

Forecasting and Perception of Average and Latest Hours Worked by On-Call Anesthesiologists

Franklin Dexter, MD, PhD*

Richard H. Epstein, MD†

Richard L. Elgart, MD†

Johannes Ledolter, PhD‡

BACKGROUND: We studied the value of providing information to anesthesia providers about the length of time typically worked during on-call shifts. The mean time at which a shift ends can be used for purposes of trades, payments, or reverse auctions, because the mean is proportional to the total time. The 80th percentile (with a suitable upper confidence limit for uncertainty due to limited sample sizes) can be used for judging the earliest time by which after-work activities reasonably can be planned.

METHODS: (A) Three years of operating room (OR) information system data were analyzed. Dependent variables were the earliest times when the numbers of ORs running were always ≤ 6 ORs, ≤ 4 ORs, and ≤ 2 ORs. We progressively built linear regression models for each of the three dependent variables using day of the week, scheduled number of cases, scheduled hours of cases (including turnovers), and linear time trend. Calculations were repeated after excluding residuals. Calculations were repeated using regression trees. (B) Anesthesiologists were surveyed about their perceptions of the mean and 80th percentiles.

RESULTS: (A) For the three thresholds and two end points (mean and 80th percentile), differences among days of the week were as large as 45 min. Differences between end points for the same weekdays were as large as 245 min. Comparatively, additional knowledge about the number or hours of cases provided in the late afternoon on the working day before surgery reduced the mean absolute error by only 4.1–6.0 min. Results were insensitive to a variety of analytic methods. Information available more days before the day of surgery (e.g., 1 wk) would have had even less incremental predictive value. (B) The mean absolute error of anesthesiologists' estimates for 80th percentiles was 60 min, principally because of underestimation of the 80th percentiles. More than half (69%, $P = 0.0003$) of anesthesiologists' estimates for 80th percentiles had error >30 min, whereas errors of this magnitude were less for the mean (44%, $P = 0.0004$).

CONCLUSIONS: Historical data from OR or anesthesia information management systems, or from anesthesia billing systems, can be used months before staff scheduling to provide insight to anesthesia providers on respective calls. The data are useful because experience provides limited intuition. Updates on scheduled workload available closer to the day of surgery provided only marginal increases in knowledge over the use of historical data.

(Anesth Analg 2009;109:1246–52)

Months before the day of surgery, an anesthesia group can use statistical methods to determine its staffing needs.¹ Often, the numbers of operating rooms (ORs) to be run at certain times on workdays are specified contractually with hospital administration, ideally based on statistical assessments of historical workload.^{1,2} Staff scheduling is then done in

accordance with the group's staffing obligations.^{3–6} When anesthesiologists are given the choice to work different numbers of hours for different compensation, they choose to work significantly different total hours.⁷ Therefore, if information about hours worked when on-call were provided to anesthesia providers when they make their scheduling decisions, such information would be useful if new and accurate.

In this article, we study the potential value of providing information to anesthesia providers about

From the *Division of Management Consulting, Departments of Anesthesia and Health Management and Policy, University of Iowa, Iowa; †Department of Anesthesiology, Jefferson Medical College, Philadelphia, Pennsylvania; and C. Maxwell Stanley Professor, Department of Management Sciences, University of Iowa, Iowa.

Accepted for publication May 22, 2009.

Franklin Dexter is the 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.

The University of Iowa performs statistical analyses for hospitals and anesthesia groups. Dr. 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. RHE is President of Medical Data Applications, Ltd., whose CalculatOR™ software includes the analyses considered in this article.

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

Copyright © 2009 International Anesthesia Research Society
DOI: 10.1213/ane.0b013e3181b0ffcc

the length of time typically worked during available shifts. The mean time at which a shift ends can be used for purposes of trades, payments, or reverse auctions,⁸ because the mean is proportional to the total time. The 80th percentile (with a suitable upper confidence limit for uncertainty due to limited sample sizes) can be used for judging the earliest time by which other activities reasonably can be planned (e.g., attending a child's school activity or asking one's spouse to wait for dinner).

One way we assessed the value of providing the mean and 80th percentile was to survey anesthesiologists at a department to learn whether they have accurate (≤ 30 min) perceptions of the mean and the 80th percentile. If this were the situation, then providing this information would not be of value. We hypothesized that most people would have reasonably good insight about the mean but poor insight into the 80th percentile of the time that they work on different shifts.

