

## **COMPARING TWO OPERATING-ROOM-ALLOCATION POLICIES FOR ELECTIVE AND EMERGENCY SURGERIES**

Yann Ferrand  
Michael Magazine  
Uday Rao

Department of Quantitative Analysis and Operations Management  
University of Cincinnati  
2925 Campus Green Drive, Cincinnati, OH 45221, USA

### **ABSTRACT**

When organizing the operating theatre and scheduling surgeries, hospitals face a trade-off between the need to be responsive to emergency cases and to conduct scheduled elective surgeries efficiently. We develop a simulation model to compare a flexible and a focused resource-allocation policy. We evaluate these two policies on patient and provider outcome measures, including patient wait time and physician overtime. We find that the focused policy results in lower elective wait time and lower overtime, which leads to the conclusion that electives benefit more from the elimination of emergency disruptions than what they lose from the reduced access to operating rooms. Emergency patient wait time, however, increases significantly as we shift from the flexible to the focused policy. The sensitivity analysis showed that average emergency wait time can decrease as the processing time variability increases. The trade-off between efficiency and responsiveness calls for additional research on other operating-room-allocation policies.

### **1 INTRODUCTION**

Hospitals oversee a vast amount of costly resources to provide healthcare services, and in the light of mounting healthcare costs, hospitals face constant pressure to manage their resources more efficiently. At the center of attention is the operating theatre, which is the major source of both cost and revenue for hospitals (Macario et al. 1995). The planning and scheduling of activities in operating rooms is a complex task that has fostered an extensive body of research, which has been reviewed periodically (Magerlein and Martin 1978; Blake and Carter 1997; Cardoen, Demeulemeester, and Belin 2009).

One long lasting difficulty in organizing the operating theatre is the presence of an intrinsic managerial conflict between unpredictable and urgent surgery cases, which demand hospitals responsiveness, and elective surgeries, which are scheduled to achieve a high level of efficiency, high resource utilization, short turn-around time, and short patient-wait time. The unpredictable nature of emergencies and the urgency of their needs not only destabilizes the organization of the operating theatre, but the induced variations also perturb the schedule of elective surgeries and more generally they can have a detrimental effect on healthcare quality, access, and cost.

In this paper we use simulation to evaluate patient flow in the operating rooms and investigate the effects of two distinct resource-allocation policies. Simulation is well suited for this environment where it is crucial to understand the impact of organizational changes on patient outcome before the changes are carried out. Using simulation to conduct such evaluations also benefits from being well regarded and fairly common in modern hospitals. Further, relevant approximations from other models, such as queuing theory, do not apply due to interactions between customer classes, log-normal uncertainty, and transience resulting from daily system renewal potentially using overtime.

A local hospital has considered a change in the way it handles scheduled elective surgeries and unscheduled emergency surgeries. Originally, the operating rooms were combined into one pool that would handle both types of surgeries. Emergency surgeries would have priority over elective surgeries and would need to be fit in between scheduled surgeries. This organization led to disruptions in the

schedule, resulting in elective patient wait time and staff overtime. To overcome this issue, hospital management decided to separate its surgical suites into two distinct subsets and dedicate hospital resources to each subset, which would enable the hospital to focus on ensuring an efficient handling of scheduled elective surgeries while providing responsiveness to emergencies. They also increased their total number of operating rooms. This initiative stems from the recommendations of Litvak (2003, 2009) to separate the streams of scheduled and unscheduled patients to reduce system variability and smooth patient flow. This change in resource-allocation policy essentially represents a reduction in flexibility as resources become dedicated to one of two types of surgeries.

The research question we address is whether the new focused policy allows the hospital to better handle the trade-off between efficiency and responsiveness than the former flexible policy, without increasing the number of operating rooms. To address this question, we study the performance of the focused policy when different number of operating rooms are dedicated to each type of surgery, compare the efficiency and responsiveness of the two policies, and study the sensitivity of the two policies to the volume of emergency surgeries and variance in surgery processing time. Specifically, some of the questions addressed include: For the focused policy with capacity allocated to patient segments, how many rooms should we dedicate to electives and to emergencies? Are elective surgery patients better off experiencing emergency disruptions due to a flexible (shared-capacity) policy or are they better off having exclusive access to less capacity via a focused policy? How does policy choice affect performance metrics such as waiting times in each patient segment, and operating room overtime?

We find that electives' average wait time is shorter with the new focused policy, while that of emergencies increases. This trade-off calls for additional research on other operating-room-allocation policies. The sensitivity analysis showed that average emergency wait time can decrease as the processing time variability increases.

Section 2 reviews pertinent literature, followed by a presentation of our methodology in section 3. We then report the findings from our computational results in section 4. Finally, we conclude and discuss possible future work in section 5.

