

Modeling Procedure and Surgical Times for Current Procedural Terminology-Anesthesia-Surgeon Combinations and Evaluation in Terms of Case-Duration Prediction and Operating Room Efficiency: A Multicenter Study

Pieter S. Stepaniak, MSc*

Christiaan Heij, PhD†

Guido H. H. Mannaerts, MD,
PhD‡

Marcel de Quelerij, MDS

Guus de Vries, PhD*

BACKGROUND: Gains in operating room (OR) scheduling may be obtained by using accurate statistical models to predict surgical and procedure times. The 3 main contributions of this article are the following: (i) the validation of Strum's results on the statistical distribution of case durations, including surgeon effects, using OR databases of 2 European hospitals, (ii) the use of expert prior expectations to predict durations of rarely observed cases, and (iii) the application of the proposed methods to predict case durations, with an analysis of the resulting increase in OR efficiency.

METHODS: We retrospectively reviewed all recorded surgical cases of 2 large European teaching hospitals from 2005 to 2008, involving 85,312 cases and 92,099 h in total. Surgical times tended to be skewed and bounded by some minimally required time. We compared the fit of the normal distribution with that of 2- and 3-parameter lognormal distributions for case durations of a range of Current Procedural Terminology (CPT)-anesthesia combinations, including possible surgeon effects. For cases with very few observations, we investigated whether supplementing the data information with surgeons' prior guesses helps to obtain better duration estimates. Finally, we used best fitting duration distributions to simulate the potential efficiency gains in OR scheduling.

RESULTS: The 3-parameter lognormal distribution provides the best results for the case durations of CPT-anesthesia (surgeon) combinations, with an acceptable fit for almost 90% of the CPTs when segmented by the factor surgeon. The fit is best for surgical times and somewhat less for total procedure times. Surgeons' prior guesses are helpful for OR management to improve duration estimates of CPTs with very few (<10) observations. Compared with the standard way of case scheduling using the mean of the 3-parameter lognormal distribution for case scheduling reduces the mean overreserved OR time per case up to 11.9 (11.8–12.0) min (55.6%) and the mean underreserved OR time per case up to 16.7 (16.5–16.8) min (53.1%). When scheduling cases using the 4-parameter lognormal model the mean overutilized OR time is up to 20.0 (19.7–20.3) min per OR per day lower than for the standard method and 11.6 (11.3–12.0) min per OR per day lower as compared with the biased corrected mean.

CONCLUSIONS: OR case scheduling can be improved by using the 3-parameter lognormal model with surgeon effects and by using surgeons' prior guesses for rarely observed CPTs. Using the 3-parameter lognormal model for case-duration prediction and scheduling significantly reduces both the prediction error and OR inefficiency.

(Anesth Analg 2009;109:1232–45)

The operating room (OR) is a major production unit in every hospital. For hospitals, the main 2 operational risks of ORs consist of high idle times (i.e., underutilized OR time) and work outside regular hours (i.e.,

overutilized OR time). Frequent work beyond scheduled hours not only leads to overtime costs but also to intangible costs resulting from dissatisfaction and reduced motivation of staff. Overtime work is one of the primary reasons for nurses to terminate their employment,¹ and scheduling conflicts are a major cause of nursing staff turnover.² Therefore, efficient OR management should aim for maximal use of available OR time while preventing frequent overtime work.³ OR schedules depend crucially on estimated case durations, and statistical models may help to improve these estimates to support management in the cost-efficient use of expensive surgical resources.

Herein, we provide a brief review of some relevant results in the literature on case-duration distributions

From the *Institute of Health Policy and Management, Erasmus University Rotterdam; †Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam; and Departments of ‡General Surgery, and §Anesthesiology, St. Franciscus Hospital, Rotterdam, The Netherlands.

Accepted for publication June 23, 2009.

Address correspondence and reprint requests to Pieter Stepaniak, MSc, PO Box 1738, 3000 BA Rotterdam, The Netherlands. Address e-mail to stepaniak@bmg.eur.nl.

Copyright © 2009 International Anesthesia Research Society

DOI: 10.1213/ANE.0b013e3181b5de07

and case scheduling. Early results show that OR waiting times follow a 2-parameter lognormal distribution,⁴ and that OR operation times follow a distribution that is normal⁵ or lognormal.⁶ Knowledge of the probability distributions of case durations has advanced markedly in the past decade.⁷⁻⁹ The single most important source of variability in surgical procedure times is surgeon effect. Type of anesthesia, age, gender, and American Society of Anesthesiologists risk class were additional sources of variability.⁷ In another study, Strum et al.⁸ tested surgeries with 2-component procedures. The conclusion is that dual Current Procedural Terminology (CPT) surgeries were better modeled by the lognormal distribution than by the normal distribution. Surgical procedure times are frequently distributed with nonzero start times that require a lognormal model with a shifted parameter for best model estimates.⁹ Decision rules based on the skewness and coefficient of variation of the data can be used to identify the correct alternative 78% of the time, but do not do any better than a single rule based on the skewness.⁹ The way in which the lognormal location parameter is estimated affects the ability of goodness-of-fit tests to correctly recognize the model and the accuracy of percentile point values derived from the estimated model.¹⁰

An empirical study¹¹ has shown that surgical time and total procedure time are lognormal distributed. Surgical procedure time fits the lognormal distribution for 93% of all CPT codes, whereas surgical time fits normal distribution for about 80% of all CPT codes studied.

For some of the scheduled cases, there are few or no data available, making statistical modeling difficult. These cases can disproportionately affect decision making under uncertainty because no sufficient data-driven recommendations could be obtained. Several studies have tried to solve this problem of few or no cases.^{12,13} Dexter and Ledolter¹² validated a practical way to calculate prediction bounds and compared the OR times of all cases, even those with few or no historic data for surgeon and the scheduled procedure(s). The conclusion of this study is that when historic data are available, they should be used in combination with the scheduled OR time. Historic data provide value in estimating the proportional variation in OR time. Finally, the scheduled OR time alone is nearly as good a predictor of the expected mean OR time of a new case as the Bayesian method.

In another study,¹⁴ elective case scheduling at hospitals and surgical centers at which surgeons and patients choose the day of surgery, cases are not turned away, and anesthesia and nursing staffing are adjusted to maximize the efficiency of use of OR time. In this study, 2 patient-scheduling rules are investigated: Earliest Start Time or Latest Start Time. In this study, the achievable incremental reduction in overtime by having perfect information on case duration versus using historical case durations was only a few

minutes per OR. The differences between Earliest Start Time and Latest Start Time were also only a few minutes per OR. There are cases that have a high probability of taking longer than scheduled. Increasing the case's scheduled duration could then reduce overutilized OR time.¹⁵ Dexter et al.¹⁵ studied surgeons' and schedulers' case scheduling behavior to evaluate whether such a strategy would be useful. The impact of inaccurate, scheduled case duration on staffing costs and unpredictable work hours can be reduced by allocating appropriate total hours of OR time (i.e., staffing) for the cases that will get done, regardless of the inaccuracy of the scheduled durations of those cases.

There are many other studies related to optimally scheduling cases.¹⁶⁻²³ All these studies contribute to optimizing the use of scarce and costly ORs.

Based on the above-mentioned studies, we can conclude that gains in OR scheduling efficiency may be obtained by using accurate statistical models to predict surgical and procedure times. Therefore, the 3 main contributions of this article are the following: (i) the validation of the results of Strum et al. on the statistical distribution of case durations, including surgeon effects, using OR databases of 2 European hospitals, (ii) the use of expert prior expectations to predict durations of rarely observed cases, and (iii) the application of the proposed methods to predict case durations, with an analysis of the resulting OR efficiency.

