Prospective Trial of Thoracic and Spine Surgeons' Updating of Their Estimated Case Durations at the Start of Cases

Elisabeth U. Dexter, MD, FACS,* Franklin Dexter, MD, PhD,† Danielle Masursky, PhD,† and Kimberly A. Kasprowicz, RN, MS‡

BACKGROUND: Surgeon estimates of case durations are important for operating room (OR) management decision making because many cases are rare combinations of procedures with few or no historical data. Thoracic and spine surgeons updated their scheduled OR times on the day of surgery just before the "time out" in the OR.

METHODS: All elective (scheduled) general thoracic ($n = 39$) and spine surgery ($n = 48$) cases at 1 hospital were studied over 3-month and 1.5-month periods, respectively.

RESULTS: Among cases with a change in predicted duration, most changes were made based on updates to the surgical or anesthetic procedures (thoracic 85%, spine 86%). For thoracic surgery, there was overall no significant median reduction in absolute prediction error (median 0 minutes, 95% confidence interval [CI] 0–0 minutes). Among the 37% of cases with changed predicted durations, there was a significant reduction in absolute error (median 38 minutes, 95% CI >7.5 minutes). For spine surgery, there was overall no reduction in the absolute error (median 0 minutes, 95% CI 0–0 minutes). Among the 29% of cases with changed predicted durations, absolute error was no worse, but not significantly better (point estimate of median reduction 34 minutes, 95% CI >0 minutes). Secondary observations made were no effect of updates on bias, frequent rounding of scheduled durations to the nearest half hour, and increased predictive error caused by decisions that reduced expected overutilized OR time.

CONCLUSIONS: A systematic program of routinely and/or always asking for updated case duration predictions will not substantively improve OR management decision making. However, when a change in surgical approach, surgical procedure, or anesthetic procedure is identified (e.g., at the intraoperative briefing before case start), the updated estimate of case duration should be used, because such updates are not worse and often better than original estimates. (Anesth Analg 2010; 110:1164–8)

Surgeon estimates of case durations are important for operating room (OR) management decision making because many cases are rare combinations of procedures with few or no historical data.[1,2] Furthermore, for cases with moderate historical data, surgeon estimates combined with historical case duration data are more accurate than either alone.[1,3] Recently, we performed a systematic literature review to determine what factors affect OR time for general thoracic surgery. In order of importance, they were planned anatomic procedure, planned method and surgical approach, surgical technique, composition of the surgical team, and type of anesthetic.[4] Our recommendation was to obtain an updated estimate of the case duration from the surgeon close to the day of surgery using the most recent information available.

From the *Department of Surgery, State University of New York Upstate Medical University, Syracuse, New York; †Departments of Anesthesia, and Health Management and Policy, University of Iowa, Iowa City, Iowa; and ‡Department of Anesthesiology, State University of New York Upstate Medical University, Syracuse, New York.

Accepted for publication November 19, 2009.

Franklin Dexter is section Editor of Economics, Education, and Policy for the Journal. This manuscript was handled by Steve Shafer, Editor-in-Chief, and Dr. Dexter was not involved in any way with the editorial process or decision.

Address correspondence and reprint requests to Franklin Dexter, MD, PhD, Department of Anesthesia, University of Iowa, Iowa City, IA 52242. Address e-mail to Franklin-Dexter@UIowa.edu or Web site www.FranklinDexter.net.

DOI: 10.1213/ANE.0b013e3181cd6eb9

Updating scheduled case durations on the day of surgery has value if the updates are more accurate than original estimates. More accurate duration prediction can (1) reduce surgeon waiting from scheduled start times among those surgeons with afternoon start times, (2) reduce delays in waiting for equipment being used in 1 OR when needed in another OR, and (3) decrease case handovers to other personnel.[5,6] These decisions are generally made on the day of surgery.[7–9] Because changes in scheduled durations should not affect whether a case is scheduled to be performed on a certain day (see Discussion),[4,5] there is no potential downside to surgeons or patients in providing more accurate estimates on the day of surgery.

