is pursued. To do this, we aggregate the first two surgeries into one, and assume that this aggregated surgery still follows a Gamma distribution. In other words, S_{a_2} represents the aggregated surgery of both surgeries S_1 and S_2. The mean time and its variance of surgery S_{a_2} are defined by C_2 and V_2 obtained in the Sect. 4.1. Then parameters α_{a_2} and β_{a_2} of surgery S_{a_2} can be obtained.

Lemma 3 *Under assumptions (1)–(8) with* $N = 3$, *the aggregated parameters of the first two surgeries can be evaluated as:*

$$\alpha_{a_2} = \frac{\zeta(\alpha_1,\alpha_1)(\beta_1(\alpha_2+\beta_2\tau_1)\Gamma(\alpha_1)+\beta_2[-\alpha_1\zeta(\alpha_1,\alpha_1)+\zeta(1+\alpha_1,\alpha_1)])^2}{\Gamma(\alpha_1)\big(\alpha_2\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1,\alpha_1)+\beta_2^2[-\zeta^2(1+\alpha_1,\alpha_1)+\zeta(\alpha_1,\alpha_1)\zeta(2+\alpha_1,\alpha_1)]\big)}, \tag{20}$$

$$\beta_{a_2} = \frac{\beta_1\beta_2\zeta(\alpha_1,\alpha_1)(\beta_1(\alpha_2+\beta_2\tau_1)\Gamma(\alpha_1)+\beta_2[-\alpha_1\zeta(\alpha_1,\alpha_1)+\zeta(1+\alpha_1,\alpha_1)])}{\alpha_2\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1,\alpha_1)+\beta_2^2\big(-\zeta^2(1+\alpha_1,\alpha_1)+\zeta(\alpha_1,\alpha_1)\zeta(2+\alpha_1,\alpha_1)\big)}. \tag{21}$$

Proof See the "Appendix". □

Using S_{a_2} (with parameters α_{a_2} and β_{a_2}) and S_3, the mean completion time of three surgeries C_3, and the variance V_3, can also be evaluated.

Proposition 2 *Under assumptions (1)–(8) with* $N = 3$, *the mean completion time* C_3 *and the variance* V_3 *of the first two surgeries can be calculated as:*

$$C_3 = \frac{\zeta(1+\alpha_{a_2}, \beta_{a_2}\tau_{s_2}) - \beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})} + \tau_{s_2} + \tau_3, \tag{22}$$

$$V_3 = \sigma_3^2 + \frac{\zeta(2+\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2}^2\Gamma(\alpha_{a_2})} - \frac{\zeta^2(1+\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2}^2\Gamma(\alpha_{a_2})\zeta(\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}. \tag{23}$$

where

$$\tau_{s_2} = \tau_1 + \tau_2. \tag{24}$$

Proof See the "Appendix". □

However, C_3 cannot be used to evaluate idle time and waiting time directly, since they depend on many combinations of finishing times of S_1 and S_2. For instance, there will be both idle and waiting times if S_1 finishes before its scheduled finishing time, but S_2 ends after its scheduled finishing time. Thus, we need to compare the completion time with the scheduled finishing time.

To derive the idle time of the second surgery and waiting time for the third surgery, we use the two-surgery formula by replacing S_1 with S_{a_2}.

Lemma 4 Under assumptions (1)–(8) with $N = 3$, the average idle time of surgery S_2 and average waiting time of surgery S_3 can be calculated as

$$T_{e_2} = \frac{(\beta_{a_2}\tau_{s_2} - \alpha_{a_2})\Gamma(\alpha_{a_2}) - \beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2}, \beta_{a_2}\tau_{s_2}) + \zeta(1+\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})}, \tag{25}$$

$$T_{w_3} = \frac{\zeta(1+\alpha_{a_2}, \beta_{a_2}\tau_{s_2}) - \beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})}. \tag{26}$$

Proof See the "Appendix". □

Then the total waiting time and idle time can be obtained by summing up the waiting time of surgeries S_2 and S_3, and by summing up the idle times of surgeries S_1 and S_2, respectively (see 3). Therefore, the final expressions of T_w and T_e are obtained.

Corollary 1 *Under assumptions (1)–(8) with* $N = 3$, *the total average waiting and idle times can be calculated as:*

$$T_w = \frac{\zeta(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2}) - \beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})} + \frac{\zeta(1+\alpha_1,\beta_1\tau_1) - \beta_1\tau_1\zeta(\alpha_1,\beta_1\tau_1)}{\beta_1\Gamma(\alpha_1)}, \tag{27}$$

$$\begin{aligned} T_e = &\frac{\zeta(1+\alpha_1,\beta_1\tau_1) - \alpha_1\zeta(\alpha_1,\beta_1\tau_1)}{\beta_1\Gamma(\alpha_1)} + \frac{\zeta(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})} \\ &+ \frac{(\beta_{a_2}\tau_{s_2})\Gamma(\alpha_{a_2}-\alpha_{a_2}) - \beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})}. \end{aligned} \tag{28}$$

In order to provide formulas to evaluate the performance of four surgeries, again by aggregating surgery S_{a_2} and surgery S_3 into one, and assuming Gamma distribution, we obtain the parameters α_{a_3} and β_{a_3} for the aggregated surgery S_{a_3}.