METHODS

In this section, we first present our database, then describe our methods.

Data

We retrospectively reviewed all recorded surgical cases from 2 large teaching hospitals from 2005 to 2008 (total 85,312 cases). Because there were differences in case duration based on type of anesthetic used, we classified the CPT codes by type of anesthesia: general, local, and regional.^{8,24} Monitored anesthesia care is not a type of anesthesia used in the hospitals under study. We use the following definitions: Surgical Time = the time from incision to closure of the wound; Procedure Time = time when patient enters the operating suite until the patient leaves the OR. To detect the influence of sample size on the Shapiro-Wilk* test, we divided the sample size into very small ($n < 10$), small ($10 \leq n < 30$), medium ($30 \leq n < 200$), and large ($n \geq 200$). In Table 1, we present the dataset for Hospital A. For every case-frequency interval, the number of CPT codes, the number of cases, and the total hours spent for these cases in the period 2005–2008 is shown.

There was a total of 44,223 cases, of which 289 (0.7%) cases were omitted because of incomplete data.

*For further information concerning the Shapiro-Wilk test, we refer readers to the Statistics section.

Table 1. Operating Room Hospital A and Hospital B

Case frequency	Hospital A			Hospital B		
	CPT-codes	Cases	Hours	CPT-code	Cases	Hours
$n \geq 200$	53	22,512	20,248	35	20,417	18,324
$30 \leq n < 200$	347	16,388	19,973	287	16,587	19,184
$10 \leq n < 30$	454	4012	6266	419	3271	4641
<10	318	1004	1717	201	1121	1746
Total	1172	43,916	48,204	942	41,396	43,895
	1 CPT-code			1 CPT-code		
$n \geq 200$	53	22,512	20,248	35	20,417	18,324
$30 \leq n < 200$	253	11,621	13,707	198	13,984	15,187
$10 \leq n < 30$	431	3174	4518	337	2912	3681
<10	287	541	823	189	762	1117
Total	1024	37,848	39,296	759	38,075	38,308

The Classification of the Cases by Type of Anesthesia (All Cases and by Cases with Frequency ≥ 10)

Number of CPT-Anesthesia combinations	Hospital A		Hospital B	
	CPT	Cases	CPT	Cases
Anesthesia				
General	685	17,561	579	15,664
Local	147	4310	134	4408
Regional	340	22,045	229	21,324
Total	1172	43,916	942	41,396
1 CPT-Anesthesia combinations (case frequency ≥ 10)				
General	433	15,007	294	14,783
Local	157	4050	85	2543
Regional	147	18,250	191	19,987
Total	737	37,307	570	37,313

In 15 cases (0.03%), the operation was canceled, although the patient received anesthesia, and in 3 cases, a donor procedure was performed. In our analyses, we have 43,916 cases (1172 CPT-anesthesia combinations), with hours totaling 48,204.

There were 37,848 cases (39,296 h) with 1 CPT-anesthesia combination, 5177 cases (7312 h) with 2 CPT codes, and 891 cases (1596 h) with more than 2 CPT codes. The average number of cases per year with 2 CPTs was 1294 (median 1305, min 1165, and max 1401). For CPTs with more than 2 codes, the average was 222 cases (median 221, min 201, and max 247).

To eliminate a potential confounding factor† in our study, we considered only surgical procedures with a single CPT code. Therefore, we confined our analysis to 37,307 cases with a case frequency of ≥ 10 (737 CPT-anesthesia combinations).

We broke down the CPTs according to the various surgeons (Table 2). There were 30 surgeons and 6349 CPT-anesthesia-surgeon combinations (43,916 cases, 48,204 h, Table 3). If we differentiate combinations with at least 10 cases per surgeon and 1 CPT-anesthesia code, 1341 CPT-anesthesia-surgeon combinations remain (32,347 cases, 34,512 h). Regardless of

the number of CPT codes, of the 1172 CPT codes, there are 318 CPT-anesthesia combinations (1004 cases, 1717 h), which were performed <10 times in a period of 4 yr. Of the 43,916 cases scheduled, for 46 cases (0.1%), the actual procedure code was different than the scheduled code. In 132 cases (0.3%), the actual surgeon was different than the scheduled surgeon.

In Table 1, the dataset for Hospital B is presented. There was a total of 41,916 cases, of which 520 (1.2%) cases were omitted because of incomplete data. The analysis was limited to 41,396 cases (942 CPT-anesthesia combinations, 43,895 h). There were 38,075 cases (38,308 h) with only 1 CPT code, 2707 cases (4531 h) with 2 CPT codes, and 614 cases (1056 h) with more than 2 CPT codes. The average number of cases per year with 2 CPTs was 676 (median 687, min 634, and max 699). For CPTs with more than 2 codes, the average was 153 cases (median 151, min 143, and max 166).

As in Hospital A, we considered only cases with 1 CPT code and each CPT-anesthesia combination with a case frequency of 10 or more. We confined our analysis to 37,313 cases (570 CPT-anesthesia codes).

There were 24 surgeons (Table 2) and 4473 CPT-anesthesia-surgeon combinations (41,396 cases, 43,895 h Table 3). If we differentiate combinations with at least 10 cases per surgeon, 1147 CPT-anesthesia-combinations remain (30,274 cases, 32,927 h). Of the 41,396 cases

†In our analyses, we use only surgeries with 1 CPT code (as in Strum et al.¹¹) to avoid possible confounding factors. Procedures with, for example, 2 CPT codes (CPT1 and CPT2) can be performed in different ways. First CPT1 and then CPT2, or vice versa. The sequence can then be a confounding factor.

Table 2. Number of Surgeons

Hospital A	
Number of surgeons	
Eye surgery	3
Orthopedic	4
Ear-nose-throat	2
Neurosurgery	2
Urology	3
General surgery	6
Obstetrics and gynecology	6
Jaw surgery	2
Plastic surgery	2
Total	30
Number of anesthesiologists	9
Hospital B	
Number of surgeons	
Eye surgery	3
Orthopedic	4
Ear-nose-throat	2
Urology	3
General surgery	6
Obstetrics and gynecology	6
Total	24
Number of anesthesiologists	10

Table 3. Number of Unique CPT Code-Anesthesia Surgeon Combinations

	Hospital A	Hospital B
Eye surgery	61	48
Orthopedic	970	1141
Ear-nose-throat	71	59
Neurosurgery	41	0
Urology	174	137
General surgery	2108	1244
Obstetrics and gynecology	2420	1844
Jaw surgery	47	0
Plastic surgery	457	0
Total	6349	4473
Cases	43,916	41,396
Hours	48,204	43,895

scheduled, in 28 cases (0.07%), the actual procedure code was different than the scheduled code. In 89 cases (0.2%), the actual surgeon was different than the scheduled surgeon.

Next, we describe in detail what we have studied and how the study was performed.

Fitting the Normal and 2- and 3-Parameter Lognormal Models for 1-CPT-Anesthesia-Surgeon Combinations with Case Frequency ≥ 10

We repeated the work of Strum et al.⁹⁻¹¹ for the normal and 2- and 3-parameter lognormal modeling of surgical procedure times. Repeating the work of Strum et al. is important scientifically because replication of research is a way to refine our understanding of modeling surgical cases. The 3-parameter lognormal model is of interest because surgical procedure times are frequently distributed with nonzero start times that require a lognormal model with a shift parameter for best model estimates.^{9,11} A nonzero start time means that minimum surgical procedure

times, even for the simplest procedures, are strictly positive. As is assumed,¹¹ the percentage of cases that fit the lognormal model can be even higher when segmented by the factor surgeon. Therefore, we validated whether performed procedure times and surgical times of CPT-anesthesia-surgeon combinations fit a normal, 2-parameter, or 3-parameter lognormal distribution.

