Making Management Decisions on the Day of Surgery Based on Operating Room Efficiency and Patient Waiting Times

Franklin Dexter, M.D., Ph.D.,* Richard H. Epstein, M.D.,† Rodney D. Traub, Ph.D.,‡ Yan Xiao, Ph.D.§

The authors review the scientific literature on operating room management operational decision making on the day of surgery. (1) Some decisions should rely on the expected (mean) duration of the scheduled case. Other decisions should use upper prediction bounds, lower prediction bounds, and other measures reflecting the uncertainty of case duration estimates. One single number cannot be used for good decision making, because durations are uncertain. (2) Operational decisions can be made on the day of surgery based on four ordered priorities. (3) Decisions to reduce overutilized operating room time rely on mean durations. Limited additional data are needed to make these decisions well, specifically, whether a patient is in each operating room and which cases are about to finish. (4) Decisions involving reducing patient (and surgeon) waiting times rely on quantifying uncertainties in case durations, which are affected highly by small sample sizes. Future studies should focus on using real-time display of data to reduce patient waiting.

1. Introduction

The primary focus of the operating room (OR) management team is to complete cases on the day of surgery, regardless of differences in healthcare systems, various methods of making OR allocations, and alternative practices in the scheduling of elective cases. Add-on cases must be scheduled.1–6 Gaps in the schedule must be filled.7,8 Cases must be moved.9,10 Staff must be assigned.11 Limited resources and personnel must be prioritized.12,13 Patients must be prepared.14 In addition, urgent cases must be sequenced.6

We review the scientific literature on OR management operational decision making on the day of surgery. We describe the paradigm of considering ordered priorities for decision making based on maximizing OR efficiency and reducing patient waiting time. We present illustrative examples of the relevant concepts and show why the use of other approaches, such as first scheduled, first served, results in inconsistent decisions. Finally, we use the scenarios to illustrate the impact of uncertainty in case durations. In the appendices, we list the equations that are needed to implement the methods.

Applying scientific principles in OR management does not solve staff’s frustration with working later than planned. There will be disruptions in OR schedules as long as there are ruptured abdominal aortic aneurysms presenting at 9:00 AM. After reading this article, managers and clinicians will understand better how to decide what cases to postpone, what cases to move, and so forth in response to such perturbations in OR schedules.

2. Definitions

Surgical service refers to a group of surgeons who share allocated OR time. An individual surgeon, a group, a specialty, or a department can represent a surgical service. Surgical service simply refers to the unit of OR allocation.

Scenario 1. All of the neurosurgeons practicing at a hospital are allocated OR time. Neurosurgery is a service.

Scenario 2. A busy surgeon is personally allocated 8 h of OR time every Wednesday. This surgeon is a service, from the perspective of OR time allocation.

All cases are scheduled on a first-come, first-served basis at some facilities. Those facilities have one or more surgical services. If all OR nurses and anesthesia providers can care for all patients, there would be only one service. Otherwise, the surgical services would be the...
combinations of specialties matching the staff skill mixes.

Allocated OR time is an interval of OR time with specified start and end times on a specified day of the week that is assigned by the facility to a service for scheduling cases.

Scenario 3. An orthopedics group is allocated OR time from 7:00 AM to 3:00 PM on Monday through Friday. These are scheduled hours. This does not mean that the orthopedic surgeons only perform cases until 3:00 PM. They systematically underestimate their case durations. Case duration is the time from when a patient enters an OR until he or she leaves the OR.

Turnover time is the time from when one patient exits an OR until the next patient enters the same OR on the same day. Turnover times include cleanup times and setup times but not delays between cases.

Overutilized OR time is the positive difference between the total hours of cases including turnover times performed by the service and its allocated OR time.

Scenario 4. An OR is staffed from 7:30 AM to 4:00 PM. The last case of the day in the OR ends at 6:30 PM. There are 2.5 h of overutilized OR time.

Underutilized OR time is the positive difference between allocated OR time and the total hours of cases including turnover times performed by the service.

Scenario 5. Operating room time is allocated from 8:00 AM to 4:00 PM. The last case of the day ends at 2:00 PM. There are 2 h of underutilized OR time.

Inefficiency of use of OR time equals the sum of two products: hours of underutilized OR time multiplied by the cost per hour of underutilized OR time plus hours of overutilized OR time multiplied by the cost per hour of overutilized OR time.

Operating room efficiency is the value that is maximized when the inefficiency of use of OR time has been minimized.

In the next section, we consider the relation of this definition to the common perception of OR efficiency as being related to working quickly.

Scenario 6. Urology performs 16 h of cases every Thursday. OR time is allocated in 8- and 10-h increments. Urology would be allocated two ORs each for 8 h on Thursdays. Urology would not be allocated one OR for 10 h, because there would be approximately 6 h of overutilized OR time. The allocation for Urology that maximizes expected OR efficiency would not be one OR for 8 h and the other for 10 h, because there would be approximately 2 h of underutilized OR time.

An upper prediction bound for the duration of a case is the value that will be exceeded by the next randomly selected case of the same type at the specified rate. There is a 10% chance that the duration of a case will be longer than its 90% upper prediction bound. There is a 5% chance that the duration of a case will be briefer than its 5% lower prediction bound.

Scenario 7. Dr. Johnson is scheduled to excise an isolated hepatic mass from 7:00 AM to 12:00 noon. After incision, he discovers omental metastases. The case ends unexpectedly at 9:00 AM. The next case cannot start early, because Dr. Henderson is busy in her clinic. The result would be a gap in the schedule of 3 h. However, Dr. Holmes has a case that he would like to perform today. He is available until 12:00 noon. The 90% upper prediction bound for the duration of Dr. Holmes’ case is 2.3 h. The case can be done in the open time with a low (< 10%) risk of causing an increase in overutilized OR time from Dr. Henderson’s case and reducing OR efficiency.

