Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

HEURISTICS FOR BALANCING OPERATING ROOM AND POST-ANESTHESIA RESOURCES UNDER UNCERTAINTY

Jill H. Iser
Brian T. Denton
Russell E. King

Edward P. Fitts Department of Industrial & Systems Engineering
North Carolina State University
Raleigh, NC 27695, U.S.A.

ABSTRACT

The Post-Anesthesia Care Unit (PACU) is a shared resource in the hospital where patients recover from surgery. It is fed by a set of Operating Rooms (OR's) often spanning several surgical services. It is insufficient to determine the best surgery schedule for any single OR without considering available PACU capacity. We model this as a two-stage process where the first stage is surgery and the second, post-anesthesia recovery. An interesting aspect of the second-stage process is that it begins as soon as the first stage has concluded even if a PACU bed is not available. In this case, the OR continues to house the recovering patient until a PACU bed is available. We analyze the structure of the problem, evaluate several heuristics based on competing performance measures for surgical suite efficiency, and present results of numerical experiments and insights that can be derived from them.

1 INTRODUCTION

With healthcare costs rising in the US, hospitals are anxiously pursuing ways to reduce costs. Surgery is also a crucial part of the healthcare system, as it accounts for more than 40% of hospital revenues (HFMA 2005). Developing efficient schedules for surgical suite resources can be difficult due to variability of the length of surgical procedures and anesthesia recovery times. In addition, the types of surgeries performed changes from day to day.

We consider the problem of finding a scheduling heuristic to minimize both OR overtime and PACU bed staffing hours required for the day. We consider some heuristics based on the experiments of Marcon and Dexter (2003) including the Random, Johnson's, MIX and HIHD heuristics. The Random heuristic sequences surgery cases (or cases for short) in the order they are generated. Johnson's is based upon Johnson's Scheduling Rule (Marcon 2006). MIX and HIHD use expected OR surgery times to sequence cases. MIX sequences the shortest case first, then the longest case, second shortest, second longest, and so on alternating between short and long cases. HIHD sequences the longest case first, the second longest case last, and works its way inward from the ends of the schedule as cases get shorter.

In addition, we propose two new heuristics that consider the effects of surgery scheduling on the downstream PACU resource. The first heuristic, referred to as Alternating Johnson's, staggers the inflow of cases to the PACU by using Johnson's Rule for odd-numbered ORs and the opposite of the rule for even-numbered ORs (see Section 5 for a more detailed description of all the heuristics). Thus, it tends to generate schedules with a diverse set of case completion times such that patients flow more smoothly from the ORs to PACU.

The second heuristic, referred to as block time minimization (BTM), tries to minimize blocking time by analyzing the schedule iteratively each time a case is added. Once a case is added to the schedule it is fixed. However, all remaining unscheduled patient surgeries in each OR are compared before selecting the next patient to add. A patient is added to each OR, in this manner, round-robin style.

We compare the performance of each heuristic based on two criteria: OR overtime and PACU hours used. The ORs are assumed to be available for a fixed 8 hour block. OR overtime is the sum of the time each OR is open past the 8-hour block allotted. The OR isn't considered closed until the turnover is complete after the last case scheduled for that particular OR (*turnover* is the time it takes to clean up from a previous case and set up for the next case). The OR is unavailable for surgery during this time, though the patient has been released for recovery. We assume that all patients scheduled on a given day will eventually have their surgery that day in their designated OR, i.e., there are no case cancellations. We also assume that patients don't arrive until the time their surgery is scheduled to start.

978-1-4244-2708-6/08/$25.00 ©2008 IEEE

We tested 500 sets of cases using each of the six heuristics, with 10,000 random variations of each resulting schedule, and collected statistics including the number of OR hours and PACU bed hours used to complete each schedule, the amount of overtime that resulted, the percent of days that experienced PACU blocking, and how long this blocking lasted.

We found that different heuristics seem to dominate when considering different objectives. However, the BTM heuristic was, surprisingly, often worse than the other much simpler heuristics. This indicates that PACU blocking, though it is often seen as an impediment in the OR, has some benefits, and therefore should not be minimized in all cases.

In this paper we include a review of some relevant literature, describe the problem in question in more detail, describe our discrete event simulation model for testing these heuristics, and compare and contrast the heuristic performances. Our numerical experiments are largely based on the design of Marcon and Dexter (2006).

2 LITERATURE REVIEW

We review two related bodies of work. The first is the scheduling literature on two-stage blocking problems. The second is surgery scheduling literature. For the latter we focus on papers that consider recovery resources within a surgical suite.

