# Bayesian Prediction Bounds and Comparisons of Operating Room Times Even for Procedures with Few or No Historic Data 

Franklin Dexter, M.D., Ph.D.,* Johannes Ledolter, Ph.D. $\dagger$


#### Abstract

Background: Lower prediction bounds (e.g., for fasting), upper prediction bounds (e.g., to schedule delays between sequential surgeons), comparisons of operating room (OR) times (e.g., when sequencing cases among ORs), and quantification of case uncertainty (e.g., for sequencing a surgeon's list of cases) can be done accurately for combinations of surgeon and scheduled procedure(s) by using historic OR times. The authors propose that when there are few or no historic data, the predictive distribution of the OR time of a future case be centered at the scheduled OR time, and its proportional uncertainty be based on that of other surgeons and procedures. When there are a moderate or large number of historic data, the historic data alone are used in the prediction. When there are a small number of historic data, a weighted combination is used. Methods: This Bayesian method was tested with all $\mathbf{6 5 , 6 6 1}$ cases from a hospital. Results: Bayesian prediction bounds were accurate to within $2 \%$ (e.g., the 5\% lower bounds exceeded $4.9 \%$ of the actual OR times). The predicted probability of one case taking longer than another was estimated to within $0.7 \%$. When sequencing a surgeon's list of cases to reduce patient waiting past scheduled start times, both the scheduled OR time and the variability in historic OR times should be used together when assessing which cases should be done first. Conclusions: The authors validated a practical way to calculate prediction bounds and compare the OR times of all cases, even those with few or no historic data for the surgeon and the scheduled procedure(s).


OPERATING room (OR) operational decisions on the day before and on the day of surgery rely on the uncertainty in the estimates of OR times, particularly those decisions that affect patient and surgeon waiting times. ${ }^{1}$
For example, the decision as to when a patient should stop drinking and be ready on the day of surgery relies on the shortest expected time(s) for the preceding case(s) in the patient's OR. ${ }^{1,2}$ A lower prediction bound for the OR time of a case is the value that will be exceeded by the next randomly selected case of the same type at the

[^0]specified rate. There is a $5 \%$ chance that the OR time of a case will be briefer than its 5\% lower prediction bound. ${ }^{1,2}$
For example, suppose that two surgeons are scheduled in the same OR on the same day, and the OR workday is not filled. To reduce the expected tardiness of the start of the second surgeon, a delay can be scheduled between the cases, thereby reducing the expected waiting time of the second surgeon if the first surgeon were to finish late. An appropriate delay relies on the longest time the cases are likely to take (e.g., their $90 \%$ upper prediction bounds). ${ }^{1,3}$
For example, a surgeon is doing the first case of the day in OR 1, and a different surgeon is doing the first case in OR 2. Several add-on cases have been submitted for the day, the longest of which needs the microscope being used by the surgeon in OR 1 . There are no other cases in OR 2. The OR 2 is scheduled to finish before any other OR and has staffing planned for the same hours. To reduce expected overutilized OR time, the longest add-on case (i.e., the one with the microscope) should be ${ }^{1}$ the add-on case performed in OR 2. That decision is good provided the case in OR 1 will finish before the first case of the day in OR 2. That probability depends not only on the expected OR time of the cases, but also on the uncertainty of the estimates. ${ }^{1,4}$
For example, consider the sequencing of a surgeon's list of cases in the same OR on the same day. To reduce the mean tardiness of cases, cases with short predictable OR times should be performed before cases that are longer or have large uncertainty in OR times. ${ }^{1,5}$ A brief mediastinoscopy should be performed before esophagectomy. The unpredictability of the preceding case in an OR can be quantified by its probability of finishing more than 1 h late. ${ }^{6}$

For these four and other common probabilistic problems, when there are historic data available for the surgeon and the procedure(s) scheduled, the previously developed statistical methods are very accurate. The percentages of cases exceeded by lower and upper prediction bounds match those specified within $1 \%{ }^{1}$ For example, if the $90 \%$ upper prediction bound for the OR time of a newly scheduled case is 3.0 h , the actual probability of that case having an OR time longer than 3.0 h is likely somewhere between $90 \%$ and $91 \%$. The calculated probability of one case lasting longer than another case matches the actual probability within $1 \%{ }^{1}$ For example, if the calculated probability that one case
will take longer than another case is $67 \%$, the actual probability is likely somewhere between $66 \%$ and $68 \%$.
However, these methods cannot be applied to the approximately $22 \%$ of cases of a procedure(s) that the surgeon has not scheduled at least twice before. ${ }^{6}$ This limitation is commonplace. Among outpatient cases in the United States with an anesthesia provider, $20 \%$ were of a procedure(s) performed four times or less per workday nationwide, and $36 \%$ of cases were of a procedure(s) likely performed an average of less often than once per facility per year. ${ }^{7}$ At private hospitals, there can be more than 5,600 surgeon preference cards (i.e., at least 5,600 combinations of surgeon and procedure(s)). ${ }^{8}$ At academic hospitals, there can be more than 13,000 surgeon preference cards. ${ }^{9}$ More than $75 \%$ of procedure(s) can be performed just once or twice annually. ${ }^{10}$ More than half of cases can be of a procedure(s) scheduled by the surgeon less than three times per year. ${ }^{11}$
The $22 \%$ of cases with few or no data disproportionately affect decision making under uncertainty. These cases tend to be distributed randomly among cases at facilities. If any one such case is part of a series of cases, then the time to complete that series of cases cannot be estimated using the previously described ${ }^{2,3,6,4}$ methods. The more cases performed per OR per day, the less useful the existing science is.
Previous efforts to address this challenge of cases with few or no historic data focused on using data classified by procedure(s) alone, or with type of anesthesia. ${ }^{2,12,13}$ Despite some encouraging results, ${ }^{2,12,13}$ these methods have problems in implementation.
First, pooling procedure(s) among surgeons or facilities ${ }^{10}$ provides only modest gains, because the problem is generally not that a surgeon schedules a procedure(s) that usually is performed by another surgeon at the facility, but rather that many procedure(s) are rare. ${ }^{10,14}$ There are thousands of Current Procedural Terminology or International Classification of Diseases' procedure codes, and surgeons persistently produce new technologies.
Second, performing analyses with data both pooled and not pooled by surgeon can result in inconsistent decisions, making instructions for clerks confusing. For example, consider a mean of 3.0 h and $90 \%$ upper prediction bound of 5.5 h for 11 cases scheduled by 5 surgeons versus a mean of 2.0 h and $90 \%$ upper prediction bound of 4.0 h for 2 cases of the surgeon scheduling the new case. Qualitative instructions cannot reconcile such results.
Third, studies show consistently that the surgeon is a strong predictor of OR time, second only to the procedure(s). ${ }^{15}$ Some surgeons are consistently slower or faster than others. ${ }^{15}$ Therefore, excluding classification by surgeon appropriately reduces face validity of recommendations.
In this article, we propose and validate a solution: Bayesian statistical methods. When there are no historic data, the predictive distribution of the OR time of a future case is centered at its scheduled OR time. In
addition, its proportional uncertainty is based on that of other surgeons and procedures. When there are a moderate or large number of historic data, those data alone are used in the prediction. When there are a small number of historic data, a weighted combination of the scheduled OR time and historic data are used.