3. OR Efficiency on the Day of Surgery

At most surgical facilities, OR nurses are full-time hourly or salaried employees. Therefore, on the day of surgery, the increment in nursing labor cost from 1 h of underutilized OR time is negligible. In managerial accounting, the cost of providing nursing care during scheduled hours is referred to as a sunk cost. Finishing cases before the end of scheduled hours does not substantively reduce labor costs. The same concept applies to nurse anesthetists who are employees of the surgical facility or anesthesia group.

Few anesthesiologists and nurse anesthetists can earn enough money to cover the cost of their salary plus benefits unless they are scheduled to care for whatever patients may need urgent surgery along with patients having elective, scheduled surgery. The incremental revenue lost on the day of surgery in having an hour of underutilized OR time is negligible (i.e., the opportunity cost is zero).

Therefore, on the day of surgery, the cost of an hour of underutilized OR time is effectively equal to zero. The implication is that, on the day of surgery, the inefficiency of use of OR time is minimized by minimizing hours of overutilized OR time.

Scenario 8. An anesthesiologist is assigned to an OR allocated from 8:00 AM to 5:00 PM, but with 1 expected hour of overutilized OR time. He works quickly. Two patients need fiberoptic intubations. He does both in less than 10 min. Because of rapid work, the cases finish at 5:00 PM. One hour of overutilized OR time has been prevented, and OR efficiency has been increased.

Scenario 9. An anesthesiologist is assigned to another OR allocated from 8:00 AM to 5:00 PM, but with 8 h of scheduled cases. This physician works equally quickly, resulting in cases finishing at 3:00 PM instead of at 4:00 PM. However, overutilized OR time was not reduced, and consequently, OR efficiency was not increased.

Operational OR management decisions based on maximizing OR efficiency usually are the same as those based on minimizing labor cost, but not always. We use decision
making based on OR efficiency for two reasons. First, based on whose labor cost should decisions be made? Should it be labor cost of the hospital, the hospital and anesthesia providers, the hospital and physicians, or society? Decisions based on OR efficiency are invariant to the perspective, whereas those based on labor costs are not. Second, labor costs vary depending on staff scheduling and staff assignment, whereas OR efficiency is independent of scheduling and assignment decisions. If labor costs were used, distributed decision making would no longer be consistent depending on the perspective of who makes the decision. For example, if one OR nurse works overtime to cover for another OR nurse who has called in sick, that would affect decisions on the day of surgery based on labor costs but not decisions based on OR efficiency.

4. Making OR Management Decisions Based on Ordered Priorities

Operating room management decisions can be made based on four ordered priorities: first, maintain patient safety; second, provide surgeons with open access to OR time on the workday that they and their patients choose; third, maximize OR efficiency; and fourth, reduce patient waiting times. A lower priority is considered if it does not violate a more important priority.

When the scenarios are presented below, most readers will likely consider the priorities intuitively reasonable. That matters because it supports the supposition that the following nonintuitive scientific result is useful. The priorities are sufficient to specify how OR time is allocated and how staffing is planned, how cases are scheduled, how OR time is released, how cases are moved, how staff are assigned, and how cases are sequenced. Readers interested in operations research may want to refer to two references in particular: Dexter and Traub and Dexter et al.

The scenarios below are designed to show how the priorities are used. However, we do not report the derivations and computer simulations on which they are based.

First Priority: Patient Safety

Concerns about patient safety often cause increases in both overutilized OR time and patient waiting times, and prompt operational OR management decision making on the day of surgery. Patient safety factors that can influence possible decisions include availability of specific ORs in which it is safe to do the case; surgeons, anesthesia providers, and OR nurses with skills for the procedure; equipment necessary for the procedure, and so forth.

The first priority of patient safety includes satisfying medical deadlines for the time by which urgent cases need to start. From the date and time of the event, disease, or symptoms requiring surgery, a medical deadline can be calculated by adding the disease onset time and an estimate of how soon in hours the case must start to avoid increasing the risk of morbidity or mortality. This information generally is the limit of what is known from observational studies in surgical journals.

Scenario 10. A vascular surgeon informs the surgical suite at 6:40 AM about a patient who has had a pulseless leg for 3 h requiring thrombectomy. The medical deadline provided is to start the case before 7:40 AM. Every OR is scheduled to start elective cases at 7:00 AM. The start of a scheduled elective case should be delayed because the first priority is patient safety.

Second Priority: Open Access to OR Time

On the day of surgery, the second priority of open access to OR time simply means only canceling a case if it cannot be done safely. For non-US hospitals and US hospitals with a fixed annual budget (e.g., Veterans Affairs), performing a scheduled case reduces total costs to the physicians, the hospital, the patient, and society, even if overtime is required. Most US hospitals receive predominantly fee-for-service payment (e.g., Medicare). The difference between the incremental reimbursement for each case and its variable costs at two multiple-specialty surgical suites averaged $1,430 and $1,700 per OR hour. Therefore, even if staff were paid more than triple time for working late, not canceling the case would make economic sense.

Scenario 11. At 7:00 AM, a case is submitted for vitrectomy and fluid–gas exchange for retinal detachment. ORs are allocated from 7:00 AM to 3:00 PM. OR 3 is expected to finish at 12:00 noon. All other ORs are expected to finish close to 3:00 PM. Ophthalmology can be performed only in OR 4 with its suspended microscope. Patient safety is a higher priority than is maximizing OR efficiency. The case is performed in OR 4, even though that results in overutilized OR time.