The general formula for the lognormal model can be described as follows:

$$f(x) = \frac{e^{\frac{[-\ln(x-a)-\mu]^2}{2\sigma^2}}}{(x-\theta)\sigma\sqrt{2\pi}}$$

for $x > \theta$

where θ = shift parameter for duration data $\theta > 0$.

The case where $\theta = 0$ is called the 2-parameter lognormal model.

For the 3-parameter lognormal model, we estimated the shift parameter by using a modified version of the approach of Spangler et al.¹⁰ The shift parameter describing the location or origin of the random variable is important for decision making, because it provides a lower bound on values of the random variable.¹⁰ First, for every CPT-anesthesia (surgeon) combination, we calculated the natural logarithm of surgical time and procedure time. We then used the bisection method to estimate the shift parameter, so that we could estimate 3 parameters for each combination of surgeon(s) and procedure(s). The bisection method we used is as follows:

Set LOWER = 0.

Set UPPER = smallest observed value.

Initial GUESS = (LOWER + UPPER)/2.

Subtract GUESS from all observed values, take the logarithm, and estimate the mean and sds, then recalculate the Shapiro-Wilk P value ($= P_{\text{new}}$). We repeated this iteratively using bisection to find the shift parameter that results in the largest value of the P value. We chose to stop the iteration if $(P_{\text{new}} - P_{\text{old}})/P_{\text{new}} \times 100\% < 1\%$ or if $P_{\text{new}} < P_{\text{old}}$. If the final P value was larger than 0.05, we did not reject the hypothesis of the normal model.

Estimation with Specialist Prior Guess

If very few data are available ($n < 10$), it may help to use prior information to obtain more reliable estimates of the time distribution. Therefore, we present a method to estimate the mean procedure time from prior and actual data for procedures with < 10 cases. It is well known that Bayes theorem provides a mechanism for combining a prior probability distribution for the states of nature with sample information to provide a revised (posterior) probability distribution about those states of nature. These posterior probabilities are then used to make better decisions. Our approach differs from that of Dexter et al.^{12,25} in the way

that we used the surgeon's prior statement on the distribution in terms of quantiles of the operation time.

To obtain the prior information required, we asked surgeons to make a prior statement on the distribution of the procedure time for cases with a frequency <10. For a given procedure, we asked surgeons in the period October–December 2008 before they started the scheduled case to make an estimation in terms of quantiles (25%, 50%, 75%, and 95%) of the time distribution of a procedure. With this information, we were able to update our uncertainty because of new evidence. In the analyses, we used the 2-parameter lognormal model in which the mean and variance were calculated from a weighted mean of the actual data and the prior data. Furthermore, we assumed that the specialists do not remember the previous operation times, so that all calculated times (past and current) can be treated as containing similar information. Next, we explained our model for using prior information in mathematical terms.

Let T denote the procedure time and let $\ln(T)$ be its natural logarithm. Assuming a 2-parameter lognormal model for the procedure time, it follows that $\ln(T)$ is normal with mean μ and variance σ^2 . We then had to combine the prior and actual data information to estimate the mean μ . Let m be the prior mean and s the prior variance of $\ln(T)$. Let the 25%, 50%, 75%, and 95% quantiles of $\ln(T)$ be denoted by T_{25} , T_{50} , T_{75} , and T_{95} , then the normal distribution implies that:

$$\begin{aligned} T_{25} &= m - 0.675 s \\ T_{50} &= m \\ T_{75} &= m + 0.675 s \\ T_{95} &= m + 1.645 s \end{aligned} \quad (1)$$

For example, with 2 prior estimates made by Specialists 1 and 2, we estimated the following model:

$$\begin{pmatrix} T_{25}(\text{specialist 1}) \\ T_{50}(\text{specialist 1}) \\ T_{75}(\text{specialist 1}) \\ T_{95}(\text{specialist 1}) \\ T_{25}(\text{specialist 2}) \\ T_{50}(\text{specialist 2}) \\ T_{75}(\text{specialist 2}) \\ T_{95}(\text{specialist 2}) \end{pmatrix} = m \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + s \begin{pmatrix} -0.675 \\ 0 \\ 0.675 \\ 1.645 \\ -0.675 \\ 0 \\ 0.675 \\ 1.645 \end{pmatrix}$$

where $[-0.675 \ 0 \ 0.675 \ 1.645]$ is the vector with corresponding z-values. This vector provides the regressor needed to estimate location and scale of the lognormal prior distribution corresponding to the quantiles. The vector is repeated for each specialist.

When we have data for j specialists, hence $4j$ times, we get $4j$ equations with given values on the left-hand side and with unknown values of m and s . This can be

seen as a regression model with 2 unknown parameters, m (the constant term) and s (the slope). By applying regression, the constant term m will be the sample mean of the $\ln(\text{quantiles})$ values, and the slope s can also be computed quite easily. The prior mean of the operation time is $\exp(m + 0.5 s^2)$, and the prior variance is $(\exp[s^2] - 1) \times \exp(2m + s^2)$. The prior sd is $\sqrt{\text{prior variance}}$. We take the value of m as the prior mean xs^* and the value of $1/s^2$ as τ . The posterior mean is then given by the formula:

$$\mu^* = xs^* \frac{\tau}{\tau + n_k} + \bar{x}_k \frac{n_k}{\tau + n_k} \quad (2)$$

The resulting weight is $w = \tau/(\tau + n)$ and the posterior mean is equal to:

$$\mu^* = wxs^* + (1 - w)\bar{x} \quad (3)$$

Note that this is the posterior mean of the log-times. The mean of the actual times is given by $\exp(\mu^* + 0.5\sigma^{*2})$, where σ^{*2} is the posterior variance. The prior variance is s^2 and the data variance is σ^2 .

An intuitive method is to weight these 2 values in the same way as was done for the mean, so that

$$\sigma^{*2} = ws^2 + (1 - w)\sigma^2 \quad (4)$$

Combining these results, we get: the posterior for the operation times is lognormal with mean μ^* and variance σ^{*2} . The mean of the operation times is then given by:

$$\exp(\mu^* + 0.5\sigma^{*2}) \quad (5)$$

Improving Coupling Between Estimates of Scheduled Time and the Actual Procedure Time

When reserving OR time for a procedure, the OR management needs to balance the costs of reserving too much time against the costs of reserving too little.²⁵ If too much time is allocated to a case, expensive OR capacity is likely to be wasted, leading to a decrease in OR utilization.^{12–14,16,17,21–23,27} With too little allocated capacity to a surgical case, the OR schedule must be modified, resulting in idle OR times and increased demand for anesthesiologists, nurses, and support staff. Improving coupling between estimates of scheduled time and the actual time reduces the prediction error of a scheduled surgical case. By using a simulation, we compared the effect on the prediction error of scheduling cases when applying 3 different case-modeling methods. The first method of estimating scheduled case durations is based on taking the trimmed mean time of the last 10 case durations. The second method uses the bias-corrected scheduled OR time. This method is based on the following linear regression based on data from 2005 to 2007: actual OR time = intercept + slope \times (scheduled OR time). This regression shows how much better it is for purposes of choosing how long to schedule a

case (when compared with lower/upper prediction bounds or times remaining in cases) to use statistically based methods compared with simple adjustment of the scheduled OR time. The last method uses the mean of the 3-parameter lognormal model.