Corollary 2 *Under assumptions (1)–(8) with* $N = 3$, *the aggregated parameters are:*

$$\begin{aligned} \alpha_{a_3} = &\big(\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})[\beta_{a_2}(\alpha_3+\beta_3\tau_{s_2})\Gamma(\alpha_{a_2}) + \beta_3(-\beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2}) \\ &+ \zeta(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2}))]^2\big) \Big/ \Big(\Gamma(\alpha_{a_2})[\alpha_3\beta_{a_2}^2\Gamma(\alpha_{a_2})\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2}) \\ &+ \beta_3^2(-\zeta^2(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2}) + \zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})\zeta(2+\alpha_{a_2},\beta_{a_2}\tau_{s_2}))]\Big), \end{aligned} \tag{29}$$

$$\begin{aligned} \beta_{a_3} = &\big(\beta_{a_2}\beta_3\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})[\beta_{a_2}(\alpha_3+\beta_3\tau_{s_2})\Gamma(\alpha_{a_2}) + \beta_3(-\beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2}) \\ &+ \zeta(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2}))]\big) \Big/ \Big(\alpha_3\beta_{a_2}^2\Gamma(\alpha_{a_2})\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2}) \\ &+ \beta_3^2[-\zeta^2(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2}) + \zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})\zeta(2+\alpha_{a_2},\beta_{a_2}\tau_{s_2})]\Big). \end{aligned} \tag{30}$$

Similar to the 2-surgery case, the monotonicity holds in 3-surgery scenario as well. As one can see, C_3 is obtained through τ_{s_2} and τ_3, so that C_3 is monotonically increasing with respect to τ_3 and τ_{s_2}. But τ_{s_2} is calculated by aggregation of τ_1 and τ_2. It can be proved that the monotonicity holds in the aggregation by the chain rule. Then monotonicity of C_3 with respect to $\tau_i, i = 1,2,3$, can be justified. Analogously, the monotonicity of C_3 with respect to $\sigma_i, i = 1,2,3$, still holds as well. Using similar arguments, it can be shown that V_3 is monotonically increasing with respect to τ_i and $\sigma_i, i = 1,2,3$.

4.3 *N* surgeries

Using the similar approach, we can extend the study to *N* surgeries. Specifically, the following iteration procedure is proposed:

- *Step 1*: Aggregate surgeries S_1 and S_2 into S_{a_2}. Calculate its mean surgery time C_2 and variance V_2, idle time T_{e_1} and waiting time T_{w_2}, and obtain parameters α_{a_2} and β_{a_2}.

- *Step 2*: Aggregate surgeries S_{a_2} and S_3 into S_{a_3}. Using parameters α_{a_2} and β_{a_2} to calculate C_3, V_3, T_{e_2}, and T_{w_3}. Obtain new parameters α_{a_3} and β_{a_3}.
- *Step 3*: Repeat this process until surgery N, i.e., aggregate surgery S_{a_i} (using parameters α_{a_i} and β_{a_i} from previous step) and surgery S_{i+1} into a new aggregated surgery $S_{a_{i+1}}$, with new parameters $\alpha_{a_{i+1}}$ and $\beta_{a_{i+1}}$, till $i = N - 1$. Calculate $C_{i+1}, V_{i+1}, T_{e_i}$ and $T_{w_{i+1}}$.
- *Step 4*: Calculate surgery completion time C_N and variance V_N, and the total idle time T_e and total waiting time T_w.

An illustration of such a procedure for a four-surgery schedule is shown in Fig. 3. A formal expression of such a procedure is shown in (31)–(37).

Procedure 1 *Under assumptions (1)–(8),*

$$T_{e_i} = \frac{(-\alpha_{a_i} + \beta_{a_i}\tau_{s_i})\Gamma(\alpha_{a_i}) - \beta_{a_i}\tau_{s_i}\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i}) + \zeta(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i})}{\beta_{a_i}\Gamma(\alpha_{a_i})}, \tag{31}$$

$$T_{w_{i+1}} = \frac{-\beta_{a_i}\tau_{s_i}\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i}) + \zeta(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i})}{\beta_{a_i}\Gamma(\alpha_{a_i})}, \tag{32}$$

$$C_{i+1} = \tau_{s_i} + \frac{\alpha_{i+1}}{\beta_{i+1}} + \frac{\zeta(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}) - \beta_{a_i}\tau_{s_i}\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})}{\beta_{a_i}\Gamma(\alpha_{a_i})}, \tag{33}$$

$$V_{i+1} = \frac{-\zeta^2(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}) + \zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})\zeta(2 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i})}{\beta_{a_i}^2\Gamma(\alpha_{a_i})\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})} + \frac{\alpha_{i+1}}{\beta_{i+1}^2}, \tag{34}$$

$$i = 1, 2, \ldots, N - 1,$$

where

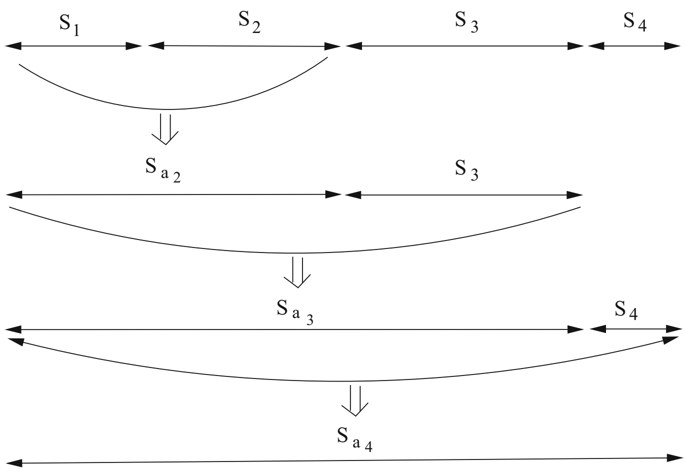

Fig. 3 Illustration of aggregation procedure

$$\tau_{s_i} = \sum_{j=1}^{i} \tau_i = \sum_{j=1}^{i} \frac{\alpha_j}{\beta_j}, \tag{35}$$

$$\begin{aligned}\alpha_{a_{i+1}} = &\Big(\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})[\beta_{a_i}(\alpha_{i+1} + \beta_{i+1}\tau_{s_i})\Gamma(\alpha_{a_i}) + \beta_{i+1}(-\beta_{a_i}\tau_{s_i}\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i}) \\ &+ \zeta(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}))]^2\Big) \Big/ \Big(\Gamma(\alpha_{a_i})[\alpha_{i+1}\beta_{a_i}^2\Gamma(\alpha_{a_i})\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i}) \\ &+ \beta_{i+1}^2(-\zeta^2(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}) + \zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})\zeta(2 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}))]\Big),\end{aligned} \tag{36}$$

$$\begin{aligned}\beta_{a_{i+1}} = &\Big(\beta_{a_i}\beta_{i+1}\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})[\beta_{a_i}(\alpha_{i+1} + \beta_{i+1}\tau_{s_i})\Gamma(\alpha_{a_i}) \\ &+ \beta_{i+1}(-\beta_{a_i}\tau_{s_i}\zeta(\alpha_{a_i}, \beta_{i+1}\tau_{s_i}) + \zeta(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}))]\Big) \Big/ \\ &\Big(\alpha_{i+1}\beta_{a_i}^2\Gamma(\alpha_{a_i})\zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i}) + \beta_{i+1}^2[-\zeta^2(1 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i}) \\ &+ \zeta(\alpha_{a_i}, \beta_{a_i}\tau_{s_i})\zeta(2 + \alpha_{a_i}, \beta_{a_i}\tau_{s_i})]\Big).\end{aligned} \tag{37}$$