Third Priority: Maximizing OR Efficiency

On the day of surgery, the third priority of maximizing OR efficiency means minimizing overutilized hours (see scenarios 8 and 9). Maximizing OR efficiency is a lower priority than surgeon open access to OR time, because otherwise no case would be performed that would be expected to result in any overutilized OR time.

Scenario 12. An anesthesiologist medically directs nurse anesthetists in two ORs. Allocated hours are 8:00 AM to 5:00 PM in both ORs. He needs to decide which of the two ORs to start first. OR 1 is scheduled from 8:00 AM to 6:30 PM versus OR 2 from 8:00 AM to 3:00 PM. Patient safety is not affected by the decision. No case would be canceled because of the decision. Expected overutilized OR time is 1.5 h in OR 1 and 0 h in OR 2. The anesthe-
Theesthesiologist starts OR 1 first, aiming to reduce expected overutilized OR time and thus increase OR efficiency. The same priorities apply to the decision of whether a housekeeper should first clean one OR or another, whether a holding area nurse should first prepare one patient or another, whether an anesthesia assistant should first bring OR supplies to one OR or another, whether the postanesthesia care unit should delay the admission of a patient from one OR or another, or whether central sterilization should first set up a case cart for one add-on case or another.

Scenario 13. Operating rooms are allocated from 7:30 AM to 3:30 PM. OR 1 finishes its last case of the day at 1:30 PM. Because OR 2 is running behind, its last case scheduled from 2:00 PM to 3:30 PM will not start until 5:00 PM. The thrombectomy is performed in OR 2. Choosing which OR the case is allocated from 7:00 AM to 3:30 PM. The thrombectomy started if a surgeon is not available to do the case. Surgeon availability is a patient safety issue (i.e., a case is not doing so is expected to reduce overutilized OR time and thereby increase OR efficiency. Appendix 1 describes that to maximize OR efficiency, one should consider add-on cases in descending sequence of case duration (i.e., longest cases first).

Scenario 14. Operating room time is allocated from 7:00 AM to 3:30 PM. Dr. Jacoby will be using a microscope for his first case in OR 3, scheduled to take 2.5 h. Turnover times are 0.5 h. Dr. Lou has one 4-h case scheduled in OR 4. He wants to follow with a new 4-h case requiring the microscope. However, several other short add-on cases could be done in OR 4 by other services. Because Dr. Jacoby will finish with the microscope before Dr. Lou has finished his first case, no delay will result that causes overutilized OR time. Because Dr. Lou’s add-on case is the longest and will result in no underutilized or overutilized OR time, it is scheduled to follow in OR 4.

Fourth Priority: Reducing Patient Waiting Time

The fourth ordered priority refers to reducing waiting time after a scheduled start time for elective cases. For urgent cases, this refers to the time interval from when the patient is available to when the case finishes. Surgeon availability is a patient safety issue (e.g., a case is not started if a surgeon is not available to do the case).

Continuation of Scenario 10. Operating room time is allocated from 7:00 AM to 3:30 PM. The thrombectomy plus turnover is estimated to take 1.5 h. The case can be done safely in any OR. The expected times that the last cases of the day will end in each OR are 1:00 PM for OR 1 and OR 2, and close to 3:30 PM in the other ORs. From the third priority of maximizing OR efficiency, the case will be done in OR 1 or OR 2. Choosing which OR depends on patient waiting times. OR 1 has four cases. OR 2 has one long case. Scheduling the thrombectomy in OR 1 would result in an average patient waiting time of 1.0 h, where 1.0 h = (4 patients each waiting 1.5 h/6 patients). Scheduling the thrombectomy in OR 2 would result in an average patient waiting time of 0.25 h, where 0.25 h = (1 patient waiting 1.5 h/6 patients). Therefore, the thrombectomy is performed in OR 2.

Some organizations try to use the scheduling paradigm of first scheduled, first served instead of minimizing waiting times as its fourth ordered priority. However, it makes no sense to base the decision in the preceding

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Consider decision options based on patient safety. Consider each possible case scheduling decision, and exclude those that cannot be done safely (e.g., from lack of equipment). When scheduling urgent cases as in scenario 15, data needed include medical deadlines based on the patients’ medical conditions (see sections 4 and 10).</td>
</tr>
<tr>
<td>2.</td>
<td>Exclude decision options based on impact of open access to OR time. On the day of surgery, this means simply that the cases will only be cancelled if doing so would reduce patient safety, as explained in section 4.</td>
</tr>
<tr>
<td>3.</td>
<td>Exclude decision options that would result in worse OR efficiency. On the day of surgery, OR efficiency means the impact on overutilized hours, as explained in section 3. Using each OR and its on-going cases, use methods of appendices 2, 3, and 4 to calculate expected underutilized and overutilized OR time in each OR. As in scenario 15, often the calculations can be skipped as the ORs with the most expected underutilized OR time are obvious without calculation because the ORs are finishing their lists of cases when the urgent cases are being scheduled.</td>
</tr>
<tr>
<td>4.</td>
<td>Evaluate patient waiting times for each remaining decision choice. For this decision, this would refer to the time from when the patient is ready until the case starts, as explained in section 4. Choose the decision providing the shortest expected average patient waiting time, as shown in equations in appendix 7. Example of using the calculations is given in the continuation of scenario 10 and in scenario 15.</td>
</tr>
</tbody>
</table>

OR = operating room.
scenario on how many weeks before the day of surgery that the patients in OR 1 and OR 2 were originally scheduled for surgery. First scheduled, first served is sufficient for sequencing urgent cases but not sufficient when combining urgent cases and elective cases, as often occurs in practice. The next scenario shows that first scheduled, first served is also insufficient for staff assignment decisions, in contrast to the priority of reducing patient waiting times.

