Validity and usefulness of a method to monitor surgical services' average bias in scheduled case durations

[Validité et utilité d'un monitorage du biais moyen des services chirurgicaux dans la durée prévue des cas réglés]

Franklin Dexter MD PhD,* Alex Macario MD MBA,† Richard H. Epstein MD,‡ Johannes Ledolter PhD§

Purpose: Unbiased prediction of case durations is an integral part of matching operating room (OR) staffing to workload. Monitoring systematic bias in surgeons' scheduled case durations can identify those services with estimates sufficiently inaccurate that statistical analysis of historical data may be useful in preference to the surgeons' estimates. We describe a method to monitor surgical services' average bias in scheduled case durations.

Methods: Actual case duration, predicted (scheduled) case duration, and service were obtained for all 58,291 cases during 39 four-week periods at an academic hospital. For each four-week period, a ratio was computed for each service. The numerator for each service equalled the sum of the differences in minutes between actual case duration and scheduled case duration. The denominator equalled the sum in hours of the actual durations of all of the service's cases. The ratio was multiplied by eight hours to yield the number of minutes of underestimated case duration per eight hours of OR time during the four-week period.

Results: The ratios followed a normal distribution for each service. Using the Student's t distribution, the 95% lower confidence bounds for the average underestimate of case duration ranged from three to 65 min per eight hours of used OR time.

Conclusions: To reduce over-utilized OR time, we recommend monitoring each service's 95% lower confidence bound of the bias in scheduled case durations. For services consistently underestimating their case durations, schedule their cases using statistical estimates of case durations based on their historical data, and disregard their own estimates.

Objectif : *La prédiction non biaisée de la durée des cas fait partie de l'attribution du personnel à la salle d'opération (SO). La mesure du biais systématique de la durée des opérations réglées peut indiquer les services dont les estimations sont suffisamment imprécises pour qu'une analyse statistique des données historiques soit plus utile que les prédictions des chirurgiens. Nous décrivons une méthode pour mesurer le biais moyen de la durée des opérations réglées.*

Méthode : *La durée réelle, la durée prévue des cas réglés et le service ont été obtenus à partir des 58 291 cas vus pendant 39 périodes de quatre semaines dans un hôpital universitaire. Pour chaque période, un ratio a été calculé. Le numérateur pour chaque service égalait la somme des différences en minutes entre la durée réelle et prévue. Le dénominateur égalait la somme en heures des durées réelles de tous les cas du service. Le ratio a été multiplié par huit heures pour avoir le nombre de minutes sous-estimées par huit heures de temps de SO pendant la période.*

Résultats : *Les ratios présentaient une distribution normale pour chaque service. Selon la distribution t de Student, la valeur la plus faible de l'intervalle de confiance de 95 % pour la moyenne estimée de la durée des cas allait de trois à 65 min par 8 h d'usage du temps de SO.*

Conclusion : *Pour réduire le dépassement de temps de SO, nous recommandons le monitorage de la valeur inférieure de l'intervalle de confiance de 95 % du biais dans la durée des cas réglés pour chaque service. Si la sous-estimation est régulière il vaut mieux utiliser des prédictions statistiques de la durée fondées sur les données historiques plutôt que des estimés personnels.*

UNBIASED prediction of case durations is an integral part of matching operating room (OR) staffing to workload. For example, of the 28 ORs at a hospital's main and ambulatory facilities, 23 ORs were allocated until 3:30 p.m. Two ORs were allocated until 6 p.m., and two ORs were allocated until 11 p.m. One OR was planned for trauma. OR allocations dictated staffing for nurses and anesthesia provider teams. Nonetheless, during the late afternoons, many more ORs were in use than were allocated (Figure). The primary cause was elective cases taking longer than scheduled, as 95% of cases were elective between 7 a.m. and 6 p.m. Thus, surgeons were underestimating their case durations to get their cases to "fit" into their allocated OR time. Resulting over-utilized OR time (e.g., after 3:30 p.m. for 23 ORs) increased direct labour costs, as well as costs from frustrated employees leaving due to the unpredictability of the workday.