Finally, using the similar arguments in 3-surgery case, the monotonicity of C_N and V_N with respect to τ_i and $\sigma_i, i = 1, \ldots, N$, can be justified.

5 Model validation

To validate the model, extensive simulation experiments have been carried out using a commercial software *SIMUL8* (Hauge and Paige 2002). In all simulations, each experiment simulates schedules for 60 days with 1000 replications. The system parameters are randomly selected from:

$$\begin{aligned} N &\in \{2, 3, 4, 5\}, \\ \tau_i &\in \{40, 80, 120\}, \quad i = 1, \ldots, N, \\ cv_i &\in \{0.2, 0.4, 0.6, 0.8\}, \quad i = 1, \ldots, N, \end{aligned}$$

where cv_i denotes the coefficient of variation (CV) of surgery time for S_i. In all experiments, the 95 % confidence intervals are within 1.5 % of the performance measures.

5.1 Accuracy of aggregation procedure

To validate the accuracy of aggregation procedure, first we assume that all surgical times are characterized by Gamma distributions. The analytical formulas approximate the aggregated surgeries also by Gamma distributions. To validate the effectiveness of such an approximation, the results obtained in Sect. 4 are compared with simulation results where in each experiment all surgeries follow Gamma distributions. If the differences are small, then it verifies that the aggregation approximation has sufficient accuracy.

Let T_w^{sim}, T_e^{sim}, and T_w^{mod} and T_e^{mod} denote the average waiting time and idle time obtained by simulation and analytical model, respectively. Introduce

$$\delta_w = \frac{|T_w^{sim} - T_w^{mod}|}{T_w^{sim}} \times 100\,\%,$$

$$\delta_e = \frac{|T_e^{sim} - T_e^{mod}|}{T_e^{sim}} \times 100\,\%.$$

Based on five dozen experiments, the average δ_w is 1.32 %, with the maximal one within 5 %. The average δ_e is 1.94 %, while the maximum is less than 7 %. An illustration example is shown in Fig. 4, where four surgeries are scheduled. The mean duration of each surgery is selected from the set of {40, 80, 120} min. The first surgery has $cv_1 = 0.6$, while the second surgery has $cv_2 = 0.4$. In Fig. 4, δ_w and δ_e are plotted as functions of cv_3. As one can see, As one can see, the δ_ws are consistently around 1–2 %, and δ_es vary around 2–4 %.

Similar accuracy is observed for completion time and its variance. These results suggest that the aggregation approach introduced here has sufficient accuracy to estimate the performance of surgery schedules.

5.2 Accuracy of analytical model

Next, we validate the effectiveness of analytical model by using the data collected on the orthopedic surgery operating rooms at UW Health. This will also justify the assumption that Gamma distribution can be used to approximate the surgical time. The data is collected from the events in 2012. Ten surgical types from more than 5200 events are included in the study and summarized in Table 1.

As one can see, the surgery time varies substantially with surgical types, and the variation of surgical time in each type also varies with CV between 0.32 and 0.78.

Fig. 4 A four-surgery example

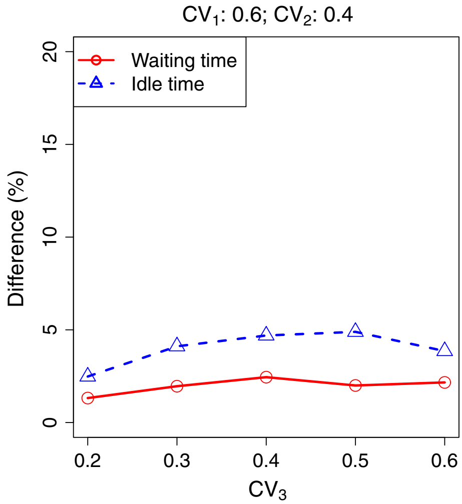

Table 1 10 surgical types

Type	Number	Mean	STD	CV
Hand upper extremity	738	70.65	50.69	0.7175
Spine	531	171.2	92.51	0.5383
Joint	944	149.44	47.97	0.3210
Sports medicine	1202	107.76	59.47	0.5519
Foot and ankle	73	165.99	60.67	0.3655
General/plastics	257	138.14	107.23	0.7762
General/bariatric	946	123.35	57.03	0.4623
General/hernia	122	155.65	118.05	0.7584
Urology	218	106.85	74.14	0.6939
Spine neuro	238	186.45	122.24	0.6556

For each surgery type, the collected data will be fitted into a distribution using the Stat Fit package in *SIMUL8*. The resulting distribution types include: Gamma, Weibull, Log-normal, Triangular, Beta, Pearson V, and Pearson VI. In each experiment, we randomly select number of surgeries, and randomly select the surgical type from Table 1. The fitted distribution for the selected surgery type will be used in simulation model. The mean and standard deviation are provided to the analytical model. Then δ_w and δ_e are evaluated using the same approach as in Sect. 5.1.