We performed a clinical trial of a thoracic surgeon's updates of scheduled case durations on the day of surgery. Four spine surgeons participated as controls for their elective cases, because they were interested in improving the scheduling of their "to-follow add-on" cases. Simultaneously, we evaluated what information the surgeons used to update their scheduled times.

METHODS

This study was performed as a quality improvement project for case duration prediction. However, the data reported in this study were recorded prospectively for publication. The IRB of the State University of New York Upstate Medical University approved the project.

All elective (scheduled) general thoracic and spine surgery cases at 1 hospital were studied for 3 months and 1.5

Table 1. Reasons for Changing Predicted Case Duration

Cause	Thoracic ($N = 35$)	Spine ($N = 48$)	Examples
Surgical procedure or approach	9	12	Not placing gastrojejunostomy feeding tube—patient decided No just before surgery Scheduled right-sided thoracotomy, laparotomy, and revision of gastric pull-up before all imaging results were obtained
Anesthesia procedures in OR	3	2	Thoracic epidural being placed in OR
Either of above	11	14	
Planned wait for pathology	1	0	May add frozen section
Surgical team	1	0	Regular assistant is ill (i.e., not available)
Anesthesia team	1	0	
Anesthesia events other than where procedures done	0	0	
Equipment	0	0	
Nursing team	0	0	

The categories were created from our published systematic literature review[4] and an informal preliminary study during which the thoracic surgeon recorded (voice-mail) qualitative reports after most cases. These categories were complete (e.g., there were no anesthesia events resulting in changed predicted OR time other than where the procedures were performed). The preliminary study suggested the categories of (i) anesthesia procedures performed in versus out of the operating room (OR) used for the case and (ii) planned wait for pathology. For example, because most of the thoracic cases were performed videoscopically, there were only 4 cases for which there was a request for an epidural catheter preoperatively, but among these 4, only 1 was placed while the patient was in the holding area representing a change in practice pattern contemporaneously to the study.

months, respectively. Thoracic surgery was the focus after our systematic review of the factors resulting in inaccurate predictions of OR times for those cases.[4] The median duration of thoracic cases was 3.5 hours (25th percentile 2.0 hours, 75th percentile 4.7 hours). For example, a case with a 3.5-hour OR time included the procedures of bronchoscopy, video-assisted thoracoscopy, and video-assisted right middle lobectomy with mediastinal lymphadenectomy. Elective spine surgery was chosen as a control because the specialty has the smallest percentage prediction error[10] and the spine surgeons at the hospital were concerned with the handling of their add-on cases. The median duration of spine cases was 4.8 hours (25th percentile 3.8 hours, 75th percentile 6.5 hours). For example, a case with a 4.9-hour OR time included the procedures of C6-7 hemicorpectomy, C6-7 anterior cervical fusion with instrumentation, iliac crest bone graft, and intraoperative neurological monitoring.

The scheduled case duration (specifically, estimated time in the OR) was chosen by schedulers weeks in advance of the day of surgery and was used to decide whether a case would be performed on a specific day and in a specific OR. The information system provided the scheduler with the trimmed mean of the most recent 10 cases of the same combination of procedures. Schedulers routinely adjusted this value based on feedback from the surgeon. On the day of surgery, before or while scrubbing, the surgeon was informed by the research nurse of the previously assigned scheduled duration and the assigned personnel. The surgeon was asked to change the predicted duration, if necessary, and if changed, to attribute the change to one or more categories (Table 1). All interviews were performed when the surgeon was out of hearing range of surgical, nursing, and anesthesia team members so that any estimates for change in duration due to one or more team members could be stated without stigma.

A priori, we did not have estimates for the percentage of cases for which the surgeon would change the scheduled duration or for the SD of the resulting changes in the absolute prediction errors. Therefore, 1 author with no involvement in data collection calculated both statistics after the first 20 thoracic cases were completed. The 8 cases with an updated duration had a SD of the change in absolute error of 38 minutes. For an 80% statistical power to detect a 30-minute reduction in the mean absolute error with $\alpha = 0.05$, there needed to be 13 pairwise comparisons. From that value, 32 cases were estimated, where 32 = (13 comparisons with change in duration) × (20 cases/8 changes). It took 2 months to obtain the first 20 elective (scheduled) cases at the study hospital. Thus, the study was continued for 1 month. The final N turned out to be 13 comparisons, as planned, but with a slightly larger (42 minutes) SD of the change in absolute error.