## **2 LITERATURE REVIEW**

To tackle the question of how to effectively handle the interferences of emergencies with the schedule of elective surgeries, research has been conducted on capacity planning and management, on patient scheduling, and on the intersection of both fields. In capacity planning, there is no consensus on whether to dedicate rooms based on patient type. Besides Litvak (2003, 2009), others have studied the implications of using dedicated operating rooms. Li (2008) compares emergency patient wait time and staff overtime under different number of rooms dedicated to emergencies, but does not measure elective patient wait time. Bhattacharyya et al. (2006) dedicate one operating room to orthopaedic cases, which are less urgent, so as to ensure that fewer cases have to wait a long time for the higher priority cases to be completed. In doing so, they also find fewer occurrences of disruptions in the schedule of electives by more urgent cases. Bowers and Mould (2004) dedicate rooms to emergency surgeries but allow elective cases to be performed in these rooms under certain conditions. Other investigators have argued against operating room dedication and advocated flexibility. Wullink et al. (2007) conduct a simulation study to compare the performance of two policies for assigning slack time in a set of 12 surgery rooms. They report better room utilization, lower overtime and lower emergency patient wait time when slack time is inserted in all the rooms than when one room is dedicated to emergency cases. They do not account for elective patient wait time however, nor do they consider the possibility of dedicating more than one room to emergencies.

Research on patient scheduling also tackles questions related to the management of patients with multiple priorities. One question is how much slack should be inserted between elective surgeries, and where to insert it. Van der Lans et al. (2008) seek to find a sequence of elective surgeries that would minimize the amount of time an emergency would have to wait to find the next available surgery room. They note however that as the day elapses and emergencies arrive, the benefits of the optimized sequence are lost. Gupta (2007) review typical surgical-suite organizational problems, and alludes to the complication that emergency cases pose to the capacity-allocation and scheduling decisions. Gerchak, Gupta, and Hening (1996) consider a stochastic dynamic program to decide how many extra elective cases to schedule on a particular day to best serve demand from elective patients while avoiding turning away emergency cases. Lamiri et al. (2008) describe a stochastic-programming model to decide over a two-week period which elective surgery to do on each day to minimize the expected cost of performing electives and the cost of overtime, while emergencies arise randomly and

disrupt the schedule. Gupta and Wang (2008) also study the question of how to handle patients with different priorities in scheduling for a primary-care clinic to maximize revenue while considering patient preferences.

Finally, a stream of research has addressed the problem of managing patient demand for a medical resource, where patients have different priorities. Green, Savin, and Wang (2006) look for dynamic ways to choose the next patient to access the medical resource, given an existing schedule of elective outpatients and uncertain demand from inpatients and emergencies. Patrick, Puterman and Queyranne (2008) use approximate dynamic programming to devise an optimal schedule of patients with different priority levels and uncertainty in demand for the medical resource.

### **3 METHODOLOGY**

#### **3.1 Organization of the Operating Theatre**

The system we simulate is the operating theatre of a local hospital, composed of 20 operating rooms. Patients undergo surgery in one of the operating rooms and then move to a step-down unit, typically a bed in the post-anesthesia care recovery unit. We measure the effect of focused and flexible resource allocation policies on patient wait time and hospital-staff overtime. To avoid the potential effect of step-down unit unavailability, we assume that all patients can leave the operating room once the surgery is completed and we restrict our model to the flow of patients through the operating room.

We simulate the eight-hour operating shift, plus possible overtime. When the shift begins, the system is empty and idle, and once all scheduled electives are completed the shift ends. Under these circumstances, it is appropriate to model this system as a terminating simulation. It is also apparent that over the course of the operating shift, the tally and time persistent statistics we collect relate to the transient behavior of the system, making simulation an appropriate approach to evaluate the performance of the system.

The schedule of elective surgeries is outside the scope of our research, and we consider it as one of our input parameters. Electives are scheduled to arrive in batches with a fixed inter-arrival time during the course of the operating shift, the first batch arriving at time zero. Consistent with the actual setting, we assume that (i) there are no no-shows, and (ii) patients arrive on schedule. To keep our comparisons fair, we fix the number of electives scheduled. Emergencies arrive during the operating day according to a Poisson process, and we assume that effectively no emergencies arrive during the overtime period, because an alternative provider team and set of resources handle these emergencies.

#### **3.2 The Resource-allocation Policies**

Figure 1 provides a graphic representation of the assignment of patients to the operating rooms under the two policies. Each elective surgery is assigned to one of the operating rooms, consistent with the reality that elective surgery patients typically pick a specific provider for their surgery and each provider team is pre-assigned to a specific operating room. Upon arrival into the system, the elective patient has to wait in an individual queue corresponding to its assigned room if another patient occupies the room. Under the flexible policy, electives are distributed across all twenty operating rooms in a cyclical fashion: the first elective is assigned to the first operating room, the second elective to the second room, and similarly until the twentieth patient has been assigned to the twentieth room, at which point the following patient is assigned to the first room again. When an emergency patient arrives for surgery, they join a single queue where all emergencies wait on a first come first serve basis for the next operating room to become available. As a room becomes available, the first emergency in the queue is assigned to that room and the surgery is conducted there. Note that there could be elective patients waiting for that surgery room, and even if no elective is waiting, the next scheduled elective could find the room busy when it enters the system.