Scenario 15. Operating rooms are allocated from 7:00 AM to 5:00 PM for elective cases, except for one OR allocated for urgent cases 24 h/day (table 1). Two extra OR teams work overtime until all but one OR is finished. Today, at 5:00 PM, several ORs have just finished cases. There are three more cases to be done, all with medical deadlines to start within 2 h. One patient has multiple orthopedic fractures, with an estimated case duration of 5 h. A woman will undergo exploratory laparotomy and partial salpingectomy, with an estimated duration of 1.3 h. The third case is incision and drainage of a penetrating leg wound, also with an estimated duration of 1.3 h. The cases are considered in descending sequence of case duration (appendix 1). The 5-h case is scheduled into the OR time allocated for urgent cases. Based on medical deadlines, the other two cases cannot wait until the 5-h case has been completed. Overutilized OR time is the same whether one OR team on call stays late and does both cases or both teams stay and each does one case. The average patient waiting time will be less if both teams stay to each finish one case. Consequently, both cases are started right away.

5. Impact of Case Duration Prediction on Overutilized OR Time (OR Efficiency)

The above scenarios show that the decision-making process affects OR efficiency substantially. In the remainder of the article, we consider the impact of uncertainty in case durations and turnovers on decisions. Uncertainty is quantified by the statistical distributions for the time to complete a given case or turnover.

Appendix 2 summarizes results in predicting the expected (average) case duration using historic data, surgeons’ estimates, or both.

Excess overutilized OR time from inaccuracy in predicting the duration of add-on cases is small. For example, a studied ambulatory surgery center had mean case durations of 1.6 h. Case durations were estimated by taking the mean duration of previous cases of the same surgeon and scheduled procedure(s). Even with no updates on the day of surgery, excess minutes of overutilized OR time versus knowing case durations perfectly were only 1.0 ± 0.1 min (SE) per OR per workday. Likewise, at a hospital surgical suite with mean case durations of 3.6 h, there were 5.4 ± 0.3 min of excess overutilized OR time per OR per workday. These are the largest possible reductions in overutilized OR time achievable through improved add-on case scheduling by calling ORs to learn when cases are finishing, looking at closed circuit cameras, monitoring case progress with computerized real-time patient tracking systems, examining graphical airport-style displays, and/or using more sophisticated statistical algorithms that consider uncertainty in estimates.

Why Excess Overutilized OR Time from Inaccuracy in Predicting the Duration of Add-on Cases Is Small

Excess overutilized OR time is small not because case durations are estimated accurately. They are not estimated accurately. Specifically, the above ambulatory surgery center had an absolute error of 0.4 h, and the hospital surgical suite had an absolute error of 0.8 h. Good statistical methods are only slightly more accurate at estimating case durations than are surgeons at providing case duration estimates without being prompted with their historic data. Simplicitic algorithms perform worse. There are large errors because of substantial inherent uncertainty around any particular estimate chosen. Even if the number of previous cases available to estimate case durations were increased from 1 (tiny) to 39 (large), the average patient waiting time would be reduced only by 2 min at the ambulatory surgery center and 4 min at the hospital.

Excess overutilized OR time is small, in part because errors in case duration average out over the time course of the day in an OR. This is particularly important for surgical suites with many cases in each OR each day. This is also important economically for surgical suites at which not all overutilized OR time means overtime (e.g., because employees are salaried or receive compensatory time off when they work late).

Excess overutilized OR time is small also, because the principal determinant of overutilized OR time is day-to-day variation in OR workload. People work late mostly because of the case itself, not errors in predicting its duration. When OR time is allocated based on OR efficiency, at least 62% of ORs will have no overutilized OR time, including the add-on cases. The value is more than 50% because overutilized OR time costs more than underutilized OR time. Errors in predicting the durations of add-on cases have a small effect on overutilized OR time, in part because there should be no overutilized OR time for so many ORs.

6. Predicting Time Remaining in Cases for Purposes of Staff Assignment and Moving Cases Based on OR Efficiency

Staff assignment in the afternoon, subject to the ordered priorities, depends on predicting times remaining.
in on-going cases (appendix 3). The ORs relieved should be those with the most expected overutilized OR time. Decisions involving more than two ORs are combinations of pairwise comparisons. Conceptually, if the historic duration of a case is 3 h and a patient has been in an OR for 1 h, the expected time remaining is slightly longer than 2 h. When a preliminary decision is made, before relieving staff in a planned OR, a check is done to ensure that the OR to be relieved is not close to finishing. That can be done by calling the OR, having someone walk to it and look in the window, glancing at a closed-circuit camera, or relying on real-time computerized information.

Using cases for which the surgeon had scheduled the procedure(s) at least twice before, the excess overutilized OR time from inaccuracy in predicting the time remaining was less than 1.4 min per OR per workday. The smallness of this value suggests that improving data flow on the day of surgery is unimportant compared with how the data are used to make staff assignment decisions.

For moving cases, the principle is the same. In scenario 13, the last case of the day in OR 2 is moved to OR 1 to reduce overutilized OR time. If an initial decision were made to move the case based on the estimated time remaining, OR 2 would be contacted to ensure that the preceding case is not about to finish. Excess overutilized OR time versus if case durations were known with perfect knowledge was less than 1.6 min per OR per workday.

