■ ECONOMICS

Estimating Times of Surgeries with Two Component Procedures

Comparison of the Lognormal and Normal Models

David P. Strum, M.D., * *Jerrold H. May, Ph.D.,*† *Allan R. Sampson, Ph.D.,*‡ *Luis G. Vargas, Ph.D.,*§
William E. Spangler, Ph.D.‖

Background: Variability inherent in the duration of surgical procedures complicates surgical scheduling. Modeling the duration and variability of surgeries might improve time estimates. Accurate time estimates are important operationally to improve utilization, reduce costs, and identify surgeries that might be considered outliers. Surgeries with multiple procedures are difficult to model because they are difficult to segment into homogenous groups and because they are performed less frequently than single-procedure surgeries.

Methods: The authors studied, retrospectively, 10,740 surgeries each with exactly two CPTs and 46,322 surgical cases with only one CPT from a large teaching hospital to determine if the distribution of dual-procedure surgery times fit more closely a lognormal or a normal model. The authors tested model goodness of fit to their data using Shapiro-Wilk tests, studied factors affecting the variability of time estimates, and examined the impact of coding permutations (ordered combinations) on modeling.

Results: The Shapiro-Wilk tests indicated that the lognormal model is statistically superior to the normal model for modeling dual-procedure surgeries. Permutations of component codes did not appear to differ significantly with respect to total procedure time and surgical time. To improve individual models for infrequent dual-procedure surgeries, permutations may be reduced and estimates may be based on the longest component procedure and type of anesthesia.

Conclusions: The authors recommend use of the lognormal model for estimating surgical times for surgeries with two component procedures. Their results help legitimize the use of log transforms to normalize surgical procedure times prior to hypothesis testing using linear statistical models. Multiple-procedure surgeries may be modeled using the longest (statistically most important) component procedure and type of anesthesia.

SURGICAL scheduling is complicated by variability inherent in the duration of surgical procedures. Modeling that variability, in turn, provides a mechanism to gener-

ate time estimates that are important operationally to improve operating room utilization and decrease surgical costs.[1] Time estimates are important for building simulation models of surgical environments[2] and for decision analysis based on such simulations.

We studied *individual* and *aggregate* models in this article for surgeries that have two Current Procedural Terminology (CPT) codes associated with them. Individual models fit a mathematical distribution to surgeries segmented by dual procedure code, surgeon, or type of anesthesia. Modeling in this manner permitted each dual procedure to have its own unique probability distribution and is most useful if there is reason to believe that different dual procedures should have different models. When fitting individual models to subsets of the data, it is important to reduce coding permutations (whenever possible) to avoid unnecessary segmentation and thus excessively small samples with reduced estimation precision.

We also studied aggregate models that use the records from all dual-procedure surgeries to derive expected time estimates. Aggregate models make implicit assumptions that the variability in each of the data subsets is similar. Using aggregate models, all dual-procedure surgeries are modeled simultaneously to provide time estimates that best describe all the surgeries.

Modeling of surgeries involving only one CPT was investigated previously.[3,4] Times for these surgeries were well modeled by the lognormal distribution. We wanted to test the hypothesis that the lognormal distribution is also the best model for estimating times for two procedures performed in the same surgical session.

Dual-procedure surgeries are difficult to model individually because they are performed infrequently. It is logical to assume that time estimates for multiple-procedure surgeries might be constructed using the component procedures. It is unlikely that these estimates would reflect only the simple arithmetic sum of the component estimates.

Coding permutations (order-dependent combinations of the same codes) are observed for multiple procedures. Additional research is needed to determine if coding permutations represent statistically different surgeries with respect to duration estimates. We investigated dual-procedure surgeries to determine if the lognormal distribution[5] was superior to the normal

* Associate Professor of Anesthesiology and Associate Professor of Business, Department of Anesthesiology, Queens University. † Professor of Operations, Decision Sciences, and Artificial Intelligence and of Intelligent Systems, § Professor of Operations, Decision Sciences, and Artificial Intelligence, The Joseph M. Katz Graduate School of Business, ‡ Professor, Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania. ‖ Assistant Professor of Information Technology, J.F. Donahue Graduate School of Business, Duquesne University, Pittsburgh, Pennsylvania.