## Methods

## Data Set Used

The data set included the OR times of all 65,661 cases performed between January 1, 1996, and December 31, 1999, at an academic hospital's tertiary surgical suite and ambulatory surgery center. Procedures were as defined by the Current Procedural Terminology code(s). There were 19,838 different combinations of surgeon, scheduled procedure(s), and presence or absence of an anesthesia provider. We used data analyzed previously, so results ${ }^{2,3,6,4,10}$ can be compared. Scheduled ( $v s$. actual) procedure code(s) were used, because for a future case for which a prediction bound is being calculated, only the scheduled procedure(s) would be known. ${ }^{16}$

## Lower and Upper Prediction Bounds

Let the random variable $X_{k}$ refer to the natural logarithms of OR times classified by ${ }^{15}$ the $k$ th combination of surgeon, scheduled procedure(s), and anesthetic, $k=1$, $2, \ldots, p$. For brevity, we henceforth refer to each of the $p$ combinations as "surgeon and procedure(s)." The $n_{k}$ previously observed (historic) OR times in hours for the $k$ th combination of surgeon and procedure(s) are $\exp \left(x_{k 1}\right), \exp \left(x_{k 2}\right), \ldots, \exp \left(x_{k n_{k}}\right)$. The sample mean of the $n_{k}$ historic data $x_{k 1}, x_{k 2}, \ldots, x_{k n_{k}}$ equals $\bar{x}_{k}$, and the sample variance equals $\hat{\sigma}_{k}^{2}$. The logarithm of the scheduled OR time in hours for the next case is $x s_{k}^{*}$. The objective is to predict the OR time of the next case, $\exp \left(X_{k}^{*}\right)$. We use * to represent the next case throughout the article.

From equations 8 and 17 in the appendix, the $5 \%$ lower Bayesian prediction bound equals

$$
\begin{equation*}
\exp \left(\mu_{k}^{*}+t_{0.05,2 \alpha_{k}^{*}} \cdot \sqrt{\frac{\beta_{k}^{*}}{\alpha_{k}^{*}} \cdot \frac{1+\tau+n_{k}}{\tau+n_{k}}}\right) \tag{1}
\end{equation*}
$$

where

$$
\begin{gather*}
\mu_{k}^{*}=x s_{k}^{*} \cdot \frac{\tau}{\tau+n_{k}}+\bar{x}_{k} \cdot \frac{n_{k}}{\tau+n_{k}},  \tag{2}\\
\alpha_{k}^{*}=\alpha+\frac{n_{k}}{2}  \tag{3}\\
\beta_{k}^{*}=\beta+\frac{1}{2} \sum_{i=1}^{n_{k}}\left(x_{k i}-\bar{x}_{k}\right)^{2}+\frac{n_{k} \tau\left(x s_{k}^{*}-\bar{x}_{k}\right)^{2}}{2\left(\tau+n_{k}\right)}, \tag{4}
\end{gather*}
$$

and $t_{0.05,2 \alpha_{k}^{*}}$ is the 0.05 percentile of the $t$ distribution
with $\alpha_{k}^{*}$ degrees of freedom (e.g., $t_{0.05,2 \alpha_{k}^{*}}=-1.65$ for large $\alpha_{k}^{*}$ ). The $90 \%$ upper bound is calculated with $t_{0.90,2 \alpha_{k}^{*}}$
The Bayesian setup allows us to estimate prediction bounds (equation 1) for cases with no or very few historic data. Furthermore, the setup shows how to combine optimally prior information with available historic data. The values of $\alpha, \beta$, and $\tau$ are obtained (see appendix) by using all historic data, without regard to the combination of surgeon and procedure(s). The prior estimates of $\alpha$ and $\beta$ are revised as data on $n_{k}$ OR times become available for the $k$ th surgeon and procedure(s). Equations 3 and 4 show how the sample variability revises the prior estimates of $\alpha$ and $\beta$, which affect the variance component in the predictive distribution (i.e., the square root term in equation 1). In addition, the parameter $\tau$ expresses how the prior (logarithm of the scheduled OR time) $x s_{k}^{*}$ and the historic mean OR time $\bar{x}_{k}$ should be combined. Equation 1 shows that $x s_{k}^{*}$ has a substantial influence for $n_{k} \leq \tau$, but that its influence diminishes with larger sample sizes. For large sample sizes, the Bayes approach of equation 1 converges to the result previously reported and given in equation 8 .

## Testing for the First New Case Taking Longer Than the Second New Case

The probability that one case will last longer than another is estimated using equations $18-20$ in the appendix.