Based on five dozens experiments, we observe that the average difference in waiting time is 3.40 %, and all differences are within 6 %. The average difference in idle time is 5.34 %, and all are within 9 %. Such results indicate that the developed model has acceptable accuracy in estimating patient waiting time and room idle time for a given surgery schedule. In addition, since only Gama distribution is assumed in the analytical model, while more than half dozen different distribution types have been used in fitting the simulation model, we hypothesize that when the CV of surgery time is small (i.e., less than 1), the system performance is practically independent of the distribution type, but mainly depends on the mean and CV. In other words, the analytical model is suitable for a general distribution of surgery times. Similar properties have been observed in healthcare clinics (Reynolds et al. 2010; Wang et al. 2014; Zhong et al. 2016a, b), as well as in manufacturing systems (Li and Meerkov 2009).

To further investigate the accuracy with respect to number of surgeries and CVs of surgeries, we carry out extensive simulation experiments, each with 1000 replications for 60 days. We observe that the error will accumulate as the number of surgeries increases. However, such an increase in error is relatively slow. When the number of surgeries is increased from 2 to 5, the errors in waiting time and idle time are increased from 2.1 to 2.6 and 2 to 3 %, respectively. For the impact of CVs, similar experiments are carried out, and we do not find significant differences. The errors in waiting time and idle time are always between 1.5 to 2.7, and 1 to 2 %, respectively, for CVs from 0.3 to 0.6.

In summary, the presented method provides an effective tool to evaluate the performance of OR schedules in orthopedic surgery. As the calculation can be carried out in a fraction of a second, it provides an effective alternative of simulations and can help reduce the computation intensity substantially in OR scheduling optimization.

6 Software illustration

Using the above results, a Java-based computer program has been developed. The software enables the user to create any surgery types, input their mean and coefficients of variation based on historical data. Then the user can select the surgeries to be scheduled, and submit for evaluation.

For example, as shown in Fig. 5, the user selects six surgeries, `HAND_UPPER_IN`, `HAND_UPPER_OUT`, `JOINT_IN`, `JOINT_OUT`, `FOOT_ANKLE_IN`, and `FOOT_ANKLE_OUT`, to schedule in one operating room.

The software can quickly evaluate different combinations and provide three best scenarios in patient waiting time and room idle time for the scheduler to select. Since typically only a limited number of surgeries will be scheduled in one room per day, finding an optimal sequence can be easily achieved. Figure 6 shows five of

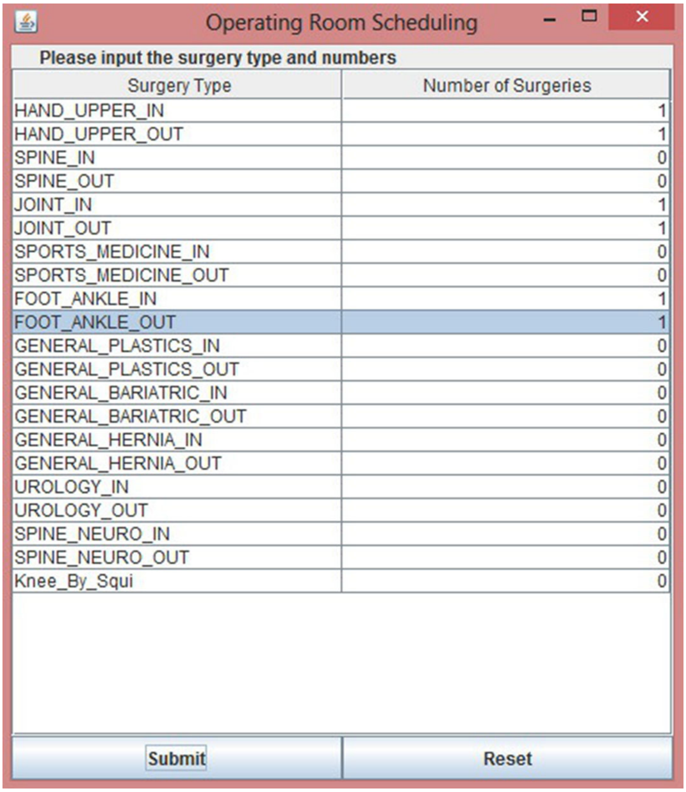

Fig. 5 Program illustration: selection

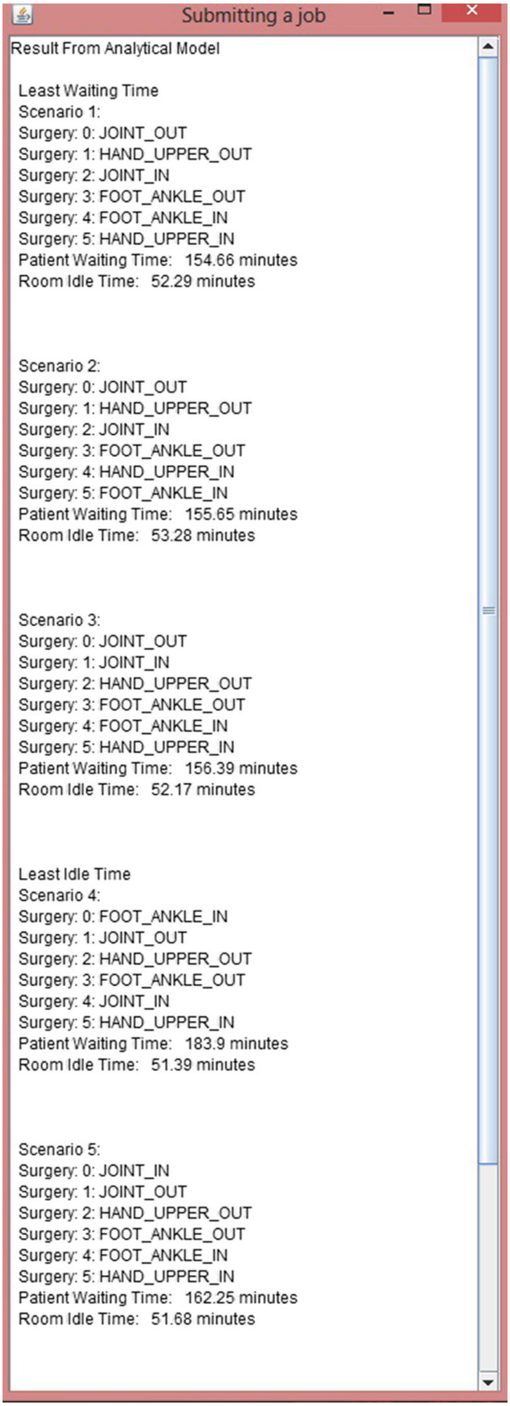

Fig. 6 Program illustration: results

them in terms of shortest patient waiting time or room idle time. The reason to provide multiple results rather than an optimal one is that the scheduler may have other constraints to consider so that he/she can pick up the one satisfying his/her specific concerns with reasonable good performance.