The other way we assessed the value of the statistics was to test whether information available from the OR scheduling system can provide managerially useful updates of the mean and 80th percentile. If this were the situation, then providing the mean and 80th percentile just before staff scheduling would be of less value, and vice versa. The additional information available about OR scheduling that we studied were those known when the date of surgery is selected for each case, specifically the total numbers of cases scheduled and the total hours of cases scheduled. These data have previously been shown to be valid predictors of total workload of individual facilities for trends over time and for variation among workdays.⁹⁻¹¹ Information on assignment of cases to specific ORs was not used because at the tertiary facilities for which our study is relevant, case assignments are often not made until the afternoon before surgery and even thereafter are often changed. For example, at one such facility, $25.7\% \pm 0.2\%$ (standard error) of cases were moved or added-on after the "final" schedule was posted the day before surgery,¹² and those cases affect the hours worked by the people on-call.

METHODS

Statistical Analysis of Historical Data

The times at which cases ended were obtained from the study hospital's OR information system for its main OR suites, which exclude non-OR anesthetics, labor and delivery, and gastrointestinal endoscopy procedure rooms. The dates studied were 39 4-wk periods from Monday December 5, 2005, to Friday November 28, 2008. Holidays and days with (planned) exceptionally low caseloads (e.g., between Christmas and New Year) were excluded. Cardiac and hepatic transplantation cases were excluded, because the studied hospital has separate anesthesiologists and OR nurses who care for these patients.

Table 1. Mean and 80th Percentile of Earliest Times When Hospital has Specified Numbers of Operating Rooms in Use

ORs	Day of the week	Mean	95% UCL of 80th%ile	80th%ile
≤ 2 ORs	Monday	7:45 PM	9:45 PM	9:10 PM
	Tuesday	8:10 PM	9:45 PM	9:30 PM
	Wednesday	7:35 PM	9:30 PM	9:00 PM
	Thursday	8:25 PM	10:00 PM	9:30 PM
	Friday	8:15 PM	10:00 PM	9:45 PM
≤ 4 ORs	Monday	6:30 PM	7:45 PM	7:30 PM
	Tuesday	6:50 PM	8:00 PM	8:00 PM
	Wednesday	6:20 PM	7:30 PM	7:15 PM
	Thursday	7:00 PM	8:15 PM	8:00 PM
	Friday	6:45 PM	8:15 PM	7:45 PM
≤ 6 ORs	Monday	5:45 PM	6:30 PM	6:15 PM
	Tuesday	6:05 PM	7:15 PM	6:45 PM
	Wednesday	5:30 PM	6:30 PM	6:15 PM
	Thursday	6:05 PM	7:00 PM	6:45 PM
	Friday	5:55 PM	6:45 PM	6:45 PM
≤ 16 ORs	Monday	3:40 PM	4:30 PM	4:15 PM
	Tuesday	3:45 PM	4:45 PM	4:30 PM
	Wednesday	3:15 PM	4:15 PM	4:00 PM
	Thursday	3:40 PM	4:30 PM	4:30 PM
	Friday	3:35 PM	4:30 PM	4:15 PM

The statistical computations above were made directly from the raw data (i.e., not involving the linear regressions in Tables 2 and 3). Although we rounded the mean to the nearest 5 min in the table distributed to the attending anesthesiologists at the study hospital, for the calculations in the Results, we report the trimmed mean to the nearest minute. For our implementation at the study hospital, we reported the UCL as the "80th percentile," with the details given in a footnote explaining the uncertainty due to the sample sizes. In Table 1, we additionally included the point estimate of the 80th percentile as comparison to match Figure 1. $N = 144$ Mondays, 151 Tuesdays, 155 Wednesdays, 153 Thursdays, and 152 Fridays. The overall $N = 755$ listed in Table 1 = $144 + 151 + 155 + 153 + 152$ workdays. In the last rows, 16 ORs are studied as a control for the average end of the day at the hospital, with its 32 ORs (see Discussion).

ORs = operating rooms; mean = trimmed mean, with 10% trimmed from the lower and upper values before the mean was calculated¹³; 95% UCL of 80th%ile = refers to the 95% upper confidence limit (UCL) of the 80th percentile.