Using historical data only for scheduling cases can be counter-productive. The goal of this study was to measure surgical services' average bias in scheduled case durations. Bias indicates whether the estimates are consistently too high or consistently too low. Monitoring of bias in scheduled case durations can identify those services for which the OR manager should consider having case durations chosen by analysis of historical data. In the Discussion, we explain the rationale for using this management intervention selectively.
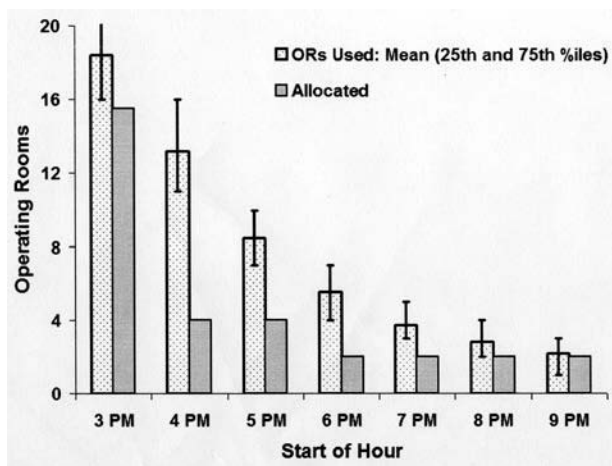

FIGURE Operating rooms allocated *vs* used in afternoons Mondays to Fridays. The data match those used to generate data in the Table. The figure shows the mean daily number of operating rooms in use for at least part of the hour starting at the listed times. The error bars show the 25th percentiles and 75th percentiles of the daily number of operating rooms in use, respectively.

Methods

Definitions

The following definitions were used throughout our paper, and determined precisely what was monitored.

Service refers to a group of surgeons who share allocated OR time, whether an individual surgeon, a group, a specialty, a division, or a department. For example, if a busy surgeon were personally allocated eight hours of OR time every Monday, the surgeon would be a service. At the hospital studied in this paper, each of the 13 surgical services was an academic division with three to dozens surgeons.

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

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 scheduled delays between cases.

OR workload is total hours of cases including turnover times.

Over-utilized OR time is the positive difference between a service's OR workload and its allocated OR time.[1] For example, an OR was allocated from 7 a.m. to 3 p.m. The first patient entered the OR at 7 a.m. The OR was busy until finishing at 5 p.m. The OR workload was ten hours. There were two hours of over-utilized OR time.

Under-utilized OR time is the positive difference between allocated OR time and OR workload.[1] For example, OR time is allocated from 7 a.m. to 4 p.m. The last case of the day ends at 2 p.m. There are two hours of under-utilized OR time.

Inefficiency of use of OR time equals the sum of two products: hours of under-utilized OR time multiplied by the labour cost per hour of under-utilized OR time plus hours of over-utilized OR time multiplied by the labour cost per hour of over-utilized OR time.[1]

OR efficiency is the value that is maximized when the inefficiency of use of OR time has been minimized. This paper considers only OR allocations and case scheduling based on OR efficiency. A review was published recently.[2]

Statistical method for monitoring bias in scheduled case durations

The actual case duration, scheduled case duration, and service were obtained for all 58,291 cases performed between September 1, 2001 and August

31, 2004. An analysis was performed for each service. Methodological design is considered in the Discussion.

There were 39 four-week periods of data. For each four-week period, a ratio was computed for each service. The numerator in minutes equalled the sum of the differences between actual case duration and scheduled case duration for each of the service's cases during the four-week period. The denominator in hours equalled the sum of the durations of all cases performed by the service during the four-week period. The ratio was multiplied by eight hours. The result was the four-week period's bias in the service's scheduled case durations reported per eight hours of used OR time (see Discussion).

Validity of using this method for confidence intervals relies on knowing the statistical distribution. Applying Lilliefors test (StatXact-6, Cytel Software Corporation, Cambridge, MA, USA) to the three-year data set let us test whether the statistical distributions of the estimates were consistent with a normal distribution. We expected a normal distribution to be valid, because the central limit effect applies, in that each estimate was a standardized sum of many individual differences.

Usefulness of the method was evaluated by determining whether the method detects underestimates of case duration, when expected. This could be evaluated at the study hospital, because there were more ORs in use in the afternoons than were allocated (Figure). We calculated the mean and standard deviation of the 39 estimates for the minutes of bias in scheduled case durations reported per eight hours of OR time. From them, we applied Student's t distribution to calculate 95% lower and upper confidence limits.

Results

Systematic underestimation of case durations can contribute to unnecessary hours of over-utilized OR time. The 95% lower confidence bounds for the average bias in scheduled case durations were underestimates ranging from three to 65 min per eight hours of used OR time (Table). This range shows that the method can be useful, in that differences from zero were detected, and those differences varied among services.