## Probability That a New Case Will Finish More Than 1 b Late

From equation 17 in the appendix, the probability that a new case will finish more than $L$ hours late equals

$$
\begin{equation*}
P\left[t\left(2 \alpha_{k}^{*}\right) \geq \frac{\ln \left(\exp \left(x s_{k}^{*}\right)+L\right)-\mu_{k}^{*}}{\sqrt{\frac{\beta_{k}^{*}}{\alpha_{k}^{*}}} \cdot \frac{1+\tau+n_{k}}{\tau_{+n_{k}}}}\right], \tag{5}
\end{equation*}
$$

where $t\left(2 \alpha_{k}^{*}\right)$ refers to the $t$ distribution with $2 \alpha_{k}^{*}$ degrees of freedom. For testing, we use $L=1 \mathrm{~h}$.

## Mean Operating Room Time

From the posterior distributions in the appendix and equations 2-4, the point estimate of the mean OR time can be estimated as

$$
\begin{equation*}
\exp \left(\mu_{k}^{*}+\frac{1}{2} \cdot \frac{\beta_{k}^{*}}{\alpha_{k}^{*}}\right) \tag{6}
\end{equation*}
$$

The mean is relevant economically for scheduling, because it is proportional to the total OR time.

## Testing the Accuracy and Usefulness of the Estimates and Predictions

To test the validity of lower prediction bounds from equation 1, several hundred thousand samples were
taken with replacement from the population of 65,661 cases. For each sampled case, the following six steps were followed:

1. The logarithm of the scheduled OR time $\left(x s_{k}^{*}\right)$ was calculated.
2. The case's combination of surgeon and procedure(s) (i.e., $k$ ) was determined.
3. All other cases of that same combination of surgeon and procedure(s) were determined and used as historic data. Therefore, $n_{k}$ equaled the number of cases in the data set of the $k$ th combination minus the one case selected at random.
4. From the first and third steps, $\mu_{k}^{*}, \alpha_{k}^{*}$, and $\beta_{k}^{*}$ were determined from equations 2,3 , and 4 , respectively.
5. Using the posterior parameters from the fourth step, the prediction bound was estimated by applying equation 1 .
6. If the actual OR time of the selected case was less than the value of the lower prediction bound from the fifth step, an indicator value was set equal to 1 . Otherwise, the indicator value was set equal to 0 .

The relative frequency of 1 s and 0 s from the sixth step was determined. The samples were drawn with replacement until the width of the $95 \%$ confidence interval for the proportion of cases exceeded by the prediction bounds was calculated to within $0.12 \%$. The process was repeated for the upper prediction bounds. The process was also repeated after limiting the data to the 44,120 cases not used to estimate $\alpha, \beta$, and $\tau$ in equations 2-4.
To evaluate the accuracy of the process of comparing two future OR times using equations 18-20, two cases were selected at random with replacement from the 65,661 cases. The case with the longest scheduled OR time was tested for the probability of its taking longer than the case with the shorter of the scheduled OR times. Whether the first case was truly longer than the second case was recorded as a 1 for yes and 0 for no. The process was repeated 2 million times. The results were stratified by the estimated probability rounded to the nearest 0.05 . The average of the probabilities for each bin of a width of 0.05 was compared to a ratio, with the numerator being the sum of the count of occurrences for the first case truly being longer than the second case and the denominator being the number of simulations with entries in the bin. ${ }^{4}$
To compare the mean absolute errors of scheduling a new case using either equation 6 or the scheduled OR time, samples were drawn with replacement from the 65,661 cases. The process was continued until the mean difference had been estimated to within 0.1 min .

## Results

## Lower and Upper Prediction Bounds

The Bayesian prediction bounds were accurate to within $2 \%$ (see appendix). The $90 \%$ upper bounds were


Fig. 1. Gamma distribution probability plot of the inverse of the variance of operating room times, one value for each of the 302 combinations of surgeon and scheduled procedure(s) with at least 30 cases. If the data were from a gamma distribution, the plotted values would fall along a straight line extending from the lower left corner of the plot toward the upper right corner. Using the method of moments, the corresponding parameters of the resulting inverse gamma distribution were $\alpha=2.32$ and $\beta=0.142$.
exceeded by $9.7 \%$ of the actual OR times (i.e., were conservative by $0.3 \%$ ). The $5 \%$ lower bounds exceeded $4.9 \%$ of the actual OR times. Excluding the $33 \%$ of cases used to estimate the parameters (i.e., only studying surgeon and procedure(s) combinations with fewer than 29 historic cases), the $90 \%$ upper bound accuracy was $10.2 \%$, and the $5 \%$ lower bound accuracy was $5.5 \%$.
A heuristic for upper and lower prediction bounds is to add or subtract half of the scheduled OR time. The heuristic was reasonable for the studied hospital because, in the absence of historic data for the combination of surgeon and procedure(s), the Bayesian $90 \%$ upper prediction bound equaled $149 \%$ of the scheduled OR time and the $5 \%$ lower prediction bound equaled $57 \%$ of the scheduled OR time. Use of this heuristic would reduce patient and surgeon waiting versus trying to use the scheduled OR time itself plus or minus some value (e.g., 1.5 h ). ${ }^{2}$ Nevertheless, it would be a poor choice to skip use of the historic data and instead simply take a percentage of the scheduled OR time. Figure 1 shows that the variance differs among combinations of surgeon and procedure(s). Using the same percentages for all combinations results in an overall accurate coverage rate, but a rate that is too high or low for many combinations. Figure 2 shows this result in a histogram of the ratio of the $90 \%$ upper Bayesian prediction bound to the scheduled OR time for all cases.

## Testing for the First New Case Taking Longer Than

 the Second New CaseThe Bayesian method was used to compare the OR times of $100 \%$ of the pairwise combinations of cases. The predicted probability of one case being longer than another was estimated to within $0.7 \%$ (table 1). This means, for example, that if the predicted probability is $67 \%$, likely the actual probability is also close to $67 \%$.


Fig. 2. Histogram of the ratio of the $90 \%$ upper Bayesian prediction bound to the scheduled operating room (OR) time for all cases. From the 65,661 cases, 25,000 random samples were taken with replacement. Equation 1 was applied. The histogram shows the 24,004 for which the $\mathbf{9 0} \%$ upper prediction bound was between $75 \%$ and $225 \%$ of the scheduled OR time. If the historic data were not used and only the scheduled OR time (i.e., $n_{k}=0$ in equations $2-4$ ), all cases would have a $90 \%$ upper prediction bound equal to $149 \%$ of the scheduled OR time. There is a peak with many cases at that percentage, because this best overall percentage is used for those cases from combinations of surgeons and procedure(s) with little or no historic data.