The hospital has considered avoiding these disruptions to the elective schedule, which increase elective-patient wait time and possibly induce staff overtime, by dedicating a subset of rooms to each patient type. Under this focused policy, elective and emergency patients are restricted to use only the rooms dedicated to them. We make no change to the elective schedule and the arrival pattern is exactly the same, but we assign these patients only to the elective dedicated rooms, thus assigning more patients per room. Note that it could be best to assign elective patients differently under the flexible and the focussed policies, but we did not make such a change to avoid confounding issues. Emergencies still wait in a single queue and go to the next available surgery room among those dedicated to emergencies. Under this policy, a decision has to be made as to how many rooms should

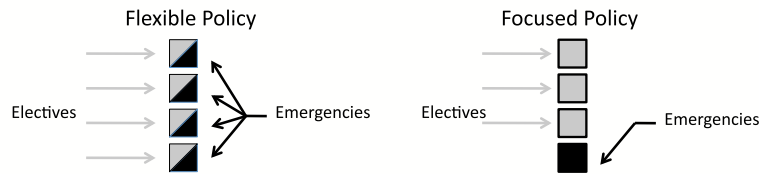


Figure 1: Assignment of patients to the operating rooms under the two policies.

be assigned to each patient segment. We test several combinations of rooms-to-segment assignment to identify the one that results in the best balance of elective and emergency wait time.

### 3.3 Analysis of Historical Data

To form our base case scenario, we analyzed 16 weeks of historical data on the number of cases per week for electives and emergencies, and the total workload per week at the local hospital. From these data we derived the average number of elective and emergency cases per eight-hour day, and average processing time per surgery. There was little variation in the number of cases and total workload across the 16 weeks, reinforcing the fact that these two averages were appropriate inputs in simulating the system and would be representative of the actual conditions of the system under study. We assumed that the arrival rate of emergencies was constant during the 8-hour shift, as the historical data did not indicate otherwise. Resulting from this analysis, we scheduled 75 elective surgeries and an average of 12 emergencies per day. The hourly arrival rate of emergencies is 1.5.

Given the 8-hour shift, the inter-arrival time for batches of electives is chosen so as to spread the electives throughout the shift. Since 75 electives are scheduled per day, we conducted our analyses with a batch size of 15 electives, allowing us to keep the batch size constant. An inter-arrival time of 90 minutes allows the 75 electives to be spread out throughout the 8-hour shift. Note that 90 minutes is also close to the average processing time of electives. We fixed the batch size at 15 to avoid confounding effects of elective batch scheduling and room allocation policy on elective patient wait times. Yet part of our investigation is to identify how many rooms should be dedicated to each patient type under the focused policy. The result may depend on the batch size of electives, and in practice the hospital may want to have a batch size that equals the number of rooms dedicated to elective surgeries. To account for this possibility, we simulate the focused policy with a batch size of 14 and 16, keeping the inter-arrival time at 90 minutes, and adjusting the size of the last batch so that the total adds up to 75 elective patients.

For the processing times, we used a lognormal distribution, which is consistent with other data from the hospital and with studies in the literature (e.g. Zhou and Dexter 1998). From the 16 weeks of data, the mean processing time we derived is 93 minutes and 125 minutes for elective and emergency surgeries respectively. Since we do not have data on the actual surgery time, we had no information on the variance of the processing times, and neither could we fit a distribution. Therefore, we used the same coefficients of variations as those in the study of Wullink et al. (2007). The processing times are assumed identically distributed, and all the surgery times are drawn from these lognormal distributions upon arrival of the entities in the system, and are all independent of each other.

### 3.4 Output Measures and Sensitivity Analysis

The simulation is modeled in Arena (version 10.0), using information from the reference Arena textbook by Kelton, Sadowski, and Swets (2009). After model verification, we did several validation test runs, such as running the simulation with a single patient, or increasing the number of patient arrivals, to make sure the simulation model was providing the expected results. Based on the half-width of the confidence intervals in initial runs, we found that we needed 400 replications to reduce the largest half-width of the 95% confidence intervals to values that were acceptable to the decision makers (half-widths were either under 5% of the estimated output value, or under 1 minute when the output value is small). To increase the precision of our comparisons, we use common random numbers by assigning a separate stream of random numbers to each source of randomness in the model, which are the arrival times of emergencies, and the surgery durations of electives and emergencies. For

each scenario, we collect average and maximum patient wait time (in queue) for both electives and emergencies.