A *P*-value < 0.05 in the Lilliefors test would indicate lack of consistency with a normal distribution. This was an issue for just one of 13 services, with a *P*-value of 0.04 (Table). This *P*-value close to 0.05 is adequate, because confidence intervals for the mean are robust to small deviations from normality. With Bonferroni correction, all *P* values would be > 0.05, further supporting use of a normal distribution.

Discussion

One of the two components of inaccurate case scheduling is the bias (i.e., the systematic underestimation or overestimation of case durations). The other component is the absolute error (below). We developed and validated a method to measure the bias in scheduled case durations per eight hours of used OR time.

When and how to use the monitoring method

Near zero bias provides for a near zero difference between actual and forecasted OR workloads. OR allocations are based on OR workload forecasted using historical workload data.[1–4] Since the cost per hour of over-utilized OR time exceeds the cost per hour of under-utilized OR time, OR allocations should exceed average OR workloads.[1] Case scheduling should then aim to achieve an unbiased estimator

TABLE Average underestimate of case duration (min) per eight hours of used operating room time for each of 13 surgical divisions at an academic hospital

Service	Lower 95% confidence interval	Mean	Standard deviation	Upper 95% confidence interval	Consistent with normal distribution P-value
L	65	72	19	78	0.190
M	52	56	11	59	0.290
D	50	57	22	64	0.620
B	46	55	27	64	0.960
G	38	47	30	57	0.960
C	34	40	17	45	0.470
H	34	39	16	44	0.580
I	32	41	26	49	0.490
E	32	46	41	59	0.044
J	25	36	35	47	0.830
F	18	26	24	34	0.970
K	11	24	41	37	0.950
A	3	11	25	19	0.690

of the forecasted OR workload used to create the allocations.[5] Although this argument applies universally, our method's usefulness is limited to facilities with systematic underestimation of case durations resulting in over-utilized OR time.

If bias is present, the Figure shows that, first and foremost, OR allocations should be matched to the existing workload, thereby increasing OR efficiency (see definitions). Clearly, and importantly, the facility studied has the staffing for larger OR allocations without increasing capacity, because the cases were, in fact, performed.

We used the 95% lower confidence bound for the mean bias in scheduled durations, rather than the mean, so that decision-making would not be spurious simply due to expected statistical variation. Although we used 39 four-week periods to establish validity of the method, in practice, smaller sample sizes (e.g., six four-week periods) may be used so that estimates are based on more recent behaviour of services. The sum of the differences between actual and scheduled case duration would be calculated, as well as the minutes of OR time. From the mean and standard deviation of the six ratios of the sums, Student's t distribution would be used to find the lower confidence bound for the mean bias in scheduled case durations per eight hours of OR time. If this value were large, then for that service we recommend using statistical analysis of their historical data for scheduling case duration (below). A reasonable maximum value for "large" is 15 min per eight hours of OR time. Such a cost in time exceeds that of commonly used OR management interventions (e.g., reducing turnover times,[6] improving add-on case scheduling with closed circuit cameras and real-time patient tracking systems,[2] and adjusting timing of the release of allocated OR time).[7]

Near zero bias can be achieved in practice through the use of historical case duration data for case scheduling and/or having schedulers and surgeons motivated to be accurate. We recently reviewed methods for case duration prediction.[2]

The problem of systematic bias in scheduled case durations could be avoided altogether by scheduling all cases solely based on statistical estimates of historical case duration data. There are four reasons why we do not recommend this approach. Each reason (below) is related to the resulting detrimental impact on the precision of the estimate as indicated by the absolute error (i.e., magnitudes of the absolute values of the errors of the estimates). An increase in absolute error can reduce OR efficiency, and result in increased patient and surgeon waiting time on the day of surgery.[2] Thus, we recommend monitoring the 95%

lower confidence of the bias. Use historical data alone only for services that are consistently underestimating their case durations.

First, although statistical methods can be near perfect for preventing bias in scheduled case durations (below), they still have large absolute errors. For example, in a previous study we found that an ambulatory surgery facility had a mean absolute error per case of 0.4 hr, and a hospital 0.8 hr per case.[2,5]

Second, statistical methods have particularly large absolute errors when there are few historical case duration data for the scheduled procedure(s) of the case. At tertiary hospitals, more than half of cases are of procedures scheduled by the surgeon less than three times during the preceding year.[8] More than 20% of cases are of procedures or combinations of procedures performed previously only once or twice.[8,9] Results are similar for ambulatory surgery centres.[10] In our experience having inadequate historical data on which to base the prediction is a poorly appreciated issue.