The moving cases decision differs from other OR management decisions on the day of surgery in that the former decision does not have to be made. Whereas urgent cases must be performed and staff must be assigned, there is not a medical necessity to move a case. A threshold can be chosen for the potential saving in overutilized OR time to occur to move a case. The threshold seems to be a judgment call without one right answer. Moving cases for potentially small reductions in overutilized OR time may reduce the educational value to staff who studied the literature in anticipation of doing the case, may create a disincentive for OR staff to report that cases are finishing, and so forth. A survey of physician directors of ORs revealed a median response of 1 h for how much overutilized OR time should be expected to be reduced to justify the move. Importantly, 45% chose less than 45 min or more than 1.25 h. Therefore, uncertainties in case durations and in times remaining in cases have negligible effects on the decision versus variations among individuals in their opinions of how the data should be used.

7. Sequencing Cases to Increase OR Efficiency and Reduce Patient Waiting Times

Scenario 14 described a situation that, in practice, would depend on judging the risk that a case scheduled to take 2.5 h in one OR would finish before a case scheduled to take 4 h in another OR. Comparing the durations of cases affects both OR efficiency and patient waiting times. When each surgeon has scheduled his or her procedure(s) at least twice before, the probability that one case will take less time than another can be estimated to within an accuracy of 1.5% (appendix 5).

The statistical method provides insight into how to make the decision. Calculations use not only the means of each of the pairs of cases, but also the SDs of those cases and the number of historic cases. The latter are used so that uncertainties in the estimated durations are included in the calculation. Determining the probability that one case will last longer than another is similar to using the Student t test to determine whether one mean exceeds another, except that the uncertainty is larger because two cases are being compared instead of two means (appendix 5). At tertiary surgical suites, many cases are of procedure(s) that the surgeon has not previously scheduled. Approximately half of the pairs of cases had at least one case that was not of a procedure or combination of procedures scheduled by its surgeon five times or less within 2 yr at the ambulatory surgery and tertiary surgical suites studied above. These findings show that for decisions involving probabilities (i.e., decisions that affect patient and thus surgeon waiting times), additional data before or on the day of surgery or both will be valuable. These findings differ from those above for decisions to increase OR efficiency.

8. Upper Prediction Bounds for Reducing Patient (and Surgeon) Waiting Times

Scenario 7 describes using an upper prediction bound to insert a case into a gap in the schedule. Waiting times are reduced, while assuring a low risk of overutilized OR time. A delay in the OR schedule can be used to reduce patient waiting without creating overutilized OR time.

Scenario 16. A four-OR ambulatory surgery center has no add-on cases. In each of three of the four ORs, a surgeon has a list of cases for the day. In OR 4, Dr. Zachary has three cases, followed by Dr. Hu with one case. OR time is allocated from 7:00 AM to 3:30 PM. At 8:00 AM, Dr. Zachary’s first case has only just started. The OR manager knows that Dr. Hu has a busy clinic in the morning. Dr. Hu’s patient can be reached by phone to come later. Dr. Zachary’s cases were originally scheduled to end at 10:30 AM. The updated end time is 11:30 AM considering the 1-h late start. The upper 90% prediction bound on the duration of Dr. Hu’s case is 3 h. Dr. Hu can be given an updated start time of 12:30 PM without
risking overutilized OR time and thereby reduced OR efficiency.

These upper prediction bounds for the duration of the next one case are accurate to within 1% (appendix 6).\textsuperscript{7,8} Despite their accuracy, upper prediction bounds can be so much longer than the mean historic duration as to be useful only to the extent of showing that the historic case duration data alone are insufficient to know how long the next case will take. This occurs when there are few historic data as described in the preceding section.

Scenario 16 shows why the fourth priority of reducing patient waiting is a lower priority than the third priority of maximizing OR efficiency. Otherwise, scheduled delays would be planned between all cases, limited only by how late staff could work safely.

9. Lower Prediction Bounds to Reduce Patient Waiting Times

The time to have patients arrive and be ready for surgery can be updated on the day of surgery. Patient waiting times would be very long if all patients were ready before any cases started. If all patients are ready when their cases are scheduled to start, there is marked overutilized OR time. There seems to be no one perfect balance. The determination of when to have patients arrive on the day of surgery is equivalent to specifying quantitatively the relative cost of patients’ waiting time compared with the cost of the staff’s idle time.\textsuperscript{14,34} A societal cost perspective used median annual compensations in the United States of patients versus one anesthesiologist, one general surgeon, two OR nurses, and a full-time equivalent housekeeper. Patients would arrive early enough that they waited 95% of the time because the OR was not ready, whereas the staff waited 5% of the time with empty ORs.\textsuperscript{14}

**Scenario 17.** The third case in OR 8 has a scheduled start time of 2:00 PM (table 2). All three cases are being done by the same surgeon. The third patient has been asked to arrive by 1:00 PM but to call before leaving home for an updated start time. At 12:00 noon, she calls. Three hours is the updated 5% lower prediction bound for the sum of the time remaining in the first case, the duration of the second case, and the duration of two turnover times. Her updated arrival time is 2:00 PM for a start after 3:00 PM, where 3:00 PM = 12:00 noon + 3 h.

The 5% lower prediction bounds are accurate to within 0.3% (appendix 7).\textsuperscript{14}

Lower prediction bounds cannot be estimated accurately by using the expected (mean) duration and subtracting a safety factor.\textsuperscript{14} For example, at one studied hospital, if patients were asked to arrive sufficiently early to be ready for surgery 1.5 h before the scheduled end of the preceding case in the OR, the overall risk of OR staff waiting for the patient would be 5%. Among patients with a preceding case longer than 3.5 h, the risk that the staff would wait for the patient would be 14.3 ± 0.1%.\textsuperscript{14}

10. Topics Not Covered

This review article covered the scientific literature on OR management operational decision making on the day of surgery. Decisions on the day of surgery are important to the extent that they must be made and that they are visible to everyone working in surgical suites. Nonetheless, there are other important topics in OR management. The following are examples.