Address reprint requests to Dr. Strum: Queen's University Department of Anesthesiology, KGH, 76 Stuart Street, Kingston, Ontario, K7L 2V7, Canada. Address electronic mail to: dstrum1@cogeco.ca. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

distribution for modeling surgical and total times. Finally, we also investigated several factors (CPT1, CPT2, surgical subspecialty, type of anesthesia, age, and emergency status) associated with variability in time estimates for dual-procedure surgeries.

Methods

We reviewed all recorded surgical cases from a large academic health sciences center performed over a 7-yr period from 1989 to 1995. Use of anonymous patient records was approved by the human subjects review committee of the institution that collected the data. Data were collected using a previously described computerized system.[6] Variables included in the initial data included total procedure time (TT), defined as the time from entry into the operating suite until emergence from anesthesia, surgical procedure time (ST), defined as the time from incision to closure of the surgical wound, age, American Society of Anesthesiologists physical status classification, type of anesthesia, CPT codes, emergency status (Emerg), and surgical specialty category (CAT) as defined using main headers from the CPT classification.[7]

Detailed Description of the Data

Of 60,643 total case records in the initial database, 779 were omitted from analysis due to incomplete data, leaving 59,864 surgeries that included between one and three component CPTs. There were 46,322 surgeries with only one CPT code, 10,740 with exactly two different CPT codes, and 2,802 patients with three CPT codes. To reduce confounding factors, we confined our analyses of multiple procedures to those described by exactly two CPT codes. Dual-procedure surgeries in our database were named (jointly) for both provider-ordered procedures, one code designated first (CPT1) and the other code second (CPT2). For example, when modeling initially, CPT1 = 52000, CPT2 = 53899 and CPT1 = 53899, CPT2 = 52000 were considered *different* dual-procedure surgeries. TT for the former was 69 ± 48 min, n = 138 surgeries, and TT for the latter was 96 ± 63 min, n = 21 surgeries (mean ± SD). CPT 52000 was cystoscopy and CPT 53899 was urological surgery. Basic statistics were summarized for dual-procedure surgeries (designated CPT1-2) for TT and ST.

Individual Probability Models

CPT1-2 values for TT and ST were investigated for lognormality using log transformations and normal probability plots. Preliminary analyses[4] indicated that to obtain a better individual model fit, data should be subdivided into (more) homogeneous subgroups by CPT and type of anesthesia (general, local, regional, monitored) prior to being fit to a distribution. To determine the best model for estimating procedure times, samples were

segmented by dual-procedure surgery (CPT1-2) and the normal and lognormal models fit to each. Permutations of the component codes were assumed (initially) to represent different surgeries, and each was fit separately. Samples were not segmented initially by type of anesthesia to avoid reducing excessively the sparse number of surgeries to be fit. Other issues related to general lognormal modeling of surgical times have been discussed elsewhere.[3,8,9]

We performed Shapiro-Wilk (SW) goodness-of-fit tests[10,11] to determine whether a data sample was consistent with a normal distribution. The SW test for normality was applied to the logs of the data values, thereby creating a test for lognormality. When using the SW test, one assumes that the null hypothesis is that the model describes the data. Hence, a large P value indicates that it is not reasonable to reject the null hypothesis, *i.e.*, the data fit the model well. We tested TT and ST for all dual-procedure surgeries with case frequencies of five or more.

We cross-tabulated the SW test results for all CPT1-2 combinations by P value of the SW tests to compare goodness-of-fit tests for the normal and lognormal models. To detect the influence of sample size on the SW tests, we divided the sample arbitrarily into small (n ≤ 30)- and medium (n > 30)-sized samples. Because commonly used levels of significance for hypothesis testing are between 1% and 10% for a test like SW, a frequently used rule of thumb is to regard a P value of at least 0.10 as leading to retention of the null hypothesis (the model fits well) and a P value less than 0.01 as always leading to its rejection (the model fits poorly). We interpreted values between 0.01 and 0.1 as a mediocre fit for the model.

