Maximizing Operating Room and Recovery Room Capacity in an Era of Constrained Resources

Suzanne M. Sokal, MSPH; David L. Craft, PhD; Yuchiao Chang, PhD; Warren S. Sandberg, MD, PhD; David L. Berger, MD

Hypothesis: Three parallel processing operating rooms (ORs) (concurrent induction and turnover) with a dedicated 3-bed mini–recovery room (mini–postanesthesia care unit [PACU]) will optimize patient throughput and main PACU workload when compared with 4 traditional ORs or 4 parallel processing ORs.

Design: Statistical and mathematical models projected the impact of parallel processing on case throughput and PACU use.

Setting: Academic medical center with 48 traditional ORs using serial induction and turnover and 1 experimental OR, the operating room of the future, with parallel processing.

Participants: All surgical cases from October 2002 through March 2004 (N=49,887).

Interventions: A statistical model projected the duration of induction, surgery, turnover, and PACU stay for cases performed in a traditional OR (n=48,667) based on the operating room of the future (n=1220) experience. A fluid queuing model compared each combination using specific probability density functions.

Main Outcome Measures: Each OR configuration was evaluated for case throughput and minutes of work sent to the PACU.

Results: Although all cases save OR time with parallel processing, only select surgeon-case combinations translate time saved into additional cases per day (26%). Without additional PACU slots, output from 4 parallel processing ORs further stresses the PACU. Three parallel processing ORs and a mini-PACU balances incremental volume by offsetting PACU utilization in 84% of cases.

Conclusion: In a PACU-constrained environment, 3 parallel processing ORs with a mini-PACU configuration offers increased throughput and decreased PACU workload.

Arch Surg. 2006;141:389-395

OPERATING SUITES ARE INCREASINGLY EXPERIENCING capacity constraints created by a combination of growing surgical volume, fixed physical space, and human resource shortages. The impact of this supply-demand mismatch can result in longer wait lists, forced overtime, and off-hours surgery. To accommodate incremental volume within space and human resource limitations, significant improvements in operating room (OR) efficiency must be made.

Previous work in OR efficiency demonstrated that improvements in a single step in the perioperative flow is insufficient to accommodate additional cases.1,2 Operating rooms maintain a production model whereby each step in the perioperative flow occurs in series as follows: patient induction, surgery, room turnover. In an environment built and staffed to support serial processing, accommodating additional cases is difficult without adding ORs or working longer hours. To perform more surgeries per block without decreasing surgical time, a transformation in workflow is required.

Massachusetts General Hospital, Boston, established a different model for managing workflow within the operative theater as part of the Center for Integration of Medicine and Technology Operating Room of the Future (ORF) Program. The ORF demonstrated improved throughput by moving from a serial processing to a parallel processing production model.3 This was accomplished by performing portions of the clinical process flow simultaneously, specifically, anesthesia induction and OR setup (Figure 1). Additionally, anesthesia is tailored so that emergence occurs simultaneously with the end of surgery or patients are promptly removed to an adjacent early