Individuals may not want to follow systematic decision making. For example, some surgeons may alter medical deadlines (see section 4) to obtain a higher priority for their cases. Methods to monitor for gaming strategies have been developed but would be used retrospectively in meetings, not on the day of surgery when making a decision for a specific patient.

Operational decision making on the day of surgery based on maximizing OR efficiency means minimizing hours of overutilized OR time.\textsuperscript{7} This also applies to case scheduling and to releasing allocated OR time.\textsuperscript{5,20} However, OR allocation relies not only on the hours of overutilized OR time, but also on the hours of underutilized OR time (see section 2).\textsuperscript{16,18,19,21} Such operational decisions made before the day of surgery have a much larger

---

**Table 2. Steps in Choosing Updated Time of Arrival of a Patient as in Scenario 17**

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Evaluate impact of decision on patient safety. For example, is the arrival time based on fasting in a patient with type I diabetes? In Scenario 17, this criterion was irrelevant.</td>
</tr>
<tr>
<td>2.</td>
<td>Evaluate impact of decision on open access to OR time. On the day of surgery, this means the potential impact on case cancellation, as explained in section 4. Generally, this criterion is irrelevant to the decision, as in scenario 17.</td>
</tr>
<tr>
<td>3.</td>
<td>Evaluate impact of decision on OR efficiency. On the day of surgery, this means the impact on overutilized hours, as explained in section 3. A low chance should be maintained for the surgical suite having to wait for the patient as considered in the first paragraph of section 9. A 5% lower prediction bound can be used for the time required to complete the preceding cases in the patient’s OR.</td>
</tr>
<tr>
<td>4.</td>
<td>Evaluate impact on patient waiting time. For this decision, this would refer to waiting from the updated scheduled start time, as explained in sections 4 and 9. Choose the latest possible arrival time that is appropriate for the patient using the equations in appendix 7. An example of using the calculations is given in scenario 17.</td>
</tr>
</tbody>
</table>

OR = operating room.
effect on OR efficiency than do the decisions considered in this article.5,10,18,20,21,30

Tactical and strategic decisions invariably involve money. Surgical suites with substantial resources can have extra surgical equipment and substantial underutilized OR time.21 Therefore, surgeons, anesthesia providers, OR nurses, and patients are affected less by any perturbation in the OR schedule (e.g., from an urgent case). The plan for how much extra equipment and excess OR capacity to have is a financial one. How to make such decisions, approximately a year before the day of surgery, is well understood.21–35 Such decisions likely have a much larger effect on patient waiting times than do the decisions considered in this article.5,10,18,20,21,30

11. Conclusions

Four principal advances have been made during the past decade in the science of OR management decision making on the day of surgery.

First, some decisions should rely on the expected (mean) duration. Other decisions should use upper prediction bounds, lower prediction bounds, and other such measures reflecting the uncertainty of case duration estimates. If case durations were known precisely, for each case, only one estimate would be needed for its duration. Because case durations are uncertain, one single number cannot be used for accurate decision making.

Second, operational OR management decisions on the day of surgery can be made based on four ordered priorities. First scheduled, first served is not a viable scheduling paradigm under many circumstances. The best ways to display data to facilitate decision making based on the ordered priorities remains unknown.

Third, decisions involving increasing OR efficiency by reducing overutilized OR time rely on mean durations. Excess overutilized OR time from inaccuracy in predicting the duration of add-on cases is negligible. Limited data are needed to make good decisions on the day of surgery, specifically whether a patient is in each OR and whether the case is about to finish. How the decision is made is likely more important than obtaining additional data.

Fourth, decisions involving reducing patient (and surgeon) waiting times rely on quantifying uncertainty in case durations. A major cause of the uncertainty results from small sample sizes. Real-time patient tracking systems and other technologies are likely to be beneficial for reducing waiting times. However, how best to use them for such probabilistic decision making is unknown.

Appendix 1: Scheduling Add-on Case Based on Minimizing Overutilized OR Time

Whenever possible, postpone the decision of confirming start times until as close as possible to the time that the first case will start.3 If impractical, assign the case to the OR that (1) has no restrictions on staff skills to prevent the case from being done safely, (2) has sufficient expected underutilized OR time to complete the case, and (3) will leave the least amount of remaining underutilized OR time when the case is finished.1 If no OR has sufficient expected underutilized OR time to complete the case, schedule the case into the OR with the most expected underutilized OR time.5

If the decision can be postponed until close to when the first add-on case is started, overutilized OR time is reduced.1 Sort the add-on cases in descending order based on estimated case duration (i.e., longest cases first). In that descending sequence, apply the method of the preceding paragraph.1–4 Typically, either zero or one add-on case can be scheduled into OR time that would otherwise be underutilized.1,5,20 Then, this algorithm is optimal.1

The performances of both algorithms are robust to rounding case durations to the nearest 15 min, which makes estimated start times more convenient.1 Restrictions into which OR a case can be assigned for purposes of safety has essentially no impact on the OR efficiency achieved by these algorithms versus other algorithms studied.1

Appendix 2: Expected Case Duration

Use the mean duration of previous cases performed by the surgeon for the same scheduled procedure(s) and anesthetic.25,50,57 If information on the scheduled anesthetic is not available, use only the surgeon and scheduled procedure(s).25,50,57 If the surgeon has not previously scheduled the procedure(s), use the mean of other surgeons’ durations for cases of the same scheduled procedure(s).57 If the procedures have not been scheduled in combination before, use the largest of the means of the individual component procedures.20 Often, the estimates are calculated at regular intervals and stored in the OR information system for lookup when a new case is scheduled.40 Having the surgeon adjust the statistically derived estimate up or down to reflect case complexity can decrease the absolute error.20,41 Including other patient factors does not significantly improve predictive accuracy.20