## Probability That a New Case Will Finish at Least

## 1 b Late

A heuristic to sequence a surgeon's list of cases to reduce waiting past scheduled start times (tardiness) is to schedule the shortest case(s) first. ${ }^{1,5}$ Theoretical justification for use of this heuristic for ORs has not previously been described. When historic data are absent or ignored, $n_{k}=0$ in equations $2-4$. Equation 5 shows that, then, each increase in the scheduled OR time results in an increase in the probability that the case will finish more than 1 h late. Figure 3 shows the result graphically. The equations apply for any other interval desired (e.g., case finishing 15 min late).
The most common scheduled OR time in the studied data was 2.5 h . All such cases were considered. Historic

Table 1. Pairwise Comparison of Operating Room Times

| Range of Predicted <br> Probabilities That Case with <br> Longer Scheduled OR Time Will <br> Take Longer Than Case with <br> Shorter OR Time | Difference between <br> Predicted and <br> Observed Percentage, \% |
| :---: | :---: |
| $50 \% \leq$ probability $<55 \%$ | 0.2 |
| $55 \% \leq$ probability $<60 \%$ | -0.1 |
| $60 \% \leq$ probability $<65 \%$ | -0.4 |
| $65 \% \leq$ probability $<70 \%$ | -0.1 |
| $70 \% \leq$ probability $<75 \%$ | -0.6 |
| $75 \% \leq$ probability $<80 \%$ | -0.5 |
| $80 \% \leq$ probability $<85 \%$ | -0.0 |
| $85 \% \leq$ probability $<90 \%$ | 0.0 |
| $90 \% \leq$ probability $<95 \%$ | 0.5 |
| $95 \% \leq$ probability $<100 \%$ | 0.7 |

[^1]

Fig. 3. Predicted probability of a case finishing more than $1 \mathbf{h}$ late based only on the scheduled operating room (OR) time. The calculation was made using equation 5 and the parameters estimated in the appendix, with $n_{k}$ set equal to 0 in equations $2-4$. The results can be compared with figure 4.
data were ignored (i.e., $n_{k}$ set equal to 0 , while our estimated values for $\alpha$ and $\beta$ were used). The predicted percentage of cases finishing more than 1 h late was $13 \%$. The actual percentage was $15 \%$ (fig. 4).
When historic data are available, they can be used to make a Bayesian prediction of the probability of a case finishing more than 1 h late. Figure 1 shows that different combinations of surgeon and procedure(s) have different variances. Consequently, among cases with the same scheduled OR time, depending on the historic data, there can be a wide range of Bayesian predicted probabilities that each case will finish more than 1 h late (fig. 4, vertical axis). Actual proportions accurately track these observed probabilities, as does the management endpoint of tardiness ( 0 if finishes early, otherwise minutes finished late).
The implication is that for sequencing surgical cases,


Fig. 4. Differences in actual operating room times among cases scheduled to last 2.5 h . The choice of 2.5 h was made because this was the most common (modal) scheduled operating room time. There were 4,219 such cases. For each case, equation 5 was used to estimate the probability that the case would finish more than 1 h late. These probabilities were binned into four nonoverlapping intervals, rounding down at category breaks. The actual probability of finishing more than 1 h late is shown on the horizontal axis of the left pane. The average tardiness is shown on the right pane. The tardiness is the minutes late that the case actually finished, but set equal to zero if it finished earlier than scheduled. Although the probability of a case finishing late is simple conceptually and has an analytical solution, decisions should be made on the day of surgery based on reducing mean tardiness. ${ }^{1}$ The error bars show the $95 \%$ confidence intervals.
historic OR times should be used when they are available, in addition to the scheduled OR times.

## Mean Operating Room Time

The above results apply to decision making under uncertainty. When scheduling an OR case based on the efficiency of use of OR time, the mean OR time is relevant. ${ }^{1,17,18}$

The Bayesian estimate of the mean in equation 6 and the scheduled OR times were compared with the actual OR time. With respect to bias, whereas the Bayesian method underestimated the actual OR time by an average of 3.2 min , the scheduled OR time underestimated it by an average of 8.6 min . With respect to precision, the mean absolute error of the Bayesian estimate was 3.0 min less than that of the scheduled OR times. Although the Bayesian method was significantly ( $P<0.0001$ ) more accurate on both criteria, the differences were too small to be of practical advantage when focusing on the mean OR time.

## Discussion

We proposed and validated a practical way to calculate prediction bounds and compare the OR times of cases, even when there are few or no historic data for the surgeon and the scheduled procedure(s). As summarized in the introduction, the latter situation arises frequently in both tertiary surgical suites and outpatient surgery centers. Cases without data have a disproportionately large impact on OR management decision making, because any one case without data previously meant that no data-driven recommendation could be obtained. The Bayesian method can be used for (1) deciding when patients should be ready on the day of surgery, (2) adding or filling holes in the OR schedule, (3) sequencing of cases when faced with a constraining resource (e.g., the surgeon or an expensive piece of equipment), and (4) sequencing a surgeon's list of cases.
We also applied our Bayesian formulation to study how and when historic data can improve OR management decision making. When historic data are available, they should be used in combination with (not in lieu of) the scheduled OR time. The results show that this is not because the historic data substantially improve the estimate for the average OR time. Rather, historic data provide value in estimating the proportional variation in OR time.