At the time of the power analysis, spine operations had been studied for 2 weeks and 18 cases. We expected that the average absolute errors would be similar for the 2 specialties because general thoracic surgery has larger absolute percentage errors than elective spinal surgery,[10] but shorter average OR times. Given the rate of accrual, the limiting factor for sample size would be the thoracic cases.

Planned and final statistical analysis was based on ranks because there are outliers, probability distributions are mixtures from our studying multiple surgical procedures, and managerial cost is correlated but not proportional to reductions in absolute prediction errors. The permutation test was used to calculate P values for differences in absolute prediction error using all cases of each specialty (StatXact-8, Cytel Software Corp., Cambridge, MA).[11] This test is similar to the Wilcoxon signed-rank test, but includes ties resulting from the cases for which predictions of case durations were unchanged. The ties neither count in favor of the treatment nor in favor of the null hypothesis, but are included in the effective sample size. These and all other P values were 2-sided and calculated using exact methods. If we had limited consideration to the 1-sided test of improvement, then our inclusion of ties would have been in favor of the null hypothesis of no improvement. The corresponding 95% confidence intervals (CIs) for medians were calculated using the Clopper-Pearson method.[12,13] Such CIs are conservative (i.e., wider than necessary), but can be calculated including the ties. When analyzing only those cases for which the predicted durations were changed, the Wilcoxon signed-rank test was used. The

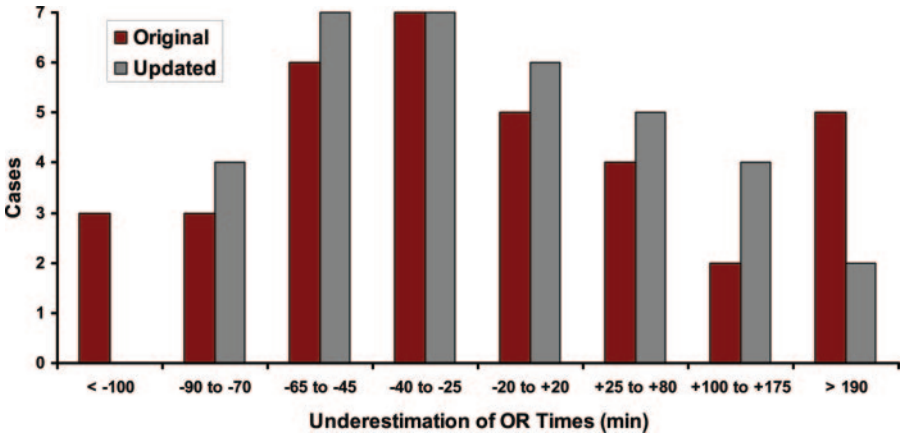

Figure 1. Thoracic surgery cases' original scheduled operating room (OR) times versus updated estimates. The histogram shows underestimates of actual case durations.

corresponding CIs for median differences were calculated using Hodges-Lehmann estimation.[11] Incidences of different reasons for changing predicted durations were compared with 50% using the binomial test, with 50% corresponding to "most cases."

RESULTS

Among the cases with a change in predicted duration, most changes were made based on different surgical or anesthetic procedures (thoracic 85%, $P = 0.022$; spine 86%, $P = 0.013$). Such changes would be known at the intraoperative briefing performed before case start.

For thoracic surgery, there was a statistically significant reduction in the absolute error in case duration prediction ($P = 0.045$, Fig. 1). However, the effect was modest, with median reduction of 0 minutes (95% CI 0–0 minutes). Whereas the P value of 0.045 reflected results for all cases, including the 37% (13 of 35) with changes in predicted durations on the morning of surgery, the median of 0 minutes was influenced only by the majority of cases that were unchanged (63%, 95% CI >48%). Among cases with a change in predicted duration, there was a statistically significant reduction in absolute error ($P = 0.039$, point estimate of median 38 minutes, 95% CI >7.5 minutes).