To make it possible to compare the outcome of the 3 methods, only procedures with a case frequency of 10 or more were used, with 1 CPT-anesthesia code and fitting the 3-parameter lognormal model. We used the data available (from the sample). Historical data from 2005 to 2007 were used and then the window was expanded to include predictions made on each day in 2008 using data from 2005 to 2007 and from 2008 until the day before making the prediction. The originally scheduled sequence of cases was not changed. For instance, when scheduling an inguinal hernia repair (Lichtenstein) on January 2, 2008, only historical data up to and including January 1, 2008 were used. The actual time on January 2, 2008 was used for scheduling this procedure for January 4, 2008.

The difference between the actual OR time of a procedure is compared with the scheduled procedure time as calculated by each of the 3 methods. If the actual procedure time is larger than the scheduled time, that procedure is underreserved. Otherwise it is overreserved. For each method, the number of under- or overestimated procedures is counted as the mean under- and overestimated time per case. Differences in the mean under- and overestimated time per case between the 3 methods were tested with a paired *t*-test.

OR Inefficiency

OR inefficiency was defined as the sum of underutilized OR time and overutilized OR time, multiplied by the relative costs of overtime.^{16,23,26} Underutilized time was hours of staffed operating time at straight-time wages, but not used for surgery, setup, or cleanup of the OR. Overutilized time was hours after OR time, staffed at overtime. The relative cost of overtime in our study was 1.50. The cost per hour of overutilized OR time includes indirect costs, intangible costs, and retention and recruitment costs incurred on a long-term basis as a result of staff working late. Due to fixed OR capacity in our hospital (8 AM–4 PM), the short-term objective in maximizing OR efficiency is to reduce overutilized OR time.¹⁵ In Hospital A for example, the mean end time of all ORs running after 4 PM is 4.19 (± 17) min.³

We analyzed the effect of the different methods of case-duration prediction on OR efficiency. In the first method, we used the trimmed mean of the last 10 case durations; in the second method, we used the bias-corrected scheduled OR time; and the mean of the 3-lognormal model in the last method. Case scheduling with original cases in 2008 was used. For each method, add-on elective cases with their concomitant

turnover times were scheduled daily. Best Fit Descending was used, which is an off-line algorithm in which add-on elective cases are sorted based on longest to shortest with fuzzy constraints. Cases were considered in the order specified by the algorithm. If no OR had sufficient open time available for the case, and if sufficient open time was available in the OR with the most remaining time provided, the scheduled duration of the case was shortened by ≤ 15 min, then the case was assigned to the OR with the most remaining time.²⁸

For all cases (2 or more CPTs, and procedures with case frequency < 10) that are not meeting the criteria, we used the actual case duration as the scheduled duration (i.e., perfect retrospective knowledge).

After scheduling the cases and knowing the actual OR times of these same cases, the mean overutilized OR time was calculated considering each OR-day to be independent of all others. Differences in the mean overutilized OR time between the 3 methods were tested with a paired *t*-test.

Statistics and Software

The null hypothesis of the Shapiro-Wilk test statistic (*W*) is that a sample is from a normally distributed population. Thus, $P < 0.05$ for *W* rejects this supposition of normality. Most authors agree that this is the most reliable test for nonnormality for small to medium-sized samples.^{29–37} To perform the Shapiro-Wilk test, we used StatsDirect statistical software and also SPSS15, Excel 2007, and COBOL. Normal probability plots were examined visually for those CPT-anesthesia-(surgeon) combinations that were not well fitted by either the normal or lognormal models. We analyzed Q-Q-P-P and box plots to confirm the results of the Shapiro-Wilk test. Examination of the calculated skewness and kurtosis, and of the histogram, box plot, and normal probability plot for the data may provide clues as to why the data failed the Shapiro-Wilk. In our database, start and end of anesthesia time, surgical time, and procedure time are recorded exactly (to the minute). D'Agostino²⁹ indicated that the Shapiro-Wilk test can be affected by rounding.

RESULTS

Fitting the Normal and 2- and 3-Parameter Lognormal Models

In some of the procedures, we found outliers. In the database, there is a so-called "remark field" in which unexpected events during an OR are entered. The outliers we encountered were attributable to logistical problems (16 times) in the OR, surgeon arriving late (12 times), and OR team not ready (4 times). These outliers can be seen as incidental, so we removed these data. Table 4 shows the results of fitting CPT-anesthesia groups to the normal and the 2- and 3-parameter lognormal models for both hospitals separately.

If we look at Hospital A for the CPT-anesthesia combinations, then procedure times fit the normal

model 37.9% and surgical time 52.5%. The fits for the 2-parameter lognormal model ($P \geq 0.05$) are 57.7% and 69.6%, respectively. For the 3-parameter lognormal model, the fits for procedure time ($P \geq 0.05$) are 80.6% and 84.1% for surgical time. If we differentiate CPT-anesthesia-surgeon combinations, then procedure times fit the 2-parameter lognormal model ($P \geq 0.05$) in 70.4% of the combinations. The results for surgical times are 79.6% (Table 5). For the 3-parameter lognormal model, the fits for the procedure time are 87.6% and 90.7% for surgical time. The results for Hospital B are approximately in line with those for Hospital A.

We tried to understand why surgical time fits the normal and 2- and 3-parameter lognormal models better than procedure time. Procedure time consists of 3 main activities: administering anesthesia, preparing the patient for surgery, and performing the actual surgery. For the cases under study, the proportion of surgery time is on average 75% of the total procedure time. Preparation time and anesthesia time are 18% and 7%, respectively. While preparing the patient for surgery, relatively more OR staff members are involved in various activities and protocols compared with administering anesthesia and surgery. To better understand this, for every CPT-anesthesia code, we tested both the anesthesia time and preparation time for the 2-parameter lognormal model. With $P \geq 0.05$, 92.5% of anesthesia time is lognormally distributed, whereas 17.6% of the preparation time shows a fit to the lognormal model. Hence, preparation time is poorly modeled compared with anesthesia time. This could explain why procedure time is less well modeled for the lognormal model than surgical time.

Table 6 is a paired comparison of the 2- and 3-parameter lognormal models and the normal model using the Friedman test. We compared the normal model with the 2-parameter lognormal model and the 3-parameter model. The 2-parameter lognormal model was superior to the normal model for modeling procedure time and surgical time. The 3-parameter lognormal was superior to the 2-parameter lognormal model and normal model. Surgical time was estimated better than procedure time when modeling with both the 2- and 3-parameter lognormal models and the normal model.

Estimation with Specialist Prior Guess

In the Results section, we focused (arbitrarily) on the total thyroidectomy procedure (Table 7). The results of other procedures are found in Table 8. The 2 procedure times (261 and 198 min) are calculated after combining the prior statements of the specialists with the previously calculated times. Because we have data for 2 specialists, and therefore 8 time estimates, we get 8 equations with given values on the left-hand side, the values in the column "ln(quantiles) (Table 7)," and with unknown values of m and s .

In Table 7, we show the output of SPSS. The R^2 of this regression is 0.85, indicating a good fit. The outcomes are $m = 4.947$ and $s = 0.287$. In other words, the prior statements of the specialists can be translated as a lognormal model with a mean of 4.947 and sd of 0.287. The prior mean of the operation time is 147 min, and the prior variance is 1.847. The prior sd is 43 min. We took the value of $m = 4.947$ as the prior mean $x s^*$ and the value of $1/s^2 = 1/0.287^2 = 12.14$ as τ . The resulting weight was $w = 0.574$, and the posterior mean (Eq. 3, Methods) was equal to 5.162.