## Benefits

The benefit of the Bayesian method is in reducing physician and patient waiting, but probably not a direct reduction in overutilized OR time. The OR management decisions aimed at reducing expected overutilized OR time include scheduling cases before the day of surgery,
scheduling add-on cases, and releasing allocated OR time. ${ }^{1,17,18}$ Provided OR staffing and allocations have been made based on finding the optimal balance of underutilized and overutilized OR time, such decisions can then reasonably be made using the expected mean OR time of each case. ${ }^{1,17,18}$ When there are at least two historic cases of the same combination of surgeon and procedure(s), the incremental reduction in overutilized OR time has been shown to be negligible from increasing the accuracy of the prediction of the OR time of a new case versus using the mean of the OR times of the historic cases. ${ }^{1}$ The reason is that decisions to reduce overutilized OR time are affected much more by the day-to-day variability in the total hours of cases than by variability in OR time prediction. ${ }^{1}$ People generally work late because of extra cases, not underestimation in the time to complete each case, because rarely does the underestimation change whether and when the case is performed. ${ }^{19}$ In the current article, we show that the scheduled OR time alone is nearly as good a predictor of the expected mean OR time of a new case as is the Bayesian method. Therefore, the incremental reduction in overutilized OR time from further improvements in predicting OR times is likely negligible.
Several features of the Bayesian method can be important for implementation. First, because the scheduled OR time affects whether a case is scheduled into an OR allocation on a specific day, ${ }^{1,18}$ some surgeons resist reduction in their autonomy in choosing the value. Such organizational resistance to change is irrelevant to the Bayesian method, because the scheduled OR time alone can still be used for that purpose. Second, the premise of relying on expert judgment when there are few data versus predominantly historic data when there are many data has face validity to hospital audiences without statistical training. Third, the information system using the Bayesian method can always provide an estimate or decision recommendation, eliminating the need to educate scores of clerks, physicians, and nurses as to the interpretation of absence. Fourth, equations 2-4 show that the information system need not use the raw data for calculation, only the sample size, mean, and variance (i.e., one record for each combination of surgeon and procedure(s)). Finally, we speculate about the advantage of combining the Bayesian method with new real-time methods of collecting OR time data (e.g., from vital signs). ${ }^{20}$ An information system can now function autonomously, providing recommendations and progressively increasing its usefulness as its historic data increases.

## Limitations

The Bayesian method relies on the logarithms of the scheduled OR times not being consistent underestimates of the logarithms of OR times. At some facilities, surgeons and schedulers may systematically underestimate OR times to get cases onto the OR schedule. ${ }^{21}$ Without


Fig. 5. Histogram of natural logarithm of operating room (OR) times for a single surgeon's strabismus surgery cases. The 105 cases are those cases that the surgeon scheduled as Current Procedural Terminology 67311, "Strabismus surgery, recession or resection procedure; one horizontal muscle."
adjustment, the Bayesian method may be inaccurate, and the argument of the preceding paragraph will be incorrect. Bias can be monitored and incorporated into the equations in the Methods simply by changing each listed $x s_{k}^{*}$ to $\left(x s_{k}^{*}+\delta\right)$, where $\delta$ is the overall or specialtyspecific proportional bias. Feedback can alternatively be provided immediately at the time when the case is scheduled if the Bayesian estimate for the expected mean OR time differs substantially from the scheduled OR time.
Our work is limited in providing one Bayesian method, in no way necessarily the best. We only considered statistical methods that require, in practice, no mathematical calculation other than some arithmetic and exponentials. That way, the method (equations 1-20) can be implemented as an SQL database query and/or run from a Web page. The latter is what we have been doing in practice. Bayesian methods with different distributional assumptions generally have computational solutions requiring numerical integration, which we have found challenging for widespread hospital implementation.

## Appendix

## Previous Studies Based on Assumption of Log Normal Distribution

Assume that $X_{k}$, the natural logarithm of OR time, follows a normal distribution:

$$
\begin{equation*}
X_{k} \sim N\left(M_{k}, \sigma_{k}^{2}\right) \tag{7}
\end{equation*}
$$

where $M_{k}$ is the unknown mean and $\sigma_{k}^{2}$ is the unknown variance. This assumption has been validated previously. ${ }^{13,22}$ For example, figure 5 shows the natural logarithms of OR times for a surgeon's 105 strabismus surgery cases with one muscle (chi-square test of normality, $P=$ 0.71 ; Lilliefors test of normality, $P=0.69$ ). Figure 6 shows the corresponding probability plot. Figures 5 and 6 of Dexter and Traub ${ }^{2}$ show data for laparoscopic cholecystectomy, and dependence of the skewness of these distributions on $n_{k}$.

From the assumption of equation 7 , the $5 \%$ lower prediction bound for the OR time of a new case, that is for $\exp \left(X_{k}^{*}\right)$, can be obtained accurately ${ }^{1,2}$ by taking


Fig. 6. Normal probability plot of natural logarithm of operating room (OR) times for a single surgeon's strabismus surgery. The 105 cases are those cases that the surgeon scheduled as Current Procedural Terminology 67311, "Strabismus surgery, recession or resection procedure; one horizontal muscle." If the data were from a normal distribution, the plotted values would fall along a straight line extending from the lower left corner of the plot toward the upper right corner.

$$
\begin{equation*}
\exp \left(\bar{x}_{k}+t_{0.05, n_{k}-1} \cdot \sqrt{\hat{\sigma}_{k}^{2} \cdot \frac{1+n_{k}}{n_{k}}}\right) \tag{8}
\end{equation*}
$$

where $\bar{x}_{k}$ and $\hat{\sigma}_{k}^{2}$ are the sample estimates from the $n_{k}$ historic data. The $90 \%$ upper bound is calculated by substituting $t_{0.90, n_{k}-1}$ into equation 8 (e.g., $t_{0.90, n_{k}-1}=1.28$ for large $n_{k}$ ). ${ }^{1,3,6}$ If $n_{k}<2$, the prediction bounds cannot be calculated, because $\hat{\sigma}_{k}^{2}$ cannot be estimated from 0 or 1 cases. Furthermore, although the prediction bounds are valid for small $n_{k},{ }^{2,3,6}$ they are often not useful being so wide.