2.1 Blocking Problems

We first distinguish between blocking problems (which we consider here) and no-wait problems. A no-wait problem requires that entities (in our case patients) do not wait between stages. This concept is important when considering chemical and other processes where timing is crucial and production of a product cannot be stopped from the time it starts until it finishes. Blocking is different in that pausing between processes is allowed. However, there is no buffer space between processes to release the current resource(s) for the production of the next piece. Blocking is consistent with surgery scheduling where the patient stays in the OR until a PACU bed is available.

Hall and Sriskandaraja (1996) thoroughly review known problems considering either blocking or no-wait assumptions. They consider numerous combinations of machine configurations, objective functions, and other restrictions and classify the complexity of the problems (if known). They investigate in greater detail a few of the configurations. One that has received attention (both the deterministic and stochastic cases) is the $F_m|blocking|C_{max}$ problem. This considers *m* machines in a flowshop (or sequential) setup, blocking is allowed, and has the objective of trying to minimize the largest completion time. This is close to our problem, in that we would describe the OR-PACU relationship as a two-stage flowshop; as we have mentioned previously, we are allowing blocking; and we can compare C_{max} to the time that the last patient leaves a PACU bed. However, the difference is that we consider a surgical suite consisting of several ORs that operate in parallel, though their jobs cannot be interchanged. This is more similarly related to the $S_2|no\text{-}wait|C_{max}$ problem which denotes a two-stage process shop where each job requires two operations, each restricted to a set of machines. In our case the first-stage set consists of only one OR, and the second-stage set consists of all PACU beds. The authors point out that this problem is currently solvable if only one machine is available at both stages, but is strongly NP-complete if either stage has two or more machines.

2.2 Surgery-specific problems

Gupta (2007) presents a general overview of the types of problems that arise in healthcare. One problem in particular, the booking horizon problem, is aimed at predicting staffing needs based on case scheduling. He points out that models can be used to gain insight to a problem to help create insightful algorithms that can then be tested in a detailed discrete even simulation. He also discusses a model for surgery sequencing. The model focuses on minimizing waiting times, OR idling times, and tardiness. The author points out that this problem is combinatorially hard, though they do discuss a solution that considers a single OR at a time with only two cases to be scheduled. Additionally, they propose that cases with smaller variances should be scheduled first with the intent of having the least effect on subsequent cases.

Hsu et al. (2003) consider scheduling with blocking in multi-stage processes focusing more specifically on the no-wait version of the problem. The authors consider a Mixed Integer Program for an Ambulatory Surgical Center with the objective to minimize the number of nurses needed in PACU recovery. One limitation to their analysis is that they have only applied this using a deterministic approach, though it is well established that there is actually a large amount of variation in the process (Denton 2007). Marcon et al. (2003) develop a discrete event simulation model to determine the number of resources needed for a surgical suite. They focus on staffed PACU beds and porters (who transport patients from anesthesia recovery to inpatient units in the hospital or to checkout of the hospital). They find that the number of porters staffed has more of an impact on the system performance than PACU length of stay. We can draw a comparison between the function of the porters and downstream in-patient units (where patients staying overnight are transferred to complete their overall recovery) to conclude that if there isn't space/availability in processes downstream of PACU (i.e. in-patient units, etc.) they will have the same effect as a shortage of porters. Additionally, they found that the shorter the surgical procedure was, the higher the ratio of PACU beds to ORs required to accommodate flow.

The aforementioned do not consider the stochastic two-stage blocking problems we study in this article. Marcon and Dexter (2006) test the performance of several simple heuristics for a similar problem. Their heuristics focus on sequencing of cases in a single OR. We extend this by designing and evaluating heuristics that consider blocking and the effect of limited PACU resources directly with the goal of coordinating the sequencing of cases across multiple ORs. Our model includes variation from the expected times at every processing stage (surgery, recovery, and turnover).

3 PROBLEM DESCRIPTION

We focus on the relationship between the ORs and the PACU. Each surgeon has a list of cases for the day that must be performed. We assume that a surgeon is assigned to a given OR for the entire day and only one surgeon is assigned to any given OR. After surgery, patients from all ORs go to the same PACU area and compete for the next available bed (see Figure 1). PACU capacity is dependent on two factors: (1) the number of physical beds available and (2) the number of nurses staffed in the PACU. In our model we are only concerned with the number of physically available beds and how long each is open. This is reasonable since PACU hours are directly related to staffing needs and costs. If all beds in the PACU are full, or no staffed beds are available, a patient completing surgery begins anesthesia recovery in their OR. This blocks the OR, delaying turnover and the start of the next scheduled surgery on the case list. The patient either completes recovery in the OR or moves to a PACU bed when one becomes available for their remaining recovery time. The effects of blocking are demonstrated in Figure 2. The first instance of blocking occurs when both PACU beds are occupied, and case 7's surgery is complete. This patient begins its recovery in the OR until a PACU bed becomes available, and then spends the remainder of its recovery time in PACU (where it also eventually contributes to blocking case 8).