The dependent variables studied were the earliest times when a member of the anesthesia group's call team could leave the hospital (Table 1). These were defined in terms of the earliest times when there were:

1. ≤ 2 ORs, when the 2nd call person can leave.
2. ≤ 4 ORs, when the OR director who starts work at 11 AM can leave.
3. ≤ 6 ORs, when the 3rd call person working since 7 AM leaves.

To obtain the dependent variables, the numbers of cases running during each 15 min increment were calculated including turnovers, limiting consideration to turnovers 60 min or less. Turnovers were included because anesthesia providers are working during these intervals to transfer care of the prior patient and to setup for the next case. Considering each 15 min interval backward from midnight, we determined for each day the time after which no more ORs than specified by a threshold were in use. Figure 1 shows an example of the time in hours after 12 noon when the number of ORs running was equal to or less than the threshold of two ORs.

Two summary statistics of the three dependent variables were calculated (see Introduction, Table 1).

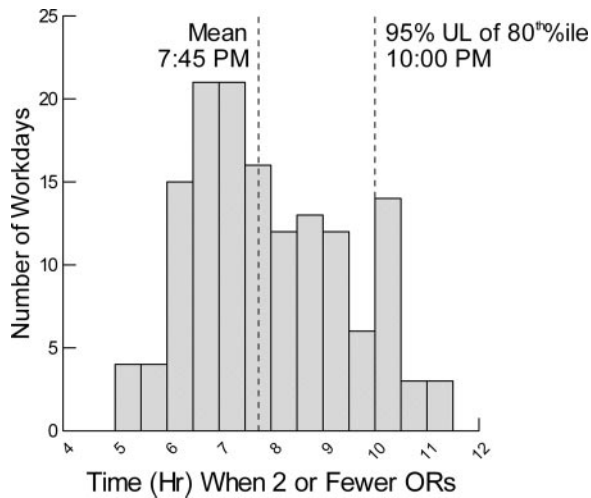


Figure 1. Histogram for the earliest time (number of hours after 12 noon) after which the hospital always had ≤ 2 operating rooms (ORs) in use on Mondays. The figure shows no particular shape and little smoothness, explaining our calculation of percentiles by the empirical method. The $n = 144$ Mondays' values were sorted in ascending sequence. The "mean" of 7:45 PM refers to the trimmed mean, with the 10% lowest and largest values trimmed before the mean was calculated.¹³ For the data in Figure 2, the difference between the trimmed mean and the mean was just 1 min. The 95% upper confidence limit (bound) of the 80th percentile was the 125th value, which corresponds to the 87th percentile.^{14,15}

The mean is proportional to total hours worked and is thus relevant to auctioning,⁸ paying for, selling, and/or trading shifts. However, because the sample mean was sensitive to the outliers present for some end points on some weekdays, we report the trimmed mean, with 10% trimmed from each end.¹³ To incorporate statistical error for the 80th percentile, we report the 95% upper confidence limit (bound), as calculated using the Clopper-Pearson method.^{14,15} We studied the 80th percentile for two reasons. First, the corresponding incremental cost of working later than this time corresponds to $4 \times$ regular hourly rate,¹⁶⁻¹⁸ which is very high. Second, the 80th percentile was simple to explain intuitively as it implies that the anesthesiologists assigned to the shift will work later than this time only once every 5 days (i.e., once a week). Note that with our 3 years of data, the 95% upper prediction limit for the 80th percentile was essentially the same as reporting the 90th percentile (Fig. 1, Table 1).

We progressively built linear regression models for each of the three dependent variables using day of the week, scheduled number of cases, scheduled hours of cases including turnovers, and linear time trend (Table 2) (Systat 12, Systat Software, San Jose, CA). Our model used only the independent variables on the working day before surgery, because by then almost all elective cases had been scheduled (Fig. 2) and at that time maximum information was available (see Discussion). To determine the predictive value of the independent variables, we calculated the deleted residuals (Table 3). These are the differences between

Table 2. Linear Statistical Model for Earliest Times After 12 Noon at Which the Study Hospital has the Specified Ranges of Operating Rooms Still in Use