From the assumption of equation 7 , the comparison of the OR times of two new cases is addressed by the Behrens-Fisher problem (i.e., Student $t$ test with unequal variances). ${ }^{1,4,23}$ We refer to the first new case with subscript 1 and the second new case with subscript 2 , whichever combinations of surgeons and procedures(s) they are from. To test for the first new case being longer than the second new case, that is for $\exp \left(X_{1}^{*}\right) \geq \exp \left(X_{2}^{*}\right)$, let ${ }^{23}$

$$
\begin{equation*}
c=\frac{\hat{\sigma}_{1}^{2} \cdot \frac{1+n_{1}}{n_{1}}}{\hat{\sigma}_{1}^{2} \cdot \frac{1+n_{1}}{n_{1}}+\hat{\sigma}_{2}^{2} \cdot \frac{1+n_{2}}{n_{2}}} \tag{9}
\end{equation*}
$$

and

$$
\begin{equation*}
t_{\text {test }}=\frac{\bar{x}_{1}-\bar{x}_{2}}{\sqrt{\hat{\sigma}_{1}^{2} \cdot \frac{1+n_{1}}{n_{1}}+\hat{\sigma}_{2}^{2} \cdot \frac{1+n_{2}}{n_{2}}}} \tag{10}
\end{equation*}
$$

Then, ${ }^{23}$
$P\left[\exp \left(X_{1}^{*}\right) \geq \exp \left(X_{2}^{*}\right)\right]=P\left[X_{1}^{*} \geq X_{2}^{*}\right]$

$$
\begin{align*}
& =P\left[X_{1}^{*}-X_{2}^{*} \geq 0\right] \\
& =c P\left[t\left(d f_{1}\right) \leq t_{\text {test }}\right]+(1-c) P\left[t\left(d f_{2}\right) \leq t_{\text {test }}\right] \tag{11}
\end{align*}
$$

where

$$
\begin{equation*}
d f_{1}=\frac{n_{1}-1}{c} \text { and } d f_{2}=\frac{n_{2}-1}{1-c} \tag{12}
\end{equation*}
$$

Because the values of $d f$ are not integers, we calculate the probabilities of the $t$ distribution from its relation with the Incomplete Beta Function. ${ }^{24}$ Press et al. ${ }^{24}$ provide the computer code. The primary limitation to equations $9-12$ is that to calculate $\hat{\sigma}_{1}^{2}$ and $\hat{\sigma}_{2}^{2}, n_{1} \geq 2$ and $n_{2} \geq 2$.

## Two Additional Assumptions for Inference and

## Prediction in the Absence of Historic Data

Suppose that no surgeon- and procedure(s)-specific historic OR time data are available. To predict the OR time of a new case and to


Fig. 7. Histogram of the logarithm of the ratio of the scheduled to actual operating room (OR) time for the 18,381 cases of the 15,916 combinations of surgeon and procedure(s) for which there were either one or two cases. Sixteen cases of the 18,381 are not shown, because they fall outside the shown range.
calculate the prediction bounds, we make two assumptions, one about the prior distribution of $\sigma_{k}^{2}$ and the other about $M_{k}$.

Assume that $\sigma_{k}^{2}$ follows an inverse gamma distribution:

$$
\begin{equation*}
\sigma_{k}^{2} \sim \mathrm{IG}(\alpha, \beta) \tag{13}
\end{equation*}
$$

where $\alpha$ and $\beta$ are unknown parameters common to all combinations of surgeons and procedure(s) at the facility. Following Strum et al. ${ }^{13,22}$ in their investigations of the statistical distributions of OR times, we considered those combinations of surgeon and procedure(s) that were performed a moderate to large $\left(n_{k} \geq 30\right)$ number of times. There were 302 such combinations of surgeon and procedure(s) corresponding to 21,541 cases. The probability plot in figure 1 shows that the assumption of the inverse gamma distribution is reasonable (chi-square test, $P=0.51$; Kolmogorov-Smirnov, $P=0.88$ ).

Assume that the conditional prior distribution of $M_{k}$ given $\sigma_{k}$ is normal with prior mean $\mu_{k}$ and variance $\sigma_{k}^{2} / \tau$ :

$$
\begin{equation*}
M_{k} \left\lvert\, \sigma_{k} \sim \mathrm{~N}\left(\mu_{k}, \frac{\sigma_{k}^{2}}{\tau}\right)\right. \tag{14}
\end{equation*}
$$

The validity of equation 14 is considered in the next section. The $\tau$ represents the ratio of the variance of an individual observation to the variance of the mean. If $\tau$ were a small value (e.g., $\tau=1$ historic case), our prior information about $M_{k}$ would be vague, making the incremental value of any historic data relatively influential.

A consequence of equations 7 and 14 together is that the conditional prior distribution of $X_{k}$ given $\sigma_{k}$ is normal with prior mean $\mu_{k}$ and variance $\sigma_{k}^{2}+\left(\sigma_{k}^{2} / \tau\right)$. Combining terms, the variance equals $\sigma_{k}^{2} \cdot((\tau+1) / \tau)$. Bringing in the inverse gamma distribution for $\sigma_{k}^{2}$ from equation 13, the predictive prior distribution of the logarithm of the OR time of the next case, $X_{k}^{*}$, is a scaled Student $t$ distribution ${ }^{25}$ :

$$
\begin{equation*}
\sqrt{\frac{\alpha}{\beta} \cdot \frac{\tau}{1+\tau}}\left(X_{k}^{*}-\mu_{k}\right) \sim t(2 \alpha) . \tag{15}
\end{equation*}
$$

## Prior Values for Use in Inference and Prediction in the Absence of Historic Data

We obtained the prior values for $\alpha$ and $\beta$ using the data in figure 1 . Using the method of moments, the corresponding parameters of the inverse gamma distribution were $\alpha=2.32$ and $\beta=0.142$.