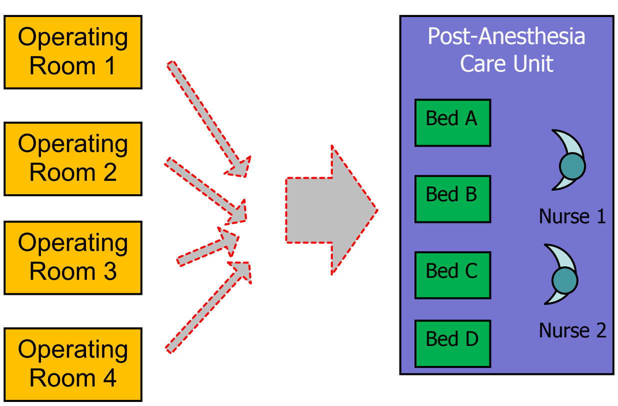

Figure 1: OR-PACU Flow

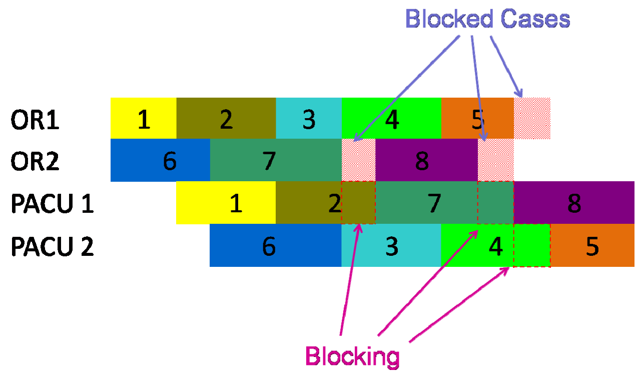

Figure 2: An example of a schedule that illustrates blocking in the OR: If surgical cases complete, and no PACU beds are available to take the patient for recovery this causes the completed case in the OR to be blocked by the cases currently being recovered in the PACU.

As is common among hospitals, surgeons or groups of surgeons (called surgical services) are assigned blocks (not to be confused with blocking) of OR time that may repeat only on a weekly and/or monthly basis. Therefore case lists, i.e. number of cases and mix of case types can be very different on a day to day basis. Even if the planned cases for the day are known, individual cases have variable surgery lengths, and patients have variable anesthesia recovery rates. For example, total knee replacement surgery is considered to be among procedures that have a more predictable surgery duration. Figure 3 illustrates the empirical probability density functions for surgery times and recovery times for a total knee replacement demonstrating considerable uncertainty. Results are based on a single surgeon during a 1-year period.

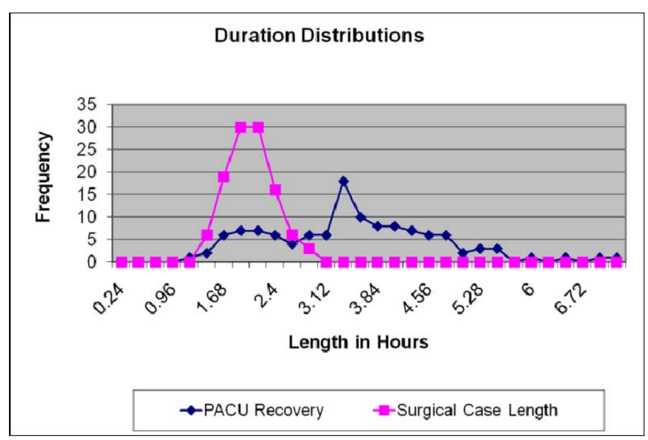

Figure 3. Total Knee Replacement surgery and anesthesia recovery time distribution for a single surgeon at a North Carolina Teaching Hospital.

There are several things a decision-maker may consider when it comes to evaluating a particular schedule.

We consider the costs associated with staffing additional nurses to make additional PACU resources available. Also, we evaluate the cost of those nurses per hour that they are employed. We also consider the variable cost of OR overtime.

Other performance measures that we explore include the total number of beds used. This gives a general idea of how many beds we would like to allow space for in the PACU in an original design. We also collect statistics on the total time the OR's are open. This indicates if the current block length is sufficient, or if we should consider extending the day on a regular basis to prevent recurring overtime costs. One last statistic we collect is the percentage of days the PACU is "on hold." This is something the teams in the OR use to describe when any OR is blocked because all PACU beds are currently full. This is perceived as a negative by the surgical personnel, so we are keeping track of how many days we experience this discomfort.

There are several possible objectives, and tradeoffs to be considered. We focus on minimizing a schedule's expected resulting cost, as measured by the total OR overtime and the sum of the PACU nurse-hours.