As one can see, by considering shortest patient waiting time, the schedule of surgeries JOINT_OUT, HAND_UPPER_OUT, JOINT_IN, FOOT_ANKLE_OUT, FOOT_ANKLE_IN, and HAND_UPPER_IN results in the shortest average patient waiting time 154.66 min, and the average room idle time is 52.29 min. The next schedule has slightly longer waiting time and idle time (155.65 and 53.28 min) by switching FOOT_ANKLE_IN and HAND_UPPER_IN. The third schedule switches HAND_UPPEER_OUT and JOINT_IN from the first one and obtains more waiting time (156.39 min) but less idle time (52.17 min). If shortest room idle time is the objective, then the schedule of FOOT_ANKLE_IN, JOINT_OUT, HAND_UPPER_OUT, FOOT_ANKLE_OUT, JOINT_IN, and HAND_UPPER_IN is the best schedule, with 51.39 min idle time. But its waiting time is extended to 183.9 min. The second choice is JOINT_IN, JOINT_OUT, HAND_UPPER_OUT, FOOT_ANKLE_OUT, FOOT_ANKLE_IN, and HAND_UPPER_IN. It only has minimum longer waiting time (51.68 min), but much shorter waiting time (162.25 min). Thus, the scheduler may select this one.

In addition, the software provides the freedom for the scheduler to construct his/her own schedule and then compare the performance with the suggested ones. As the calculation can be carried out in a fraction of a second, it provides an effective alternative of simulations and can help reduce the computation intensity substantially. Such a program presents a quantitative tool for the scheduler to sequence orthopedic surgeries. Currently the software is used in a pilot study at UW Health.

7 Conclusions

This paper introduces an analytical model to evaluate the performance of OR schedules in orthopedic surgery. Since the variability is small in orthopedic surgery, and the schedule performance is practically independent of distribution type, these enable us to use Gamma distribution to approximate surgical time. Then the room idle time and patient waiting time can be calculated for two-surgery schedule. Using an aggregation approach, every two surgeries can be represented by an aggregated one. Through an iterative procedure, multiple surgery schedules can be evaluated and the idle time and waiting time can be estimated. In addition, the completion time and its variance can be obtained as well. Using the data collected on the hospital floor in UW Health, numerical experiments have shown that such a method results in acceptable accuracy and can be effectively used for performance evaluation of surgery schedules. A preliminary version of a Java software suite to implement the algorithms has been developed and deployed in a pilot study at UW Health. Such a model provides the healthcare professionals a quantitative tool for evaluation of OR schedules.

In future work, we expect to extend the work in the following directions:

- extending the model to include surgery cancellation and delay;
- considering preoperative procedures, i.e., activities before surgery, in the model;
- generalizing the study to other types of surgeries, particularly, the surgeries that have larger variations;
- developing methods for surgery scheduling optimization, especially with multiple surgical rooms;
- including constraints in surgery schedules, e.g., a specific surgery needs to be sequenced at a given time interval or follow a given order;
- investigating the fundamental properties in surgery scheduling, derive insights and principles, and finally,
- applying the model in daily scheduling activities on the hospital floor.

Acknowledgments The authors thank C. Zhao and X. Zhong at ISyE Department of UW Madison for their helpful comments and suggestions, and the OR staff at UW Health for their valuable support on this project. This work is supported in part by National Science Foundation Grants Nos. CMMI-1233807, CMMI-1536987 and a Grant from UW Health.

Appendix: Proofs

Proof of Proposition 1 From Lemmas 1 and 2, by conditioning both scenarios where the first surgery S_1 finishes earlier or later than the scheduled time, the average completion time and the variance of two surgeries are calculated as:

$$\begin{aligned} C_2 &= C_{e_2}p_{e_1} + C_{w_2}p_{w_1} = \frac{\zeta(1+\alpha_1, \beta_1\tau_1) - \beta_1\tau_1\zeta(\alpha_1, \beta_1\tau_1)}{\beta_1\Gamma(\alpha_1)} + \frac{\alpha_1}{\beta_1} + \frac{\alpha_2}{\beta_2}, \\ &= \frac{\alpha_1}{\beta_1} + \frac{\alpha_2}{\beta_2} + \frac{\zeta(1+\alpha_1, \alpha_1) - \alpha_1\zeta(\alpha_1, \alpha_1)}{\beta_1\Gamma(\alpha_1)}, \\ V_2 &= V_{e_2}p_{e_1} + V_{w_2}p_{w_1} = \frac{-\zeta^2(1+\alpha_1, \beta_1\tau_1) + \zeta(\alpha_1, \beta_1\tau_1)\zeta(2+\alpha_1, \beta_1\tau_1)}{\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1, \beta_1\tau_1)} \\ &= \frac{-\zeta^2(1+\alpha_1, \alpha_1) + \zeta(\alpha_1, \alpha_1)\zeta(2+\alpha_1, \alpha_1)}{\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1, \alpha_1)} + \frac{\alpha_2}{\beta_2^2}. \end{aligned}$$