We considered the logarithm of the scheduled OR time of the new case, $x s_{k}^{*}$, to be the prior value for $\mu_{k}$ and the prior prediction of $X_{k}^{*}$. Figure 7 shows a histogram of the prediction errors, $\left(x_{k}^{*}-x s_{k}^{*}\right)$, for the 18,381 cases for which $n_{k} \leq 2$. The symmetric distribution around zero confirms that $x s_{k}^{*}$ is a good prior value, because it provides an unbiased prediction of the logarithm of the actual OR time in hours. This implies that $\exp \left(x s_{k}^{*}\right)$ provides an unbiased prediction of the 50 th percentile of $\exp \left(X_{k}^{*}\right)$, which is slightly less than the expected (mean) value of $\exp \left(X_{k}^{*}\right)$. We would therefore expect that the scheduled OR time would slightly, but significantly, underestimate the actual OR


Fig. 8. Normal probability plot for the differences between the means of the logarithms of operating room times and the logarithm of scheduled operating room times for 100 cases selected at random from among all 10,562 cases with at least 30 cases of their combinations of surgeon and procedure(s) and with variance of the logarithms of operating room times between 0.050 and 0.099 . The figure shows that the functional (normal) relation of equation 14 is reasonable. If the data were from a normal distribution, the plotted values would fall along a straight line extending from the lower left corner of the plot toward the upper right corner. However, as studied by Strum et $a l .,{ }^{22}$ the rounding of scheduled operating room times to large (in this circumstance 15 min ) increments results in disruptions to normality. The use of 100 cases was sufficiently small that points did not overlap so much as to impair viewing of the undulation around the straight line.
time. This was the finding reported in the final paragraph of the Results. In those analyses, the expected (mean) value averaged the 52nd percentile of $\exp \left(X_{k}^{*}\right)$, with lower and upper quartiles of $51 \%$ and $52 \%$, respectively.
We estimated the prior value for $\tau$ from the variance of the prediction errors. Using the 21,541 cases for which we estimated $\alpha$ and $\beta$ above, $\operatorname{Var}\left(x_{k}^{*}-x s_{k}^{*}\right)=0.120$. The variance of a $t$ distribution with $2 \alpha$ degrees of freedom equals ${ }^{26} 2 \alpha /(2 \alpha-2)$. From the result of equation 15 ,

$$
\begin{align*}
\operatorname{Var}\left(x_{k}^{*}-s_{k}^{*}\right) & =\frac{\beta}{\alpha} \cdot \frac{\tau+1}{\tau} \cdot \frac{2 \alpha}{2 \alpha-2} \\
& =\frac{\beta}{\alpha-1} \cdot \frac{\tau+1}{\tau} \\
& =0.120 . \tag{16}
\end{align*}
$$

Substitution of the prior values for $\alpha=2.32$ and $\beta=0.142$ from two paragraphs above, $\tau=8.68$ cases.

The assumption in equation 14 was validated by studying the distribution of $\left(\bar{x}_{k}-x s_{k}^{*}\right)$, with $\bar{x}_{k}$ and $x s_{k}^{*}$ taking the places of $M_{k}$ and $\mu_{k}$, respectively. When the variances from figure 1 were rounded to the nearest 0.05 , the interval $0.05 \leq \hat{\sigma}_{k}^{2}<0.10$ included the most cases, 10,562 , as well as many (156) different values of $k$. Figure 8 shows a normal probability plot for $\left(\bar{x}_{k}-x s_{k}^{*}\right)$. The functional (normal) form of equation 14 is reasonable. From Strum et al., ${ }^{22}$ the undulation around the straight line in figure 8 was the expected consequence of our data's $x s_{k}^{*}$ being in $15-\mathrm{min}$ intervals, resulting in $\left(\bar{x}_{k}-x s_{k}^{*}\right)$ taking a small set of values.

## Updating the Estimates and Predictions Using Historic Data

Our prior distributions in equations 13 and 14 have the advantage that for any sample size and for any values of the observations sampled from the model of equation 7 , the posterior distributions of $\sigma_{k}^{2}$ and $M_{k}$ belong to the same family. ${ }^{25}$ The Bayesian literature refers to such prior distributions as conjugate priors. Specifically, the posterior distribution of $\sigma_{k}^{2}$ is again an inverse gamma distribution. The posterior of $M_{k}$ and the posterior predictive distribution of $X_{k}^{*}$ are scaled $t$ distributions,
with revised parameters that update the prior values with the sample information:

$$
\begin{equation*}
\sqrt{\frac{\alpha_{k}^{*}}{\beta_{k}^{*}} \cdot \frac{1+\tau+n_{k}}{\tau+n_{k}}} \cdot\left(X_{k}^{*}-\mu_{k}^{*}\right) \sim t\left(2 \alpha_{k}^{*}\right) \tag{17}
\end{equation*}
$$

with parameters given in equations 2-4.
From the posterior predictive distribution of equation 17, equations 9-12 are modified to provide the Bayesian comparison of the OR times of two cases:

$$
\begin{equation*}
c=\frac{\frac{\beta_{1}^{*}}{\alpha_{1}^{*}} \cdot \frac{1+\tau+n_{1}}{\tau+n_{1}}}{\frac{\beta_{1}^{*}}{\alpha_{1}^{*}} \cdot \frac{1+\tau+n_{1}}{\tau+n_{1}}+\frac{\beta_{2}^{*}}{\alpha_{2}^{*}} \cdot \frac{1+\tau+n_{2}}{\tau+n_{2}}} \tag{18}
\end{equation*}
$$

and

$$
\begin{equation*}
t_{\text {test }}=\frac{\mu_{1}^{*}-\mu_{2}^{*}}{\sqrt{\frac{\beta_{1}^{*}}{\alpha_{1}^{*}} \cdot \frac{1+\tau+n_{1}}{\tau+n_{1}}+\frac{\beta_{2}^{*}}{\alpha_{2}^{*}} \cdot \frac{1+\tau+n_{2}}{\tau+n_{2}}}}, \tag{19}
\end{equation*}
$$

where

$$
\begin{equation*}
d f_{1}=\frac{2 \alpha_{1}^{*}}{c} \text { and } d f_{2}=\frac{2 \alpha_{2}^{*}}{1-c} \tag{20}
\end{equation*}
$$

## Assessment of Sensitivity of Estimates to Prior Values of $\alpha, \beta$, and $\tau$

Using only cases from 1996 and 1997, there were 124 combinations of surgeon and procedure(s) with $n_{k} \geq 30$ versus 302 combinations with all the data. The estimated $\alpha=1.70$ and $\beta=0.089$, versus $\alpha=2.32$ and $\beta=$ 0.142 with all the data. From these 7,405 cases, $\tau=14.9$ cases, versus from all 21,541 cases, $\tau=8.68$ cases. Applying the $\alpha, \beta$, and $\tau$ from 1996 and 1997 to all of the data, the $90 \%$ upper bounds were exceeded by $10.0 \%$ of cases. The $5 \%$ lower bounds exceeded $5.1 \%$ of cases. Therefore, the Bayesian method was insensitive to the choice of $\alpha, \beta$, and $\tau$. Furthermore, repeating with $\tau=1.49$ cases (an order of magnitude less), the results were the same, within $0.2 \%$.