4 DESCRIPTION OF HEURISTICS

Each of the heuristics overviewed in section 1 are now described in more detail. The input to each heuristic is a set of cases assigned to each OR (generation of this data is described in the next section). Let n is the number of cases assigned to a given OR, p_{jk} is the expected stage k time for case j, U is a set representing the unsequenced cases, and S is the case sequence. The cases in each OR are sequenced according to each of the following algorithms.

4.1 Random

For the Random heuristic, we use the original randomly generated case order.

4.2 Johnson's

Johnson's rule is applied directly based on the assumption that each OR has a dedicated PACU bed. The expected surgery time for each case represents the first stage processing time while the expected recovery time represents the second stage time. Each OR is sequenced independently.

4.3 MIXed OR time (MIX)

$$Step\ 0: U = \{1,2,3,\ldots n\}, S = \emptyset, k = 1$$
$$Step\ 1: t \leftarrow \underset{j \in U}{\operatorname{argmin}}\, p_{j1}, S_{[k]} = t, U \setminus \{t\}, k = k + 1$$
$$Step\ 2: t \leftarrow \underset{j \in U}{\operatorname{argmax}}\, p_{j1}, S_{[k]} = t, U \setminus \{t\}, k = k + 1$$

Repeat steps 1 and 2 until $U = \emptyset$.

4.4 Half Increase in OR time and Half Decrease in OR time (HIHD)

$$Step\ 0: U = \{1,2,3,\ldots n\}, S = \emptyset, k = 1, m = n$$
$$Step\ 1: t \leftarrow \underset{j \in U}{\operatorname{argmax}}\, p_{j1}, S_{[k]} = t, U \setminus \{t\}, k = k + 1$$,

increment i

$$Step\ 2: t \leftarrow \underset{j \in U}{\operatorname{argmax}}\, p_{j1}, S_{[m]} = t, U \setminus \{t\}, m = m - 1$$

Repeat steps 1 and 2 until $U = \emptyset$.

4.5 Alternating Johnson's

For odd-numbered OR's we use the Johnson's Rule described in section 5.2. For even-numbered OR's we use the opposite logic: if the shortest processing time is a surgery time, we put it last; if it is a recovery time, we put it first.

4.6 Block Time Minimization (BTM)

The BTM heuristic tries to minimize blocking time and, as a secondary objective (if there is no blocking), to spread out surgical case end time and therefore maximize the time between arrivals to the PACU. It then evaluates the expected blocking at each step of the schedule-building process (see Figure 4). First, we assign the shortest surgery to OR 1. Then we assign the longest case to OR 2. This maximizes the time between the case end times (this meets the secondary objective, because with this configuration, we have the same number of PACU beds as OR's, so there will never be blocking in the first cases assigned to each OR). Then for OR 3 (and consecutive ORs) we choose the case that maximizes the time between expected end times among all 3 ORs.

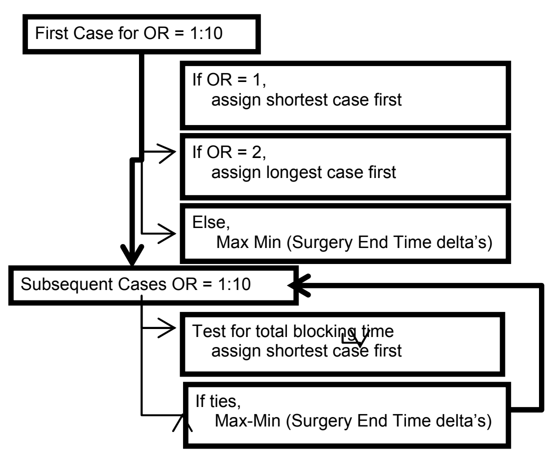

Figure 4: BTM Heuristic Flow Diagram

Once each OR has its first case of the day assigned, we assign second (and subsequent) cases by evaluating the expected blocking time as a result of using each specific case, and assign the case with the minimum expected time. If no blocking occurs with any case option, we resort to the secondary objective and choose the case which most evenly spaces out the case end times. We continue assigning the cases round robin-style.

5 EXPERIMENTATION AND EVALUATION

In order to evaluate the performance of the heuristics we create a set of 500 caseload scenarios for each OR for the day, based on the distributions described in Marcon and Dexter (2006). For each scenario, the case sequences for each OR is determined using each of the heuristics. For each case in each OR, the surgery and recovery times represent expected values from given underlying distributions.

Since the surgery and recovery times are uncertain, we then create 10,000 random instances for each caseload scenario. This is done by multiplying expected surgery time, turnover time and anesthesia recovery time by a random variable distributed normally with a mean of 1 and standard deviation of 0.25. Then, for each random instance, the OR case sequences from each heuristic is evaluated to determine performance. We then collect statistics of the performance of each heuristic and compare the results.