Third, reduced absolute error is achieved when the surgeon adjusts the statistically-derived estimate up or down to reflect case complexity.[11,12]

Fourth, good statistical methods have slightly less absolute error than do surgeons who are providing case duration estimates without being prompted with their historical data.[11] However, simplistic methods built into some commercial software perform poorly compared to surgeons who are motivated to be accurate.[11,13]

Rationale for methodological design

Both elective and urgent cases were analyzed. To understand why, consider a hypothetical surgical suite with several urgent cases that can wait without increasing patient risk. The suite's policy is to sequence urgent cases from shortest to longest scheduled duration, thereby reducing average patient and surgeon waiting times.[2] Assume that one service does not specify case duration estimates for its non-elective cases. Since historical case duration data are used by the OR information system, the average bias in scheduled case durations for those cases is zero minutes. Assume that a second service routinely schedules its cases with durations that are one hour too brief. That service's surgeons underestimate their case durations to get their cases a higher priority in the queue for urgent cases. This example shows that bias in scheduled case durations should be analyzed not just for elective cases, but also urgent cases.

Totals for each four-week period were analyzed, for three reasons. First, the analysis was not applied to each case, because there were some outliers in differences between scheduled and actual durations. Second,

the analysis was not applied to each day, because the OR workloads of services varied markedly from day to day. On some days, certain services did not operate at all, regardless of whether just elective or non-elective cases were considered. Yet, all services had many hours of cases for each four-week period. Third, the principal determinant of OR workload by service is day of the week.[1] Vacations, meetings, etc., are often two weeks long. The use of four-week periods has consistently been sufficiently long to eliminate these effects.[6,14,15]

Bias was reported per eight hours of OR time, not per case, as the latter would not capture the economically relevant issue. To illustrate why, we use the following example. Two services both underestimate their case durations by an average of 15 min per case. One service performs one very long case a day in its OR, resulting in an average of 15 min of underestimated time per eight hours of allocated OR time. The other service performs 12 short cases a day in its OR, resulting in three hours of underestimated time per eight hours of allocated OR time. Even though the allocated OR time should exceed the average workload, three hours of underestimated time will surely result in over-utilized OR time.

Bias was not reported as minutes per day, as explained by the following hypothetical example. Two services both underestimate their case durations by an average of 15 min per day. One service is a group of three surgeons who fill an OR each afternoon for four hours. The other service is a department that fills five ORs for ten to 12 hr each day. An extra 15 min per day is quite different for the two services, because excess labour costs are proportional to baseline staffing.

Conclusions

We established the validity and usefulness of monitoring the lower 95% confidence bound of the mean bias (usually underestimate) in scheduled case durations per eight hours of OR time. For services consistently underestimating their case durations, cases can be scheduled using statistical estimates of case durations based on their historical data.

References

1 *Strum DP, Vargas LG, May JH.* Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. Anesthesiology 1999; 90: 1176–85.

2 *Dexter F, Epstein RH, 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.

3 *Dexter F, Traub RD, Macario A.* How to release allocated operating room time to increase efficiency: predicting which surgical service will have the most underutilized operating room time. Anesth Analg 2003; 96: 507–12.

4 *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.

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: 933–42.

6 *Dexter F, Abouleish AE, Epstein RH, Whitten CW, Lubarsky DA.* Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. Anesth Analg 2003; 97: 1119–26.

7 *Dexter F, Macario A.* When to release allocated operating room time to increase operating room efficiency. Anesth Analg 2004; 98: 758–62.

8 *Zhou J, Dexter F, Macario A, Lubarsky DA.* Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late. J Clin Anesth 1999; 11: 601–5.

9 *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.

10 *Dexter F, Macario A.* What is the relative frequency of uncommon ambulatory surgery procedures in the United States with an anesthesia provider? Anesth Analg 2000; 90: 1343–7.

11 *Wright IH, Kooperberg C, Bonar BA, Bashein G.* Statistical modeling to predict elective surgery time. Comparison with a computer scheduling system and surgeon - provided estimates. Anesthesiology 1996; 85: 1235–45.

12 *Shukla RK, Ketcham JS, Ozcan YA.* Comparison of subjective versus data base approaches for improving efficiency of operating room scheduling. Health Serv Manage Res 1990; 3: 74–81.

13 *Sorge M.* Computerized O.R. scheduling: is it an accurate predictor of surgical time? Can Oper Room Nurs J 2001; 19: 7–18.

14 *Dexter F, Epstein RH.* Optimizing second shift OR staffing. AORN J 2003; 77: 825–30.

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.