Author Affiliations: Center for Clinical Effectiveness in Surgery and Departments of Surgery (Drs Sokal and Berger), Radiation Oncology (Dr Craft), Medicine (Dr Chang), Anesthesia and Critical Care (Dr Sandberg), Massachusetts General Hospital, and Harvard Medical School (Drs Sandberg and Berger), Boston.
exception is Friedman7 who was able to successfully located recovery room (as active operating rooms, and serves as a dedicated PACU space yielded a 4-room design. In a 4-room configuration, called the pod, 3 rooms function as active operating rooms, and 1 serves as a dedicated recovery area for emergence and extubation.3 Hanss,4 Torkki,5 and Smith6 were also able to complete more cases per OR block by overlapping induction and turnover. However, all published research evaluating the impact of parallel processing on throughput to date have used additional staff to accommodate the incremental volume.34 One exception is Friedman7 who was able to successfully leverage the efficiencies of parallel processing without incremental staffing allocations. The ORF, which like Torkki’s5 and Smith’s6 work focused on a single-room model, requires additional staffing to function properly. In a production-based environment, a design that balances resource requirements with throughput goals is needed. A multiroom version of the ORF may minimize the cost associated with staffing a parallel processing production model while reaping its efficiency benefits. The multiroom configuration allows for the realignment of staff and postanesthesia care unit (PACU) resources to support concurrent anesthesia induction and OR setup, as well as emergence from anesthesia and PACU handoff. Integrating the staffing needs (Table 1) of a high-throughput area with the desire to have additional proximity, dedicated PACU space yielded a 4-room design. In a 4-room configuration, called the pod, 3 rooms function as active operating rooms, and 1 serves as a dedicated recovery room (Figure 2). The proximity of the mini-PACU improves efficiency and safety of patient transfer. This design maintains the standard circulating nurse and scrub technician configuration for the 3 operating rooms and uses 2 recovery room nurses in the mini-PACU. Anesthesia staffing remains unchanged with 2 staff anesthesiologists covering 4 residents. The net staffing exchange would be substituting 1 recovery nurse for 1 scrub technician. To determine if the pod is a superior configuration, throughput capability and the effect it would have on main recovery room workload must be assessed and compared with both 4 standard ORs and 4 parallel processing ORs.

Table 1. Staff to Room Ratios for Each Configuration

<table>
<thead>
<tr>
<th>Staff</th>
<th>4 Standard ORs</th>
<th>The Pod</th>
<th>4 Parallel Processing ORs*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon</td>
<td>4:4</td>
<td>3:3</td>
<td>4:4</td>
</tr>
<tr>
<td>Anesthetists</td>
<td>2:4</td>
<td>2:3</td>
<td>4:4</td>
</tr>
<tr>
<td>Residents/CRNA</td>
<td>4:4</td>
<td>4:4</td>
<td>4:4</td>
</tr>
<tr>
<td>Nurses/scrub tech</td>
<td>8:4</td>
<td>8:4</td>
<td>10:4</td>
</tr>
</tbody>
</table>

Abbreviations: CRNA, certified registered nurse anesthetist; OR, operating room.

*Staff ratios for 4 parallel processing ORs were extrapolated based on institutional experience with staffing the OR of the future.

Figure 1. Illustration of parallel processing. Compared with serial processing, whereby surgery is followed by turnover and then induction, in parallel processing induction and turnover occur simultaneously.

Figure 2. Flowchart of mini-PACU vs. main PACU. The mini-PACU uses 2 recovery nurses and staffed in parallel processing vs. 1 recovery nurse and staffed in serial processing.

Evaluation of each OR configuration required the creation of statistical and mathematical models. A detailed process flowchart of a standard OR was developed and used to identify critical time intervals in the perioperative flow. Data for each time interval were acquired from the Nursing Perioperative Record for all cases completed in the main OR and same-day surgery unit between October 2002 and March 2004 (N=49887). The data were aggregated into the following key time intervals: induction, surgery, turnover, and PACU length of stay (Table 2). The time intervals were used to develop a statistical model of a parallel processing environment, which in turn fed the mathematical models used to evaluate each configuration.

The statistical analysis compared 49887 cases performed in the standard OR environment with 1220 (2.4%) cases performed in a parallel processing environment with a proximate early recovery space (the ORF). This comparison allowed estimates of the impact of parallel processing on process flow to be calculated. The experience from the ORF was used to project the length of each interval for cases completed in a standard OR as follows: (1) predictors for each interval were identified; (2) a mixed-effects model was used to estimate the effect of each predictor in and out of the ORF; (3) the predicted value of each interval was calculated as if the case were performed in the ORF. The difference between the observed time and the projected time was considered the time savings for each case.

Output from the statistical model was used to generate a revised set of historical data for the durations of sequential steps in a case, including duration of the induction of a patient, duration of the surgical procedure, and time spent in the PACU for each surgeon-case combination. This data served as input for a fluid queuing model of the pod. This algebraic model allowed for the calculation of both throughput and PACU use.