The prior variance, s^2 , is 0.0824, and the data variance is 0.1459. Weighing these 2 values as was done for the mean (Eq. 4, Methods) gives a value of 0.109 for σ^{*2} . Combining these results, we get the posterior for the operation times, which is lognormal distributed with mean $\mu^* = 5.162$ and variance $\sigma^{*2} = 0.109$. The mean of the operation times is then 184 min. Note that the prior mean was 147 min, and the data average time was 249 min. The posterior mean of 184 lies closer to the prior mean than to the data mean. This is because the prior distribution has a relatively small sd (43 min) as compared with that of the data (90 min) and because the number of data points (9) is small.

If we wish to determine, for instance, a 95% upper bound for the operation time, then this is done by estimating the 95% bound for the log-times. In our example, the log-time has normal posterior with $\mu^* = 5.162$ and variance $\sigma^{*2} = 0.109$, so that $\sigma^* = 0.330$. The 95% upper bound for the log-time is then $\mu^* + 1.645\sigma^* = 5.705$. The bound for the time itself is then $\exp(5.705) = 300$ min.

Table 8 presents the results for the data mean (sd), prior [mean time, (sd)] and posterior [mean time, sd] for 30 procedures. From this table, we see that the posterior mean is a weight of the data mean and the prior mean. The variance of the posterior mean always lies between the data variance and prior variance.

Improving Coupling Between Estimates of Scheduled Time and the Actual Procedure Time

In Hospital A (Table 9), under the standard method, the average overreserving per case is 22.9 min (21.4 min), whereas the average underreserving is 21.6 min (18.8 min).

The result of the regression is: actual OR time = $18.16 + 0.88 \times (\text{scheduled OR time})$ with standard error of the constant 0.30 and slope 0.04 ($P < 0.0001$), R^2 0.55.

Applying the biased regression, then the average overreserving per case is 16.3 min (9.4 min), whereas the average underreserving is 12.6 min (7.6 min). For the 3-lognormal model, the results are 12.9 (8.4) overreserving and 9.6 (5.4) underreserving. The average overreserving and underreserving among the 3 methods is significant ($P < 0.001$). The results for Hospital B are in line with Hospital A (Table 8).

Table 4. Tabular Comparison of Sample Size and Shapiro-Wilk Goodness-of-Fit *P* Values for the Normal Model, 2-Parameter Lognormal Model, and 3-Parameter Lognormal Model for Procedure Time and Surgical Time (1 CPT-Anesthesia Combinations)

Normal model						
Hospital A						
Procedure time						
Normal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	271	36.8%	160	21.7%	431	58.5%
Medium, 30 ≤ <i>n</i> < 200	153	20.8%	100	13.6%	253	34.3%
Large, <i>n</i> ≥ 200	34	4.6%	19	2.6%	53	7.2%
Total	458	62.1%	279	37.9%	737	100.0%
Surgical time						
Normal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	204	27.7%	227	30.8%	431	58.5%
Medium, 30 ≤ <i>n</i> < 200	124	16.8%	129	17.5%	253	34.3%
Large, <i>n</i> ≥ 200	22	3.0%	31	4.2%	53	7.2%
Total	350	47.5%	387	52.5%	737	100.0%
Hospital B						
Procedure time						
Normal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	194	34.0%	143	25.1%	337	59.1%
Medium, 30 ≤ <i>n</i> < 200	138	24.2%	60	10.5%	198	34.7%
Large, <i>n</i> ≥ 200	34	6.0%	1	0.2%	35	6.1%
Total	366	64.2%	204	35.8%	570	100.0%
Surgical time						
Normal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	146	25.6%	191	33.5%	337	59.1%
Medium, 30 ≤ <i>n</i> < 200	100	17.5%	98	17.2%	198	34.7%
Large, <i>n</i> ≥ 200	15	2.6%	20	3.5%	35	6.1%
Total	261	45.8%	309	54.2%	570	100.0%
2-parameter lognormal model						
Hospital A						
Procedure time						
2-parameter lognormal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	179	24.3%	252	34.2%	431	58.5%
Medium, 30 ≤ <i>n</i> < 200	112	15.2%	141	19.1%	253	34.3%
Large, <i>n</i> ≥ 200	21	2.8%	32	4.3%	53	7.2%
Total	312	42.3%	425	57.7%	737	100.0%
Surgical time						
2-parameter lognormal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	121	16.4%	310	42.1%	431	58.5%
Medium, 30 ≤ <i>n</i> < 200	89	12.1%	164	22.3%	253	34.3%
Large, <i>n</i> ≥ 200	14	1.9%	39	5.3%	53	7.2%
Total	224	30.4%	513	69.6%	737	100.0%
Hospital B						
Procedure time						
2-parameter lognormal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	101	17.7%	236	41.4%	337	59.1%
Medium, 30 ≤ <i>n</i> < 200	86	15.1%	112	19.6%	198	34.7%
Large, <i>n</i> ≥ 200	29	5.1%	6	1.1%	35	6.1%
Total	216	37.9%	354	62.1%	570	100.0%
Surgical time						
2-parameter lognormal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	84	14.7%	253	44.4%	337	59.1%
Medium, 30 ≤ <i>n</i> < 200	64	11.2%	134	23.5%	198	34.7%
Large, <i>n</i> ≥ 200	9	1.6%	26	4.6%	35	6.1%
Total	157	27.5%	413	72.5%	570	100.0%
3-parameter lognormal model						
Hospital A						
Procedure time						
3-parameter lognormal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	86	11.7%	345	46.8%	431	58.5%
Medium, 30 ≤ <i>n</i> < 200	41	5.6%	212	28.8%	253	34.3%
Large, <i>n</i> ≥ 200	16	2.2%	37	5.0%	53	7.2%
Total	143	19.4%	594	80.6%	737	100.0%
Surgical time						
3-parameter lognormal	<i>P</i> < 0.05		<i>P</i> ≥ 0.05		CPT codes	
Small, 10 ≤ <i>n</i> < 30	67	9.1%	364	49.4%	431	58.5%
Medium, 30 ≤ <i>n</i> < 200	37	5.0%	216	29.3%	253	34.3%
Large, <i>n</i> ≥ 200	13	1.8%	40	5.4%	53	7.2%
Total	117	15.9%	620	84.1%	737	100.0%

(Continued)

Downloaded from http://journals.lww.com/anesthesia-analgia by BMDMS5P1HKav1zEoun1QIN4a+kULhEZgbsl
Ho4M0hCwCkX1AMvYQp/IOHrHD3BD0DfY7TVSfAC3VC4/OA/pDDa8KKGK/v0ym+78= on 07/31/2024

Table 4. Continued

Hospital B						
Surgical time						
3-parameter lognormal	$P < 0.05$		$P \geq 0.05$		CPT codes	
Small, $10 \leq n < 30$	38	6.7%	299	52.5%	337	59.1%
Medium, $30 \leq n < 200$	35	6.1%	163	28.6%	198	34.7%
Large, $n \geq 200$	8	1.4%	27	4.7%	35	6.1%
Total	81	14.2%	489	85.8%	570	100.0%
Procedure time						
3-parameter lognormal	$P < 0.05$		$P \geq 0.05$		CPT codes	
Small, $10 \leq n < 30$	52	9.1%	285	50.0%	337	59.1%
Medium, $30 \leq n < 200$	31	5.4%	167	29.3%	198	34.7%
Large, $n \geq 200$	19	3.3%	16	2.8%	35	6.1%
Total	102	17.9%	468	82.1%	570	100.0%

OR Inefficiency

In Hospital A 12,138 cases were scheduled. The mean overutilized OR time (min) per OR per day for the standard method is 23.4 (22.7–24.0), for the biased corrected mean time 16.6 (16.1–17.2) and the 3-lognormal 6.6 (6.2–6.9). For Hospital B 8,794 cases were scheduled. The mean overutilized OR time per OR per day for the standard method is 30.6 (29.6–31.5), for the bias-corrected mean time 22.2 (21.4–22.9) and for the 3-lognormal model 10.6 (10.1–11.2).