Another important measure of wait time often used in hospitals and emphasized in the literature (Patrick and Puterman 2008) is the proportion of patients of a given priority class who wait longer than an established acceptable limit, which can vary based on patient priority. At the hospital that motivated this research, the maximum tolerable wait time prior to starting an emergency surgery is 30 minutes, for the most emergent cases. To this effect, we record the number of patients who wait more than 30 minutes, and we also get the average wait time for these patients only, to estimate how much over the 30 minutes mark their wait time is. Note that we also record this metric for elective surgeries as a way to compare the policies and evaluate their sensitivity. Aside from patient wait time, the surgeons and the hospital staff time are also critical, and because of processing-time variability and uncertainty in emergency arrivals, some surgeries have to be completed after the end of the eight-hour shift, accumulating overtime. To compute overtime, we record the amount of time a patient was in the system after the end of the shift and averaged it across all these patients. We also record the average number of overtime patients, which is the number of patients who were still the system after the end of the shift. This metric informs us on the magnitude of the consequences of the congestions due to uncertainties and variability throughout the shift. Finally, we record the utilization of each operating room during the 8-hour shift. Note that even though the simulation runs may have some overtime, we restrict the measurement of utilization to the 8-hour shift for two reasons. First, during overtime an operating room is busy. Second, it allows us to keep the denominator common when computing the utilization, thus allowing fair comparisons across scenarios.

To evaluate the sensitivity of the flexible and focused policies to the volume of emergencies and to the processing-time variability, we increase the number of emergencies from 1.5 to 2.25 per hour. We decided on this number by adding 3 standard deviations to the average of 12 emergencies that we obtained from the historical data. To assess the sensitivity to processing-time variability, we increase the coefficient of variation of both elective and emergency surgeries to 1.5, a process with relatively high variability. With both of these changes, we again search for the best allocation of rooms to each type of surgery under the focused policy, and then compare it to the flexible policy, and we compare the output of both policies to the output under the base-case input conditions.

## **4 COMPUTATIONAL RESULTS**

### **4.1 Focused Policy: How Many Rooms Should We Dedicate to Electives and to Emergencies?**

For the focused policy, we vary the number of dedicated rooms assigned to electives and to emergencies, and we run the simulation under the base case for processing time, number of scheduled electives, and arrival rate of emergencies. Five different allocations are tried, and when considering the average wait time for both types of patients, depicted by the solid line in Figure 2, it appears that dedicating 5 rooms to emergencies and 15 to electives gives a well-balanced result, at least under the assumption that electives are scheduled to arrive in batches of 15. We reach this conclusion because by assigning one more room to electives reduces the average wait time (half-width) from 20 (0.52) to 18 (0.48) minutes, while it increases the average wait of emergencies from 5 (1.09) to 15 (2.34) minutes. Similarly assigning one more room to emergencies results in an increase from 20 to 37 minutes for electives, while for emergencies the wait decreases from 5 to 1.5 minutes. To test the robustness of dedicating 5 rooms to emergencies and 15 to electives, we also ran experiments with elective arrivals in batches of 14 and 16 every 90 minutes, with the last batch suitably modified to keep the total number of electives scheduled to be constant at 75. In Figure 2 we use dashed lines to show the corresponding average elective wait times for 4, 5, and 6 Dedicated Emergency Operating Rooms (DEOR). While the emergency wait times are unaffected by the batching and only depends on the number of DEOR, the elective wait times change and this change is larger for 6 DEOR. Despite the change in elective wait time values, we observe that the 5 DEOR solution still appears to effectively balance elective and emergency wait times. This suggests that, under static elective batching, the number of DEOR to use may depend more on the workload imposed and the resulting resource utilizations than on the elective batch size, although a small change in the number of DEOR is possible with changes in elective batching. In fact, one can argue from a steady state queuing analysis that 4 DEOR would be required for our data to prevent explosion of emergency patients (which does not occur in the simulation even with 3 DEOR because the system uses overtime to clear all queues.) We also note that the elective batching may be effectively matched with the DEOR choice, in that for 6 DEOR with 14 rooms dedicated to electives we use Batch 14, for 5 DEOR we use Batch 15, and for 4 DEOR we

use Batch 16, yielding a slightly different (and improved) tradeoff curve. Within the realm of elective scheduling, the simulation modeling also permits exploration of the impact of other factors such as the interplay between processing times and patient interarrival times, with the latter adjusted during the day to incorporate emergencies.