For spine surgery, there was no significant reduction in the absolute error in case duration prediction ($P = 0.109$, median 0 minutes, 95% CI 0–0 minutes). Most cases did not have a change in predicted case duration (71%, 34 of 48, 95% CI >59%). Among the 29% of cases with changes in predicted durations on the morning of surgery, there was a trend toward reduction in absolute error ($P = 0.093$, point estimate of median 34 minutes, 95% CI >0 minutes). As for thoracic surgery, updated predictions were not worse than originally scheduled durations (see Discussion).

Secondary Observations

Overall, bias (e.g., persistent underestimation) was unaffected by updates (i.e., the number of minutes for which cases took longer or shorter than predicted were the same) (Fig. 1). For thoracic surgery, there was no significant reduction overall ($P = 0.879$, point estimate of median reduction in underestimation 0 minutes, 95% CI 0–0 minutes) or among cases with changes in predicted duration ($P = 0.894$, median 0 minutes, 95% CI −52 to +45 minutes). For spine surgery, the results were the same overall ($P = 0.346$, median 0 minutes, 95% CI 0–0 minutes) and among cases with changes in predicted duration ($P = 0.317$, point estimate 15 minutes, 95% CI −30 to +60 minutes).

The updated predictions of case duration could have been provided in any combination of hours and minutes (e.g., 3:57 if a surgeon wanted). Nonetheless, all of the 35 thoracic cases were originally scheduled for full hours or in increments of half hours (i.e., with times ending in":00" or ":30" and 31 of the 35 were scheduled this way when updated). The other 4 updated times were scheduled for increments of quarter hours (i.e., with times ending in ":15" or ":45," such as 4:45). Among the spine cases, all 48 of the updated times were recorded as full hours or half hours (i.e., with times ending in ":00" or ":30"). For each case, we calculated the difference of the actual OR time and the time rounded down to the nearest whole hour. The differences followed a uniform distribution between 0 and 59 minutes (Kolmogorov-Smirnov test, $P = 0.387$). Thus, surgeons' behavior of choosing times in 30-minute intervals resulted in an average absolute error of 7.5 minutes. This value can be compared with the thoracic cases' revised estimates' median absolute error of 54 minutes (95% CI 41–79 minutes) and the spine cases' median absolute error of 57 minutes (95% CI 47–80 minutes). Thus, rounding was substantial, but alone accounted for only approximately 13% of the absolute error.

As explained in the Discussion, we would expect that a common reason for a surgeon changing the estimate of OR time before the intraoperative "time out" would be the unexpected placement of regional anesthetic catheters, etc., in the OR. This was as expected and accounted for 17% (5 of 29) of the changes in predicted duration (Table 1). At facilities where block or holding areas are never used for procedures, the incremental benefit of updating case duration predictions would be slightly less.

DISCUSSION

Our results show that a systematic program of routinely asking for updated predictions of case durations will not substantively improve OR management decision making. However, when a change in surgical approach, surgical procedure, or anesthetic procedure is identified (e.g., at the intraoperative briefing or "time out" before case start), the

updated estimate of case duration should be used, because such updates are not worse and often better than original estimates. Because such cases are ongoing (i.e., the patient is in the OR), the estimates can be useful for predicting the times remaining in ongoing cases.[2] For such updates to guide managerial decision making, every 5 minutes or so[14] the estimated times remaining need to be recalculated automatically, which *does not* mean progressive subtractions of 5 minutes.[2]

Improving case duration prediction on the day of surgery for thoracic and spine surgery matters principally because inaccuracy results in excess overutilized OR time and thus influences add-on case scheduling.[5,15,16] Therefore, the objective is not to reduce absolute error per se but to reduce overutilized OR time. Ideally, organizations would use updated estimates of case duration to make decisions that reduce overutilized OR time,[2,5,17,18] and in practice they do.[19,20] For example, suppose that there was only 1 procedure area for regional blocks, arterial lines, etc., to be placed before induction. That area should be used preferentially for cases for which reducing OR time would reduce overutilized OR time.[5,17,21] Such selective choices can substantially affect OR throughput because, if staffing is being planned appropriately, only approximately one-third of ORs should have overutilized OR time daily.[22–24] Therefore, we expected that a common reason for a surgeon changing the estimate of OR time before the intraoperative "time out" would be unexpected placement of regional anesthetic catheters, etc., in the OR. This was precisely what was observed (Table 1).