DISCUSSION

Modeling the distribution of OR cases is one of the key steps in a planning process. In our study, the focus is more on decision making before the day of surgery. In other studies^{12,16,25} the focus is toward decisions on the day of surgery. These do not involve average OR times, but rather lower prediction bounds, upper prediction bounds, and especially times remaining in cases. Both focuses are helpful in the effective scheduling and efficient use of expensive surgical resources. We find that the percentage of cases fitting the normal and 2- and 3-parameter lognormal models is higher for surgical time than for total procedure time (the opposite was true for Strum et al.¹¹). The evidence supports the idea that type of surgery is the most important single source of variability among surgeries.⁷ Using the bisection method and applying the 3-parameter lognormal model fits procedure time and surgical time better than the 2-parameter lognormal model without shift parameter. This can be explained by the fact that the 2-parameter lognormal model is a limitation of the 3-parameter lognormal model. When segmenting to the factor surgeon, the fits are even higher for the 2- and 3-parameter lognormal models. One could ask why the fits are better with CPT-anesthesia-surgeon segmentations. Offering an *a priori* hypothesis, Strum et al.⁷ suggest that this may be attributable to surgeon work rates. If Strum et al. are correct, then segmentation into surgeon-specific groups should result in more homogeneous work rates and thus a better fit to the lognormal. Another reason is that, because of further segmentation, the number of available cases reduces and, because of

this reduction of cases, the P values will increase. This could also explain why the lognormal model fits for the CPT-anesthesia-surgeon combinations are higher. We confirm as in other studies that small groups have a better fit than the medium and large groups. This lack of discrimination relates to the design of the statistical tests. D'Agostino, Shapiro and Wilk, and others^{29,31–37} discuss the fact that goodness-of-fit tests become more discriminating as the sample sizes increase. Conversely, it may be obvious that samples with $n < 10$, for example, may indiscriminately fit almost any model.

If few data are available, the use of prior information given by the surgeon may lead to a better estimation of the case duration. Because the posterior distribution contains all the information we need to make statistical decisions, we can use it for predicting case durations and case scheduling. The uncertainty of the posterior data is less than when using only the data without prior information. On the other hand, if the amount of historical data for a specific procedure increases, the usefulness of the prior information will decrease. This is because with an increasing number of observations, the sample mean will determine the outcome. Our approach differs in some respects from the classical one as discussed, for instance, by Dexter et al.¹² This is caused by the fact that we have prior data that are quite informative and that can be translated in terms of a lognormal prior distribution. In the classical approach, the prior on the 2 parameters μ and σ^2 consists of 3 parts:

1. For given σ , the (conditional) prior for μ is normal. The prior for σ is inverted γ .
2. The (unconditional, marginal) prior for μ is a t -distribution.
3. The (unconditional, marginal) posterior for μ is (another) t -distribution.

Our prior information is not directly related to mean and variance, but can be translated to mean and variance of the normal distribution (of the log-times). Therefore, we combined a normal prior with a normal distribution of the observed data. However, in applying the calculation rules to get the posterior, we used

Table 5. Tabular Comparison of Sample Size and Shapiro-Wilk Goodness-of-fit *P*-values for the Normal Model, 2-Parameter Lognormal Model, and 3-Parameter Lognormal Model for Procedure Time and Surgical Time (1 CPT-Anesthesia-Surgeon Combinations)

Normal model						
Hospital A						
Procedure time						
Normal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	329	24.5%	322	24.0%	651	48.5%
Medium, $30 \leq n < 200$	371	27.7%	310	23.1%	681	50.8%
Large, $n \geq 200$	8	0.6%	1	0.1%	9	0.7%
Total	708	52.8%	633	47.2%	1341	100.0%
Surgical time						
Normal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	228	17.0%	423	31.5%	651	48.5%
Medium, $30 \leq n < 200$	329	24.5%	352	26.2%	681	50.8%
Large, $n \geq 200$	7	0.5%	2	0.1%	9	0.7%
Total	564	42.1%	777	57.9%	1341	100.0%
Hospital B						
Procedure time						
Normal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	318	27.7%	226	19.7%	544	47.4%
Medium, $30 \leq n < 200$	301	26.2%	295	25.7%	596	52.0%
Large, $n \geq 200$	6	0.5%	1	0.1%	7	0.6%
Total	625	54.5%	522	45.5%	1147	100.0%
Surgical time						
Normal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	219	19.1%	325	28.3%	544	47.4%
Medium, $30 \leq n < 200$	261	22.8%	335	29.2%	596	52.0%
Large, $n \geq 200$	5	0.4%	2	0.2%	7	0.6%
Total	485	42.3%	662	57.7%	1147	100.0%
2-parameter lognormal model						
Hospital A						
Procedure time						
2-parameter lognormal		$P < 0.05$		$P \geq 0.05$		CPT codes
small, $10 \leq n < 30$	201	15.0%	450	33.6%	651	48.5%
medium, $30 \leq n < 200$	189	14.1%	492	36.7%	681	50.8%
large, $n \geq 200$	7	0.5%	2	0.1%	9	0.7%
Total	397	29.6%	944	70.4%	1341	100.0%
Surgical time						
2-parameter lognormal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	101	7.5%	550	41.0%	651	48.5%
Medium, $30 \leq n < 200$	167	12.5%	514	38.3%	681	50.8%
Large, $n \geq 200$	5	0.4%	4	0.3%	9	0.7%
Total	273	20.4%	1068	79.6%	1341	100.0%
Hospital B						
Procedure time						
2-parameter lognormal		$P < 0.05$		$P \geq 0.05$		CPT codes
small, $10 \leq n < 30$	221	19.3%	323	28.2%	544	47.4%
medium, $30 \leq n < 200$	127	11.1%	469	40.9%	596	52.0%
large, $n \geq 200$	4	0.3%	3	0.3%	7	0.6%
Total	352	30.7%	795	69.3%	1147	100.0%
Surgical time						
2-parameter lognormal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	124	10.8%	420	36.6%	544	47.4%
Medium, $30 \leq n < 200$	146	12.7%	450	39.2%	596	52.0%
Large, $n \geq 200$	3	0.3%	4	0.3%	7	0.6%
Total	273	23.8%	874	76.2%	1147	100.0%
3-parameter lognormal model						
Hospital A						
Procedure time						
3-parameter lognormal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	77	5.7%	574	42.8%	651	48.5%
Medium, $30 \leq n < 200$	84	6.3%	597	44.5%	681	50.8%
Large, $n \geq 200$	5	0.4%	4	0.3%	9	0.7%
Total	166	12.4%	1175	87.6%	1341	100.0%
Surgical time						
3-parameter lognormal		$P < 0.05$		$P \geq 0.05$		CPT codes
Small, $10 \leq n < 30$	57	4.3%	594	44.3%	651	48.5%
Medium, $30 \leq n < 200$	64	4.8%	617	46.0%	681	50.8%
Large, $n \geq 200$	4	0.3%	5	0.4%	9	0.7%
Total	125	9.3%	1216	90.7%	1341	100.0%

(Continued)