For many decisions, the objective is not to minimize the absolute errors in the time to complete a case but of a series of cases in the same OR on the same day.25 Assume that the time to complete each case and turnover time is independent of the time to complete others in a series of cases. Then, the sum of the means for each case is an unbiased estimator for the expected duration of the time to complete a series of cases. Formally, linear programming can explicitly minimize the absolute error for a series of cases. However, this method achieved an absolute error of 1.63 ± 0.04 h versus 1.31 ± 0.03 h using the mean when surgeon-specific data were available.25 In addition, linear programming failed for the 49 ± 0.4% of the series of cases containing at least one case that was of a combination of procedures that the surgeon did not otherwise schedule within the 3-yr period.25

A statistic has not yet been identified that performs better than the mean. When surgeon-specific data were used, the 10% trimmed mean or L-estimator42 provided values similar to the sample mean, because 95 ± 0.04% of the combinations of surgeon and scheduled procedure(s) had fewer than 10 occurrences in 3 yr.25 R-estimators with coefficients43 of 0.5, 0.9, or 2.0 did not achieve significantly smaller mean absolute errors in the times to complete series of cases. When the surgeon had not previously scheduled the procedure(s), none of 15 combinations of robust estimators and shrinkage methods significantly reduced the absolute errors of single cases from 1.1 ± 0.04 h.57

Appendix 3: Assigning Cases

Let the case duration of the kth case Xk follow a two-parameter log normal distribution with parameters μk and σk. Represent the initial
period of time for which a case has been ongoing by \( d_k \). Then, the expected value for the time remaining equals

\[
E[X_k - d_k | X_k > d_k] = \frac{\exp(\mu_k + \alpha_k^2/2)}{1 - \Phi\left(\frac{\ln(d_k) - \mu_k - \alpha_k^2/2}{\alpha_k}\right)} - d_k.
\]

Let there be \( N_k \) previous cases of the same scheduled procedure and surgeon, with durations \( x_{k1}, x_{k2}, \ldots, x_{kN_k} \). \( N_k \geq 2 \). Let \( \bar{x}_k \) and \( s_k \) denote the sample mean and SD of \( \ln(x_{k1}), \ln(x_{k2}), \ldots, \ln(x_{kN_k}) \). The maximum likelihood estimators for \( \mu_k \) and \( \sigma_k^2 \) are \( \bar{x}_k \) and \( s_k^2 \).\(^{45}\)

Without loss of generality, suppose that

\[
E[X_2 - d_1 | X_2 > d_1] > E[X_2 - d_2 | X_2 > d_2].
\]

Then, the first case is relieved unless it is discovered during a check that the case will finish within 15 min. For 0.5 h \( \leq d_1 = d_2 \leq 1.5 \) h, there were between 3.4 and 3.8 excess minutes of underutilized OR time per pair of cases \( u_k \) if case durations were known with perfect knowledge.\(^{11}\) If OR times were

\[
t(\text{ttest}), c
\]

0.5 h from Student distribution.

\[X2]\] 2. Let \( x_k \) denote the sample mean of \( x_{k1}, x_{k2}, \ldots, x_{kN_k} \). The studied rule was to move the case if \( \max(0, x_{k1} - d_1 - U_1) + \max(0, x_{k2} - d_2 - U_2) + P_{\text{move}} < \max(0, m_1 + m_2 + x_k - d_1 - U_1) \), and the first case will not finish within 20 min. Depending on the scenario, there were 2.0 - 4.3 excess minutes of underutilized OR time per OR with underutilized OR time \( v_k \) if case durations were known with perfect knowledge.\(^{10}\) Following appendix 3, the excess minutes of underutilized OR time would then be less than 1.6 min per OR per workday, where \( 1.4 \approx 38\% \) of 3.8 min.

Appendix 4: Moving Cases

The first case is ongoing in OR 1 and is scheduled to be followed by the second case, resulting in underutilized OR time. Consider the inter-vening turnover to be the third case. Without loss of generality, assume that OR 2 into which the second case would be moved is just now available for it. Let \( U_1 \) represent the underutilized OR time in OR 1 if the case were moved. Let \( U_2 \) represent the underutilized OR time in OR 2 if the case were not moved. Let \( P_{\text{move}} \) represent the penalty cost in units of underutilized OR time for moving the case. Let \( m_k \) denote the sum of \( x_{k1}, x_{k2}, \ldots, x_{kN_k} \). The studied rule was to move the case if \( \max(0, m_k - d_k - U_1) + \max(0, m_k - d_k - U_2) + P_{\text{move}} < \max(0, m_k - d_k - m_k - d_k - U_1) \), and the first case will not finish within 20 min. Depending on the scenario, there were 2.0 - 4.3 excess minutes of underutilized OR time per OR with underutilized OR time \( v_k \) if case durations were known with perfect knowledge.\(^{10}\) Following appendix 3, the excess minutes of underutilized OR time would then be less than 1.6 min per OR per workday.