Single-case deterministic durations were used to build the mathematical model of the pod. In addition, the following assumptions were made about the model: (1) patient emergence from anesthesia occurred simultaneously with surgery completion, or patients were removed promptly to the adjacent mini-PACU for emergence, as is currently the case in the ORF; (2) all time spent in the mini-PACU counted toward required PACU time; (3) the mini-PACU was packed efficiently (recovery in mini-PACU = surgery time + [OR turnover time - mini-PACU turnover time]); (4) when the main PACU was full, operations ceased; and (5) the relationship between work sent to the main PACU and congestion time of the main PACU was linear.

To compare the 3 configurations on PACU burden, the relationship between work sent to the PACU and the number of hours of PACU congestion was analyzed. A linear relationship (meaning the more work sent to the PACU the more often it is congested) was assumed. Historic data showed a positive correlation between daily work sent to the PACU and the fraction of time the PACU was congested. With the relationship between PACU congestion and workload estab-
lished, each of the OR configurations were compared, starting with 4 parallel processing ORs.

RESULTS

Cases completed in a parallel processing environment are always faster than those completed using serial processing. However, not every surgeon-case combination benefits equally from parallel processing. While some surgeon-case pairings accumulate sufficient time savings to accommodate additional volume, others are only able to eliminate overtime or end the day earlier. Because the pod configuration would convert an OR into a recovery room, only cases capable of realizing the highest level of efficiency gains from parallel processing, ie, additional volume, were identified as potential work for the pod.

Of the nearly 50,000 cases analyzed, 26% reached the required level of efficiency. Simply put, parallel processing enabled 1 in 4 cases to be completed much faster that additional case volume could be completed within a standard 9-hour OR day. In a hospital environment with unconstrained PACU capacity and sufficient case volume, 4 parallel processing ORs would be optimal because this configuration offers the highest throughput potential. However, the study organization has severely limited PACU capacity. To that end, an analysis of PACU workload generated by 4 parallel processing ORs was completed. Results showed insufficient capacity in the main PACU to support 4 parallel processing ORs. Subsequent analyses focused on comparing 4 standard ORs with the pod configuration.

Workload for the main recovery room was evaluated for both the 4 standard ORs and pod configurations. In the pod, time spent in the mini-PACU reduces main-PACU workload. A case was considered a good fit for the pod if enough main-PACU workload was eliminated to compensate for the additional case volume associated with parallel processing. Main-PACU workload was reduced in 84% of cases if performed in the pod.

Combining the evaluation criteria, 17% of cases resulted in an increased number of cases that could be completed per block and also reduced main-PACU workload. In the pod, time spent in the mini-PACU reduces main-PACU workload. A case was considered a good fit for the pod if enough main-PACU workload was eliminated to compensate for the additional case volume associated with parallel processing. Main-PACU workload was reduced in 84% of cases if performed in the pod.

Figure 2. Schematic and sample floor plan of the pod configuration and patient flow diagram for the pod configuration. Three operating room (OR) suites with adjacent induction rooms (IRs) are supported by a 3-bed mini–postanesthesia care unit (PACU) constructed from a once functional OR. Arrows denote patient flow through the pod; Rec, receiving room.
together identifying 72 hours of good case time per day. Completing these same cases in a parallel processing environment would consume just 46 hours. Based on these projections, there is enough volume to fill 3 active parallel processing ORs in the pod and also empty an extra OR per day. This is particularly important as the model cannibalizes an OR.