Contrary to our hypothesis from systematic literature review,[4] and findings for laparoscopic surgery that more nursing handoffs are associated with longer OR times,[25] we found no prospective predictive ability of surgical, nursing, or anesthesia teams on scheduled duration (Table 1). This result highlights that studies of case durations with team data available in retrospect can be useful for policy (e.g., reduce nursing handoffs), while not for decision making on the day of surgery. Results might differ if add-on cases done by different teams were included.

Studying updates on the day of surgery, compared with days or weeks ahead, was not a limitation. Before the day of surgery, the economically important decisions involving case duration predictions involve whether to perform a certain case on a certain day in a certain OR.[5,16] Provided estimates are unbiased, or have been adjusted,[6,26,27] accuracy of estimation of case duration has little effect on that decision, especially for specialties such as thoracic and spine surgery with long case durations.[5,15,16,23,28,29] Briefly, the reason for this result is that for elective (scheduled) cases, the sum of the underestimation of OR time among all cases is attributable predominantly to small amounts of underestimation for nearly all cases, not to the few cases with large underestimation of case duration.[27] Consequently, it is irrelevant for the decision of the OR in which a case will be performed whether the case will be scheduled for 3.5, 4.0, or 4.5 hours.[5,24]

Although our study was limited to thoracic and spine surgery, and included only a few surgeons, the results were effectively the same for the 2 specialties, even though their proportional predictive errors are very different.[10] The surgeons, including the first author of this report, were motivated to participate in the project to improve their case duration predictions. Nonetheless, despite potential bias to reduce absolute error more than would be observed routinely, the updated predictions were overall (i.e., including cases without changes) no worse. If our results include the influence of a positive performance bias and/or Hawthorne effect, then our results would reinforce our conclusion that routinely trying to obtain updated case duration predictions is unlikely to be of economic value.

Based on these findings, we recommend that researchers in OR management and clinical OR managers focus their efforts on decision support to assure meaningful responses to the uncertainty on the day of surgery.[2,5,17,18] This recommendation to focus on response rather than prevention is important because by definition it involves reliance on the demonstrated utility of updated forecasts of the time remaining in ongoing cases.[2]

AUTHOR CONTRIBUTIONS

All authors were involved in study design and review of the manuscript. EUD, DM, and KAK helped in data acquisition; FD helped in data analysis; and EUD, FD, and DM helped in manuscript preparation.

DISCLOSURE

The University of Iowa assists hospitals and anesthesia groups with statistical analyses of operating room information system data. However, Franklin Dexter receives no funds personally other than his salary from the State of Iowa, including no travel expenses or honoraria, and has tenure with no incentive program.

ACKNOWLEDGMENTS

The authors appreciate the assistance of the State University of New York Upstate spine surgeons: Drs. Eric Deshaies, William Lavelle, Michael Sun, and Richard Tallarico.

REFERENCES

1. Dexter F, Ledolter J. Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historical data. Anesthesiology 2005;103:1259–67
2. Dexter F, Epstein RH, Lee JD, Ledolter J. Automatic updating of times remaining in surgical cases using bayesian analysis of historical case duration data and "instant messaging" updates from anesthesia providers. Anesth Analg 2009;108:929–40
3. Wright IH, Kooperberg C, Bonar BA, Bashein G. Statistical modeling to predict elective surgery time. Anesthesiology 1996;85:1235–45
4. Dexter F, Dexter EU, Masursky D, Nussmeier NA. Systematic review of general thoracic surgery articles to identify predictors of operating room case durations. Anesth Analg 2008; 106:1232–41
5. Dexter F, Epstein RD, Traub RD, Xiao Y. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. Anesthesiology 2004;101:1444–53
6. Wachtel RE, Dexter F. Reducing tardiness from scheduled start times by making adjustments to the operating room schedule. Anesth Analg 2009;108:1902–9
7. Moss J, Xiao Y. Improving operating room coordination: communication pattern assessment. J Nurs Adm 2004;34: 93–100
8. Nemeth C, O'Connor M, Klock PA, Cook R. Discovering healthcare cognition: the use of cognitive artifacts to reveal cognitive work. Org Stud 2006;27:1011–35