The average room idle time can be derived as follows:

$$
\begin{aligned}
T_{e_1} &= p_{e_1}\int_0^{\tau_1}(\tau_1 - x)g(x|x<\tau_1)dx = \int_0^{\tau_1}(\tau_1 - x)g(x)dx \\
&= \tau_1 - \frac{\alpha_1}{\beta_1} - \frac{\tau_1\zeta(\alpha_1,\beta_1\tau_1)}{\Gamma(\alpha_1)} + \frac{\zeta(1+\alpha_1,\beta_1\tau_1)}{\beta_1\Gamma(\alpha_1)} \\
&= \frac{-\alpha_1\zeta(\alpha_1,\alpha_1) + \zeta(1+\alpha_1,\alpha_1)}{\beta_1\Gamma(\alpha_1)} \\
&= \frac{-\frac{\tau_1^2}{\sigma_1^2}\zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right) + \zeta\left(1+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1^2}\right)}{\frac{\tau_1}{\sigma_1}\Gamma\left(\frac{\tau_1^2}{\sigma_1^2}\right)}.
\end{aligned}
$$

Similarly, the average patient waiting time can also be calculated:

$$
\begin{aligned}
T_{w_2} &= p_{w_1}\int_{\tau_1}^{\infty}(x-\tau_1)g(x|x>\tau_1)dx = \int_{\tau_1}^{\infty}(x-\tau_1)g(x)dx \\
&= \frac{-\beta_1\tau_1\zeta(\alpha_1,\beta_1\tau_1) + \zeta(1+\alpha_1,\beta_1\tau_1)}{\beta_1\Gamma(\alpha_1)} = \frac{-\alpha_1\zeta(\alpha_1,\alpha_1) + \zeta(1+\alpha_1,\alpha_1)}{\beta_1\Gamma(\alpha_1)} \\
&= \frac{-\frac{\tau_1^2}{\sigma_1^2}\zeta\left(\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1}\right) + \zeta\left(1+\frac{\tau_1^2}{\sigma_1^2},\frac{\tau_1^2}{\sigma_1}\right)}{\frac{\tau_1}{\sigma_1}\Gamma\left(\frac{\tau_1^2}{\sigma_1^2}\right)}.
\end{aligned}
$$

□

Proof of Lemma 3 From (2), the aggregated parameter α_{a_2} should be determined by the ratio of mean square and variance of the first two surgeries. Thus,

$$
\begin{aligned}
\alpha_{a_2} = \frac{C_2^2}{V_2} &= \Big[\zeta(\alpha_1,\beta_1\tau_1)[\beta_1(\alpha_2+\beta_2\tau_1)\Gamma(\alpha_1) + \beta_2(-\beta_1\tau_1\zeta(\alpha_1,\beta_1\tau_1) \\
&\quad + \zeta(1+\alpha_1,\beta_1\tau_1))]^2\Big] \Big/ \Big[\Gamma(\alpha_1)[\alpha_2\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1,\beta_1\tau_1) + \beta_2^2(-\zeta^2(1+\alpha_1,\beta_1\tau_1) \\
&\quad + \zeta(\alpha_1,\beta_1\tau_1)\zeta(2+\alpha_1,\beta_1\tau_1))]\Big] \\
&= \frac{\zeta(\alpha_1,\alpha_1)[\beta_1(\alpha_2+\beta_2\tau_1)\Gamma(\alpha_1) + \beta_2(-\alpha_1\zeta(\alpha_1,\alpha_1) + \zeta(1+\alpha_1,\alpha_1))]^2}{\Gamma(\alpha_1)[\alpha_2\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1,\alpha_1) + \beta_2^2(-\zeta^2(1+\alpha_1,\alpha_1) + \zeta(\alpha_1,\alpha_1)\zeta(2+\alpha_1,\alpha_1))]}.
\end{aligned}
$$

Again from (2), the aggregated parameter β_{a_2} also depends on the ratio between the mean and then variance. Therefore, we obtain

$$\beta_{a_2} = \frac{C_2}{V_2} = \Big[\beta_1\beta_2\zeta(\alpha_1,\beta_1\tau_1)[\beta_1(\alpha_2+\beta_2\tau_1)\Gamma(\alpha_1)+\beta_2(-\beta_1\tau_1\zeta(\alpha_1,\beta_1\tau_1) + \zeta(1+\alpha_1,\beta_1\tau_1))]\Big] \Big/ \Big[\alpha_2\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1,\beta_1\tau_1)+\beta_2^2[-\zeta^2(1+\alpha_1,\beta_1\tau_1) + \zeta(\alpha_1,\beta_1\tau_1)\zeta(2+\alpha_1,\beta_1\tau_1)]\Big]$$
$$= \frac{\beta_1\beta_2\zeta(\alpha_1,\alpha_1)[\beta_1(\alpha_2+\beta_2\tau_1)\Gamma(\alpha_1)+\beta_2(-\alpha_1\zeta(\alpha_1,\alpha_1)+\zeta(1+\alpha_1,\alpha_1))]}{\alpha_2\beta_1^2\Gamma(\alpha_1)\zeta(\alpha_1,\alpha_1)+\beta_2^2[-\zeta^2(1+\alpha_1,\alpha_1)+\zeta(\alpha_1,\alpha_1)\zeta(2+\alpha_1,\alpha_1)]}.$$

□

Proof of Proposition 2 The scheduled finishing time of surgeries S_1 and S_2 will be $\tau_1+\tau_2$, i.e., the scheduled finishing time of the aggregated surgery is

$$\tau_{s_2} = \tau_1+\tau_2 = \frac{\alpha_1}{\beta_1}+\frac{\alpha_2}{\beta_2}.$$

Then, α_{a_2} and β_{a_2}, the parameters of the aggregated surgery S_{a_2}, can be obtained from Lemma 3. Using Proposition 1, by replacing τ_1 with τ_{s_2}, α_1 with α_{a_2}, and β_1 with β_{a_2}, and replacing α_2 with α_3, β_2 with β_3, the completion time C_2 and the variance V_3 can be obtained. □

Proof of Lemma 4 By replacing τ_1 with τ_{s_2}, α_1 with α_{a_2}, β_1 with β_{a_2}, the room idle time of surgery S_2 and the patient waiting time for surgery S_3 can be calculated. First, calculate the probability that the aggregated surgery S_{a_2} has probabilities $p_{e_{a_2}}$ and $p_{w_{a_2}}$ to be earlier or later than the aggregated scheduled time τ_{s_2}, respectively.