**COMMENT**

Improving OR efficiency in an era of constrained resources is of utmost importance. Achieving this goal requires a new way of thinking about this problem, given that previous studies\(^1\)^\(^,\)^\(^2\) have not demonstrated the ability to accommodate additional cases through improvements to a single step in patient flow. Parallel processing of induction and turnover has been shown to work,\(^3\)\(^-\)\(^7\) however most of the tests of parallel processing have added additional staff to facilitate the work.\(^3\)\(^-\)\(^6\) The pod is an attempt to accommodate additional volume through a combination of parallel processing portions of the process flow without adding staff. The modeling demonstrates that there are more than enough case hours to fill the pod. The pod configuration should be able to process more cases than 4 standard ORs. In fact, the pod is capable of processing enough cases to free an additional OR per day, further improving main OR capacity.

This study has several limitations. First and foremost, process flow projections were generated for a multiroom model based on experience with a single-room set-up. The single-room model (the ORF) actually serves very few surgeons, and by extension, very few procedure types. The statistical model used to evaluate the 3 configurations made interval-specific time savings projections based on experience in the ORF. To the extent that cases performed in the ORF are not representative of process requirements for all operative cases, the time estimates used in this analysis may prove inaccurate. Another limitation of the model is that it did not account for the correlations between each step in the process flow. It is reasonable to assume that changes in the efficiency

---

**Table 2. Description of Surgical Milestones and Critical Time Intervals**

<table>
<thead>
<tr>
<th>Time Interval Name</th>
<th>Milestones</th>
<th>Time Interval Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Induction</td>
<td>Start of anesthesia care; ready for surgical preparation</td>
<td>Anesthesia provider notes event in anesthesia information system, signaling the beginning of billed activities to patient is ready for surgeon to begin positioning and skin preparation</td>
</tr>
<tr>
<td>Surgery</td>
<td>Surgery start; surgery finish</td>
<td>Surgical instrument contacts the patient’s skin to dressing applied postprocedure or drapes removed if no dressing was used</td>
</tr>
<tr>
<td>Turnover</td>
<td>Patient out of room; next patient in room</td>
<td>Time from the prior patient out of the OR to next patient in OR time for sequentially scheduled cases</td>
</tr>
<tr>
<td>Recovery PACU</td>
<td>Time into PACU to time out of PACU</td>
<td>The time at which the anesthesiologist turns over care of the patient to the PACU until the patient is transported out of the PACU</td>
</tr>
</tbody>
</table>

Abbreviations: OR, operating room; PACU, postanesthesia care unit.

*Milestone and critical time interval definitions were adapted from Sandberg et al\(^8\) and standard terms were used from the American Association of Clinical Directors Procedural Times Glossary.\(^8\)

---

**Figure 3.** Postanesthesia care unit (PACU) congestion vs PACU work. The x-axis indicates PACU work sent (person-minutes); y-axis, number of hours of PACU congestion.

\[ y = 0.0004x - 5.5389 \]
of 1 step in the process impacts performance in subsequent steps. However, these potential interrelationships were not considered in the statistical projections.

Recognizing that parallel processing often requires additional resources, the pod configuration attempted to remain staffing neutral through the reorganization of the work facilitated by the conversion of a working OR into a recovery space. There is an exchange of 1 recovery room nurse for a scrub technician in the OR that would be converted to a mini-PACU. The anesthesia staff allocated to the converted OR would monitor the mini-PACU, facilitate transfers and reports, and also assist with patients in the remaining ORs. Although in theory this allocation seems reasonable, it may prove inadequate in a working version of the pod, thus additional staff may be required. The pod configuration leads to a slightly increased number of transfers in patient care because some patients must go from the mini-PACU to the main PACU rather than being discharged home or brought to the hospital ward. This has not been a problem in the ORF that also uses a proximate recovery room. The proximity of the mini-PACU actually improves patient safety in transfer from the OR. In the ORF, transfers from the mini-PACU are done by nurses or nursing assistants. The pod would accommodate all surgical case types across many surgical services in a tertiary hospital setting that meet operational criteria. This includes general, orthopedic, pediatric, plastic, urologic, gynecologic, thoracic, and even certain case types from neurosurgery. It may not be feasible to run the pod as a multiservice set of suites owing to equipment needs and other operational issues. The initial incarnation will most likely be limited to 2 or 3 different services. While a formal financial analysis has not been completed, the basic premise of doing more cases with the same resources implies that all parties will benefit financially.