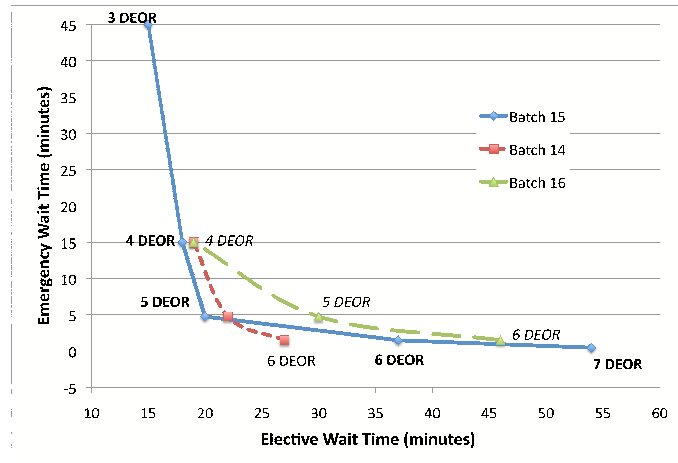


Figure 2: Average patient wait time under the focused policy with different number of rooms allocated to elective and emergency surgeries.

The other performance metrics, captured in Tables 1 and 2, show a similar pattern to the one depicted by average patient wait time. In particular we note that the average overtime is minimized with 5 rooms dedicated to emergencies and 15 to electives. Also, with fewer than 15 rooms dedicated to electives, the average number of electives who wait more than 30 minutes increases rapidly, as does the average wait time of these electives. Finally, note that for emergencies, some replications saw no emergency wait more than 30 minutes, thus reducing the average values. In Tables 1 and 2 all sensible comparisons show that when the values are different, the differences are statistically significant, given their respective half-widths.

Table 1: Overtime measures under the focused policy with different number of rooms allocated to elective and emergency surgeries.

Number of Rooms Dedicated to Emergencies	Average Number of Overtime Patients	Average Overtime	Maximum Overtime
7	19	78	659
6	15	70	659
5	12	58	659
4	11	62	659
3	11	79	703

#### 4.2 Are Electives Better Off Experiencing Emergency Disruptions or Having Access to Less Capacity?

For the elective patients, the focused policy amounts to a direct capacity reduction. The flexible policy indirectly reduces capacity by permitting emergency disruptions to the elective schedule. We seek to understand which capacity reduction is more detrimental. We do so by comparing the performance of the flexible and the focused policies to a hypothetical benchmark flexible policy with no emergencies. Aside from reducing the number of emergencies to zero in the benchmark flexible policy, we use the base case input conditions, and for the focused policy we assign 5 rooms for emergencies and 15 for electives, drawing on the findings from the previous comparisons. This benchmark flexible policy allows us to compare the performance deterioration of the flexible policy due to disruptions by emergencies to the performance deterioration due to the capacity reduction via the focused policy.

Table 2: Average number of patients whose wait was greater than 30 minutes and their average wait time, under the focused policy with different number of rooms allocated to elective and emergency surgeries.

Number of Rooms Dedicated to Emergencies	Elective (waiting more than 30 min.) Average Number	Elective (waiting more than 30 min.) Average Wait Time	Emergency (waiting more than 30 min.) Average Number	Emergency (waiting more than 30 min.) Average Wait Time
7	38	100	0.11	2
6	29	88	0.33	6
5	19	62	1	16
4	17	62	2	39
3	15	60	5	82

Typical results are summarized in Tables 3 - 5, and all sensible comparisons show that when the values are different, the differences are statistically significant, given their respective half-widths.

Table 3: Average and maximum patient wait time under the flexible and the focused policies, and with the hypothetical no-emergency scenario.

Policy Considered	Average Elective Wait	Average Emergency Wait	Maximum Elective Wait	Maximum Emergency Wait
Flexible - No Emergency	8	NA	220	NA
Flexible	27	0.28	776	22
Focused (5-15)	20	4.76	272	191

Table 4: Average and maximum overtime, and room utilization, under the flexible and the focused policies, and with the hypothetical no-emergency scenario.

Policy Considered	Average Number of Overtime Patients	Average Overtime	Maximum Overtime	Room Utilization (Min, Max)
Flexible - No Emergency	4	30	215	(0.58, 0.77)
Flexible	12	78	717	(0.61, 0.91)
Focused (5-15)	12	58	659	(0.24, 0.75) Emergency Rooms (0.91, 0.93) Elective Rooms

Considering patient wait time in Table 3, we observe the following. With the flexible policy, going from the hypothetical no-emergency scenario to the base-case scenario increases the average wait time of electives from 8 to 27 minutes. Now if we compare the flexible policy with no emergency to the focused policy, the number of rooms accessible to electives is reduced from 20 to 15 rooms, but the elective patients average wait time increases only to 20 minutes. Thus, from the perspective of elective wait time, in this case, it is better to use the focused policy, as the electives suffer less from the capacity reduction than they do from the emergency disruptions. The other metrics in Tables 4 and 5 lead to the same conclusion: shifting from the flexible policy under the hypothetical no-emergency scenario to the focused policy results in less overtime increase, lower maximum elective wait time, and fewer electives waiting more than 30 minutes than when shifting to the flexible policy under the base-case scenario. These results are noteworthy because emergency disruptions take capacity away from electives only when there are urgent emergencies whereas, by dedicating capacity to the emergencies, the static focused policy takes capacity away from electives whether there are emergencies in the system or not. Yet the fact that electives are not interrupted by emergencies in the focused policy allows electives to experience lower average wait times than in the flexible policy. A possible explanation for this result is that under the flexible policy, there is enough capacity to handle the total workload, as indicated by the utilization figures in Table 4 ranging from 0.61 to 0.91 per operating room. Once we implement the focused policy, the utilization of the rooms dedicated to electives indicates that there is still enough capacity devoted to them, as the utilization ranges from 0.91 to 0.93. From the emergency patients' perspective however, the focused policy gives less capacity to the emergencies than the flexible policy (in which emergencies can access all operating rooms), so the focused policy