	≤ 2 ORs	≤ 4 ORs	≤ 6 ORs
Day of the week			
<i>F</i> statistic ^a	18.4	23.9	35.5
<i>P</i> value	<0.0001	<0.0001	<0.0001
<i>F</i> statistic, no outliers	18.5	24.0	39.6
<i>P</i> value, no outliers	<0.0001	<0.0001	<0.0001
Scheduled numbers of cases at 6 PM the working day before surgery			
<i>F</i> statistic ^b	20.0	27.2	40.7
<i>P</i> value	<0.0001	<0.0001	<0.0001
<i>F</i> statistic, no outliers	17.3	20.8	46.2
<i>P</i> value, no outliers	<0.0001	<0.0001	<0.0001
Scheduled hours of cases and turnovers at 6 PM the working day before surgery			
<i>F</i> statistic	7.3	14.0	27.5
<i>P</i> value	0.0072	0.0002	<0.0001
<i>F</i> statistic, no outliers	7.5	17.3	27.1
<i>P</i> value, no outliers	0.0062	<0.0001	<0.0001
Four-week period (i.e., linear trend)^c			
<i>F</i> statistic	6.9	3.6	3.0
<i>P</i> value ^d	0.0087	0.056	0.086
<i>F</i> statistic, no outliers	7.4	3.5	2.0
<i>P</i> value, no outliers	0.0065	0.061	0.15

The $N = 755$ workdays without exclusion of outliers. With exclusion of outliers, the sample sizes are $N = 753, 751,$ and 750 . The (first) degrees of freedom for the *F* tests are 4 for day of the week and 1 for the other variables. The R^2 for the three columns were 0.15, 0.21, and 0.30 for $\leq 2, \leq 4,$ and ≤ 6 ORs, respectively. The R^2 is larger (0.53) for ≤ 16 ORs.

OR = Operating room.

^a For 0-2, 0-4, and 0-6 ORs including outliers, $P \geq 0.17$ for interaction of day of the week with scheduled numbers of cases. There were four degrees of freedom.

^b All *F* tests in the table are partial (Type III) tests. The $P \geq 0.11$ for addition of quadratic term.

^c All three estimates of the linear trend were positive. All estimated coefficients for the scheduled numbers and hours of cases were positive. We investigated several other models of changes over time including those with seasonal variation. Almost all terms were statistically significant ($P < 0.001$) but reduced the absolute forecasting error (Table 3) by an increment of ≤ 1 min.

^d When performed using regression trees,[†] none of the generated models included a trend component. [†] Please see Results section.

predicted and observed values, with each prediction made using the linear model without the observed value.¹⁹ For example, each single value was predicted using the linear model with parameters estimated from the other 754 numbers.*

Among the 2265 combinations of 755 dates and three end points, there were 12 with externally studentized residuals¹⁹ larger than three. We repeated all calculations excluding these 12 outliers as one of our sensitivity analyses.

As another sensitivity analysis, we repeated the regression analyses using a different type of regression, namely regression trees with a least absolute values loss function (Systat 12). Regression trees are automatically created for predicting relationships including interactions. We used least absolute values as the loss function so that the mean absolute error could be compared directly with that of the preceding classical multivariable linear regression. To compute a

*For readers who are used to using stepwise and all variables regression, the deleted residuals are those used to calculate the prediction sum of squares.

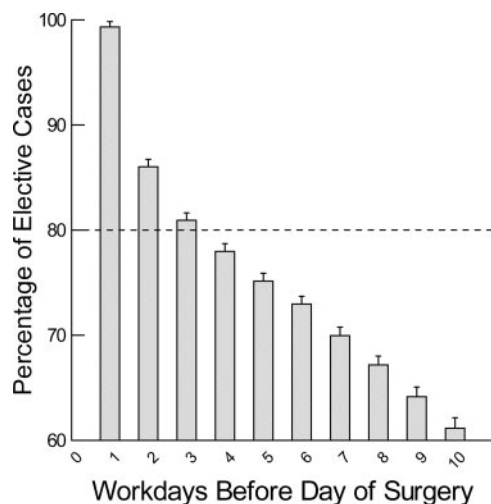


Figure 2. Percentages of elective cases scheduled on the specified numbers of workdays before the day of surgery. The mean number of cases scheduled at 6:00 PM the listed number of workdays before surgery was calculated for each 4-wk period. The ratio was taken for each 4-wk period between that value and the mean number of those cases performed on the day of surgery. The mean of the 39 ratios represents the height of the bars. The 95% confidence intervals were calculated by comparing the ratios among the 39 4-wk periods and using Student's *t* distribution.