The caseload scenarios are based on 10 OR's and 10 PACU beds. The OR's are dedicated to surgeons of three different services: Short, Medium, and Long surgical cases. OR overtime is the maximum of the last surgery's end time minus the end of the scheduled block time (8 hours in our simulation), or zero. PACU nurse-hours are the combined hours that each PACU bed is open, assuming it is available from the beginning of the block until the last patient has vacated each bed for the day.

We created the model using two types of software that interacting together. The first part of the coding was in Matlab. This part created schedules using for each caseload using each heuristic, and kept track of the resulting metrics. The second portion of the code was written in C#, and called by Matlab after each schedule was created. This part of the code evaluated each schedule over 10,000 iterations using the randomly generated (using the normal distribution mentioned previously) "actual" processing times. To test 500 sets of cases, using 10,000 iterations of each of 6 heuristics' schedules for this set took 4.8 hours on a Pentium 4 processor. The bulk of the time was spent creating schedules for the BTM heuristic.

6 NUMERICAL RESULTS

We compared the average OR overtime required to complete each set of cases in the order dictated by each heuristic and the average number of nurse-hours needed as defined above (see Figure 5). We find that Johnson's rule is dominant (Johnson's is not statistically significant from all other heuristics, except BTM, in respect to OR Overtime, but it is with respect to PACU nurse-hours). In Figure 6 we also compare OR overtime with the average percent of case sets with PACU delay using each heuristic. In this comparison, we see that the HIHD, Alt. Johnson's, and MIX heuristics dominate (they are not statistically significantly different from each other on either axis).

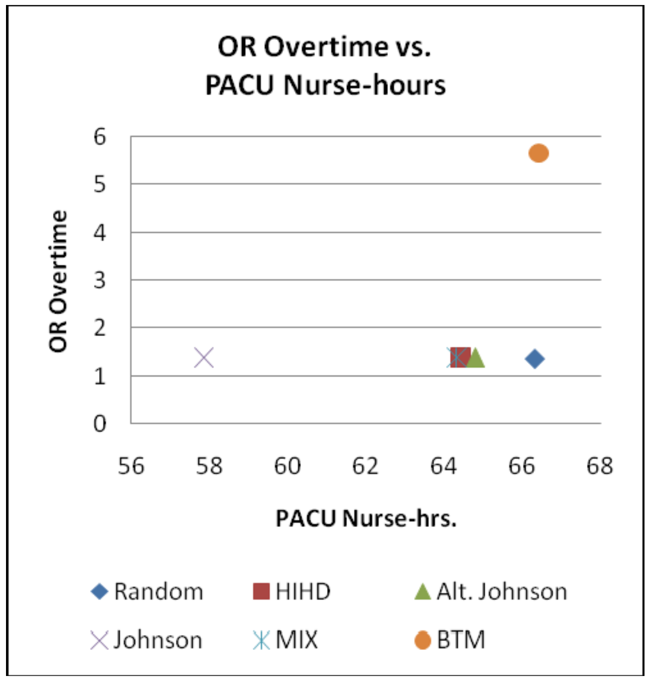

Figure 5: OR Overtime vs. PACU Nurse-Hours for 10 ORs and 10 PACU beds

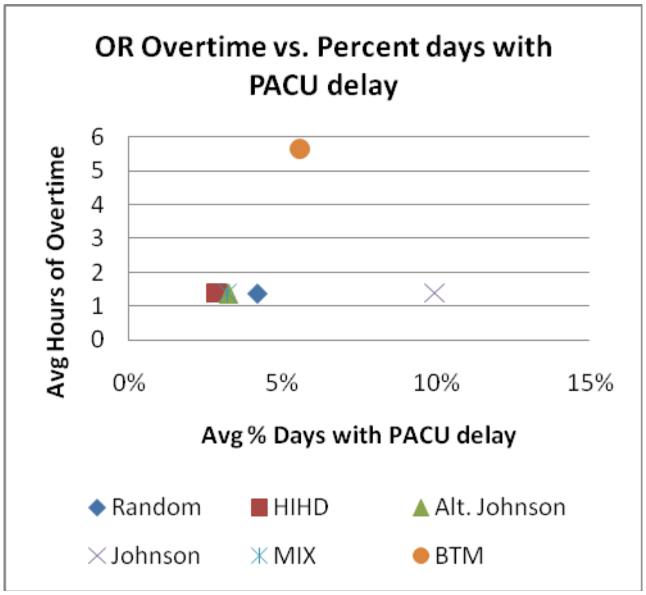

Figure 6: OR Overtime vs. Percent PACU Delay for 10 ORs and 10 PACU beds.

We next increase the number of PACU beds available to 20, which is essentially infinity for this problem. We notice (Figures 7) that we achieve similar results. However, as we would expect with infinitely available PACU beds, none of the heuristics experiences any PACU delays (Figure 8).