Table 5: Average number of patients whose wait was greater than 30 minutes and their average wait time, under the focused policy with different number of rooms allocated to elective and emergency surgeries.

Policy Considered	Average Number of Electives	Average Wait Time of Electives	Average Number of Emergencies	Average Wait Time of Emergencies
Flexible - No Emergency	8	54	NA	NA
Flexible	19	98	0	0
Focused (5-15)	19	62	1	16

results in larger emergency patient average wait time. Moreover, in terms of operating room utilization, the focused policy generates more imbalance across the rooms, as the DEOR have low utilization, ranging from 0.24 to 0.75, whereas the rooms dedicated to electives all have utilization above 0.90.

**4.3 Is the Focused Policy Sufficiently Responsive to Emergencies?**

Since the hospital must be responsive to emergency surgeries, we need to assess how the reduction in capacity caused by the focused policy affects emergency patients wait time. From Table 3, with the flexible policy, emergencies experience virtually zero wait time, as anticipated given the input conditions: their arrival rate is 1.5 per hour, there are twenty operating rooms and emergencies can access the first room that becomes available. The maximum emergency wait time of 22 minutes shows the responsiveness is robust to the actual processing-time variations and unpredictability in emergency arrivals. With the focused policy and five rooms dedicated to emergencies, average emergency wait time is just under 5 minutes, which is an acceptable level of responsiveness, but the maximum emergency wait time is 191 minutes. Moreover, the number of emergencies who wait more than 30 minutes is small enough that it does not occur in every replication. In fact, the proportion of replications when no emergency wait more than 30 minutes is 0.73. On average, 3.7 emergencies waited more than 30 minutes, or 30% of the expected 12 emergencies per shift, and they waited an average of 59 minutes. These results call for possible relaxation of the strict dedication in the focused policy, thereby allowing an emergency to use a room dedicated to electives if the wait time reaches a certain threshold.

The trade-off between efficiency and responsiveness is now apparent, as the focused policy is more efficient because of the smaller average and maximum elective-patient wait time, and shorter overtime, but is also less responsive to emergencies.

**4.4 How sensitive are these policies to an increase in emergency surgeries?**

As we increase the arrival rate of emergencies from 1.5 to 2.25 per hour, we observe the following, based on the output depicted in Tables 6 and 7, where all sensible comparisons show that when the values are different, the differences are statistically significant, given their respective half-widths. First, as indicated in Table 6, the allocation of 5 rooms to emergencies and 15 to electives under the focused policy still balances well the wait time of emergency and elective patients. Dedicating fewer than 5 rooms to emergencies would not make sense since we previously found that it only provided a minor decrease in elective patient wait time (see Table 1), and the system now needs to handle more emergencies. Dedicating 6 rooms to emergencies reduces emergency wait time, at the expense of elective wait time. Second, comparing the results of the flexible policy given the increased emergency arrival rate to those given the base-case input conditions, displayed in Table 7, we observe an increase in average elective wait time but almost none in average emergency wait time, even though the small difference between 0.28 and 0.73 is statistically significant. Lastly, the elective wait time with the flexible policy is similar to that with the focused policy and 6 rooms allocated to emergencies and 14 to electives. The maximum elective wait time, however, is much lower with the focused policy than with the flexible one, and so is the average overtime. Thus, the flexible policy is still the most responsive to emergencies, at the expense of efficiency. Moreover, given the extremely large maximum elective wait times with the flexible policy, it may be preferable to adopt the focused policy and possibly allow an emergency to use an elective room and disrupt their schedule if the wait time reaches a threshold.



Table 6: Average and maximum patient wait time and overtime under the focused policy with different number of rooms allocated to elective and emergency surgeries and increased emergency arrival rate.

Number of Rooms Dedicated to Emergencies	Average Elective Wait	Average Emergency Wait	Maximum Elective Wait	Maximum Emergency Wait	Average # of Overtime Patients	Average Over-time	Maximum Over-time
7	54	3	414	158	21	80	1504
6	37	9	378	242	17	75	1521
5	20	22	272	368	14	72	1565

Table 7: Comparison of the flexible policy under the base case and with an increased emergency arrival rate.