Table 3. Reduction in Absolute Forecast Error by Using the Full Linear Statistical Model from Table 2 Versus Simply the Mean for Each Day of the Week

	Difference of absolute values of deleted residuals (minutes) (mean \pm SE)	
	All cases	Excluding outliers
≤ 2 ORs	3.7 \pm 1.0	3.8 \pm 1.0
≤ 4 ORs	4.4 \pm 0.9	4.3 \pm 0.9
≤ 6 ORs	5.6 \pm 0.9	5.7 \pm 0.9

The absolute values of the deleted residuals represent the magnitudes of the forecast errors. Two statistical models were used. One model was the full statistical model of Table 2. The other model included only "Day of the Week" from the top of Table 2. For both statistical models, deleted residuals were calculated (i.e., each residual was calculated using a regression model fit excluding that one datum).¹⁹ Absolute values were taken for both deleted residuals. The differences were calculated between each pair. The means and the SE of the pairwise differences are given above. In the first column of results, all $N = 755$ days. For corresponding analysis of ≤ 16 ORs, the reduction in absolute errors was larger (11.5 ± 1.2 min). In the second column, the $N = 753, 751,$ and $749,$ respectively. The performance in the second column was deliberately biased to appear better than managerially appropriate because the outliers that were excluded would be of particular concern to the people working those especially long hours.

OR = operating room.

regression tree for each of the three dependent variables, and to maximize the potential predictive ability of the model, we used all independent variables in Table 1 plus all possible sums and differences of those variables. Jackknife estimation was used to calculate the standard error of the reduction in the absolute error.

As a final sensitivity analysis, we repeated the linear regression analyses using a different dependent variable, specifically, the time after which there were 16 or fewer ORs in use. We studied 16 ORs because, each day, there are eight anesthesiologists who are

assigned to work after 4 PM, if necessary, and each typically medically directs two ORs. We used this secondary end point of the earliest times when there were ≤ 16 ORs to explore our hypothesis that if the independent variables were poor predictors of the three primary dependent variables, the cause of poor prediction would be that they better predict when many anesthesiologists are working. If most of the add-on cases are performed by the three on-call anesthesiologists and most of the elective (scheduled) cases are performed by the other anesthesiologists, then the numbers and hours of elective cases would poorly predict the hours worked by the three anesthesiologists who are on-call.

Survey of Anesthesiologists

Although we obtained data on three dependent variables, for the survey of anesthesiologists we considered just the smallest (time when ≤ 2 ORs) and largest (time when ≤ 6 ORs), thereby reducing the number of questions (Table 4). The survey was performed at the study hospital during a 2-wk period. One of the anesthesia clinical directors asked 32 of 42 anesthesiologists to participate, with 10 missed due to their being on vacation, at a meeting, or working outside the main ORs. All 32 anesthesiologists invited completed the survey but with a few missing values. The two anesthesiologists who were aware of the survey design were excluded. The survey was performed by those two anesthesiologists at the study hospital as a quality improvement project to help them decide whether to post the historical data and to assist in their explaining to the department how the information can be used. The results in this article were implemented at the hospital the week after the survey was completed. The statistical analysis was comparison of observed proportions to $1/2$ (StatXact-8, Cytel Software, Cambridge, MA).

RESULTS

The study hospital has call team members leaving when the numbers of ORs in use are ≤ 6 ORs, ≤ 4 ORs, and ≤ 2 ORs. The three thresholds studied were the times when these numbers of ORs were always in use. For the three thresholds and two end points of mean and 95% upper confidence limit of the 80th percentile, differences among days of the week were as large as 45 min (Table 1). Differences between end points for the same weekdays were as large as 245 min. Comparatively little additional knowledge was available in the late afternoon on the working day before surgery, as the mean absolute error was reduced by only 4.1–6.0 min (Table 3, Fig. 3).[†] Information available more days before the day of surgery (e.g., 1 wk) would have had even less incremental predictive value.

[†]Using regression trees, the mean \pm standard error of the reductions in the absolute error were $0.00 \pm 0.00, 0.00 \pm 0.00,$ and 0.08 ± 0.01 min for ≤ 2 ORs, ≤ 4 ORs, and ≤ 6 ORs, respectively.