We compared the *overall* performance of the individual lognormal and normal models using qualitative (tabular) comparisons. To determine if one distributional model performed better on particular CPT1-2 combinations, we compared the performance of the two models on the same data sets. We used the more graphically oriented normal probability plots to examine those CPT1-2 combinations for which the formal SW tests indicated that both models were inadequate. We also compared the goodness-of-fit of the lognormal and the normal models using the Sign and Friedman tests for TT and ST.

Component Procedure Estimates

It is logical to assume that duration estimates for dual-procedure surgeries might be constructed using estimates derived from their component CPTs. To generate estimates for these component procedures, 46,322 single-CPT surgeries were summarized to provide median time estimates (MTE) for the component CPTs. MTEs were used because the median is an appropriate mea-

sure of central tendency given previous indications of lognormality[4] for single-CPT surgeries. Specific MTEs were matched to component CPTs for dual-procedure surgeries using lookup tables.

We explored the hypothesis that CPT1 (provider designated) in a dual-procedure surgery is typically a longer individual surgery than CPT2. Basic statistics were compiled for MTE1 and MTE2 for component procedures for TT and ST. To estimate how frequently MTEs for CPT1 exceeded those for CPT2, MTEs for the two component procedures were subtracted (MTE1 − MTE2), and the differences were plotted as a frequency histogram.

To investigate specialty origins of dual-procedure surgeries, component procedures (both CPT1 and CPT2) were categorized into one of 20 surgical specialty categories based on primary headers of the CPT classification and cross-tabulated by surgical specialty (CAT1 and CAT2, respectively).

Coding Permutations

Dual-procedure surgeries in our database were *provider designated* by combinations of two procedures (CPT1-2), but permutations (order-dependent combinations) were observed in which the order of the same two component CPT codes was reversed (*i.e.*, CPT1-2 coexisted with similar surgeries CPT2-1).

To detect systematic differences in lnTT and lnST natural logs of TT and ST, respectively, among permutations, we performed individual two-sample *t* tests on each dual CPT. For each dual CPT tested, one subset was coded CPT1-2, and the other was coded CPT2-1. The results of the individual *t* tests (using pooled variances) were tabulated. If all the null hypotheses were true, *i.e.*, no differences in surgical times existed among permutations, then the *P* values should behave together like a sample from a uniform (0,1) distribution. We used uniform probability plots and Kolmogorov-Smirnov tests to explore how well the *P* values were described by a uniform (0,1) distribution.

Aggregate Models

To further explore differences in surgery times among permutations, we reordered provider-assigned codes (CPT1-2) arbitrarily to eliminate permutations. The new order of CPTs (CPTA-B) was determined by ordering the CPT with the greater numerical value of the code as CPTA and that with the lesser numerical value as CPTB. Because the numeric values for CPT codes are assigned on anatomic and pathologic grounds, we considered the values of the codes arbitrary with respect to the duration of the surgical procedures. Permutations were identified by paired comparisons of the provider-designated and numeric value–ordered codes. To investigate systematic

differences among permutations with respect to their durations (lnTT and lnST), we fit an aggregate linear model of the form:

$$\text{ln Time} = \text{CPTA-B} + \text{Anes} + \text{Perm} + \text{Perm}*\text{Anes}$$
$$+ \text{Perm}*\text{CPTA-B} + \text{error} \quad (1)$$

where Perm = 0 if CPTA = CPT1 and Perm = 1 if CPTA > CPT1 numerically. Anes was a categorical variable for type of anesthesia; lnTime was lnTT or lnST; and an asterisk in an expression denoted an interaction term. Anesthesia was included as a factor in this model because it was found previously to be associated with the variability in surgical duration.[4,12] If any of the terms Perm, Perm * Anes, or Perm * CPTA-B were significant, then coding permutations have statistical impact on the true mean lnTime, by itself or in conjunction with the type of anesthesia or a particular combination of CPTs.