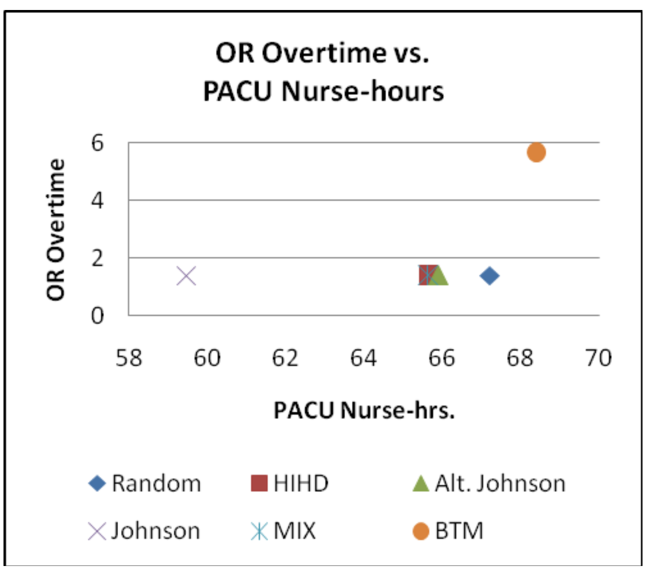

Figure 7: OR Overtime vs. PACU Nurse-Hours for 10 ORs and 20 PACU beds (sufficiently represents infinity beds in this case).

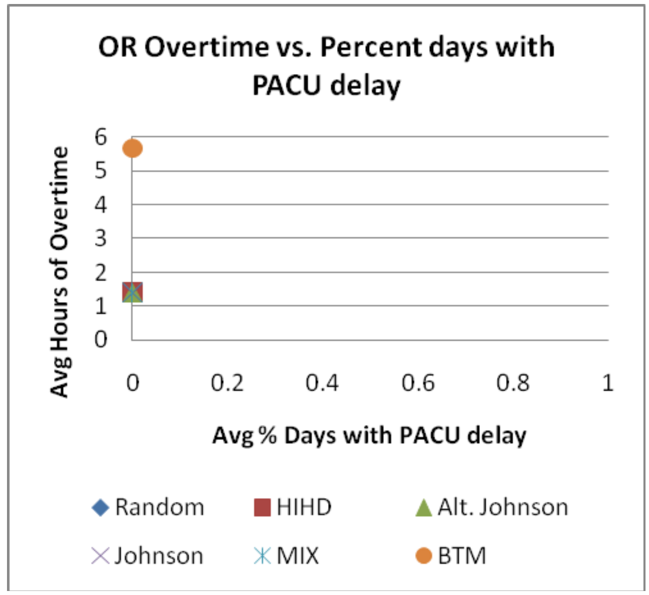

Figure 8: OR Overtime vs. Percent PACU Delay for 10 ORs and 20 PACU beds.

Finally we reduce the number available beds from the original 10 to 5. Here we see a slight frontier form. There is a small gain of additional PACU nurse-hours for a slight improvement in OR Overtime. On both axes, there is a statistically significant difference between Johnson's and the cluster of HIHD, MIX, Alternative Johnson's, and Random heuristics. Overall, we gain a large reduction in PACU nurse hours (as we only have 5 beds available), and a relatively small increase in the number of OR Overtime hours. Also, notice that with so few beds available, blocking is experienced nearly every day.

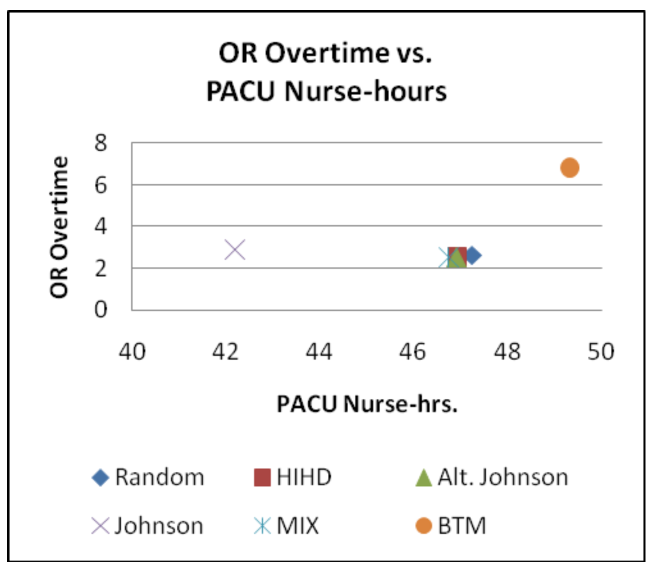

Figure 9: OR Overtime vs. PACU Nurse-Hours for 10 ORs and 5 PACU beds (sufficiently represents infinity beds in this case).