This analysis represents the first step in testing the viability of this configuration and significant work remains to be done. Subsequent analyses are required to determine if three 9-hour blocks can be filled with appropriate cases from a single surgeon on a daily basis. This analysis would require the development of surgeon-case profiles to test for fit into the pod. A multiroom model that combines a parallel processing production model with a proximate dedicated recovery room holds potential for optimizing the balance between cost and throughput. By thinking differently about deploying staff and space, ORs would be able to accommodate incremental volume using current resources.

Acknowledgment: We thank Andrew Karson, MD, Bethany Daily, MHA, and Marie Egan, RN, for their participation in feedback sessions during the development of the mathematical and statistical models used in this article, and Kate Robertson for extraction of data from hospital databases.

REFERENCES

5. Torkki PM, Marjamaa RA, Torkki MI, Kallio PE, Kirvela OA. Use of anesthesia induction rooms can increase the number of urgent orthopedic cases completed within 7 hours. Anesthesiology. 2005;103:401-405.

Robert Quinlan, MD, Worcester, Mass: First of all, I am happy that the same problems are happening in Boston as they are in Worcester. Our needs are very similar. How important was it to have an induction room? Years ago when I was in medical school at Cornell there were induction rooms for every operating room, and we have lost that effect. We have a large room where 1 anesthesiologist is responsible for preparing maybe 10 patients at a time with induction in the OR. This idea of an induction room, how important is that? Or is it really a team approach that you have a smaller number of teams that seem to relate better to a common goal? What’s more important—a team or that induction room?

Dr Berger: I think that they are both important. I think that all of this is obviated if you have an unlimited space. If you have unlimited space, then you could have 1 large recovery room and you could have all of your patients come out of your operating rooms as fast as you could and fill the space. In our situation, we do not have that. We cannot use that as a model. I think that an induction room next door allows your anesthesiologist within your group to put that patient to sleep or be completely prepared such that when the case kits are opened and the room is clean that process can begin and the patient can come into the room. If you have 1 anesthesiologist responsible for 10, that will dilute out the effect of having it, so the proximity of the induction room allows the care team that is involved to continue in their team roles and succeed, so I think it is actually fairly important.

Andrew Warshaw, MD, Boston, Mass: As your shill in the audience, let me ask a question I know that you want asked. In the OR of the future, the efficiency that it gains depends upon having surgeons who are doing short cases and who are themselves efficient, even though their time remains constant because the gains made are in the turn-over time, not in the operating room. In this model, what are the characteristics of the surgery and of the surgeon that would get you the efficiencies you want? Are there only certain kinds of cases? What other groups fall into that 17%?
In my opinion the wasted time and money are unconscionable with what goes on today, just as it did 50 years ago, regarding case turnover logistics.

At the same time, however, let me finish by emphasizing what huge advances in anesthesiology have taken place, and only because our anesthesiologists have become so expert are we able to do the complex and prolonged cases which are possible today.

Dr Berger: Thank you, Dr Hendren. That's absolutely right. You need to have a team, and the way this works is with a team. Juan Sandberg is an anesthesiologist who works with us in the OR and you have to overcome that and it is important to select your team with people who are interested in actually doing something like this. So you need everybody wanting to work the same way and getting the results from it. It is impossible to do it without anesthesia, nursing, and surgery working together, and boy that has created some problems in the orientation or creation of this system, but once you get everybody working together actually works very well.

Pardon Kenney, MD, Boston: I am the Chief of Surgery in a smaller community hospital. This is a great thing. This is the kind of thing we chiefs of surgery lie awake at night worrying about, and to convince a CEO [chief executive officer] that it's a good thing to do to take an OR out of service is really admirable. My question has to do with something you talked about in one of your last slides. As I understand it, you assume this is going to be revenue neutral as far as FTEs [full-time equivalents] are concerned. In a perfect world, could you speculate as to how many more FTEs you would really like to have to make it work even better?