Table 5. Continued

Hospital B						
Procedure time						
3-parameter lognormal	$P < 0.05$		$P \geq 0.05$		CPT codes	
Small, $10 \leq n < 30$	81	7.1%	463	40.4%	544	47.4%
Medium, $30 \leq n < 200$	95	8.3%	501	43.7%	596	52.0%
Large, $n > 200$	3	0.3%	4	0.3%	7	0.6%
Total	179	15.6%	968	84.4%	1147	100.0%
Surgical time						
3-parameter lognormal	$P < 0.05$		$P \geq 0.05$		CPT codes	
Small, $10 \leq n < 30$	57	5.0%	487	42.5%	544	47.4%
Medium, $30 \leq n < 200$	89	7.8%	507	44.2%	596	52.0%
Large, $n \geq 200$	2	0.2%	5	0.4%	7	0.6%
Total	148	12.9%	999	87.1%	1147	100.0%

Table 6. Friedman Results for Paired Comparison of the Shapiro-Wilk P -Value for the Normal, 2-Parameter Lognormal, and 3-Parameter Lognormal for Procedure Time and Surgical Times

Hospital A				
Hypothesis	$\ln 2(\text{pt}) = \ln 2(\text{st})$	$N(\text{pt}) = N(\text{st})$	$\ln 2(\text{st}) = N(\text{st})$	$\ln 2(\text{pt}) = N(\text{pt})$
Friedman test statistic	41	335	247	298
Kendall coefficient of concordance	0.015	0.228	0.051	0.095
Rank sum	1315–1847	1121–1901	1598–1279	1761–1357
P	0.0039	0.0011	0.0038	0.0117
Hospital B				
Hypothesis	$\ln 2(\text{pt}) = \ln 2(\text{st})$	$N(\text{pt}) = N(\text{st})$	$\ln 2(\text{st}) = N(\text{st})$	$\ln 2(\text{pt}) = N(\text{pt})$
Friedman test statistic	28	241	161	202
Kendall coefficient of concordance	0.151	0.625	0.094	0.597
Rank sum	895–981	1002–1354	1204–906	1364–994
P	0.0084	0.0033	0.0128	0.0094
Hospital A				
Hypothesis	$\ln 3(\text{pt}) = \ln 3(\text{st})$	$\ln 3(\text{st}) = \ln 2(\text{st})$	$\ln 3(\text{pt}) = \ln 2(\text{pt})$	
Friedman test statistic	121	501	498	
Kendall coefficient of concordance	0.011	0.064	0.054	
Rank sum	1528–1978	1705–1125	1814–1134	
P	0.0021	0.0004	0.0003	
Hospital B				
Hypothesis	$\ln 3(\text{pt}) = \ln 3(\text{st})$	$\ln 3(\text{st}) = \ln 2(\text{st})$	$\ln 3(\text{pt}) = \ln 2(\text{pt})$	
Friedman test statistic	87	435	477	
Kendall coefficient of concordance	0.131	0.101	0.314	
Rank sum	904–1041	1304–879	1415–987	
P	<0.000	<0.000	<0.000	
Hospital A				
Hypothesis	$\ln 3(\text{st}) = N(\text{st})$	$\ln 3(\text{pt}) = N(\text{pt})$		
Friedman test statistic	89	298		
Kendall coefficient of concordance	0.0047	0.0059		
Rank sum	1487–897	1948–1007		
P	<0.000	<0.000		
Hospital B				
Hypothesis	$\ln 3(\text{st}) = N(\text{st})$	$\ln 3(\text{pt}) = N(\text{pt})$		
Friedman test statistic	147	198		
Kendall coefficient of concordance	0.0089	0.021		
Rank sum	1487–879	1547–921		
P	<0.000	<0.000		

pt = procedure time; st = surgical time; N = normal model; $\ln 2$ = 2-parameter lognormal; $\ln 3$ = 3-parameter lognormal.

the classical framework, which is not fully consistent. However, the central formulas (equations 2 and 4) have a direct intuitively appealing interpretation that also applies in our framework: we take a weighted average of the prior and data information, and the weights are inversely proportional to the uncertainty involved in both types of information: proportional to $\tau = 1/s^2 = 1/(\text{prior variance})$ and to $n = 1/(1/n) = 1/(\text{data variance})$.

If we wish to more closely model the classical set-up, we need to estimate the parameters α and β of the prior (inverted γ) distribution of the sd. These 2 parameters can be estimated by considering all other types of operations and modeling the resulting set of (inverted) sample variances for all these types of operations as in the study by Dexter et al.¹² The (marginal) posterior of the mean (of the log-times) then becomes t instead of normal.

Table 7. Estimation of the Mean Procedure Time from Prior and Actual Data Procedure: Total Thyroidectomy

Actual data		Prior estimates		
Actual	ln(actual)	Specialist	Quantiles	ln(quantiles)
268	5.591	1	100	4.605
126	4.836	1	150	5.011
378	5.935	1	170	5.136
311	5.740	1	200	5.298
172	5.147	2	110	4.7
361	5.889	2	160	5.075
162	5.088	2	200	5.298
261	5.5652	220	5.394	

Descriptive Statistics of Actual Data

	N	Minimum	Maximum	Mean	SD
Time	9	126	378	248.56	89.972
ln(time)	9	4.836	5.935	5.453	0.382

Output Generated by SPSS

Coefficients				
	B	SE	t	Significance
Constant	4.947	0.047	105.804	0.000
Z vector	0.287	0.049	5.835	0.001

Dependent variable: ln(quantiles).

Table 8. Results of the Estimation of the Mean Procedure Time from Prior and Actual Data

Procedure	Historical number cases	Number prior estimates	Data mean (min)	Historic data (SD) (min)	Prior mean (min)	Prior SD (min)	Posterior mean (min)	Posterior SD (min)
Osteochondritis dissecans knee	4	2	41	16	55	12	51	14
Corticotomy chin	4	2	80	30	74	22	77	25
Strumectomy	9	2	211	53	200	46	205	48
Nephrectomy (abd/via lumb, open)	9	2	214	43	205	49	210	48
Parathyroidectomy	8	1	158	28	182	44	174	38
Nerve repair	5	3	177	61	141	39	158	47
Recto-Sigmoidresection Hartman	8	1	176	26	143	48	161	34
Extirpatie glandula submandibularis	9	2	93	22	72	22	83	22
Stomach resection with cholecystectomy	5	1	198	23	186	39	187	36
Distal tibia	4	2	128	48	81	24	86	30
PTFE loop	4	2	191	101	171	87	172	96
Liver surgery	4	2	154	61	128	39	134	44
Sigmoid resection (open)	3	2	232	62	225	46	226	51
Endoscopic staggerings lymphadenectomy	8	2	156	25	159	40	157	40
Latissimus dorsi lap upper body	4	1	275	45	279	40	277	41
Total thyroidectomy	7	2	249	90	147	43	184	62
Transversectomy	7	2	145	42	168	38	165	39
Lymph dissection	9	2	171	55	91	40	136	47
Ileostomy	3	2	149	45	91	40	111	42
Reconstruction AP colon intraperitoneaal	6	2	95	54	88	26	91	53
PKLND, staggerings lymphadenectomy	8	2	274	158	228	60	238	110
Arhrdese knee	3	1	239	45	224	34	229	39
Intracondylaie humerus	4	1	107	30	95	44	93	37
Debulking	4	2	260	74	124	51	153	70
Ureterimplantation	4	1	203	90	205	57	194	71
Nissen fundoplication	4	1	234	36	239	41	235	40
Rectosigmoidresection	4	1	231	70	172	21	173	25
Enteroanastomosis	3	1	215	175	137	37	156	57
Reconstructive surgery by Roux Y	5	2	332	134	300	61	311	78
Endoscopic hemicolectomy	8	3	224	79	216	48	219	57

PTFE = polytetrafluoroethylene.