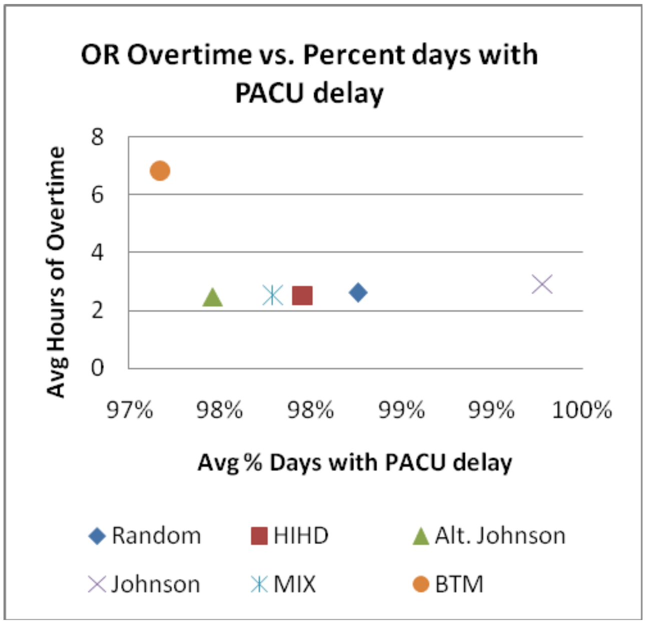

Figure 10: OR Overtime vs. Percent PACU Delay for 10 ORs and 5 PACU beds.

Notice that in all three of these comparisons, the BTM heuristic, which tries to minimize blocking, performs poorly. In contrast, the Johnson's heuristic has the highest occurrence of blocking, though it clearly dominates in OR Overtime and PACU nurse-hours. This indicates that blocking is beneficial in some way. Blocking can be of benefit if the amount of OR overtime it generates, if any, corresponds to a smaller cost than the additional nurse-hours for an additional PACU bed to be opened. See Figure 11 for an example of this. The results also point to improvements that could be made in the BTM heuristic. The round-robin style assignment doesn't sufficiently allow all ORs to be considered at once. Consider that in our model some of the ORs are assigned many shorter cases while others are assigned a fewer number of longer cases. Despite these differences, BTM still treats the first case in each OR - long or short - in the same way.

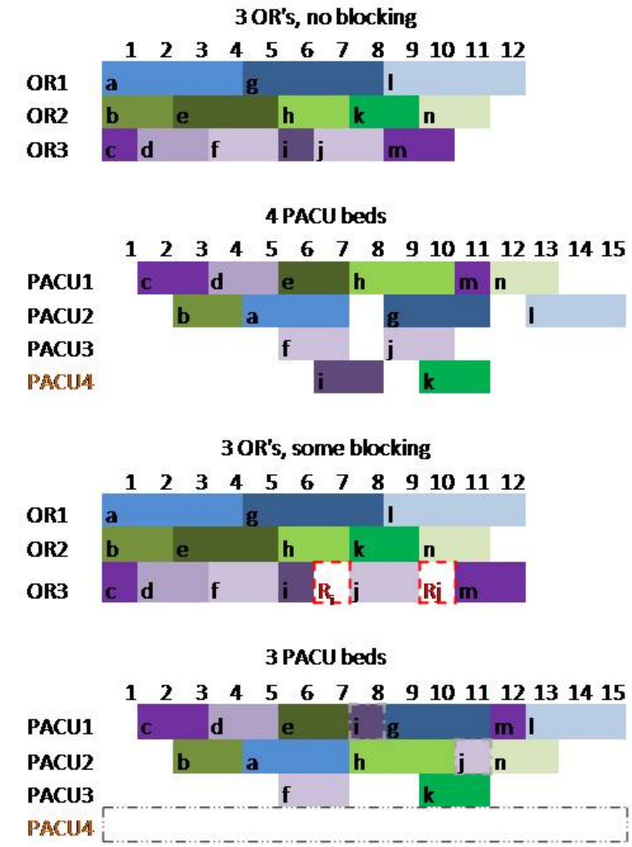

Figure 11: An illustration of the potential benefits of blocking when it prevents an additional PACU resource from being required.

Finally, it should be noted that in creating a set of cases, there is one aspect of the case set generation that may be giving the Random heuristic an advantage over totally random assignment: the requirement that OR schedules be 70% full. This means that if an OR's schedule is not at least 70% full, cases will be randomly generated until one is found that will nominally fit within the 8-hour day, and will put the OR over the 70% full mark. This means that there is likely to be a shorter case at the end of the day since for the Random heuristic we don't change the order from the order in which it was generated.