Question 2, what is the ideal type of surgery? If I understand you correctly, you are mixing a bunch of different specialties. Is there some ideal mix that you would aspire to that would make the system work even better?

Dr Berger: Sure, there absolutely is. We have run one version of this already in our outpatient setting in the day care environment with hernias that I do, where we began this. I do hernias under local with IV [intravenous] sedation, and we created one of these circular environments out there and we went from being able to do in a day 4 to 5 cases on average in a room. Now I do 8 to 10 hernias in a room, the same room, in a day and usually we take an hour break for lunch as a team, so the whole team of the surgeons here just stops and takes an hour break. That's in a revenue-neutral environment using the same number of nurses as we did before.

We have not implemented this pod configuration yet. We hope that when we run it, we can run it revenue neutral with FTEs. It is designed to be revenue neutral with FTEs. We have to run it to see. Until we spend 3 months or 6 months doing it, we will not know. The ORF right now is not revenue neutral in that regard that we have to use an extra nurse and we have to use an extra anesthesiologist. I believe that the reason for that is because we have constructed 1 operating room with 1 holding room and 1 slot as a recovery room. I think that the 4-pod configuration is what is going to overcome that FTE problem.

In terms of generating more revenue, we know we will generate more revenue for anesthesia because they are going to do more cases in 3 rooms than they would do in 4 rooms, so for anesthesia it is a net win. In terms of the hospital we know it is a net win because the hospital is going to generate more revenue in 3 rooms than they would in 4 rooms, so that is a net win. And we know for the surgeon it is a win because the surgeons not only can they get their cases done but they could possibly get finished earlier in the day and furthermore, you do it with a lot less stress. You do not have to get mad about the fact that you are sitting there and waiting for an hour for your case to turn over and everything to get done. So it turns into a net win for all 3 groups involved certainly on the smaller models that we have done and now we want to hopefully roll this out
in the next couple of months and see if we can do it on a broader scale and then of course do it across divisions which is a whole other thing. Can you do it in the same day where you have 1 room doing total hips, 1 room doing general surgery, and 1 room doing pediatric surgery, let’s say. That is going to be a whole other hurdle that needs to be taken down essentially, because that is going to cross anesthesia, that is going to cross nursing, and that is a big problem certainly at our institution. Getting everybody to work together is a challenge in and of itself, and then when you make the obstacles higher it is even more significant of a challenge.

David Butsch, MD, Montpelier, Vt: I certainly enjoyed your paper. Other than your model, do you have any suggestions for improving turn-over time which seems to be the bane of all our existence? In our hospital, it is always worse after lunch.

Dr Berger: I would say that the average turnover time in the ORF when we are operating it on Fridays with short cases is 15 minutes. That is because the system is working together. The surgeon does not leave the room or the area around the operating room. There is an expectation that everybody is going to do their job and that people are going to work together. If the surgeon needs to help move the patient to the mini-recovery, the surgeon can do that, as opposed to there being, as Dr Hendren said, a defined role. The anesthesia resident does this. The nurse does this. It breaks down the team barriers and it allows everybody to work together to get the job done. I think that is the key. If you do not want to do any of this, if you can just get the idea that the people are working together as a team to get the job done rather than the surgeons yelling at the anesthesiologists and the anesthesiologists yelling at the surgeons, and then the nurses getting caught up in the middle, I think that will work to an extent. You will not get the economy that you get as you maximize it with the technology here but you will get a good jump on it.

Announcement

On the new Calendar of Events site, available at http://pubs.ama-assn.org/cgi/calendarcontent and linked off the home page of the Archives of Surgery, individuals can now submit meetings to be listed. Just go to http://pubs.ama-assn.org/cgi/cal-submit/ (also linked off the Calendar of Events home page). The meetings are reviewed internally for suitability prior to posting. This feature also includes a search function that allows searching by journal as well as by date and/or location. Meetings that have already taken place are removed automatically.