Appendix 5: Comparing Durations of Pairs of Cases

Let \( t\text{cum} = (x_k - s_k)/\sqrt{s_k^2(1+1/N_k)} + s_k^2(1+1/N_k) \). Let \( T(t\text{cum}) \) refer to the cumulative distribution function of the Student \( t \) distribution with \( v \) degrees of freedom. The probability that the duration of the next case of the second type exceeds the duration of the next case of the first type equals \( c((N_k - 1)/c, t\text{cum}) + (1 - c)T((N_k - 1)/1 - c), t\text{cum}) \), where \( c = (s_k^2[1 + 1/N_k])/(s_k^2[1 + 1/N_k] + s_k^2[1 + 1/N_k]) \) to within an accuracy of 1.5%.\(^{12}\)

Appendix 6: Upper Prediction Bounds

Estimate the 90\% upper prediction bound for the duration of a future case\(^{7,6}\) by

\[
\exp(\tilde{x} + s \cdot \sqrt{1 + 1/N_k} \cdot T^{-1}(N_k - 1, \tau)),
\]

where \( T^{-1}(N_k - 1, \tau) \) is the \( \tau \)th percentile of the Student \( t \) cumulative distribution function with \( N_k - 1 \) degrees of freedom. Round the final value to the nearest 15 min, achieving a practical time such as 12:30 pm instead of 12:32 pm. For \( \tau = 0.90 \), the resulting bounds are at least as long as their actual duration for 90 \( \pm 0.2\% \) of cases.\(^{8}\)

For pairs of cases, make random draws (\( t_1^{\text{random}} \) and \( t_2^{\text{random}} \)) from student \( t \) distributions with \( N_k - 1 \) and \( N_k - 1 \) degrees of freedom, respectively. Calculate

\[
\exp(\tilde{x} + s_k \cdot \sqrt{1 + 1/N_k} \cdot t_1^{\text{random}} + \exp(\tilde{x} + s_k \cdot \sqrt{1 + 1/N_k} \cdot t_2^{\text{random}}). \]

Repeat thousands of times until the 99\% two-sided confidence interval\(^{45}\) for the 90th percentile for the sum is less than 5 min. Add the mean of the durations of the turnover times. With rounding, the bounds are at least as long as the actual durations for 91 \( \pm 0.6\% \) of pairs of cases.\(^{8}\)

Equation 1 and appendix 5 use two-parameter lognormal distributions classified by surgeon and scheduled procedure(s). Appendix 7 uses two-parameter lognormal distributions classified by scheduled procedure(s). The two-parameter lognormal distribution provides a better fit than normal distributions when classified by procedure and type of anesthetic.\(^{46}\) Concerns about the statistical distribution can be excluded by using the slightly conservative\(^{7} \) distribution-free methods. Sort the historic data in ascending sequence of duration, \( x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[N_k]} \). Let \( L \) equal the smallest integer satisfying \( 0.9 (N_k + 1) \leq L \leq N_k \). Then, the 90\% upper prediction bound is the \( x_{[L]} \) previous case duration.\(^{48}\) Performance is best for \( N_k \geq 19.\(^{7}\)

Appendix 7: Lower Prediction Bounds

Cases were classified based on scheduled procedure(s). The \( \tau = 0.05 \) prediction bounds achieved an actual risk of 5.3 \( \pm 0.1\% \) for the OR staff to wait for the patient.\(^{14}\) For \( N_k \geq 19 \), the achieved risk was 5.0\% \( \pm 0.1\% \). The distribution-free method of appendix 6 was no more accurate. The updated estimate for an ongoing case is

\[
\exp\left(\tilde{x} + s_k \cdot \sqrt{1 + 1/N_k} \cdot T^{-1}(N_k - 1, \tau) + (1 - \tau) \cdot T^{-1}(N_k - 1, 0.2)\right)
\]

For a range of \( d \), the risk ranged from 5.5 to 5.8%.\(^{14}\)

References

5. Dexter F, Traub RD: How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesth Analg 2002; 94:934-42
8. Dexter F, Traub RD, Lebowitz P: Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations. Anesth Analg 2001; 92:945-6
9. Dexter F, Smith TC, Titman DJ, Macario A: Physicians' perceptions of minimum time that should be saved to move a surgical case from one operating room to another: Internet-based survey of the Association of Anesthesia Clinical Directors' (AACD) members. J Clin Anesth 2003; 15:206-10
10. Dexter F: A strategy to decide whether to move the last case of the day in an operating room to another empty operating room to decrease overtime labor costs. Anesth Analg 2000; 91:925-8

Anesthesiology, V 101, No 6, Dec 2004

13. Lebowitz P: Schedule the short procedure first. AORN J 2003; 78:651–8


15. Dexter F, Macario A, Qian F, Traub RD: Forecasting surgical groups' total hours of elective cases for allocation of block time: Application of time series analysis to operating room management. ANESTHESIOLOGY 1999; 91:1501–8


17. Epstein RH, Dexter F: Statistical power analysis to estimate how many months of data are required to identify operating room staffing solutions to reduce labor costs and increase productivity. Anesth Analg 2002; 94:640–3


20. Dexter F, Traub RD, Macario A: How to release allocated operating room time to increase efficiency. Predicting which surgical service will have the most under-utilized operating room time. Anesth Analg 2003; 96:507–12

21. Dexter F, Macario A: Changing allocations of operating room time from a system based on historical utilization to one where the aim is to schedule as many surgical cases as possible. Anesth Analg 2002; 94:1272–79


23. Dexter F, Blake JT, Penning DH, Lubarsky DA: Calculating a potential increase in hospital margin for elective surgery by changing operating room time allocations or increasing nursing staffing to permit completion of more cases: A case study. Anesth Analg 2002; 94:138–42


32. Dexter F, Traub RD, Fleisher LA, Rock P: What sample sizes are required for pooling surgical case durations among facilities to decrease the incidence of procedures with little historical data? ANESTHESIOLOGY 2002; 96:1230–6


37. Macario A, Dexter F: Estimating the duration of a case when the surgeon has not recently performed the procedure at the surgical suite. Anesth Analg 1999; 89:1241–5


47. Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times: Comparison of the lognormal and normal models. ANESTHESIOLOGY 2000; 92:1160–7