Policy Considered	Average Elective Wait	Average Emergency Wait	Maximum Elective Wait	Maximum Emergency Wait	Average # of Overtime Patients	Average Over-time	Maximum Over-time
Flexible	38	0.73	1559	44	17	88	1567
Flexible (Base Case)	27	0.28	776	22	19	78	717

**4.5 How sensitive are these two policies to an increase in processing-time variance?**

As we increase the processing-time variance of elective and emergency surgeries to get coefficients of variation of 1.5, we make the following observations. First, with the focused policy, the pattern of elective and emergency patient wait time is similar to the pattern we observed under the base case. The allocation of 5 rooms to emergencies and 15 to electives still provides the best balance of the two patient wait times. And not surprisingly, average elective patient wait time and average overtime increase. Given the maximum values, it appears that the system would have to be changed in order to accommodate such high variability in procedure time.

In addition, comparing the focused and the flexible policy leads to a different conclusion than under the base case. We find that the flexible policy provides lower average wait times for elective and emergency surgeries, while the other performance measures do not deteriorate. This suggests that, as expected, the flexible policy is better suited for situations with higher variability.

Finally, comparing these results to those under the base case provides an unexpected result. Indeed it has been established that increasing processing time variability will result in poorer system performance in terms of throughput, work in progress, and entity wait time (Hopp and Spearman, 2008). This result, however, does not account for entities with distinct priority, and we observe in Table 8 that as the processing-time variances of elective and emergency surgeries increase, the average wait time of elective patients increases but that of emergencies decreases. Albeit small, the decreases in emergency patient average wait time from 0.28 to 0.05 minute with the flexible policy and from 4.76 to 2.88 minutes with the focused policy are both statistically significant according to the confidence intervals' half-widths. Thus, emergency patients benefit from higher processing-time variance, since their wait time decreases, while elective patients see their wait time increase. One possible explanation for this observation is that with increased processing time variance, there are more shorter duration surgeries and a few very long duration surgeries and emergencies can exploit the shorter elective surgery durations to obtain quicker response from the system.

Table 8: Impact of increased processing time variability on patient wait time.

Policy Considered	Average Elective Wait	Average Emergency Wait	Maximum Elective Wait	Maximum Emergency Wait
Flexible	55	0.05	5101	18
Focused (15-5)	64	2.88	4902	205
Flexible (Base Case)	27	0.28	776	22
Focused (Base Case)	20	4.76	272	191

## 5 CONCLUSIONS AND FUTURE WORK

We have compared two distinct resource-allocation policies to handle the flow of scheduled elective surgeries and unpredictable emergency surgeries. The flexible policy is the historical way of assigning patients to operating rooms. Under the flexible policy, emergency surgeries access any operating room and have priority over electives. The new policy provides the surgeon with more focus as the operating rooms are divided into two subsets and patients access the subset of rooms that corresponds to their type, either elective or emergency.

Comparing these two policies in terms of patient wait time and hospital-personnel overtime shows that the focused policy results in lower elective wait time and lower overtime, except in situations with high processing time variability. This leads to the conclusion that typically electives benefit more from the elimination of emergency disruptions than what they lose from the reduced access to operating rooms. Emergency patient wait time, however, increases significantly as we shift from the flexible to the focused policy, and in particular an average of 30% of the emergencies wait more than their target wait time of 30 minutes. Interestingly, average emergency wait time can decrease as the processing time variability increases, which implies that hospital management can be more responsive to emergencies by having a mix of long and short elective procedures.

These results call for further investigation of what would constitute an ideal resource-allocation policy. Specifically, we could investigate the effect of relaxing the focused policy by allowing electives and emergencies to access a room from the other dedicated subset if their wait time reaches their maximum tolerable wait time. We could also consider combining the two policies and have a mix of rooms dedicated to a patient type and rooms that could handle both electives and emergencies.

In addition, more extensive designs of experiments need to be conducted to fully understand the effect of all the factors involved, both those under and beyond managerial control. Particular attention needs to be given to elective scheduling approaches, and to the design of elective schedules that are well suited for the room-allocation policy. Ultimately, such an effort would yield principles that would allow us to draw conclusions in a practical setting. Yet another possible way to help provide guidance to practitioners would be to develop queueing approximations that would take into account the particularities and the transient nature of this hospital setting.

## ACKNOWLEDGMENTS

We would like to thank Dr. David Kelton for his review of an earlier version of this paper. His input has greatly improved the quality of this research.