$$p_{e_{a_2}} = \int_0^{\tau_{s_2}} g(x)dx, \quad p_{w_{a_2}} = \int_{\tau_{s_2}}^{\infty} g(x)dx.$$

Then, using α_{a_2} and β_{a_2}, we obtain the average idle time

$$T_{e_2} = p_{e_{a_2}}\int_0^{\tau_{s_2}}(\tau_{s_2}-x)g(x|x<\tau_{s_2})dx = \int_0^{\tau_{s_2}}(\tau_{s_2}-x)g(x)dx$$
$$= \int_0^{\tau_{s_2}}(\tau_{s_2}-x)\frac{\beta_{a_2}^{\alpha_{a_2}}}{\Gamma(\alpha_{a_2})}x^{\alpha_{a_2}-1}e^{-\beta_{a_2}x}dx$$
$$= \Big[-\beta_{a_2}^{-1+\alpha_{a_2}}\tau_{a_2}^{\alpha_{a_2}}(\beta_{a_2}\tau_{a_2})^{-\alpha_{a_2}}\cdot[(-\alpha_{a_2}+\beta_{a_2}\tau_{a_2})\Gamma(\alpha_{a_2}) - \beta_{a_2}\tau_{a_2}\zeta(\alpha_{a_2},\beta_{a_2}\tau_{a_2})+\zeta(1+\alpha_{a_2},\beta_{a_2}\tau_{a_2})]\Big]\Big/\Gamma(\alpha_{a_2})$$
$$= \frac{(-\alpha_{a_2}+\beta_{a_2}\tau_{s_2})\Gamma(\alpha_{a_2})-\beta_{a_2}\tau_{s_2}\zeta(\alpha_{a_2},\beta_{a_2}\tau_{s_2})+\zeta(1+\alpha_{a_2},\beta_{a_2}\tau_{s_2})}{\beta_{a_2}\Gamma(\alpha_{a_2})},$$

and the average waiting time

$$
\begin{aligned}
T_{w_3} &= p_{w_{a_2}} \int_{\tau_{s_2}}^{\infty} (x-\tau_{s_2}) g(x|x>\tau_{s_2}) dx \\
&= \int_{\tau_{s_2}}^{\infty} (x-\tau_{s_2}) g(x) dx = \int_{\tau_{s_2}}^{\infty} (x-\tau_{s_2}) \frac{\beta_{a_2}^{\alpha_{a_2}}}{\Gamma(\alpha_{a_2})} x^{\alpha_{a_2}-1} e^{-\beta_{a_2} x} dx \\
&= \Big[(\beta_{a_2}\tau_{a_2})^{-\alpha_{a_2}} [(\beta_{a_2}\tau_{a_2})^{\alpha_{a_2}} (\alpha_{a_2} - \beta_{a_2}\tau_{a_2}) \Gamma(\alpha_{a_2}) \\
&\quad + \beta_{a_2}^{\alpha_{a_2}} \tau_{a_2}^{\alpha_{a_2}} ((-\alpha_{a_2} + \beta_{a_2}) \Gamma(\alpha_{a_2}))] \Big] \Big/ [\beta_{a_2} \Gamma(\alpha_{a_2})] \\
&\quad + \Big[(\beta_{a_2}\tau_{a_2})^{-\alpha_{a_2}} [\zeta(1+\alpha_{a_2}, \beta_{a_2}\tau_{a_2}) - \beta_{a_2}\tau_{a_2} \zeta(\alpha_{a_2}, \beta_{a_2}\tau_{a_2})] \Big] \Big/ [\beta_{a_2} \Gamma(\alpha_{a_2})] \\
&= \frac{-\beta_{a_2}\tau_{s_2} \zeta(\alpha_{a_2}, \beta_{a_2}\tau_{s_2}) + \zeta(1+\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2} \Gamma(\alpha_{a_2})} \\
&= \frac{-\tau_{s_2} \zeta(\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\Gamma(\alpha_{a_2})} + \frac{\zeta(1+\alpha_{a_2}, \beta_{a_2}\tau_{s_2})}{\beta_{a_2} \Gamma(\alpha_{a_2})}.
\end{aligned}
$$

□

References

Blake JT, Carter MW (1997) Surgical process scheduling: a structured review. J Soc Health Syst 5(3):17–30

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

Cardoen B, Demeulemeester E, Belien J (2010) Operating room planning and scheduling: a literature review. Eur J Oper Res 201(16):921–932

Cayirli T, Veral E, Rosen H (2006) Designing appointment scheduling systems for ambulatory care services. Health Care Manag Sci 9(1):47–58

Chakraborty S, Muthuraman K, Lawley M (2010) Sequential clinical scheduling with patient no-shows and general service time distributions. IIE Trans 42(5):354–366

Coit D, Jin T (2000) Gamma distribution parameter estimation for field reliability data with missing failure times. IIE Trans 32(12):1161–1166

Denton B, Viapiano J, Vogl A (2007) Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manag Sci 10(1):13–24

Dexter F, Epstein RH, Traub RD, Xiao Y (2004) Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. Anesthesiology 101(6):1444–1453

Dexter F, Traub RD (2002) How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesth Analg 94(4):933–942

Erdogan SA, Denton BT (2010) Surgery planning and scheduling. In: Cochran JJ (ed) Wiley encyclopedia of operations research and management science. Wiley, Hoboken

Fei H, Chu C, Meskens N (2009) Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. Ann Oper Res 166(1):91–108

Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: a survey. Health Care Manag Sci 14(1):89–114

Guinet A, Chaabane SS (2003) Operating theatre planning. Int J Prod Econ 85(1):69–81

Gupta D, Denton B (2008) Appointment scheduling in health care: challenges and opportunities. IIE Trans 40(9):800–819

Hauge JW, Paige KN (2002) Learning SIMUL8: the complete guide. Plain Vu Publishers, Bethlingham

Health Care Financial Management Association (2003) Achieving operating room efficiency through process integration. Healthc Finance Manag 57(3):S1–S7

Jebali A, Alouaneb ABH, Ladeta P (2006) Operating rooms scheduling. Int J Prod Econ 99(1–2):52–62

Kuo P, Schroeder R, Mahaffey S, Bollinger R (2003) Optimization of operating room allocation using linear programming techniques. J Am Coll Surg 197(6):889–895
Lamiri M, Xie X, Dolgui A, Grimaud F (2008) A stochastic model for operating room planning with elective and emergency demand for surgery. Eur J Oper Res 185(3):1026–1037
Li J, Meerkov SM (2009) Production systems engineering. Springer, New York
Liu N, Ziya S, Kulkarni VG (2010) Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. Manuf Serv Oper Manag 12(2):347–364
Mancilla C, Storer R (2012) A sample average approximation approach to stochastic appointment sequencing and scheduling. IIE Trans 44(8):655–670
May JH, Spangler WE, Strum DP, Vargas LG (2011) The surgical scheduling problem: current research and future opportunities. Prod Oper Manag 20(3):392–405
Min D, Yih Y (2010) An elective surgery scheduling problem considering patient priority. Comput Oper Res 37(6):1091–1099
Muthuraman K, Lawley M (2008) A stochastic overbooking model for outpatient clinical scheduling with no-shows. IIE Trans 40(9):820–837
Pham DN, Klinkert A (2008) Surgical case scheduling as a generalized job shop scheduling problem. Eur J Oper Res 185(3):1011–1025
Reynolds J, Zeng Z, Li J, Chiang S-Y (2010) Design and analysis of a health care clinic for homeless people using simulations. Int J Health Care Qual Assur 23(6):607–620
Robinson LW, Chen RR (2003) Scheduling doctors' appointments: optimal and empirically-based heuristic policies. IIE Trans 35(3):295–307
Van Oostrum JM, Van Houdenhoven M, Hurink JL, Hans EW, Wullink G, Kazemier G (2008) A master surgical scheduling approach for cyclic scheduling in operating room departments. OR Spectr 30(2):355–3764
Wang J, Zhong X, Li J, Howard PK (2014) Modeling and analysis of care delivery services within patient rooms: a system-theoretic approach. IEEE Trans Autom Sci Eng 11(2):379–393
Wright PD, Bretthauer KM, Cote MJ (2006) Reexamining the nurse scheduling problem: staffing ratios and nursing shortages. Decis Sci 37(1):39–70
Wullink G, Houdenhoven MV, Hans EW, Oostrum JM, Lans MVD, Kazemier G (2007) Closing emergency operating rooms improves efficiency. J Med Syst 31(6):543–546
Zacharias C, Pinedo M (2013) Appointment scheduling with no-shows and overbooking. Prod Oper Manag 23(5):788–801
Zhang Z, Xie X, Geng N (2014) Dynamic surgery assignment of multiple operating rooms with planned surgeon arrival times. IEEE Trans Autom Sci Eng 11(3):680–691
Zhang B, Murali P, Dessouky MM, Belson D (2009) A mixed integer programming approach for allocating operating room capacity. J Oper Res Soc 60(5):663–673
Zhong X, Li J, Ertl SM, Hassmer C, Fielder L (2016) A systems theoretic approach for modeling and analysis of mammography testing process. IEEE Trans Syst Man Cybern Syst 46(1):126–138
Zhong X, Song J, Li J, Ertl SM, Fielder L (2016) Analysis and design of gastroenterology (GI) clinic in Digestive Health Center: a systems approach. Flex Serv Manuf J 28(1):90–119

Zexian Zeng received both his master degrees in Industrial and Systems Engineering and Computer Science from University of Wisconsin-Madison in 2014. Currently, he is working towards the Ph.D. degree at the Department of Preventive Medicine, Northwestern University Feinberg School of Medicine. His research interests are in healthcare computation, healthcare system efficiency and safety.

Xiaolei Xie is an assistant professor with the Department of Industrial Engineering at Tsinghua University. He obtained his Ph.D. from Department of Industrial and Systems Engineering at University of Wisconsin, Madison, in 2014. His research interests are healthcare analyticals and productions systems engineering. He is a member of the Institute for Operations Research and the Management Sciences (INFORMS) and the Institute of Electrical and Electronics Engineers (IEEE).

Heidi Menaker is a Senior Health Systems Engineer in the Quality, Safety, and Innovation team at the University of Wisconsin (UW) Health. In her 5 years at UW Health, Heidi has facilitated teams in improving and designing processes using Lean and other improvement methodologies. The last 3 years have been dedicated to applying Lean concepts to the design of the physical building as well as the internal processes of UW Health's newest facility, UW Health at The American Center. Prior to UW Health, Heidi taught and facilitated Lean teams at Aurora Health Care. Heidi holds a bachelor's and master's degree in Industrial and Systems Engineering from University of Wisconsin-Madison.

Jingshan Li received the B.S. degree from Department of Automation, Tsinghua University, Beijing, China, the M.S. degree from Institute of Automation, Chinese Academy of Sciences, Beijing, and the Ph.D. degree in electrical engineering-systems from University of Michigan, Ann Arbor, in 1989, 1992, and 2000, respectively. He was a Staff Research Engineer at the Manufacturing Systems Research Laboratory, General Motors Research and Development Center, Warren, MI from 2000 to 2006, and was with Department of Electrical and Computer Engineering and Center for Manufacturing, University of Kentucky, Lexington, KY from 2006 to 2010. He is now a Professor in Department of Industrial and Systems Engineering, University of Wisconsin, Madison, WI. His primary research interests are in modeling, analysis and control of manufacturing and healthcare systems. He received 2010 NSF Career Award, 2009 IIE Transactions Best Application Paper Award, 2005 IEEE Transactions on Automation Science and Engineering Best Paper Award, 2006 IEEE Early Industry/ Government Career Award in Robotics and Automation, and multiple awards in international conferences. He is a Senior Editor of IEEE Robotics and Automation Letters, Department Editor of IIE Transactions and Associate Editor of IEEE Transactions on Automation Science and Engineering, International Journal of Production Research, Flexible Service and Manufacturing, and International Journal of Automation Technology, and was an Associate Editor of Mathematical Problems in Engineering.