Furthermore, our results for CPTs with few data may potentially be useful if the data from the 2 hospitals were compared with findings in another study.¹² The latter article did not find the Bayes method to have important

value for the mean. The overall effect for every case including those with multiple CPTs would be needed.

Finally, we find that compared with the standard way of case scheduling using the mean of the

Table 9. Improving Coupling Between Estimates of Scheduled Time and the Actual Time

Method	Method			Effect	
	1 Standard	2 Regression	3 3-parameter lognormal	(1)–(2)	(1)–(3)
Hospital A [History: 2005–2007, cases used for scheduling 2008 ($n = 10,664$)]					
Total hours of overreserving	2284	2401	1041	–5.1%	54.4%
No. of cases overreserving	5992	8815	4845	–47.1%	19.1%
Average overreserving/case (min)	22.9	16.3	12.9	28.5%	43.6%
sd (min)	21.4	9.4	8.4		
Total hours of underreserving	1534	387	765	74.8%	50.1%
No. of cases underreserving	4255	1849	4780	56.5%	–12.3%
Average underreserving/case (min)	21.6	12.6	9.6	41.9%	55.6%
sd (min)	18.8	7.6	5.4		
Hospital B [History: 2005–2007, cases used for scheduling 2008 ($n = 8794$)]					
Total hours of overreserving	1867	1314	747	29.6%	60.0%
No. of cases overreserving	5031	5631	4294	–11.9%	14.6%
Average overreserving per case (min)	22.3	14.0	10.4	37.1%	53.1%
sd (min)	21.8	8.3	7.1		
Total hours of underreserving	1874	949	924		
No. of cases underreserving	3713	3163	4078	14.8%	–9.8%
Average underreserving per case (min)	30.3	18.0	13.6	40.6%	55.1%
sd (min)	26.7	10.1	8.2	62.2%	69.3%

Overreserving: scheduled time–actual time >0 (h).

Underreserving: scheduled time–actual time <0 (h).

3-parameter lognormal distribution for case scheduling reduces the mean overreserving OR time per case up to 53.1% and the underreserving OR time up to with 55.6%. Using the 3 parameter lognormal model for case scheduling causes a lower mean overutilized OR time up to 20.0 (19.7–20.3) min per OR per day as compared with the standard method and 11.6 (11.3–12.0) min per OR per day as compared with the bias-corrected scheduled OR time.

Limitations

The prior information could be misleading when the prior variance is too small, because specialists may underestimate the variance. Surgeon case durations for specific procedures may change progressively, for example, as a result of subtle changes in the demographics of a patient population.¹⁵ We asked specialists in every specialty if they were aware of these changes. None recognized that these changes had occurred in the past 4 yr. We assumed that the specialists do not remember the previous operation times, so we treated all realized times (past and current) as containing similar information. In practice, surgeons may or may not actually remember historical case durations. Although the studied procedures have a relatively low occurrence and are performed by different surgeons, we believe that there may be an effect of the memory of an individual surgeon on the results but it will be very small.

In the simulation for case-duration prediction and efficiency gains, we omitted procedures not fitting the 3-parameter lognormal mode and procedures with a case frequency <10 . Because of this, the real efficiency gains may be overestimated. Although in the hospitals

under study, 86% of all cases consist of 1 CPT code, we cannot make general conclusions or statements regarding the impact of improving case-duration prediction on the efficiency of use of OR time, but only as related to the cases under study.

CONCLUSION

OR case scheduling can be improved by using the 3-parameter lognormal model with surgeon effects and by using the surgeon's prior guesses for rarely observed CPTs. Compared with standard case scheduling practices and the bias-corrected method using the 3-lognormal model for case scheduling, both significantly reduce the average underestimated and overestimated OR time per case as well as the OR inefficiency.

REFERENCES

1. Stachota P, Normandin P. Reasons registered nurses leave or change employment status. *J Nurs Adm* 2003;33:111–8
2. Thomson TP, Brown H. Turnover of licensed nurses in skilled nursing facilities. *Nurs Econ* 2002;20:66–9
3. 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
4. Rossiter CE, Reynolds JA. Automatic monitoring of the time waited in out-patient departments. *Med Care* 1963;1:218–25
5. Barnoon S, Wolfe H. Scheduling a multiple operating room system: a simulation approach. *Health Serv Res* 1968;3:272–85
6. Hancock WM, Walter PR, More RA, Glick ND. Operating room scheduling data base analysis for scheduling. *J Med Syst* 1988;12:397–409
7. Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 2000;92:1454–66
8. 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

9. May JH, Strum DP, Vargas LG. Fitting the lognormal distribution to surgical procedure times. *Decis Sci* 2000;31:129–48
10. Spangler WE, Strum DP, Vargas LG, May JH. Estimating procedure times for surgeries by determining location parameters for the lognormal model. *Health Care Manag Sci* 2004;7:97–104
11. Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. *Anesthesiology* 2000;92:1160–7
12. Dexter F, Ledolter J. Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historical data. *Anesthesiology* 2005;103:1259–67
13. Dexter F, Traub RD. Statistical method for predicting when patients should be ready on the day of surgery. *Anesthesiology* 2000;93:1107–14
14. Dexter F, Traub R. How to schedule surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesth Analg* 2002;94:933–42
15. 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
16. Dexter F, Epstein RH, Traub R, Xiao Y. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. *Anaesthesiology* 2004;101:1444–53
17. 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
18. 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
19. Dexter F, Epstein RH, Marsh HM. A statistical analysis of weekday operating room anesthesia group staffing costs at nine independently managed surgical suites. *Anesth Analg* 2001;92:1493–8
20. Dexter F, Macario A. 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
21. Dexter F, Epstein RH. Operating room efficiency and scheduling. *Anaesthesiology* 2005;18:195–8
22. Friedman DM, Sokal SM, Chang Y, Berger DL. Increasing operating room efficiency through parallel processing. *Ann Surg* 2006;243:10–4
23. McIntosh C, Dexter F, Epstein RH. Impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: tutorial using data from an Australian hospital. *Anesth Analg* 2006;103:1499–516
24. Dexter F, Dexter EU, Masursky D, Nussmeier NA. Systematic review of general thoracic surgery articles to identify predictors of operating room case durations. *Anesth Analg* 2008;106:1232–41
25. Dexter F, Epstein R, Lee R, 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
26. Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. *Anesthesiology* 1999;90:1176–85
27. Dexter F, Macario A. When to release allocated operating room time to increase operating room efficiency. *Anesth Analg* 2004;98:758–62
28. Dexter F, Macario A, Traub RD. Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Use of bin packing algorithms and fuzzy constraints in operating room management. *Anesthesiology* 1999;91:1491–500
29. D’Agostino RB. Tests for the normal distribution. In: D’Agostino RB, Stephens MA, eds. *Goodness-of fit techniques*. New York: Marcel Dekker, Inc., 1986:367–419
30. Conover WJ. *Practical nonparametric statistics*. 3rd ed. New York: Wiley, 1999
31. Shapiro SS, Wilk MB. An analysis of variance test for normality. *Biometrika* 1965;52:591–9
32. Royston JP. The W test for normality. *Appl Stat* 1982;31:176–80
33. Royston JP. Shapiro-Wilk normality test and P-value. *Appl Stat* 1995;44:547–51
34. Madansky A. *Testing for normality, prescriptions for working statisticians*. New York: Springer-Verlag, 1988:15–31
35. Pearson ES, D’Agostino RB, Bowman KO. Tests for departure from normality: comparison of powers. *Biometrika* 1977;64:231–46
36. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52:591–611
37. Shapiro SS, Wilk MB, Chen HJ. A comparative study of various tests for normality. *J Am Stat Assoc* 1968;63:1343–72