## REFERENCES

- Bhattacharyya, T., M.S. Vrahas, S.M. Morrison, E. Kim, R.A. Wiklund, R.M. Smith, and H.E. Rubash. 2006. The value of the dedicated orthopaedic trauma operating room. *The Journal of TRAUMA Injury, Infection, and Critical Care* 60 (6): 1336-1341.
- Blake, J.T., and M.W. Carter. 1997. Surgical process scheduling: a structured review. *Journal of Health Systems* 5 (3): 17-30.
- Bowers, J., and G. Mould. 2004. Managing uncertainty in orthopaedic trauma theatres. *European Journal of Operational Research* 154: 599-608.
- Cardoen, B., E. Demeulemeester, and J. Belin. 2009. Operating room planning and scheduling: a literature review. *European Journal of Operational Research* 201(3) : 921-932.
- Gerchak Y., D. Gupta, and M. Hening. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science* 42 (3): 321-334.
- Green, L., S. Savin, and B. Wang. 2006. Managing patient demand in a diagnostic medical facility. *Operations Research* 54: 11-25.
- Gupta, D. 2007. Surgical suites operations management. *Production and Operations Management* 16 (6): 689-700.
- Gupta, D., and L. Wang. 2008. Revenue management for a primary care clinic the presence of patient choice. *Operations Research* 56 (3): 576-592.
- Hopps, W. J., and M. L. Spearman. 2008. *Factory physics*, 3rd ed. New York: McGraw-Hill, Inc.
- Kelton, D., R. Sadowski, and N. Swets. 2009. *Simulation With Arena*. 5th Ed. New York: McGraw-Hill.

- Lamiri, M., X. Xie, A. Dolgui, and F. Grimaud. 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research* 185 (3): 1026-1037.
- Litvak, E. 2003. *Optimizing patient flow: moving patients smoothly through acute care settings*, Institute for Healthcare Improvement, Cambridge, Massachusetts.
- Litvak, E. 2009. *Managing patient flow in hospitals: strategies and solutions*. 2nd Ed. E-Book. Joint Commission Resources.
- Li, Y., and W.E. Stein. 2008. A simulation analysis of dedicated emergency operating rooms. Working Paper, Texas A & M, Texas.
- Macario, A., T.S. Vitez, B. Dunn, and T. McDonald. 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.M., and J.B. Martin. 1978. Surgical demand scheduling: a review. *Health Services Research* 13: 418-433.
- Patrick, J., and M.L. Puterman. 2008. Reducing wait times through operations research: optimizing the use of surge capacity. *Healthcare Policy* 3(3): 75-88.
- Patrick, J., and M. Puterman and M. Queyranne. 2008. Dynamic Multi-Priority Patient Scheduling for a Diagnostic Resource. *Operations Research* 56 (6): 1507-1525.
- Van der Lans, M., E.W. Hans, J.L. Hurink, G. Wullink, M. van Houdenhoven, and G. Kazemier. Anticipating emergency surgery in operating room departments. Working Paper, Department of Operational Methods for Production and Logistics, University of Twente, Netherlands.
- Wullink G., M. Van Houdenhoven, E. W. Hans, J. M., Van Oostrum, M. Van der Lans, and G. Kazemier. 2007. Closing Emergency Operating Rooms Improves Efficiency, *Journal of Medical Systems* 31: 543-546.
- Zhou, J and F. Dexter. 1998. Method to Assist in the Scheduling of Add-on Surgical Cases-Upper Prediction Bounds for Surgical Case Durations Based on the Log-normal Distribution. *Anesthesiology* 89 (5): 1228-1232.

#### AUTHOR BIOGRAPHIES

**YANN FERRAND** is a PhD candidate in Operations Management in the College of Business at the University of Cincinnati. His primary research interests include health care operations and management. He has been working with Cincinnati Children Hospital Medical Center and the Mayfield Clinic & Spine Institute in Cincinnati on scheduling and patient flow projects. He has also worked and published papers on several economic studies related to the pharmaceutical industry. His email address is <[ferrany@mail.uc.edu](mailto:ferrany@mail.uc.edu)>.

**MICHAEL MAGAZINE** is Professor and Ohio Eminent Scholar of Quantitative Analysis and Operations Management. His research interests include scheduling, supply chain management and other applications of manufacturing and health systems. He has worked on the design and analysis of heuristics and applications in the production and manufacturing area and has been the holder or coholder of several research grants in this area. He has published over fifty papers in these areas. He is co-editor of *Quantitative Models for Supply Chain Management*. His email address is <[magazim@ucmail.uc.edu](mailto:magazim@ucmail.uc.edu)>.

**UDAY RAO** is an Associate Professor of Operations Management in the College of Business at the University of Cincinnati. Dr. Rao's current research interests are in quantitative models for capacity decisions (risk-based strategic planning, tactical resource allocation, and real-time utilization), periodic production systems (cyclic scheduling), remanufacturing, inventory control logistics (under uncertainty), supply chain optimization (design and coordination), telecom network planning (broadband-access), workforce planning, simulation, Radio Frequency Identification (RFID) Tags, Intelligent Agents, and Patient Flow Analysis. His email address is <[raous@ucmail.uc.edu](mailto:raous@ucmail.uc.edu)>.