9. Riley R, Forsyth R, Manias E, Iedema R. Whiteboards: mediating professional tensions in clinical practice. Commun Med 2007;4:165–75
10. Dexter F, Dexter EU, Ledolter J. Influence of procedure classification on process variability and parameter uncertainty of surgical case durations. Anesth Analg 2010;110:1155–63
11. Sprent P. Applied Nonparametric Statistical Methods. London: Chapman & Hall, 1989:15–7, 30–1, 37–40
12. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26: 404–13
13. Hahn GJ, Meeker WQ. Statistical Intervals. A Guide for Practitioners. New York: Wiley, 1991:82–4, 100–5
14. Epstein RH, Dexter F, Ehrenfeld JM, Sandberg WS. Implications of event entry latency on anesthesia information management decision support systems. Anesth Analg 2009;108:941–7
15. Dexter F, Macario A, Traub RD. Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Use of bin packing algorithms and fuzzy constraints in operating room management. Anesthesiology 1999;91: 1491–500
16. 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;94:933–42
17. Dexter F, Willemsen-Dunlap A, Lee JD. Operating room managerial decision-making on the day of surgery with and without computer recommendations and status displays. Anesth Analg 2007;105:419–29
18. Dexter F, Lee JD, Dow AJ, Lubarsky DA. A psychological basis for anesthesiologists' operating room managerial decision-making on the day of surgery. Anesth Analg 2007;105:430–4
19. Wachtel RE, Dexter F. Influence of the operating room schedule on tardiness from scheduled start times. Anesth Analg 2009;108:1889–901
20. Stepaniak PS, Mannaerts GH, de Quelerij M, de Vries G. The effect of the operating room coordinator's risk appreciation on operating room efficiency. Anesth Analg 2009;108:1249–56
21. O'Sullivan CT, Dexter F. Assigning surgical cases with regional anesthetic blocks to anesthetists and operating rooms based on operating room efficiency. AANA J 2006;74:213–8
22. Strum DP, Vargas LG, May JH, Bashein G. Surgical suite utilization and capacity planning: a minimal cost analysis model. J Med Syst 1997;21:209–22
23. McIntosh C, Dexter F, Epstein RH. Impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: tutorial using data from an Australian hospital. Anesth Analg 2006; 103:1499–516
24. Pandit JJ, Dexter F. Lack of sensitivity of staffing for 8 hour sessions to standard deviation in daily actual hours of operating room time used for surgeons with long queues. Anesth Analg 2009;108:1910–5
25. Cassera MA, Zheng B, Martinec DV, Dunst CM, Swanstrom LL. Surgical time independently affected by surgical team size. Am J Surg 2009;198:216–22
26. Olivares M, Terweisch C, Cassorla L. Structural estimation of the newsvendor model: an application to reserving operating room time. Manage Sci 2008;54:41–55
27. Dexter F, Macario A, Ledolter J. Identification of systematic under-estimation (bias) of case durations during case scheduling would not markedly reduce over-utilized operating room time. J Clin Anesth 2007;19:198–203
28. Dexter F, Macario A, Lubarsky DA, Burns DD. Statistical method to evaluate management strategies to decrease variability in operating room utilization. Application of linear statistical modeling and Monte-Carlo simulation to operating room management. Anesthesiology 1999;91:262–74
29. Dexter F, Macario A, O'Neill L. Scheduling surgical cases into overflow block time—computer simulation of the effects of scheduling strategies on operating room labor costs. Anesth Analg 2000;90:980–6