In the next set of calculations, we assumed that the prior values are updated annually (i.e., 1996-1997 prior values were applied to 1998 data, and 1996-1998 prior values were applied to 1999 data). We assumed that the lookup table with the running totals of $n_{k}, \Sigma x_{k i}$, and $\Sigma x_{k i}^{2}$ for each combination of surgeon and procedure(s), for use in equations 2-4, is updated nightly. For example, to estimate prediction bounds for cases performed on Tuesday August 4, 1998, the running totals used were for January 1, 1996, through August 3, 1998. Then, the calculated $90 \%$ upper bounds were exceeded by $8.0 \% \pm 0.1 \%$ of cases $(\mathrm{n}=32,930)$. The $5 \%$ lower bounds exceeded $6.9 \% \pm 0.1 \%$ of cases. Repeating using $\tau=1.49$ cases, results were the same, within $0.2 \%$. Therefore, improvement in accuracy would rely on modifying the assumption of equation 13 of a common $\alpha$ and $\beta$ for those combinations of surgeon and procedure(s) with little to no data, not those combinations with small to large numbers of historic cases. Those cases with little to no data are precisely those for which $\alpha$ and $\beta$ cannot be estimated directly.

## References

1. 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
2. Dexter F, Traub RD: Statistical method for predicting when patients should be ready on the day of surgery. Anesthesiology 2000; 93:1107-14
3. 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; 92:943-6
4. Dexter F, Traub RD: Sequencing cases in operating rooms: Predicting
whether one surgical case will last longer than another. Anesth Analg 2000; 90:975-9
5. Lebowitz P: Schedule the short procedure first. AORN J 2003; 78:651-8
6. Zhou J, Dexter F: Method to assist in the scheduling of add-on surgical cases: Upper prediction bounds for surgical case durations based on the log normal distribution. Anesthesiology 1998; 89:1228-32
7. 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
8. Wortmann KD, Weeks T: Successful: Preference list building. Surg Serv Manag 1998; 4:20-6
9. Porter J: Building doctor's preference cards. Surg Serv Manag 1999; 5:43-7
10. 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
11. 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
12. 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
13. Strum DP, May JH, Sampson AR, Vargas LG, Spangler WE: Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. Anesthesiology 2003; 98:232-40
14. Wachtel RE, Dexter F: Differentiating among hospitals performing physiologically complex operative procedures in the elderly. Anesthesiology 2004; 100:1552-61
15. Strum DP, Sampson AR, May JH, Vargas LG: Surgeon and type of anesthesia predict variability in surgical procedure times. Anesthesiology 2000; 92: 1454-67
16. Dexter F: Application of prediction levels to operating room scheduling. AORN J 1996; 63:607-15
17. Dexter F, Traub RD, Qian F: Comparison of statistical methods to predict the time to complete a series of surgical cases. J Clin Monit Comput 1999; 15:45-51
18. 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
19. 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
20. Xiao Y, Hu P, Hao H, Ho D, Dexter F, Mackenzie CF, Seagull FJ, Dutton R: An algorithm for processing vital sign monitoring data to remotely identify operating room occupancy in real-time. Anesth Analg 2005; 101:823-9
21. Dexter F, Macario A, Epstein RH, Ledolter J: Validity and usefulness of a method to monitor surgical services' average bias in scheduled case durations. Can J Anesth 2005; 52:935-9
22. Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times: Comparison of the log-normal and normal models. Anesthesiology 2000; 92:1160-7
23. Bhoj DS: An approximate solution to the Behrens-Fisher problem. Biometrical J 1993; 35:635-40
24. Press WH, Teukolsky SA, Vetterling WT, Flannery BP: Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd edition. Cambridge, England, and New York, Cambridge University Press, 1992, pp 207, 220-222
25. DeGroot MH: Optimal Statistical Decisions. New York, McGraw Hill, 1970, pp 42, 169, 170
26. Johnson NL, Kotz S, Balakrishnan N: Continuous Univariate Distributions, vol 2, 2nd edition. New York, John Wiley \& Sons, 1995, p 366

[^0]:    * Professor and Director of the Division of Management Consulting, Departments of Anesthesia and Health Management and Policy, † C. Maxwell Stanley Professor of Management Sciences, College of Business, The University of Iowa.

    Received from the Division of Management Consulting, Department of Anesthesia, The University of Iowa, Iowa City, Iowa. Submitted for publication May 17, 2005. Accepted for publication August 8, 2005. Support was provided solely from institutional and/or departmental sources. Dr. Dexter is Director of the Division of Management Consulting, a Division of the Department of Anesthesia. He receives no funds personally other than his salary from the State of Iowa, including no travel expenses or honorarium, and has tenure with no incentive program. This work will be presented at the Institute for Operations Research and Management Sciences Annual Meeting, San Francisco, California, November 13, 2005. David O. Warner, M.D., acted as Handling Editor for this article.
    Address reprint requests to Dr. Dexter: Division of Management Consulting, Department of Anesthesia, The University of Iowa, Iowa City, Iowa 52242. Address electronic mail to: franklin-dexter@uiowa.edu. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

[^1]:    The predicted percentage equaled the sum of the probabilities divided by the number of samples with replacement (102,691 to 783,731 ). The observed percentages equaled the number of occurrences divided by the number of samples. The reported endpoint was the difference of the predicted and observed percentages. The differences have SEs less than $0.2 \%$
    $\mathrm{OR}=$ operating room.