7 CONCLUSIONS AND FUTURE RESEARCH

For future research, we would like to consider how many PACU beds are needed based on the expected demand for the facility (which depends on the number of ORs in the facility, the services offered, the number of surgeons of each type on staff, and the available block time they have been assigned). Also, notice from Figure 12, that on average these heuristics require fewer than 9 beds to be available. If we only make 9 beds available, how often do we encounter adverse effects (such as OR overtime and blocking)? If we make more beds available, how often do they go unused? In a preliminary test, we found no more than 17 beds was needed to create a no-wait situation with 10 OR's with attributes as described in our model.

Since surgery demand can vary vastly from day to day, or over time, we would like to consider a flexible suite design. We can use mixed integer programming, queuing, and/or simulation to test configuration ideas. We can also consider more steps in the hospital flow. Downstream from the PACU are a series of recovery beds including Intensive Care, Step-down and Routine units. These units can also create blocking in upstream units including PACU and the ORs.

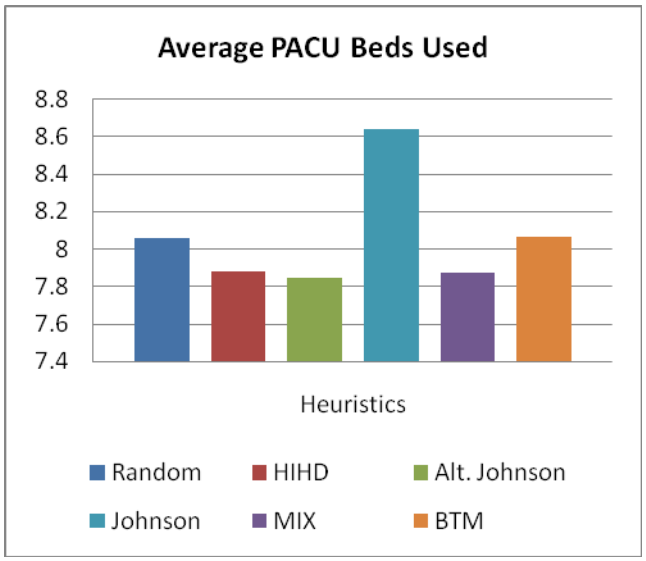

Figure 12: The average number of beds used by each heuristic when up to 20 beds are available.

ACKNOWLEDGMENTS

This project was funded in part by CMMI-0620573 (Denton) from the National Science Foundation. Also we acknowledge the help of John Telford whose C# framework we used to generate the simulation model (Telford, 2008).

REFERENCES

Denton, B.T. 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*. 10,1: 13-24.

HFMA 2005. Achieving operating room efficiency through process integration. Technical report, Health Care Financial Management Association Report.

Gupta, D. 2007. Surgical Suites' Operations Management. *Production and Operations* Management. 16, 6:689-700.

Hall, N.G. and C. Sriskandarajah. 1996. A Survey of Machine Scheduling Problems with Blocking and No-Wait in Process. *Operations Research.* 44,3:510-525

Hsu, V.N., R. de Matta and C.-Y.Lee. 2003. Scheduling Patients in an Ambulatory Surgical Center. *Naval Research Logistics*. 50:218-238.

Marcon, E., S. Kharraja, N. Smolski, B. Luquette, J.P. Viale. 2003. Determining the number of beds in the Postanesthesia care Unit: A Computer Simulation Flow Approach. *Anesthesia & Analgesia* 96:1415-23.

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

Telford, S.J. 2008. Discrete Event Simulation in C#. Master's Project Report, Integrated Manufacturing Systems, N.C. State University.

AUTHORS' BIOGRAPHIES

JILL H. ISER is Doctoral student in the Edward P. Fitts Dept. of Industrial & Systems Engineering at North Carolina State University. She previously worked as an Industrial Engineer at Michelin North America. Her primary research interests are in healthcare optimization using simulation and other IE techniques. Her email address is <jahowar2@ncsu.edu>.

BRIAN T DENTON is an Assistant Professor in the Edward P. Fitts Department of Industrial & Systems Engineering at North Carolina State University. His primary research interests are in the development and application of operations research methods to health care delivery, and his work relating to surgical scheduling won the IIE Transactions outstanding publication award in 2005. He completed his Ph.D. in Management Science at McMaster University in Hamilton, Ontario, Canada. His email address is <bdenton@ncsu.edu>.

RUSSELL KING is Professor and Director of Graduate Programs in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University (NCSU). At NCSU he also serves as the Director of the Furniture Manufacturing and Management Center and is associate faculty of the Operations Research Program and the Integrated Manufacturing and Systems Engineering Institute. He is a Fellow of the Institute of Industrial Engineers, a member of the NCSU Academy of Outstanding Teachers, and an Alumni Distinguished Graduate Professor. His research interests can be generally classified as including control and scheduling of production and logistics systems and demand activated manufacturing/sourcing systems. He received a Ph.D. in Industrial Engineering from the University of Florida. His email address is <king@ncsu.edu>.