Table 4. Survey Completed by 32 Anesthesiologists at Study Hospital

Late running operating rooms QI Survey

I am interested in your perception as to the time of day when the 2nd call and 3rd call attendings finish their last case in the OR. (Note: This is not the time they leave the hospital, but when their last case exits the OR.)

Please provide your best guess to all 8 questions below without consulting other attendings and return the survey to me.

If you are never scheduled to take 2nd or 3rd call, please check the appropriate box and skip the corresponding questions. If you are scheduled to take these calls but often trade them away, please fill in the estimated times.

Thanks.

<Name of clinical director was here>

I am never scheduled to take 2nd call

I am never scheduled to take 3rd call

1. On an average Monday, when does 2nd call finish his or her last case in the OR? _____ PM

2. On an average Thursday, when does 2nd call finish his or her case in the OR? _____ PM

3. On an average Monday, when does 3rd call finish his or her last case in the OR? _____ PM

4. On an average Thursday, when does 3rd call finish his or her last case in the OR? _____ PM

For the following 4 questions, rather than the average time, we are interested in the 80th percentile time. This corresponds to the 2nd latest time of every 5 calls (4/5 = 80%). For example, if a set of 5 ordered times were:

1) 09:00 2) 09:07 3) 09:23 4) 09:44 5) 10:15 the 80th percentile would be 09:44, since that is the 2nd latest time.

Please supply your best guess. Everyone knows that for Mondays and Thursdays the 80th percentile is later than 4:00 PM and earlier than 11:59 PM.

5. On Mondays, what is the 80th percentile time when 2nd call finishes his or her last case in the OR? _____ PM

6. On Thursdays, what is the 80th percentile time when 2nd call finishes his or her last case in the OR? _____ PM

7. On Mondays, what is the 80th percentile time when 3rd call finishes his or her last case in the OR? _____ PM

8. On Thursdays, what is the 80th percentile time when 3rd call finishes his or her last case in the OR? _____ PM

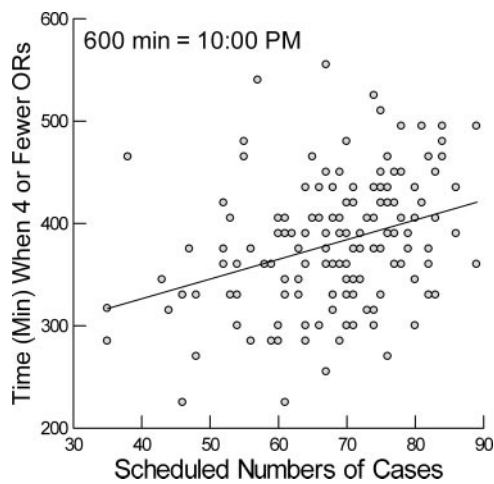


Figure 3. Example of the relationship between the time when a specified number of operating rooms (ORs) or fewer are in use and the scheduled numbers of cases. In the figure, four or fewer ORs are in use at the listed hours past 12 noon on Wednesdays. The scheduled numbers of cases are the number as of 6:00 PM the working day before surgery. The figure shows the marked scatter quantified in Table 3 by the small reduction in the mean absolute residual error. The mean \pm SE are 51.6 ± 3.7 min for the absolute values of the deleted residuals calculated using just the mean as a predictor and 47.2 ± 3.4 min using the linear model. Although the regression is statistically significant ($P < 0.00001$), the pairwise reduction in the absolute values of the residuals was only 4.4 ± 2.2 min.

The mean absolute error of anesthesiologists' estimates for 80th percentiles was 60 min, principally because of underestimation of the 80th percentiles (Table 5). As hypothesized, more than half (69%, $P = 0.0003$) of anesthesiologists' estimates for 80th percentiles had error >30 min, whereas errors of this magnitude were less for the mean (44%, $P = 0.0004$).