Primary Component Procedure

To achieve our goal of understanding the various factors that explain variability in lnTT and lnST for provider-ordered (CPT1-2) combinations, we used MTEs from component CPTs to fit an aggregate linear model with no interaction terms of the form:

$$\text{ln Time} = \text{MTE1} + \text{MTE2} + \text{Anes} + \text{CAT1} + \text{CAT2}$$
$$+ \text{Emerg} + \text{Age} + \text{error} \quad (2)$$

where MTE1 = median time estimate for CPT1, MTE2 = median time estimate for CPT2, Anes = type of anesthesia, CAT1 = surgical specialty category of CPT1, CAT2 = surgical specialty of CPT2, Emerg = emergency status (yes or no), and Age was expressed in years. CAT1, CAT2, Anes, and Emerg were categorical variables. Due to the exploratory nature of our analyses and the relatively large number of independent variables, it was not feasible to examine interaction effects.

Model 2 allowed for different permutations of CPT12, *e.g.*, for MTE1-2 *versus* MTE2-1. We did this to compare a model with provider-designated codes with another model that ordered CPTs on another criterion, such as the duration of MTEs (*i.e.*, model 3).

Longest Component Procedure

Longer procedures (CPTL) are more variable than shorter procedures (CPTS) and have a proportionately greater effect on scheduling.[12] To study the effect of modeling based on MTEs, we looked up MTEs for CPT 1 and CPT2 and designated the component CPT with the longest MTE as CPTL. This effectively identified the longest component procedure and simultaneously eliminated coding permutations. To test the ability of this duration dependent model to detect variability in lnTT

Table 1. Comparisons of Sample Size and Shapiro–Wilk Goodness-of-fit *P* Values for Log-normal Distribution

Category	$P < 0.01$	$0.01 \leq P \leq 0.1$	$P \geq 0.1$	Row Totals
Surgical time, lnST				
Not fit, $n \leq 4$	NF	NF	NF	6052 (100)
Small, $4 < n \leq 30$	13 (5.33)	19 (7.79)	212 (86.9)	244 (100)
Medium, $n > 30$	3 (18.7)	3 (18.7)	10 (62.5)	16 (100)
Column totals	16 (6.15)	22 (8.46)	222 (85.4)	260 (100)
Total time, lnTT				
Not fit, $n \leq 4$	NF	NF	NF	6052 (100)
Small, $4 < n \leq 30$	8 (3.28)	31 (12.7)	205 (84.0)	244 (100)
Medium, $n > 30$	2 (12.5)	6 (37.5)	8 (50.0)	16 (100)
Column totals	10 (3.85)	37 (14.2)	213 (81.9)	260 (100)

Tabular comparisons of sample size and Shapiro–Wilk goodness-of-fit *P* values for the log-normal distribution for TT and ST. 6,052 CPT1–2 combinations were not fit due to insufficient sample size ($n \leq 4$). Values shown are number of surgeries, and the percent of the total row appears in parentheses; n = 3,266 surgeries, 260 fit CPT1–2 combinations.

CPT = Common Procedural Terminology code; ln = the natural logarithm; NF = not fitted; ST = surgical time; TT = total time.

and lnST, we fit an additional seven-factor main effects linear model of the general form:

$$\ln \text{Time} = \text{MTEL} + \text{MTES} + \text{Anes} + \text{CATL} + \text{CATS} + \text{Emerg} + \text{Age} + \text{error} \quad (3)$$

where MTEL = median time estimate for CPTL, MTES = median time estimate for CPTS, Anes = type of anesthesia, CATL = surgical specialty category of the longest procedure, CATS = specialty category of the shortest procedure, Emerg = emergency status, and Age was expressed in years. CATL, CATS, Anes, and Emerg were categorical variables. Factors were added stepwise to the model. It was not feasible to examine interaction effects due to the exploratory nature of our analyses and the relatively large number of independent variables.

Simplified Models

We examined r^2 for all the submodels to arrive at a parsimonious model (a reasonably predictive model with as few meaningful terms as required) for predicting lnTT and lnST for models 2 and 3. In particular, we studied models that retained factors as ordered by the original MSEs in the full seven-factor model. In doing so, we computed r^2 for all the factor submodels. For reasons of

brevity, we reported only those submodels with one, two, three, or seven main effect terms.

Results

Detailed Description of the Data

The database contained 10,740 cases (dual-procedure surgeries), each comprised of exactly two component procedures. