DECISION SCIENCES
A JOURNAL OF THE DECISION SCIENCES INSTITUTE

Check for updates

Decision Sciences
Volume 45 Number 5
October 2014

Partially Flexible Operating Rooms for Elective and Emergency Surgeries*

Yann B. Ferrand†
Department of Management, Clemson University, 131A Sirrine Hall, Clemson, SC 29634, e-mail: yferran@clemson.edu

Michael J. Magazine
Department of Operations, Business Analytics, and Information Systems, University of Cincinnati, 522 Carl H. Lindner Hall, 2925 Campus Green Drive, Cincinnati, OH 45221, e-mail: magazim@ucmail.uc.edu

Uday S. Rao
Department of Operations, Business Analytics, and Information Systems, University of Cincinnati, 528 Carl H. Lindner Hall, 2925 Campus Green Drive, Cincinnati, OH 45221, e-mail: raous@ucmail.uc.edu

ABSTRACT

In hospitals, the management of operating rooms faces a trade-off between the need to be responsive to emergency surgeries and to conduct scheduled elective surgeries efficiently. Operating rooms can be configured as flexible and handle both electives and emergencies, or as dedicated to focus on either electives or emergencies. With flexible rooms, the prioritization of emergencies over scheduled electives can lead to schedule disruptions. Focused rooms can lead to imbalances between capacity and surgery workload. Whereas hospital administrators typically handle this trade-off by employing either flexible rooms (complete flexibility) or dedicated rooms (complete focus), we investigate whether a combination of flexible and dedicated rooms (partial flexibility) could be a preferable alternative. The ensuing question is what is the right combination of flexible and dedicated rooms? A versatile simulation model is developed to evaluate different resource allocation policies under various environmental parameters and performance metrics, including patient wait time, staff overtime, and operating room utilization. The main result is that partial flexibility configurations outperform both complete flexibility and complete focus policies by providing solutions with improved values of expected wait time for both emergency and elective patients. [Submitted: July 1, 2012. Revised: January 23, 2014. Accepted: January 27, 2014.]

Subject Areas: Healthcare and Capacity Management.

*We gratefully acknowledge the feedback provided by the associate editor and two anonymous reviewers. Their comments significantly improved the quality of this paper.

†Corresponding author.

INTRODUCTION

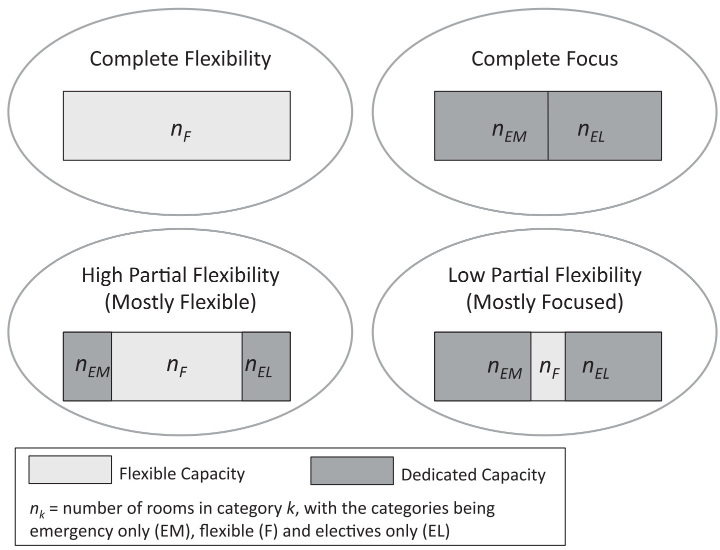
The variety and cost of the resources involved in delivering health care services pressure hospital management to be as efficient as possible, while being responsive to patient needs. Operating rooms, also commonly referred to as surgical suites, represent the largest cost center and source of revenue for hospitals (Macario, Vitez, Dunn, & McDonald, 1995) and as such, draw a lot of managerial attention. Several research papers in operations planning and scheduling have contributed to improvements in the management of surgical resources. Broad reviews on this topic have been conducted by Magerlein and Martin (1978), Smith-Daniels, Schweikhart, and Smith-Daniels (1988), Blake and Carter (1997), Cardoen, Demeulemeester, and Beliën (2010), and Guerriero and Guido (2011).

Conflicting stakeholders' objectives, numerous patient types and needs, along with multiple sources of uncertainty complicate the management of operating rooms. One widespread difficulty is the necessary handling of emergency patients who arrive randomly and need timely access to an operating room for surgery. The nature of emergencies requires that the hospital be responsive and provide surgical resources as fast as possible, which can perturb the schedule and flow of elective patients scheduled for surgery if both emergencies and elective patients have access to the same operating rooms.

To overcome the difficulties created by the conflicting needs of elective and emergency surgeries, hospital administrators have considered separating the emergency services by dedicating a set of surgical suites for emergency cases, which aim at ensuring responsiveness to emergencies, while allowing the rest of the hospital's surgical resources to concentrate on being efficient in handling scheduled electives. Haraden et al. (2003), and Ryckman et al. (2009) have recommended dedicating operating rooms to emergency cases to avoid disruptions of elective cases by emergencies, and to reduce system variability and smooth patient flow. This idea has had some success in practice (e.g., Clark, 2009; Ritchie, 2011). Ferrand, Magazine, and Rao (2010) investigated the effects of such a change, from having all the operating rooms configured as flexible (complete flexibility), handling both electives and emergencies, to having all the operating rooms dedicated (complete focus)—with some focused on electives and the remainder focused on emergencies. The results support the notion of a trade-off between efficiency toward electives and responsiveness to emergencies: shifting from complete flexibility to complete focus, with the appropriate allocation of rooms to electives and emergencies, results in a decrease of both average elective patients wait time for service and average staff overtime, while emergency patients average wait time for service increases.

This paper investigates whether a partial flexibility policy could improve a hospital's responsiveness and efficiency when compared with two extremes: complete flexibility and complete focus. In a partial flexibility policy part of the resources are flexible and can accommodate both electives and emergencies, while the rest of the resources are split between those dedicated to electives and those dedicated to emergencies. We compare two configurations of partial flexibility: (i) high partial flexibility with mostly flexible rooms and a few dedicated rooms and (ii) low partial flexibility with mostly dedicated rooms and a few flexible

Figure 1: Overview of policy choices.



rooms. These two partial flexibility policies represent a shift from either complete flexibility or complete focus, and one may be better suited than the other, so we also compare them to each other. We give an outline of these configurations in Figure 1 below.

Based on our simulation study, we formulate two observations on the benefits of partial flexibility in the presence of two priority classes and establish guidelines on how to implement partial flexibility. Our intuition is that there is a trade-off between disruptions of flexible servers and potential capacity imbalances of dedicated servers. We provide insights on how a hospital could use partial flexibility to balance responsiveness and efficiency. Specifically, if a few rooms are dedicated, we identify whether they should be dedicated strictly to emergencies or to electives. If a few rooms are flexible, we determine whether this should mean fewer rooms dedicated to emergencies or fewer rooms dedicated to electives. Since the best way to use resources might depend on the hospital's capacity, the arrival rate of emergencies, and the duration and variability of the surgeries to perform, we use computational experiments to characterize how these input parameters affect three performance measures: patient wait time, staff overtime, and room utilization.

LITERATURE REVIEW

Background: Focus and Flexibility

Since the objectives of responsiveness to emergency surgeries and efficiency in handling elective surgeries are different and conflicting, one could argue that the best approach to balance these two objectives is to separate the surgical services for these two streams of patients. In the supply chain literature, the need to adapt the supply chain to be responsive for innovative products and efficient for functional products is well established (Fisher, 1997). Hospitals have taken steps to focus some resources for specific services. For example, Hyer, Wemmerlöv, and Morris (2009) report on an empirical study to identify the potential financial and operational benefits at a hospital that transformed its trauma center to ensure that the unit would be focused in terms of resource, space, transformation processes and organization. One measure of success was that the newly focused trauma center yields lower length of stay with similar mortality levels, but the authors note that this success could have resulted from other infrastructural support initiatives, rather than the focus.

Separating surgical resources to handle demand for electives and emergencies may not yield the best possible results for patient wait times. The manufacturing literature stipulates that a company will be better able to meet uncertain demand for multiple products if it operates factories that can be flexible and produce several products to avoid capacity imbalances and reduce variability through a pooling effect. Jordan and Graves (1995) show that if each plant can produce a subset of all products, each product is produced in a subset of plants, and these subsets of products complement each other across the network of plants (a notion they refer to as chaining), the company can meet final product demand almost as well as with the more costly, completely flexible configuration where every plant can produce every product. This argument supports the idea of partial flexibility. The notion that "a little flexibility goes a long way" (Bassamboo, Randhawa, &Van Mieghem, 2010) has also been demonstrated in safety stock decisions in a multilevel system (Baker, Magazine, & Nuttle, 1986), in postponement capacity decisions for delayed product differentiation (Graman & Magazine, 2002), in manufacturing network configuration (Kulkarni, Magazine, & Raturi, 2004), in newsvendor networks (Bassamboo, Randhawa, &Van Mieghem, 2010), in multistage supply chains (Graves & Tomlin, 2003), and in parallel queuing systems (Bassamboo, Randhawa, &Van Mieghem, 2012).

The above-mentioned papers, however, do not consider the notion of product prioritization. Ata and Van Mieghem (2009) identify conditions under which it is best to serve two customer classes, express and regular demand, by either dedicating resources or using an integrated network that contains some flexible resources capable of processing either customer class. Express demand has priority over regular demand. Using Brownian models and large deviations analysis, they find that an integrated network makes it possible for firms that serve mostly regular demand to offer greater quality of service, as measured by responsiveness or delay, almost independently of the correlation between express and regular demand. In our context, this again supports using some flexible surgery rooms to serve both scheduled and emergency surgeries.

Partial Flexibility for Operating Room Planning and Scheduling

In reviewing the literature on the different ways to handle emergency interferences of scheduled elective surgeries, we found that in spite of a large body of research on operating room planning and scheduling, the majority of the research does not consider emergency surgeries. Rather, the focus is almost exclusively on electives surgeries. For example, in the literature review by Cardoen et al. (2010), only 17 of the 121 papers reviewed were classified as addressing emergent or urgent patients. Planning decisions may be classified by time frame (Gupta, 2007): (i) long term, including capacity choices on number of surgical suites and types of surgeries; (ii) medium term, including operating room allocation and use; and (iii) short term, such as scheduling. Our research focuses on capacity allocation decisions (medium term), taking inputs from long-term decisions. These capacity decisions subsequently have an effect on how the short-term scheduling decisions are made.

For capacity decisions, research has addressed the issue of balancing emergency-surgery responsiveness and elective-surgery efficiency by either dedicating some capacity (e.g., Haraden et al., 2003; Bowers & Mould, 2004; Bhattacharyya et al., 2006; Li & Stein, 2008) or by giving priority to emergencies when capacity is shared with elective surgeries (e.g., Gerchak, Gupta, & Henig, 1996; Marcon & Dexter 2006; Wullink et al., 2007; Ferrand et al., 2010; Zonderland, Boucherie, Litvak, & Vleggeert-Lankamp, 2010). Yet, in these papers capacity is either strictly dedicated or entirely shared among electives and emergencies. Only a few papers use a mix of dedicated and shared capacity (Zhang, Murali, Dessouky, & Belson, 2008; Persson & Persson, 2010). In these studies, the operating time of the shared rooms is strictly divided between electives and emergencies, so that part of the day is dedicated to electives and the rest of the day is dedicated to emergencies. As such, the effect of emergencies is different than in our setting where emergencies could access any flexible room as soon as it becomes available. We contribute to the research that considers using a combination of flexible capacity and dedicated capacity by studying how policies of partial flexibility can achieve a better balance between efficiency and responsiveness than either complete flexibility or complete focus.

For scheduling decisions, similar approaches have been studied as those for capacity decisions: either partition the operating room time into elective and emergency surgery time (e.g., Lamiri, Xie, & Zhang, 2008; Hans, Wullink, Van Houdenhoven, & Kazemier, 2008; van Oostrum et al., 2008), or accommodate emergencies in the time frame where electives have been scheduled (e.g., Van der Lans et al., 2008; Pham & Klinkert, 2008). None of the scheduling efforts consider a framework where some of the surgery room capacity is flexible and the rest of the capacity is dedicated to either elective or emergency surgeries.

Overall, the literature indicates that the question whether partial flexibility can improve a hospital's responsiveness and efficiency in handling emergency and elective surgeries is yet to be answered. Given the benefits of partial flexibility that have been documented in manufacturing and service operations (e.g., Jordan & Graves, 1995; Akçay, Balakrishnan, & Xu, 2010), we investigate whether, when, and how much a partial flexibility policy can help surgery planning in hospitals.

PARTIAL FLEXIBILITY MODEL SETUP

In this section we describe the simulated operating theatre, the resource allocation policies, the performance metrics, and the experimental design.

Characteristics of the Simulated Operating Theatre

To investigate partial flexibility, we develop a simulation model based on the operating theatre described in Ferrand et al. (2010). Our simulation model does not contain all the complexities of an actual operating theatre. Rather, it is a stylized model set up to study the effect of various room allocation policies to characterize the trade-off between efficiency and responsiveness and capture the benefits of partial flexibility. There are several complexities we are not considering:

- Fixed cost: we do not consider the costs of configuring the operating rooms to be flexible or dedicated. In practice, flexible rooms may be more costly to set up, since they need to accommodate both elective and emergency surgeries. Such costs considerations would imply a trade-off between the cost and the benefits of increased flexibility.
- Set up time: the turnaround time of the operating room is embedded in the procedure time of the surgeries, and we assume that the procedure times of elective and emergency surgeries are not affected by the room configuration.
- Patient flow: we assume that all patients can leave the operating rooms once the surgery is completed. This is to avoid the confounding effect of boarding situations, where patients have to remain in the operating room because no bed in the post-anesthesia care unit is available upon surgery completion.
- Schedule of elective patients: we keep the schedule of electives constant, to avoid the confounding effect of changing schedules in addition to room dedication/flexibility choices, and we assume that no change can be made in real time to the schedule in reaction to variability in procedure time or the presence of emergency surgeries.

There are 20 operating rooms, numbered 1 through 20, which are used daily over an 8-hour shift plus possible overtime. The system starts empty and idle every morning, and runs until all patients are cleared out of the system. We use 16 consecutive weeks of available hospital data to model arrivals and processing of emergency and elective surgeries. The data indicated that there are 75 elective surgeries scheduled over the 8-hour shift, while an average of 12 emergency surgeries are performed, with only minimal variation in surgeries per shift (coefficients of variation of 0.12 and 0.15, respectively). In our study, the goal is to spread the expected workload of the 75 elective cases fairly evenly throughout the 8-hour shift. Hence, we model elective arrivals in batches of 15 patients, with a fixed inter-arrival time of 90 minutes. The first batch arrives at the start of the operating day, and the last of the five batches is expected to be completed 30 minutes prior to the end of the shift. Late arrivals and no-shows were reportedly minimal for the operating theatre for which we got patient data, so we initially assume that scheduled electives all arrive on time for their surgeries and we do not have no-shows. We later test the effect of elective cancellations. Emergencies can arrive at any time throughout the day. Since a Poisson process is a common assumption for the arrival of random events (Gupta, 2013), we assume emergency surgery arrivals can be modeled as a Poisson process. Given 12 emergencies per day, we first use a constant arrival rate of 1.5 emergencies per hour, and later test the effect of a nonstationary arrival rate.

Historical data also provided total weekly workload for electives and emergencies, so in conjunction with the weekly number of cases we derived the average procedure time per surgery during each week (total hours/number of cases). Over the 16 weeks, the aggregate average surgery time is close to 1.5 hours for electives, and about 2 hours per emergency surgery. These weekly averages are very stable across the 16 weeks, with coefficients of variation of 0.06 and 0.13, respectively. For procedure time distributions, we used the Lognormal distribution, as it has been found to be representative of surgical time distributions (e.g., Zhou & Dexter, 1998; May, Strum & Vargas, 2000; Wullink et al., 2007). For each of the 16 weeks, we only have aggregate data on total weekly workload corresponding to a number of weekly cases, which precludes us from obtaining an accurate measure of surgery-duration variability. For procedure time variability, we used the coefficients of variations for elective and emergency surgeries from the study of Wullink et al. (2007). Hence, in the base case, procedure times for elective surgeries follow a lognormal distribution with a mean of 93 minutes and a standard deviation of 28 minutes, and a lognormal with a mean of 125 minutes and a standard deviation of 92 minutes for emergency surgeries. All times are identical and independently distributed (i.i.d.). It is fair to assume that the procedure times are independent of each other provided that the surgeons are able to perform surgeries in a similar fashion throughout the day. Average operating room utilization in the base case is approximately 80%, which is representative of other hospital settings. A utilization of 75% or greater can be considered moderate to high (Dexter, Macario, Traub, & Lubarsky, 2003), and it should not exceed 85–90% to avoid significant delays (Tyler, Pasquariello, & Chen, 2003). We conduct sensitivity analyses to test the robustness of our results to changes in the emergency arrival pattern and volume, and to changes in the elective surgery procedure time and cancellations.

The setup of this operating theatre presents a combination of factors that makes using a simulation model an attractive choice over analytical methods, such as queuing theory. A specific difficulty lies in the assignment of elective patients to operating rooms, which varies to reflect the changes in the number of dedicated and flexible rooms available to the two patient classes. Moreover, we are investigating a system that is transient in nature, thus ill-adapted to steady-state queuing theory. Finally, our investigation enables more information than existing queuing approximations, which typically focus solely on expected values. We characterize the right-hand side tail of patient wait time, which is especially critical for emergency patients. Overall, discrete event simulation is well suited to the complexities of the hospital environment where it is crucial to understand the impact of organizational changes on patients and staff before they are carried out.

Resource Allocation Policies

Under *complete flexibility*, elective and emergency surgeries share operating rooms. Electives are scheduled across all operating rooms, and emergencies can access any of these rooms. In reality, physicians are often assigned operating time in a specific operating room to conduct surgeries, a practice referred to as block scheduling. This assignment is often conducted well in advance of the surgery day. To reflect this practice in our model, electives are assigned to a specific operating room upon entry into the system without regard to the current status of the system, and wait in the queue that corresponds to this specific room. We could also envision a system where electives wait in a single queue for the next available room, which would be a distinct reality from what we model. Clearly, with electives waiting in a single queue, the disruptions from emergencies would not negatively affect electives as much as in our system's configuration. The 75 electives are assigned in a cyclic fashion through the set of twenty operating rooms: the first elective is assigned to room 20, the second to room 19, etc.; once 20 electives have been assigned, the cycle repeats and the 21st elective in then assigned to room 20, etc. Meanwhile, emergencies wait in a single queue and seize the first operating room that becomes available. Emergencies have priority over electives, but do not preempt elective surgeries that are in progress. If more than one room is available when an emergency arrives, the emergency goes to the room with the lowest number.

Under *complete focus*, emergencies and electives have access to separate subsets of rooms. Electives are assigned to an operating room in a similar fashion as in the the flexible policy, but they access fewer rooms overall. Emergencies still seize the first operating room available, but only considering the rooms included in the subset dedicated to them. Based on the computational experiment in Ferrand et al. (2010), we found that an allocation of five rooms to emergencies and 15 rooms to electives is effective in balancing the trade-off between elective and emergency patients wait time, given the input conditions on the total elective and emergency workload. In particular, dedicating one more room to emergencies resulted in a very small decrease in emergency patient wait time relative to the increase in elective patient wait time, and dedicating one more room to electives produced a similar phenomenon for elective wait time, relative to emergency wait time. Therefore, per Ferrand et al. (2010), we dedicate 5 of the 20 rooms to emergencies in the complete focus policy.

To implement *partial flexibility*, we must decide how many rooms to dedicate to emergencies (n_{EM}), how many rooms to make flexible (n_F), and how many to dedicate to electives (n_{EL}). Given this notation, we use the generic form ($n_{EM}/n_F/n_{EL}$) to refer to a room allocation combination. As such, the complete flexibility policy is (0/20/0), and the complete focus policy is (5/0/15). Since the total number of rooms is fixed at $N = n_{EM} + n_F + n_{EL}$, there are only two decision variables. We built the simulation model so that we could easily evaluate system performance with different combinations of dedicated and flexible rooms by simply entering the number of rooms to dedicate to emergencies (n_{EM}) and to electives (n_{EL}) as input parameters.

Given the number of rooms in each subset, we assign elective and emergency patients accordingly. Upon an emergency arrival, if an emergency-dedicated operating room is available we always assign the emergency to the available dedicated room with the lowest number. If no dedicated operating room is available (including if $n_{EM} = 0$), and if at least one flexible room is available, then we assign the emergency to the available flexible room with the lowest number. If all the rooms are occupied, emergencies wait in a single queue for the next available room. Since electives are assigned without regard to the current status of the system upon their arrival, we cannot decide whether or not to assign electives to flexible rooms based on the current status of the dedicated rooms. Instead, we devise an assignment rule that adapts to the number of rooms available to electives and makes reasonable use of those rooms. This assignment of electives to rooms could be improved by allowing real time adjustments, but these adjustments would add noise in our comparisons of room allocation combinations. We use the following simple heuristic to assign the electives to the $(n_F + n_{EL})$ rooms:

- If $n_{EL} <$ batch size for elective arrivals, then at each batch arrival n_{EL} electives are assigned to the dedicated rooms, the remaining electives are assigned cyclically to the n_F rooms.
- If $n_{EL} \geq$ batch size for elective arrivals, then the electives are assigned cyclically to the $(n_F + n_{EL})$ rooms.

With this heuristic, in the first case we ensure that the rooms dedicated to electives are used, while the electives that cannot be accommodated in dedicated rooms are distributed evenly across the flexible rooms. In the second case, applying the first rule would result in no elective being assigned to the flexible rooms, and implies that electives cannot get access to the flexible rooms since we do not alter the assignment of electives to rooms in real time based on the status of the system. As such, following the logic from the first case would revert to the complete focus policy, with electives only assigned to elective rooms and emergencies accessing rooms that never handle electives. This could result in no flexible room being used, unless there are emergencies, thus converting the flexible rooms to emergency-dedicated rooms. Instead, the second rule allows for some electives to access the flexible rooms and distributes electives evenly across the rooms available to them. Flexible rooms thus serve the same role as under the complete flexibility policy, with emergencies having priority in accessing those rooms, even if electives are waiting for the flexible room they have been assigned to. We recognize that with this second rule, an elective could be assigned to a flexible room even though a dedicated room is available, but this is a limitation of assigning the electives upon their arrival without regard to the current status of the system, not a limitation of our heuristic. More generally, it could happen that an elective batch is spread out over more flexible than dedicated rooms, but since we keep cycling, this will be compensated by other batches where electives are spread over more dedicated than flexible rooms. Overall, this discussion shows that the assignment of patients to rooms is not trivial, because it depends on the room allocation combination. Our rules to assign patients to rooms reflect this interdependence, and overall we seek to have consistent assignment rules across the various room allocation combinations considered, so that changes in the performance results are primarily due to the room allocation decisions.

Now, the second rule only comes into effect when the batch size is less than n_{EL}. With the batch size fixed at 15 in this study, this condition is not fulfilled in the majority of the simulation experiments; the sole exception is combination (4/1/15). Hence, our results are overall unaffected by any difference due to switches between assignment rules. For the combination (4/1/15), the cyclic assignment applies as follows: the first batch is assigned to rooms 20, 19, ..., 6 (elective rooms); the second batch is assigned to rooms 5 (flexible room), 20, 19, ..., 7; the third batch is assigned to rooms 6, 5, 20, ..., 8; etc. As such, the flexible room provides some relief to the elective rooms, and it does not affect emergencies much, because when the flexible room becomes available, emergencies have priority over electives assigned to it. Electives assigned to this flexible room are not affected much either, because there are four emergency rooms, so emergencies will not be using the flexible room much.

Performance Metrics and Experimental Design

We model the set of operating rooms with the resource allocation policies described above using Arena, version 12.0. We verify the model using well-documented techniques (e.g., Kelton, Sadowski, & Swets, 2009), including testing the model under specific input parameter values. We validate the model by comparing historical data on utilization to the utilization output from the simulated setting.

For each room allocation combination in our computational experiment, we collect the following performance metrics to capture both responsiveness and efficiency: average, 75th and 95th percentiles of the wait time in queue for elective and emergency patients, total staff overtime, average number of rooms with overtime, and average room utilization during the 8-hour shift. Total overtime captures the magnitude of overtime, while the average number of rooms with overtime gives an indication of the spread of overtime across the rooms. In practice, there may be surgeries that run so much longer than anticipated that other cases would be cancelled to avoid excessive overtime, but for comparison purposes we do not consider this eventuality. Also, note that we record utilization only during the 8-hour shift because during overtime the utilization is 100% until all the work assigned to the room is completed. Hence, this utilization measure can be compared across room allocation combinations because the denominator remains constant at 8 hours in the utilization calculation. As such, utilization gives us an indication of how busy the system is during the regular hours, which correlates negatively with the amount of overtime across room allocations. Furthermore, overtime correlates positively with elective patient wait time, since the higher the elective wait time, the more likely it is that the elective surgery is not completed during regular time.

The half-width of the 95% confidence intervals of initial runs indicated that 400 replications would narrow the half-widths enough for comparison purposes (between 5% and 20% of the estimated output value). We use common random numbers as a variance reduction technique to increase the precision of our comparisons across the design points of our experiment (Law & Kelton, 2000). We embed a Visual Basic for Applications (VBA) code in our Arena model to implement our heuristic for the cyclic assignment of electives according to the number of rooms allocated to each of the three subsets (n_{EM}, n_F, and n_{EL}).

We first test the robustness of our results by fitting a nonstationary arrival rate for emergency surgeries. We based this arrival rate on an analysis of one year of patient arrivals to a local emergency department (ED) associated with this operating theatre, and focused on the hourly arrivals during the 8-hour shift from this ED, representing approximately 12,000 arrivals. Given the previously established average of 1.5 emergency arrivals per hour for surgery, we scaled the observed hourly arrival rate to the ED accordingly and use a piecewise stationary Poisson process with a different arrival rate every hour ranging from 0.72 surgeries

Table 1: Experimental design for the three factors with the different levels considered.

Case	Emergency Arrival Process	Emergency Arrival Rate per Hour	Elective Procedure Time in Min. (Mean, *SD*)
1	Stationary	1.5	(93,28)
2	Nonstationary	1.5	(93,28)
3	Stationary (with cancellations)	1.5	(93,28)
4	Stationary	2.25	(93,28)
5	Nonstationary	2.25	(93,28)
6	Stationary	1.5	(139,42)
7	Stationary	1.5	(93,93)
8	Stationary	1.5	(70,21)

to 1.76 surgeries per hour. The variations in the hourly arrival pattern from this ED were similar to those reported for the national average by Pitts, Niska, Xu, and Burt (2008), with peaks in the late morning and early evening. Applying the hourly variations in ED arrivals to the arrivals of emergencies to the operating theatre is an approximation because these two arrival processes may not be the same. In fact, emergency arrivals to the operating theatre may not only come from the ED, but also from a referring physician, ICU patients, and inpatients.

We also consider the effect of elective surgery cancellations by factoring in a 17% cancellation rate, established by several previous studies (e.g., Hand, Levin, & Stanziola, 1990; Lacqua & Evans, 1994). We model those cancellations as no-shows upon their scheduled arrival time, thereby reducing the number of elective surgeries by 17%.

We also investigate the effects of three factors relating to patient volume, procedure time duration, and variability. First, we increase the arrival rate of emergencies by 50% to study how well the system performs under increased pressure from emergencies. We do this under both stationary and nonstationary emergency arrivals. Then we modify the elective procedure time distribution parameters, and look at two cases: we increase both the mean and the standard deviation by 50%, so that the coefficient of variation (CV) remains constant; and we increase the standard deviation only, so that the CV increases from 0.3 to 1.0. Lastly, we decrease both the mean and standard deviation of elective procedure times by 25%, so that the CV remains constant. The three factors, the different levels considered, and the design of the experiment are presented in Table 1, where Case 1 is the base case scenario.

RESULTS: BENEFITS OF PARTIAL FLEXIBILITY

Achieving High Partial Flexibility by Dedicating Operating Rooms

When comparing output performance measures across room allocation combinations, a statistically significant change between two estimates is indicated by 95% confidence intervals that do not overlap. Accordingly, in the following results, when we state that there is no change in a given performance measure, it is because the confidence intervals overlap. The results for the base case input conditions (Case 1 in Table 1) for all the different room allocations are reported in Table 2. Averages and percentiles are reported together with the half-width of their 95% confidence intervals.

Departing from the complete flexibility policy (0/20/0), we begin by dedicating a room, one at a time, to either electives or emergencies. We find that dedicating one room to emergencies (1/19/0) reduces the average wait time of electives, while average wait time of emergencies remains unchanged. We discuss the reasons for this counter-intuitive result in the section "Observations on the Benefits of Partial Flexibility." Considering the 75th and 95th percentiles of wait time for these two allocations, we find no significant change moving to a single dedicated room. However, total overtime is reduced. Conversely, dedicating one room to electives (0/19/1) does not provide any improvement, compared to the complete flexibility policy (0/20/0). Once a room has been dedicated to emergencies (1/19/0), we compare dedicating a second room, either to electives or to emergencies. In assigning the second dedicated room, we find similar results as when we dedicated the first room: dedicating the room to emergencies (2/18/0) reduces average elective wait time and total overtime without affecting average emergency wait time, while dedicating the room to electives (1/18/1) does not provide any improvement. We observe similar results when we dedicate a third room, as shown in Table 2. However, after three rooms have been dedicated to emergencies, the benefits of dedicating an additional room become insignificant, consistent with diminishing returns from dedicating rooms to emergencies. Overall, the last column of Table 2 indicates that average utilization remains stable, except for the combinations (6/14/0) and (7/13/0), and these exceptions are discussed later in this section.

Comparing high partial flexibility with the combination (3/17/0) to the complete flexibility policy (0/20/0), we find that average elective wait time decreases 23%, from 28.2 to 21.8 minutes, the 75th percentile decreases 12%, from 34.0 to 29.9 minutes, and the 95th percentile decreases 30%, from 152.5 to 106.1 minutes. Meanwhile, average emergency wait time, and the 75th and 95th percentiles do not change. The percentiles of wait time parallel the observed changes on average wait time. The percentiles not only confirm the findings from the comparison of averages, but they also magnify some of the differences across room allocation combinations. This is all the more important, since some of the differences in averages are relatively small. Total overtime decreases 13%, from 788.5 to 685.4 minutes. Lastly, we find that the average number of rooms experiencing overtime is not very sensitive to the room allocation combination, as it consistently varies between 10 and 13 rooms. In summary, high partial flexibility – with (in this case) three rooms dedicated to emergencies—can perform better than complete flexibility.

Necessarily, there is a point where we can no longer dedicate an additional room to emergencies without hurting the electives, who would not have access to enough capacity to be handled in a timely fashion. In the base case, the expected workload from the 75 scheduled elective surgeries requires that at least 15 operating rooms are available during the 8-hour operating shift to keep expected utilization of

Table 2: Performance measures under various combinations of room allocation.

Case 1			Emergency Wait Time			Elective Wait Time			Sum	
			Average	75 Percentile	95 Percentile	Average	75 Percentile	95 Percentile	Staff Overtime	Average Room
n_{EM}	n_F	n_{EL}	(min.)	(min.)	(min.)	(min.)	(min.)	(min.)	(min.)	Utilization
0	20	0	0.306 ± 0.06	0.097 ± 0.06	2.717 ± 0.49	28.21 ± 1.21	33.96 ± 1.79	152.5 ± 7.51	788.5 ± 17.5	0.798 ± < .01
1	19	0	0.271 ± 0.06	0.128 ± 0.10	2.375 ± 0.47	25.56 ± 1.17	31.12 ± 1.59	137.8 ± 7.91	747.2 ± 17.6	0.802 ± < .01
0	19	1	0.318 ± 0.07	0.171 ± 0.10	2.671 ± 0.49	27.80 ± 1.18	33.60 ± 1.76	149.2 ± 7.38	781.7 ± 17.5	0.798 ± < .01
2	18	0	0.266 ± 0.06	0.147 ± 0.09	2.232 ± 0.42	23.11 ± 1.07	29.73 ± 1.55	120.4 ± 6.86	706.0 ± 16.7	0.806 ± < .01
1	18	1	0.297 ± 0.08	0.173 ± 0.15	2.513 ± 0.48	25.16 ± 1.14	30.99 ± 1.57	133.3 ± 7.42	739.1 ± 17.1	0.803 ± < .01
3	17	0	0.355 ± 0.07	0.150 ± 0.08	3.106 ± 0.52	21.76 ± 0.91	29.93 ± 1.35	106.1 ± 5.42	685.4 ± 15.9	0.809 ± < .01
2	17	1	0.313 ± 0.07	0.218 ± 0.14	2.491 ± 0.45	23.02 ± 1.05	29.63 ± 1.50	119.4 ± 6.54	707.5 ± 16.8	0.806 ± < .01
4	16	0	0.326 ± 0.07	0.153 ± 0.10	2.929 ± 0.56	21.57 ± 0.82	31.44 ± 1.27	101.0 ± 4.56	679.4 ± 14.9	0.809 ± < .01
3	16	1	0.375 ± 0.07	0.198 ± 0.11	3.174 ± 0.52	21.76 ± 0.93	29.89 ± 1.32	106.8 ± 5.62	684.6 ± 15.9	0.809 ± < .01
5	15	0	0.297 ± 0.07	0.171 ± 0.12	2.552 ± 0.52	22.32 ± 0.72	34.53 ± 1.23	98.10 ± 3.63	691.6 ± 14.5	0.808 ± < .01
4	15	1	0.342 ± 0.07	0.154 ± 0.10	3.051 ± 0.58	21.54 ± 0.81	31.48 ± 1.25	100.2 ± 4.36	679.7 ± 15.0	0.809 ± < .01
6	14	0	0.142 ± 0.05	0.097 ± 0.12	1.342 ± 0.38	37.85 ± 0.76	63.59 ± 1.50	149.9 ± 3.63	983.8 ± 14.6	0.777 ± < .01
7	13	0	0.054 ± 0.02	0 ± 0	0.567 ± 0.22	54.05 ± 0.82	96.62 ± 1.53	169.9 ± 2.93	1289 ± 14.7	0.746 ± < .01
0	18	2	0.334 ± 0.07	0.209 ± 0.15	2.805 ± 0.50	27.29 ± 1.15	33.01 ± 1.65	145.4 ± 7.21	774.6 ± 17.2	0.799 ± < .01
0	17	3	0.345 ± 0.07	0.165 ± 0.10	2.945 ± 0.51	26.81 ± 1.10	32.48 ± 1.59	143.9 ± 6.77	761.7 ± 16.9	0.801 ± < .01
0	10	10	0.868 ± 0.19	0.745 ± 0.36	6.262 ± 1.07	24.23 ± 0.82	32.91 ± 1.36	116.4 ± 4.43	719.9 ± 15.6	0.805 ± < .01
5	14	1	0.314 ± 0.07	0.198 ± 0.13	2.619 ± 0.54	22.35 ± 0.74	34.45 ± 1.22	98.26 ± 3.71	691.6 ± 14.5	0.808 ± < .01
5	13	2	0.333 ± 0.08	0.223 ± 0.15	2.684 ± 0.55	22.32 ± 0.74	34.45 ± 1.21	98.04 ± 3.68	691.0 ± 14.5	0.808 ± < .01
5	12	3	0.369 ± 0.09	0.249 ± 0.18	3.031 ± 0.60	22.31 ± 0.74	34.36 ± 1.20	97.86 ± 3.72	691.2 ± 14.5	0.808 ± < .01
5	11	4	0.409 ± 0.10	0.272 ± 0.18	3.373 ± 0.67	22.25 ± 0.74	34.31 ± 1.21	97.42 ± 3.65	689.9 ± 14.5	0.808 ± < .01
5	10	5	0.464 ± 0.11	0.351 ± 0.21	3.724 ± 0.73	22.20 ± 0.73	34.26 ± 1.20	97.34 ± 3.75	689.3 ± 14.5	0.808 ± < .01
5	5	10	0.882 ± 0.19	0.758 ± 0.34	6.460 ± 1.15	21.96 ± 0.70	34.00 ± 1.18	96.92 ± 3.69	687.0 ± 14.5	0.808 ± < .01
5	4	11	1.070 ± 0.24	0.950 ± 0.38	7.500 ± 1.32	21.87 ± 0.68	33.82 ± 1.13	96.33 ± 3.71	686.1 ± 14.5	0.808 ± < .01
5	3	12	1.370 ± 0.31	1.383 ± 0.53	9.039 ± 1.54	21.62 ± 0.73	33.70 ± 1.12	95.38 ± 3.64	682.1 ± 14.3	0.809 ± < .01
5	2	13	1.871 ± 0.42	2.035 ± 0.77	11.84 ± 1.93	21.40 ± 0.63	33.44 ± 1.10	94.46 ± 3.49	678.5 ± 14.2	0.809 ± < .01
5	1	14	2.769 ± 0.65	3.744 ± 1.28	14.96 ± 2.50	20.98 ± 0.60	33.03 ± 1.09	91.17 ± 3.20	672.7 ± 14.1	0.810 ± < .01
5	0	15	4.758 ± 1.09	7.565 ± 2.18	21.96 ± 4.00	19.83 ± 0.52	32.30 ± 1.07	85.83 ± 2.05	658.7 ± 13.9	0.811 ± < .01
4	1	15	8.089 ± 1.35	12.97 ± 2.58	33.55 ± 3.89	20.06 ± 0.66	28.61 ± 1.03	92.01 ± 3.87	676.1 ± 16.6	0.809 ± < .01
4	0	16	14.83 ± 2.34	24.99 ± 4.40	54.41 ± 5.99	17.55 ± 0.48	27.51 ± 1.00	80.94 ± 2.13	662.6 ± 16.0	0.811 ± < .01

these rooms below 100% (75 * 93 minutes/480 minutes per room = 14.5 rooms). Hence, this is a transition point for elective wait time, as any allocation of rooms to electives that results in fewer than 15 rooms ($n_F + n_{EL} < 15$) will lead to a sharp increase in elective wait time and overtime (see, for example (6/14/0) and (7/13/0)). Meanwhile, average room utilization decreases because the utilization of the rooms dedicated to emergencies decreases as more rooms are dedicated to them. Accounting for this capacity level is crucial, as it provides an upper bound on how many rooms can be dedicated to emergencies. Besides this capacity consideration, the size of the batch arrivals of electives is also important: with electives arriving 15 at a time, having ($n_F + n_{EL} < 15$) will lead to an increase in elective wait time, as some patients will have to wait upon arrival. However, Ferrand et al. (2010) found that these transition points for electives depend more on the capacity level than on the size of the batch arrivals of electives, since the batch size had a small effect on performance measures, relative to the number of dedicated rooms.

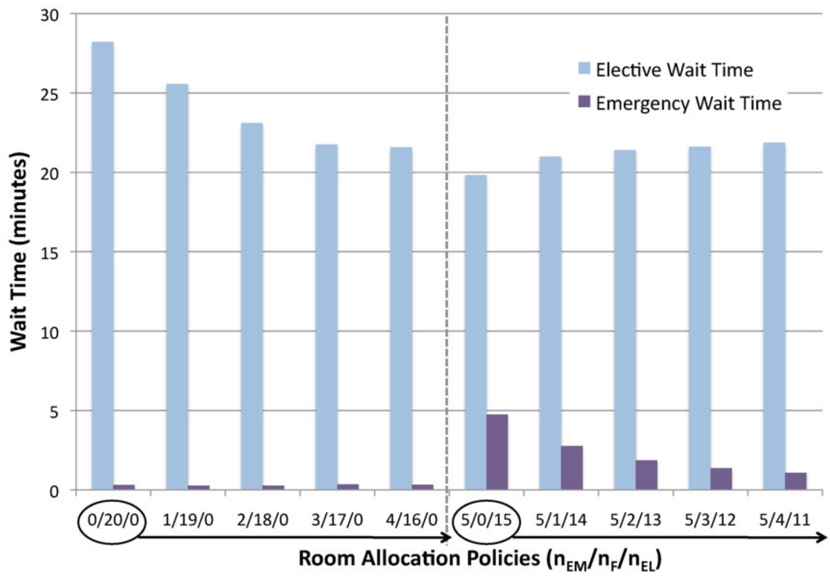
Once five operating rooms have been dedicated to emergencies, dedicating rooms to electives does not significantly reduce elective wait time averages and percentiles, or total overtime, as shown by various combinations in Table 2, except when at most one room is flexible (see combinations (5/14/1) to (5/0/15), compared to (5/15/0)). However, combinations (5/1/14) and (5/0/15) provide relatively small reductions in elective wait time and overtime, compared to the increase in emergency wait time. This suggests that once we dedicate enough capacity to emergencies, it is best to make the remaining capacity flexible.

Achieving Low Partial Flexibility by Making Operating Rooms Flexible

In a similar fashion to the previous section, we compare the effect of making rooms flexible when taking them from the pool of rooms dedicated to emergencies and from the pool of rooms dedicated to electives. Departing from the complete focus policy (5/0/15), we observe that it is beneficial for system performance to add flexible rooms by taking the rooms from those dedicated to electives. When we incrementally make one of these dedicated rooms flexible, it results in lower average emergency wait time, as well as lower 75th and 95th percentile of emergency wait time, while elective wait time is only minimally increased. Comparing low partial flexibility with three flexible rooms (5/3/12) to the complete focus policy illustrates this point: average emergency wait time is reduced 71%, from 4.8 to 1.4 minutes, the 95 percentile is reduced 59%, from 22.0 to 9.0 minutes, while average elective wait time increases 9%, from 19.8 to 21.6 minutes, and the 95 percentile wait time increases 11%, from 85.8 to 95.4 minutes. Meanwhile, total overtime is not significantly changed.

On the contrary, taking rooms from those dedicated to emergencies (e.g., (4/1/15)) deteriorates emergency wait time and does not improve elective wait time, compared to the complete focus policy. As such, policies with even fewer rooms available to emergencies (where $n_{EM} + n_F \leq 4$, e.g., (4/0/16), (3/1/16)) would deteriorate emergency wait time further. This is illustrated with the output from combination (4/0/16), which results in a three-fold increase in average emergency wait time relative to the combination (5/0/15). The consideration of

Figure 2: The benefits of partial flexibility on patient wait time.



capacity relative to the expected workload from emergencies is essential, as it gives a lower bound on the number of rooms emergencies need access to. In this case, the expected workload from the 12 emergencies requires that at least four rooms are available for emergencies (12 * 125 minutes /480 minutes per room = 3.1 rooms), so given the need for responsiveness, we can expect that dedicating less than five rooms to emergencies will result in higher emergency wait time.

Comparing High and Low Partial Flexibility

The graph in Figure 2 depicts average elective and emergency patient wait times as we move away from the complete flexibility policy (0/20/0) toward high partial flexibility (4/16/0), and from the complete focus policy (5/0/15) toward low partial flexibility (5/4/11). It indicates the benefits of moving toward partial flexibility and also allows us to compare the two types of partial flexibilities. This comparison reveals that high partial flexibility (4/16/0) is overall better than low partial flexibility (5/4/11): average emergency wait time is lower with high partial flexibility while average elective wait time is not significantly different. Comparing the 75th and 95th percentile values of the two partial flexibility policies leads to the same observation and magnifies the difference for emergencies. Meanwhile, total overtime is not significantly different. This comparison of the two ways to achieve partial flexibility indicates that having mostly flexible rooms and a few dedicated rooms to emergencies provides better system performance than having most rooms dedicated and a few flexible rooms. Furthermore, by design, high partial flexibility leads to more rooms being disrupted by emergencies than low partial flexibility. This has two implications for electives: more electives are exposed to emergency disruptions, and the electives in each room will experience fewer disruptions, given that the expected number and workload of emergencies is shared among more rooms. These two implications have opposite effects on elective wait time, and the above comparison of results shows that the net effect is insignificant.

Nonstationary Emergency Arrivals and Elective Cancellations

The average wait time resulting from the various room configurations under nonstationary emergency arrivals (Case 2), and under elective cancellations (Case 3) are displayed in Table 3 below. In our experiment, high and low partial flexibilities provide similar types of improvements under nonstationary emergency arrivals, when compared to the improvements obtained in the base case. Specifically, the comparison of high partial flexibility (3/17/0) to complete flexibility (0/20/0) indicates a reduction in elective emergency wait time without significant change in emergency wait time, and the comparisons of low partial flexibility (5/3/12) to complete dedication (5/0/15) shows a reduction in emergency wait time with only a slight increase in elective patient wait time. The magnitude of the changes is similar whether or not the arrival process of emergencies is stationary. Hence, the key difficulty with emergencies is not that their arrival process is nonstationary. Rather, it is the randomness of the arrival process, compounded by the long procedure time duration.

Cancellations have an effect on effective capacity, so we can anticipate the direction of change in wait time relative to the cancellation rate. For example, the inclusion of cancellations compared to the base case indicates a reduction in average wait time in the anticipated ways: for emergencies, the average wait drops close to zero when most rooms are flexible, and the change is insignificant when most rooms are dedicated; for electives, the average wait time is reduced between 25% and 40%, depending on the room allocation. More importantly, the results in Table 3 show that when elective cancellations are factored in, the comparisons of high partial flexibility (3/17/0) to complete flexibility (0/20/0) and of low partial flexibility (5/3/12) to complete dedication (5/0/15) indicate similar benefits of partial flexibility as in the base case.

Increase in the Number of Emergency Patients Arrivals

We also investigate whether the benefits of partial flexibility are robust to changes in the emergency stationary arrival rate and in the procedure time characteristics of elective surgeries. The results of these sensitivity analyses are reported in Table 4 below.

First, we increase the number of emergency patient arrivals (Case 4). As we shift from complete flexibility to high partial flexibility, elective wait time decreases, while emergency wait time remains unchanged, as in the base case, and for the same reason: the rooms dedicated to emergencies lead to fewer disruptions of electives. The magnitude of the improvement is approximately the same as in the base case. Shifting from the completely focused policy to low partial flexibility also provides the same type of result as in the base case, and for the same reasons: emergency wait time decreases, while that of electives increases, because emergencies

Table 3: Average wait times under nonstationary emergency arrivals, and under elective cancellations.

			Nonstationary Emergency Arrivals—Case 2		Electives Cancellations—Case 3	
n_{EM}	n_F	n_{EL}	Average Emergency Wait Time (min.)	Average Elective Wait Time (min.)	Average Emergency Wait Time (min.)	Average Elective Wait Time (min.)
0	20	0	0.395 ± 0.08	26.77 ± 1.14	0.058 ± 0.02	19.52 ± 0.95
1	19	0	0.412 ± 0.09	24.22 ± 1.09	0.064 ± 0.03	17.32 ± 0.92
0	19	1	0.404 ± 0.08	26.44 ± 1.10	0.103 ± 0.05	19.47 ± 0.93
2	18	0	0.442 ± 0.09	21.99 ± 0.96	0.045 ± 0.02	15.46 ± 0.84
1	18	1	0.453 ± 0.09	23.96 ± 1.07	0.066 ± 0.03	17.22 ± 0.87
3	17	0	0.435 ± 0.09	21.08 ± 0.83	0.050 ± 0.02	14.22 ± 0.70
2	17	1	0.467 ± 0.10	22.00 ± 0.95	0.068 ± 0.03	15.37 ± 0.85
4	16	0	0.403 ± 0.08	21.35 ± 0.77	0.095 ± 0.03	13.07 ± 0.59
3	16	1	0.477 ± 0.10	21.05 ± 0.83	0.082 ± 0.03	13.97 ± 0.67
5	15	0	0.399 ± 0.09	22.42 ± 0.73	0.091 ± 0.03	13.23 ± 0.58
4	15	1	0.417 ± 0.08	21.42 ± 0.78	0.107 ± 0.03	13.27 ± 0.60
6	14	0	0.189 ± 0.06	37.91 ± 0.75	0.052 ± 0.02	22.71 ± 0.71
7	13	0	0.098 ± 0.04	54.16 ± 0.83	0.056 ± 0.03	32.07 ± 0.83
0	18	2	0.415 ± 0.09	26.08 ± 1.07	0.089 ± 0.04	19.17 ± 0.88
0	17	3	0.419 ± 0.08	25.80 ± 1.03	0.101 ± 0.04	18.76 ± 0.85
0	10	10	1.141 ± 0.23	23.86 ± 0.81	0.553 ± 0.16	16.40 ± 0.71
5	14	1	0.419 ± 0.09	22.43 ± 0.73	0.107 ± 0.03	13.27 ± 0.59
5	13	2	0.457 ± 0.10	22.37 ± 0.72	0.128 ± 0.04	13.31 ± 0.61
5	12	3	0.522 ± 0.11	22.33 ± 0.72	0.142 ± 0.04	13.31 ± 0.61
5	11	4	0.566 ± 0.12	22.24 ± 0.71	0.157 ± 0.05	13.33 ± 0.61
5	10	5	0.622 ± 0.13	22.20 ± 0.71	0.207 ± 0.06	13.36 ± 0.61
5	5	10	1.190 ± 0.24	21.88 ± 0.68	0.659 ± 0.16	13.47 ± 0.62
5	4	11	1.403 ± 0.27	21.76 ± 0.65	0.874 ± 0.20	13.42 ± 0.61
5	3	12	1.799 ± 0.36	21.57 ± 0.64	1.247 ± 0.29	13.29 ± 0.59
5	2	13	2.535 ± 0.51	21.26 ± 0.61	1.938 ± 0.45	13.06 ± 0.55
5	1	14	3.740 ± 0.79	20.88 ± 0.59	3.257 ± 0.79	12.74 ± 0.51
5	0	15	6.150 ± 1.28	19.83 ± 0.52	6.113 ± 1.43	11.69 ± 0.39
4	1	15	9.709 ± 1.52	20.05 ± 0.66	8.671 ± 1.67	12.73 ± 0.58
4	0	16	17.94 ± 2.58	17.55 ± 0.48	17.32 ± 2.94	10.57 ± 0.38

have access to more operating rooms, and some electives share flexible rooms with emergencies. Compared to the base case, the magnitude of the improvement for emergencies is similar (−77% for average wait time, from 22.1 to 5.2 minutes), however, the deterioration for electives is greater than in the base case (+42% for average wait time, from 19.8 to 28.2 minutes).

We then test the effect of this increased emergency load in combination with the nonstationary emergency arrival pattern used for Case 2. This is Case 5 in Table 4. We observe similar patterns and similar performance measures as in Case 4, so we can draw the same conclusions as in the previous paragraph, reinforcing the attractiveness of partial flexibility policies.

Table 4: Performance measures—sensitivity analyses.

			Emergency Wait Time			Elective Wait Time			Sum	
n_{EM}	n_F	n_{EL}	Average (min.)	75 Percentile (min.)	95 Percentile (min.)	Average (min.)	75 Percentile (min.)	95 Percentile (min.)	Staff Overtime (min.)	Average Room Utilization
Sensitivity to an increase in emergency arrival rate—Case 4:										
0	20	0	0.715 ± 0.10	0.390 ± 0.17	5.045 ± 0.61	39.22 ± 1.50	51.93 ± 2.40	192.3 ± 7.65	1164 ± 23.2	0.837 ± < 0.01
1	19	0	0.715 ± 0.11	0.437 ± 0.17	5.137 ± 0.64	36.39 ± 1.51	48.71 ± 2.37	185.9 ± 9.15	1122 ± 23.0	0.841 ± < 0.01
2	18	0	0.803 ± 0.12	0.585 ± 0.23	5.602 ± 0.64	33.77 ± 1.43	44.99 ± 2.25	169.0 ± 8.32	1085 ± 22.5	0.845 ± < 0.01
3	17	0	0.772 ± 0.10	0.393 ± 0.14	5.595 ± 0.58	31.16 ± 1.33	42.13 ± 2.04	152.9 ± 7.66	1044 ± 21.6	0.849 ± < 0.01
5	3	12	5.160 ± 0.72	6.067 ± 1.13	22.77 ± 1.99	28.23 ± 1.15	37.78 ± 1.32	140.4 ± 8.58	996.7 ± 21.4	0.854 ± < 0.01
5	2	13	7.800 ± 1.09	9.796 ± 1.63	30.48 ± 2.60	27.14 ± 1.05	36.16 ± 1.21	137.4 ± 9.19	985.4 ± 21.3	0.855 ± < 0.01
5	1	14	12.41 ± 1.71	16.29 ± 2.51	43.35 ± 3.66	24.72 ± 0.81	34.47 ± 1.15	112.5 ± 7.45	965.3 ± 21.4	0.858 ± < 0.01
5	0	15	22.08 ± 2.77	31.39 ± 4.18	66.84 ± 5.59	19.83 ± 0.52	32.30 ± 1.07	85.83 ± 2.05	942.1 ± 21.9	0.860 ± < 0.01
Sensitivity to an increase in emergency arrival rate under a nonstationary arrival process—Case 5:										
0	20	0	0.738 ± 0.10	0.364 ± 0.14	5.489 ± 0.60	37.42 ± 1.48	49.09 ± 2.29	186.9 ± 7.62	1221 ± 23.7	0.831 ± < 0.01
1	19	0	0.715 ± 0.09	0.348 ± 0.13	5.642 ± 0.58	34.47 ± 1.42	45.09 ± 2.14	174.5 ± 8.14	1173 ± 23.9	0.836 ± < 0.01
2	18	0	0.916 ± 0.11	0.525 ± 0.16	6.506 ± 0.62	31.77 ± 1.29	41.69 ± 2.01	159.9 ± 7.55	1130 ± 23.0	0.840 ± < 0.01
3	17	0	0.972 ± 0.11	0.499 ± 0.14	6.903 ± 0.65	29.95 ± 1.20	40.15 ± 1.90	148.5 ± 6.80	1093 ± 22.5	0.844 ± < 0.01
5	3	12	6.859 ± 0.84	8.920 ± 1.34	28.27 ± 2.11	27.82 ± 1.02	36.87 ± 1.25	137.1 ± 7.76	1066 ± 22.5	0.847 ± < 0.01
5	2	13	9.926 ± 1.24	13.55 ± 1.91	36.47 ± 2.83	26.93 ± 0.96	35.67 ± 1.19	136.8 ± 8.80	1062 ± 22.6	0.847 ± < 0.01
5	1	14	15.85 ± 1.94	22.64 ± 2.95	50.83 ± 4.04	24.47 ± 0.74	34.14 ± 1.13	102.4 ± 5.16	1051 ± 22.9	0.849 ± < 0.01
5	0	15	27.13 ± 3.04	40.93 ± 4.74	78.44 ± 6.17	19.83 ± 0.52	32.30 ± 1.07	85.83 ± 2.05	1037 ± 23.5	0.850 ± < 0.01
Sensitivity to an increase in elective patient procedure time (mean and variance)—Case 6:										
0	20	0	4.038 ± 0.27	6.481 ± 0.56	16.71 ± 0.90	78.46 ± 1.73	120.3 ± 2.77	290.1 ± 8.13	2922 ± 27.4	0.934 ± < 0.01
1	19	0	3.115 ± 0.22	5.036 ± 0.49	14.15 ± 0.83	80.09 ± 1.74	125.5 ± 2.82	285.2 ± 8.11	2971 ± 27.3	0.929 ± < 0.01
2	18	0	2.549 ± 0.22	3.689 ± 0.49	13.13 ± 0.95	82.98 ± 1.65	132.2 ± 2.82	283.3 ± 7.21	3054 ± 26.0	0.921 ± < 0.01
3	17	0	1.755 ± 0.19	2.228 ± 0.36	10.03 ± 0.93	87.92 ± 1.51	141.5 ± 2.52	285.4 ± 6.83	3186 ± 24.2	0.907 ± < 0.01
5	3	12	1.949 ± 0.42	2.335 ± 0.79	12.00 ± 1.89	100.6 ± 1.26	161.6 ± 2.09	278.3 ± 4.09	3563 ± 21.2	0.867 ± < 0.01
5	2	13	2.510 ± 0.56	3.222 ± 1.04	14.05 ± 2.24	100.2 ± 1.23	161.1 ± 2.09	277.1 ± 4.10	3557 ± 21.2	0.868 ± < 0.01
5	1	14	3.304 ± 0.77	4.543 ± 1.40	16.92 ± 2.72	99.50 ± 1.19	159.8 ± 2.06	274.9 ± 3.57	3550 ± 21.1	0.869 ± < 0.01
5	0	15	4.758 ± 1.09	7.565 ± 2.18	21.96 ± 3.56	98.33 ± 1.15	158.5 ± 2.02	269.2 ± 3.29	3537 ± 21.0	0.870 ± < 0.01

Continued

Table 4: Continued

n_{EM}	n_F	n_{EL}	Emergency Wait Time Average (min.)	Emergency Wait Time 75 Percentile (min.)	Emergency Wait Time 95 Percentile (min.)	Elective Wait Time Average (min.)	Elective Wait Time 75 Percentile (min.)	Elective Wait Time 95 Percentile (min.)	Sum Staff Overtime (min.)	Average Room Utilization
Sensitivity to an increase in elective patient procedure time (variance only)—Case 7:										
0	20	0	0.190 ± 0.04	0.055 ± 0.05	1.856 ± 0.39	44.06 ± 1.59	52.29 ± 2.82	235.2 ± 8.71	1558 ± 31.0	0.714 ± 0.01
1	19	0	0.187 ± 0.05	0.097 ± 0.07	1.703 ± 0.41	44.40 ± 1.70	50.80 ± 3.00	242.5 ± 9.85	1560 ± 31.4	0.714 ± < 0.01
2	18	0	0.150 ± 0.04	0.052 ± 0.06	1.433 ± 0.36	44.70 ± 1.79	50.38 ± 3.39	248.5 ± 11.0	1569 ± 31.5	0.713 ± < 0.01
3	17	0	0.202 ± 0.07	0.105 ± 0.09	1.709 ± 0.43	46.41 ± 1.86	51.61 ± 3.31	263.5 ± 12.0	1595 ± 31.6	0.710 ± < 0.01
5	3	12	1.190 ± 0.35	1.298 ± 0.58	7.626 ± 1.54	51.46 ± 1.99	62.78 ± 3.57	283.0 ± 12.7	1660 ± 32.6	0.703 ± 0.01
5	2	13	1.837 ± 0.49	2.130 ± 0.88	10.86 ± 2.07	51.35 ± 1.98	62.14 ± 3.55	283.6 ± 12.7	1661 ± 32.6	0.703 ± 0.01
5	1	14	2.854 ± 0.72	3.784 ± 1.32	15.00 ± 2.65	50.96 ± 1.98	61.12 ± 3.52	282.8 ± 12.7	1658 ± 32.6	0.704 ± 0.01
5	0	15	4.758 ± 1.09	7.565 ± 2.18	21.96 ± 3.56	50.02 ± 1.98	59.66 ± 3.52	278.1 ± 12.8	1654 ± 32.6	0.704 ± 0.01
Sensitivity to a decrease in elective patient procedure time (mean and variance)—Case 8:										
0	20	0	0.051 ± 0.03	<.010 ± < .01	0.493 ± 0.21	17.38 ± 1.02	6.576 ± 1.19	121.3 ± 7.40	481.7 ± 14.3	0.650 ± < 0.01
1	19	0	0.062 ± 0.03	0.010 ± 0.02	0.592 ± 0.29	13.80 ± 0.97	3.514 ± 0.75	100.9 ± 7.87	444.0 ± 13.6	0.654 ± < 0.01
2	18	0	0.080 ± 0.03	0.016 ± 0.02	0.826 ± 0.31	9.958 ± 0.84	2.180 ± 0.64	72.67 ± 6.87	400.5 ± 13.0	0.659 ± < 0.01
3	17	0	0.154 ± 0.06	0.072 ± 0.08	1.541 ± 0.49	7.111 ± 0.64	0.885 ± 0.35	52.30 ± 5.19	365.0 ± 12.3	0.663 ± < 0.01
5	3	12	1.012 ± 0.26	1.112 ± 0.45	6.856 ± 1.32	4.033 ± 0.37	0.285 ± 0.15	31.13 ± 3.39	323.0 ± 11.4	0.667 ± < 0.01
5	2	13	1.442 ± 0.38	1.619 ± 0.66	9.078 ± 1.76	3.910 ± 0.34	0.255 ± 0.14	30.81 ± 3.44	325.6 ± 11.5	0.667 ± < 0.01
5	1	14	2.429 ± 0.63	3.479 ± 1.22	13.01 ± 2.37	3.622 ± 0.30	0.226 ± 0.13	26.82 ± 2.87	326.5 ± 11.6	0.667 ± < 0.01
5	0	15	4.758 ± 1.09	7.635 ± 2.20	22.23 ± 3.58	2.642 ± 0.14	0.176 ± 0.12	21.43 ± 1.28	324.5 ± 11.9	0.667 ± < 0.01

Change in Elective Patient Procedure Time Mean and/or Variance

When the mean and variance of elective procedure time increase (Case 6), intuitively, aside from utilization which increases, the other performance measures deteriorate. More importantly, we find that in this case shifting from a policy of complete flexibility to high partial flexibility does *not* provide a Pareto improvement. Rather, we observe a trade-off: dedicating rooms to emergencies reduces their wait time at the expense of an increase in elective patient wait time (cf. Table 4). The reasons for this difference between the results in the base case and in Case 6 are as follows. Regarding electives, in the base case, the expected aggregate room utilization generated by the workload from electives alone is 73% (75 electives with expected duration of 93 minutes, given 20 operating rooms available for 480 minutes). When we increase the mean procedure time of electives to 139 minutes, the total expected workload from elective surgeries exceeds the total capacity available during the 8-hour shift. Therefore, contrary to the previous cases, the electives now suffer more from the lack of capacity than they suffer from emergency disruptions. Regarding emergencies, their wait time increases as a result of the reduced frequency of elective surgery completion due to the increased elective procedure time. Under this increased emergency wait time, emergencies now benefit from having rooms dedicated to them. Also, when shifting from complete flexibility to high partial flexibility (3/17/0), the relative improvement for emergencies is greater than the deterioration for electives, for whom only average and 75^{th} percentile of wait time increase. Meanwhile, shifting from a policy of complete focus (5/0/15) to low partial flexibility (5/3/12) has the same benefit as before, and for the same reason: it reduces the average wait time of emergencies, while only slightly increasing that of electives, because emergencies will only use the flexible rooms sporadically, thus only slightly affect electives negatively.

When only the variance of the elective procedure time increases (Case 7), average emergency wait time tends to *decrease*, when compared to the base case (Table 2), for the combinations with high partial flexibility (3/17/0) and complete flexibility (0/20/0). This would seem to be counterintuitive, because typically an increase in procedure time variance leads to an increase in average wait time (Hopp & Spearman, 2007). The reason for this result is that the long right-hand tail of the lognormal distribution leads to a greater increase in the number of short cases relative to the increase in long cases when the variance increases. This is compounded by the fact that emergencies can access any flexible room that becomes available. From this reasoning, we can expect that emergency wait time will decrease as the variance of procedure time further increases, because the frequency of this phenomenon will increase. This decrease does not happen for the combinations with low partial flexibility because emergencies only rarely use the flexible rooms and thus are not affected as much by changes to elective surgeries. Similarly, emergency wait time remains unchanged under the complete focus policy. Meanwhile, elective patient wait time increases, as expected.

When the mean and variance of elective procedure time decrease (Case 8), we find that shifting from complete flexibility and shifting from complete focus provide the same results as in the base case. This sensitivity analysis provides evidence that the base case results are robust to a decrease in elective procedure time. With high partial flexibility (3/17/0), we notice that emergency wait time is lower than in the base case, which is expected since emergencies can access a flexible room faster when elective surgeries are shorter.

Observations on the Benefits of Partial Flexibility

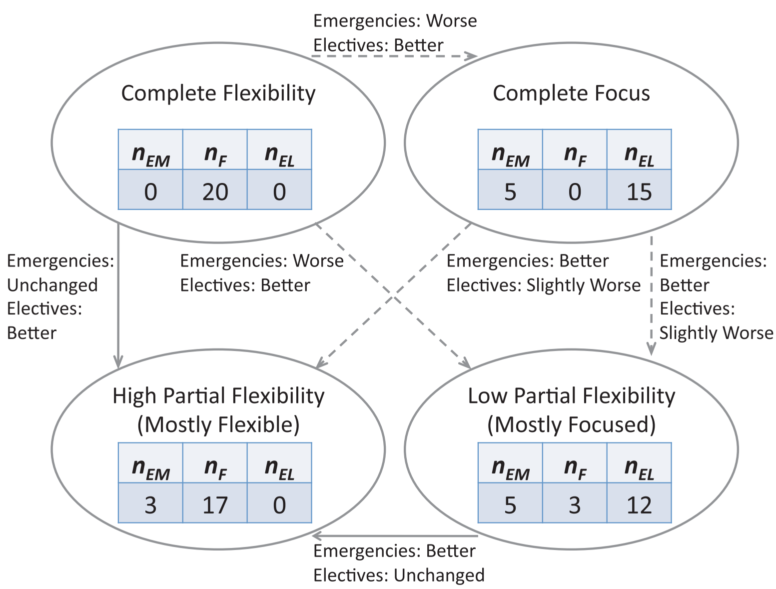
Based on our investigations, we formulate and comment on two observations that highlight how to achieve the benefits of partial flexibility. First, we define safety capacity as the total operating room time available, excluding overtime, minus the total expected operating time from the scheduled elective cases. We then have the following observations.

OBSERVATION 1: (High partial flexibility): If originally all operating rooms are flexible and there is sufficient safety capacity, then dedicating rooms to emergencies will improve the performance of the system; it will reduce the wait time of electives and will not increase the wait time of emergencies.

This observation is counter-intuitive as we may expect that dedicating rooms to emergencies would benefit emergencies at the expense of electives, yet the opposite result occurs. The reasons are as follows. For emergencies, having dedicated rooms does not provide much improvement because emergencies have higher priority in accessing flexible rooms not dedicated to them. At the same time, by dedicating rooms to emergencies, electives benefit more from the reduction of emergency disruptions than they suffer from the capacity reduction. Also, one may presume that emergency disruptions impede electives less than capacity reduction, since electives have access to less capacity all of the time with capacity reduction, while emergency disruptions imply a reduction in capacity for electives only on an as needed basis. However, since electives wait for surgery in individual queues, emergency arrivals become extremely disruptive for the electives waiting for the particular operating room where the disruption occurs. It is also counter-intuitive that dedicating rooms to electives does not improve the performance of the system. The reason is that electives wait time is decreased only if we dedicate a large number of rooms to them, which would not be viable because it would significantly increase the wait time of emergencies.

This reasoning has a caveat in that it only applies if there is sufficient safety capacity, so that at least one room could be dedicated to emergencies and the remaining rooms would still have safety capacity. This is indicated by the sensitivity analysis corresponding to an increase in elective patient procedure time mean and variance (Case 6). The first implication is that if there is not enough safety capacity to dedicate at least a room, then rooms should not be dedicated to emergencies as the reduction in capacity is more detrimental to electives than are disruptions from emergencies in this case. The second implication is that the number of rooms we can dedicate to emergencies is limited and is a function of how much safety capacity is available. This, by definition, depends on the number of rooms and the total expected workload of scheduled electives. In practice, the condition of sufficient safety capacity is not a strong limitation because it is unlikely that a hospital would decide to schedule more workload than it can accommodate with its capacity. More importantly, when there is safety capacity, dedicating even one

Figure 3: Comparing partial flexibility, complete flexibility and complete focus.



or two rooms to emergencies will improve the performance of the system, as long as emergencies can also access the rooms used for elective surgeries.

To summarize the trade-off between the capacity effect and the disruption effect, we can state that if utilization is high (i.e., low safety capacity), then the capacity effect is greater than the disruption effect on the performance of the system, and vice versa when utilization is low (i.e., high safety capacity). Characterizing "low" or 'high" would need to be determined by model experimentation or trial and error.

OBSERVATION 2: (Low partial flexibility): If originally all operating rooms are focused, then increasing flexibility will improve the performance of the system if the flexible rooms are taken from the set of rooms dedicated to electives; it will reduce the wait time of emergencies, while only slightly increasing the wait time of electives.

The flexible rooms clearly benefit emergencies, who now have access to more capacity than under the focused policy. This applies to the base case conditions as well as to the sensitivity analysis cases. Electives are only slightly negatively affected by this change because the presence of rooms dedicated to emergencies acts as a buffer against emergency disruptions, which implies that only a few emergency surgeries will be performed in the flexible rooms when all dedicated rooms are busy. As a result, if we dedicate an appropriate number of rooms to emergencies in the first place, making a few rooms flexible should leave electives almost unaffected, everything else being equal. The two observations and the effect of partial flexibility on elective and emergency wait time can be summarized in Figure 3 below.

CONCLUSIONS

With limited operating room capacity, the necessary handling of emergency surgeries with highest priority poses a recurrent challenge for hospital administrators striving to efficiently and effectively provide surgeries to scheduled elective patients. We survey several literature domains to identify the different ways that have been proposed to address this challenge, and subsequently investigate the potential benefits of partial flexibility using a stylized simulation model based on hospital data.

Our evaluation of partial flexibility indicates that it provides improvements on patient wait time and staff overtime over both completely flexible and completely focused policies. Given our findings, we formulate two observations that provide guidance to hospitals on how to implement partial flexibility: (i) high partial flexibility—when dedicating a few operating rooms, the rooms should be dedicated to emergencies, which will reduce the wait time of electives without increasing the wait time of emergencies; (ii) low partial flexibility—when making a few operating rooms flexible, the rooms made flexible should be taken from the set of rooms dedicated to electives, which will reduce the wait time of emergencies more than it will increase the wait time of electives. Overtime will be reduced with high partial flexibility and remain unchanged with low partial flexibility. By employing either form of partial flexibility, hospital managers can better handle the trade-off between responsiveness and efficiency. Determining the appropriate number of dedicated and flexible rooms can then be guided by considerating operating room capacity in relation to the expected workload from electives and emergencies, as these workload to capacity ratios give bounds on the minimum number of rooms emergencies and electives need access to, as well as the maximum number of rooms that can be dedicated to emergencies. As a starting point to set up high partial flexibility, hospital managers could dedicate to emergencies the number of rooms that corresponds to the available safety capacity, and then reduce this number if elective wait time and/or staff overtime is still too high. To set up low partial flexibility, hospital managers could start with making flexible one room that was dedicated to electives, then incrementally make additional rooms flexible if emergency wait time is still too high.

Our results suggest that the benefits of partial flexibility can be attained by more than one room allocation combination, and different combinations correspond to a different degree of capacity reduction and schedule disruption by emergencies. With partial flexibility, as one approaches complete dedication, the capacity reduction effects can be more significantly detrimental than the disruptions, while as one approaches complete flexibility the opposite is true. In addition, comparing the two forms of partial flexibility indicates that high partial flexibility tends to perform better than low partial flexibility: in general it provides much lower emergency wait time with only minor increases in elective wait time, total overtime, and room utilization. Hence, the use of low partial flexibility would need to be justified by lower operating cost or easier implementation and management. Otherwise, we recommend implementing high partial flexibility.

Furthermore, we find that the two observations on the benefits of partial flexibility hold whether the arrival rate of emergencies is stationary or not, and whether there are elective surgery cancellations or not. They are also robust to an increase in emergency arrivals, to a decrease in elective patient procedure time mean and variance, and the second observation holds when elective patient procedure time mean and variance increases. We did not test the effect of shorter or longer emergency procedures, because in practice hospitals do not have control over the type of emergency surgeries coming in, whereas they can affect the mix of long and short elective procedures through the scheduling decisions. Yet, we can expect that the longer and the more variable emergency surgeries are, the greater the negative effect these surgeries will have on electives when the operating rooms are shared, reinforcing the appeal of a partial flexibility policy.

Finally, in the light of the previous research that has considered partial flexibility, we concur that partial flexibility provides benefits. However, we point out an important distinction. When demand prioritization is not factored in, the main idea in the past literature is that partial flexibility can approach the results obtained with complete flexibility, but at a much lower cost (e.g., Jordan and Graves 1995). In our case, where patient prioritization is a key aspect, partial flexibility can provide better overall performance than complete flexibility. This result concurs with the finding from Ata and Van Mieghem (2009) who considered demand prioritization and found that when the system was serving a majority of low priority demand, as in our case, a combination of flexible and dedicated resources provided better responsiveness than did dedicated resources only.

We used our simulation model to generate a broad understanding of what type of room allocation combinations better handle the trade-off between responsiveness and efficiency, but this is not without limitations. In practice, it may be the case that switching between elective and emergency surgeries requires additional turnaround time, because of the increasing use of specialized equipment with complex setup. Our study is limited to the cases when this extra turnover time does not occur, and considering its effect would be an interesting future research topic because such extra setup time would make low partial flexibility a more appealing policy. Another limitation is that the lognormal distribution can have many cases close to zero, which may not be realistic. This limitation could be overcome with truncated lognormal distributions or by changing the location parameter. Accounting for the cost and the organizational implications for team cohesiveness of operating flexible or dedicated operating rooms would also be important future research. Studying the interactions between elective scheduling decisions and the assignment of specific surgery cases to each operating room is another interesting open research question, since the schedule of electives can affect emergency wait time. Specifically, we assumed no real time reassignment of electives to rooms to adapt to variable processing times, arrival times, and emergency disruptions. Developing dynamic prioritization schemes between emergencies and electives to account for the status of the system in real time could improve performance further. Research on analytical results with approximate flexibility measures could also guide resource allocation decisions in hospitals.

REFERENCES

Akçay, Y., Balakrishnan, A., & Xu, S. H. (2010). Dynamic assignment of flexible service resources. *Production and Operations Management*, *19*(3), 279–304.

Ata, B., &Van Mieghem, J. A. (2009). The value of partial resource pooling: Should a service network be integrated or product-focused? *Management Science*, *55*(1), 115–131.

Baker, K. R., Magazine, M. J., & Nuttle, H. L. W. (1986). The effect of commonality on safety stock in a simple inventory model. *Management Science*, *32*(8), 982–988.

Bassamboo, A., Randhawa, R. S., & Van Mieghem, J. A. (2010). Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Science*, *56*(8), 1285–1303.

Bassamboo, A., Randhawa, R. S., & Van Mieghem, J. A. (2012). A little flexibility is all you need: On the asymptotic value of flexible capacity in parallel queuing systems. *Operations Research*, *60*(6), 1423–1435.

Bhattacharyya, T., Vrahas, M. S., Morrison, S. M., Kim, E., Wiklund, R., Smith, R. M. et al. (2006) The value of the dedicated orthopaedic trauma operating room. *Journal of Trauma*, *60*(6), 1336–40.

Bowers, J., & Mould, G. (2004). Managing uncertainty in orthopaedic trauma theatres. *European Journal of Operational Research*, *154*(3), 599–608.

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*(3), 921–932.

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, *12*(4), 519–549.

Clark, K. (2009). Foreign lessons in hospital efficiency. *PRI's The World.* September 10.

Dexter, F., Macario, A., Traub, R. D., & Lubarsky, D. A. (2003). Operating room utilization alone is not an accurate metric for the allocation of operating room block time to individual surgeons with low caseloads. *Anesthesiology*, *98*(5), 1243–1249.

Ferrand, Y., Magazine, M. J., & Rao, U. (2010). Comparing two operating-room-allocation policies for elective and emergency surgeries. *Proceedings of the 2010 Winter Simulation Conference*. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc, 2364–2374.

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review* (March–April), 105–116.

Gerchak, Y., Gupta, D., & Henig, M. (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, *42*(3), 321–334.

Graman, G. A., & Magazine, M. J. (2002). A numerical analysis of capacitated postponement. *Production and Operations Management*, *11*(3), 340–357.

Graves, S. C., & Tomlin, B. T. (2003). Process flexibility in supply chains. *Management Science*, *49*(7), 907–919.

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operation theatre: A survey. *Health Care Management Science*, *14*, 89–114.

Gupta, D. (2007). Surgical suitesÕ operations management. *Production and Operations Management*, *16*(6), 689–700.

Gupta, D. (2013). Queueing models for healthcare operations. In B. T. Denton (Ed.), *International series in operations research & management science, Vol. 184. Handbook of healthcare operations management, methods and applications*. NY: Springer, 19–44.

Hand, R., Levin, P., & Stanziola, A. (1990). The causes of cancelled elective surgery. *American Journal of Medical Quality*, *5*(1), 2–6.

Hans, E., Wullink, G., Van Houdenhoven M., & Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*(3), 1038–1050.

Haraden, C., Nolan, T., Resar, R., Litvak, E. & Members of IHIs IMPACT Network and Pursuing Perfection initiatives. (2003). Optimizing patient flow: Moving patients smoothly through acute care settings. *IHI Innovation Series white paper*, Boston: Institute for Healthcare Improvement.

Hopp, W. J., & Spearman, M. L. (2007). *Factory physics* (3rd ed.). New York, NY: McGraw-Hill, Inc.

Hyer, N. L., Wemmerlöv, U., & Morris, J. A. Jr (2009). Performance analysis of a focused hospital unit: The case of an integrated trauma center. *Journal of Operations Management*, *27*(3), 203–219.

Jordan, W. C., & Graves, S. C. (1995). Principles on the benefits of manufacturing process flexibility. *Management Science*, *41*(4), 577–594.

Kelton, W. D., Sadowski, R., & Swets, N. (2009). *Simulation with arena* (5th ed.). New York, NY: McGraw-Hill Science/Engineering/Math.

Kulkarni, S. S., Magazine, M. J., & Raturi, A. S. (2004). Risk pooling advantages of manufacturing network configuration. *Production and Operations Management*, *13*(2), 186–199.

Lacqua, M. J., & Evans, J. T. (1994). Cancelled elective surgery: An evaluation. *The American Surgeon*, *60*(11), 809–811.

Lamiri, M., Xie X., & Zhang, S. (2008). Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions*, *40*, 838–852.

Law, A. M., & Kelton, W. D. (2000). *Simulation modeling and analysis* (3rd ed). New York, NY: McGraw-Hill Higher Education.

Li, Y., & Stein, W.E. (2008). A simulation analysis of dedicated emergency operating rooms. Working paper, Texas A&M, College Station, TX.

Macario, A., Vitez, T. S., Dunn, B., & McDonald, T. (1995) Where are the costs in perioperative care? Analysis of hospital costs and charges for inpatient surgical care. *Anesthesiology*, *83*(6), 1138–1144.

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

Marcon, E, & Dexter, F. (2006). Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science*, *9*:87–98.

May, J. H., Strum, D. P., & Vargas, L. G. (2000). Fitting the lognormal distribution to surgical procedure times. *Decision Sciences*, *31*(1), 129–148.

Persson, M. J., & Persson, J. A. (2010). Analysing management policies for operating room planning using simulation. *Health Care Management Science*, *13*(2), 182–191.

Pitts, S. R., Niska, R. W., Xu, J., & Burt, C. W. (2008). National Hospital Ambulatory Medical Care Survey: 2006 Emergency Department Summary. *U.S. Department of Health and Human Services*, 1–40.

Pham, D., & Klinkert, A. (2008). Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, *185*(3), 1011–1025.

Ritchie, J. (2011). Children's efficiency lauded. *Cincinnati Business Courier.* January 28.

Ryckman, F. C., Adler, E., Anneken, A. M., Bedinghaus, C. A., Clayton, P. J., Hays, K. R., et al. (2009). Cincinnati Childrens Hospital Medical Center: Redesigning perioperative flow using operations management tools to improve access and safety. In E. Litvak (Ed.), *Managing patient flow in hospitals: Strategies and solutions*, Oakbrook Terrace, IL: Joint Commission International, 97–111.

Smith-Daniels, V. L., Schweikhart, S. B., & Smith-Daniels, D. E. (1988). Capacity management in health care services: Review and future research directions. *Decision Sciences*, *19*(4), 889–919.

Tyler, D. C., Pasquariello, C. A., & Chen, C. H. (2003). Determining optimum operating room utilization. *Anesthesia & Analgesia*, *96*(4), 1114–1121.

VanBerkel, P., & Blake, J. (2007). A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Management Science*, *10*(4), 373–385.

Van der Lans, M., Hans, E. W., Hurink, J. L., Wullink, G., van Houdenhoven, M., & Kazemier, G. (2008). Anticipating emergency surgery in operating room departments. Working paper, Department of Operational Methods for Production and Logistics, University of Twente, Netherlands.

van Oostrum, J., van Houdenhoven, M., Hurink, J., Hans, E., Wullink, G., & Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, *30*(2), 355–374.

Wullink, G., Van Houdenhoven, M., Hans, E., van Oostrum, J., van der Lans, M., & Kazemier, G. (2007). Closing emergency operating rooms improves efficiency. *Journal of Medical Systems*, *31*, 543–546.

Zhang, B., Murali, P., Dessouky, M. M., & Belson, D. (2008). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, *60*(5), 663–673.

Zhou, J., & Dexter, F. (1998). Method to assist in the scheduling of add-on surgical cases-upper prediction bounds for surgical case durations based on the lognormal distribution. *Anesthesiology*, *89*(5), 1228–1232.

Zonderland, M. E., Boucherie, R. J., Litvak, N., & Vleggeert-Lankamp, C. L. (2010). Planning and scheduling of semi-urgent surgeries. *Health Care Management Science*, *13*, 256–267.

Yann B. Ferrand is an assistant professor of operations management in the Department of Management at Clemson University. He completed his Ph.D. in business administration in the Operations, Business Analytics, and Information Systems department, at the Lindner College of Business, University of Cincinnati (UC). He also holds a Master of Science in quantitative analysis and a Master of Business Administration from UC, and a Bachelor and Master degree in Business Administration from Ecole Supérieure de Commerce de Toulouse, France. His primary research interests are in providing innovative solutions in health care delivery services through better scheduling and resource allocation, handling uncertainty and variability using optimization and simulation methodologies. His research has been published in several international journals including IIE Transactions on Healthcare Systems Engineering, Interfaces, Research in Social and Administrative Pharmacy, and Economic Development Quarterly. He has worked on consulting projects with Cincinnati Children's Hospital, the Mayfield Clinic and Spine Institute, Cincinnati Health Department, and Shriners Hospital in Cincinnati.

Michael J. Magazine is currently professor of operations and business analytics and Ohio Eminent Scholar in the College of Business at the University of Cincinnati. He recently served as the Associate Dean for Faculty and Research and the Interim Dean in the College of Business at UC. He has degrees from CCNY, NYU, and the University of Florida. He has held teaching positions at North Carolina State University and at the University of Waterloo. In addition, he has had visiting appointments at PUC in Brazil, INRIA in France and Georgia Tech, MIT and the University of Michigan. His research interests include supply chain management, scheduling and other applications of manufacturing and health systems. He has served on the editorial boards of most of the major journals in management science and operations research, including Management Science, Operations Research and M&SOM. He is the coeditor of *Quantitative Models in Supply Chain Management*, and the winner of a best paper award by the Institute of Industrial Engineering in 2003. He is an INFORMS Fellow and a Fellow of the Graduate School of UC. He is the winner of the 2009 EXCEL Graduate Teaching Award. Besides teaching sports by the numbers and bracketology, he has worked on sports scheduling and analytics with several organizations.

Uday S. Rao is an associate professor of operations management in the operations, business analytics, and information systems department, at the Lindner College of Business, University of Cincinnati (UC). Before UC, he taught at Carnegie Mellon University, Cornell University, and the University of British Columbia. He has a B.Tech degree in mechanical engineering from the Indian Institute of Technology, Mumbai, and a PhD in operations research and industrial engineering from Cornell University. His research focuses on industry-motivated quantitative modeling and decision making under uncertainty applied to capacity planning, production and inventory control, supply chain configuration and coordination, telecom hub location, workforce staffing, and health care delivery improvement. His research has been published in *Decision Sciences*, *IIE Transactions, Interfaces*, *INFORMS Journal on Computing, Management Science, Manufacturing & Service Operations Management, Operations Research, and Production and Operations Management*. He has worked on consulting projects with Cincinnati Children's Hospital, Cincinnati and Hamilton County Fire Departments, Alien Technology (RFID Solutions Center), Caterpillar, Pittsburgh Plate and Glass, and Ford-Visteon.