DISCUSSION

Months in advance, anesthesia providers at hospitals frequently choose and trade shifts and make other arrangements for their call days. Anesthesiologists' intuitions were inaccurate as to the hours that they work (Table 5). Any anesthesia group that bills for its time or tracks its cases has all of the data required for the forecasts, just as they do for specialty-specific staffing.^{1,2,16} The numbers of simultaneous anesthetics (i.e., ORs in use) at each time can be calculated from the anesthesia start and end times. The results show essentially negligible value to the use of data within 1 day of surgery (e.g., number or hours of cases actually scheduled) other than historical data for the particular day of the week from an anesthesia billing,¹ OR information, or anesthesia information management system. The results match previous findings that if months ahead the durations of OR workdays are forecasted appropriately for each specialty, there is financially negligible incremental value for reducing anesthesia group costs by using additional information to even perfectly predict case durations.^{16,20-23}

Although the incremental value of additional data was strongly statistically significant (i.e., $P < 10^{-6}$) (Table 2), the absolute reductions in forecasting errors were small relative to the choice of the statistic to report (Table 3, Fig. 3). Furthermore, these absolute reductions likely overestimate the actual value of using such information in decisions. The issue is one of trust in technology. When a prediction for how late people work is given before staff schedules are made, by definition the estimate is retrospective and provides little information about what the individual will experience on one specific future evening. Providing the estimate is similar to a web site that gives the

Table 5. Survey Results

Statistic	ORs in use ("call")	Day of the week	Threshold from table 1	Absolute difference >30 min (95% lower bound) ^a	Mean underestimate ± standard error (min)	Number ^b
80th percentile	≤2 ORs (2nd call)	Monday	9:10 PM	74% (≥57%)	62 ± 11	23
	≤6 ORs (3rd call)	Thursday	9:30 PM	91% (≥77%)	81 ± 10	23
		Monday	6:15 PM	50% (≥32%)	-27 ± 12	24
		Thursday	6:45 PM	62% (≥45%)	3 ± 12	24
			69% (≥61%) ^c	29 ± 7 ^d	94	
Average	≤2 ORs (2nd call)	Monday	7:45 PM	30% (≥17%)	-8 ± 9	23
	≤6 ORs (3rd call)	Thursday	8:25 PM	48% (≥30%)	24 ± 10	23
		Monday	5:45 PM	58% (≥41%)	-36 ± 12	24
		Thursday	6:05 PM	38% (≥22%)	-29 ± 11	24
			44% (≥35%) ^e	-13 ± 6	94	

OR = operating room.

^a Absolute difference was taken between the response and the threshold in the preceding column. The column lists the percentage of anesthesiologists for whom the absolute difference exceeded 30 min.

^b Among the 32 responding anesthesiologists, there were eight who did not take 2nd call and eight who did not take 3rd call. The N = 23 for 2nd call (0-2 ORs) because one anesthesiologist who reported taking 2nd call but did not give those answers.

^c The mean absolute errors equaled 60 ± 5 min for the 80th percentile and 40 ± 4 min for the average.

^d The comparison was made with the point estimate of the 80th percentile, as in the far right-hand column of Table 1. If we had instead used the 95% upper confidence limit for the 80th percentile, then there would have been even more underestimation, and thus the absolute error would have been even larger.

^e Overall percentage was significantly higher for 80th percentile (69%) than for average (44%) (P = 0.0004, Barnard's test for the difference of proportions, StatXact-8, Cytel Software, Cambridge, MA).

average January low and high temperature in Orlando, FL. For a traveler contemplating a vacation there in January, this information may be useful. In contrast, making a prediction on January 16 for the low temperature the next day runs the risk of people perceiving psychologically that "weather forecasts are unreliable." This is because of a human bias to convert probabilities of events (40% chance of temperature <70°C) into binary perceptions (e.g., "the actual temperature was 65°C, so the meteorologist was wrong").

Our independent variables captured information about the variables relevant to the working day before surgery. By studying data at this point in time, we likely overestimated the potential value of the information practically useful for regression analysis, because some decisions may need to be made sooner. For example, when staffing (OR allocation) and case scheduling decisions are made based on maximizing the efficiency of use of OR time, most services fill their allocated time no sooner than 3 days before the day of surgery.^{16,24} At that time, 81% of cases have been scheduled (Fig. 2). By studying data from the day before surgery, we likely underestimated the predictive value relative to using information from the day of surgery. However, by the day of surgery, an anesthesiologist's desire to work or not to work late for pay is generally irrelevant as to whether the anesthesiologist works late (i.e., they are on-call and have to stay until the cases are done). We showed previously that, under such circumstances, compensation is no longer salient.²⁵ Thus, we cannot identify an argument for why providing updated information on the day of surgery related to the predicted end of shift times would be useful.

Although different anesthesia directors running the control desk make different decisions that affect how

late ORs are in use, we did not include individuals as independent variables.²⁶ Who will be scheduled to manage the desk would not be known when an anesthesiologist is choosing his or her schedule. Furthermore, we could not envision informing a person on call that tomorrow she will likely work 30 min later because the person at the control desk is Dr. Smith not Dr. Lawrence. We think that behavioral differences of individuals running a control desk show the value of managerial decision-support systems making recommendations to reduce such variability.^{23,27} We think that they also show why anesthesia providers will need and/or want to continue to be paid hourly when working late, as compared with occasionally working long scheduled shifts.²⁵

Our data analysis applies only to hospitals with add-on cases and cases being moved among ORs. If every case were known at least 1 wk in advance, there were no add-on cases, each surgeon has a list of cases for the day, and the cases are not moved among ORs, then results would be different. The survey was performed by one anesthesia group immediately before implementation. Although we have no reason to suspect that results would differ among groups, we have no supportive data.

In conclusion, a department's survey of its anesthesiologists identified a 1 h mean absolute error in estimates for the latest (80th percentile) time that they work when on-call, with more than half (69%) of estimates incorrect by >30 min. Because the two end points of mean and 95% upper confidence limit of the 80th percentile differed substantively, different decisions should rely on different statistics. Because the two end points differed among days of the week, statistics should be reported by weekday. This information can be calculated from anesthesia billing data

(or equivalent sources¹) and given to anesthesia providers months ahead before staff scheduling. Additional knowledge available on the working day before surgery (e.g., numbers of scheduled cases) had a statistically significant benefit but reduced the mean absolute error by only 4.1–6.0 min extra. The results are especially useful when interpreted along with recent behavioral studies of anesthesiologists' managerial decisions late in afternoons.^{25,26}

REFERENCES

- Dexter F, Epstein RH. Optimizing second shift OR staffing. *AORN J* 2003;77:825–30
- Dexter F, Epstein RH. Calculating institutional support that benefits both the anesthesia group and hospital. *Anesth Analg* 2008;106:544–53
- Ernst EA, Lasdon LS, Ostrander LE, Divell SS. Anesthesiologist scheduling using a set partitioning algorithm. *Comput Biomed Res* 1973;6:561–9
- Ernst EA, Matlak EW. On-line computer scheduling of anesthesiologists. *Anesth Analg* 1974;53:854–8
- Holley HS, Heller F. Computerized anesthesia personnel system. *Int J Clin Monit Comput* 1988;5:103–10
- Dexter F, O'Neill L. Weekend operating room on-call staffing requirements. *AORN J* 2001;74:666–71
- Miller RD, Cohen NH. The impact of productivity-based incentives on faculty-based compensation. *Anesth Analg* 2005;101:195–9
- De Grano ML, Medeiros DJ, Eitel D. Accommodating individual preferences in nurse scheduling via auctions and optimization. *Health Care Manag Sci*. In press. doi: 10.1007/510729-008-9087-2
- 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
- Moore IC, Strum DP, Vargas LG, Thomson DJ. Observations on surgical demand time series: detection and resolution of holiday variance. *Anesthesiology* 2008;109:408–16
- Masursky D, Dexter F, O'Leary CE, Appleget C, Nussmeier NA. Long-term forecasting of anesthesia workload in operating rooms from changes in a hospital's local population can be inaccurate. *Anesth Analg* 2008;106:1223–31
- Wachtel RE, Dexter F. Influence of the operating room schedule on tardiness from scheduled start times. *Anesth Analg* 2009;108:1889–901
- Staudte RG, Sheather SJ. Robust estimation and testing. New York: Wiley, 1990:104–6
- Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; 26:404–13
- Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17: 857–72
- 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
- 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
- Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning. A minimal cost analysis model. *Anesthesiology* 1999;90:1176–85
- Neter J, Wasserman W, Kutner MH. Applied linear statistical models. 3rd ed. Homewood, IL: Irwin, 1990:398–400, 450–1
- 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
- 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
- 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
- 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
- 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
- Masursky D, Dexter F, Garver MP, Nussmeier NA. Incentive payments to academic anesthesiologists for late afternoon work did not influence turnover times. *Anesth Analg* 2009;1622–6
- 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
- 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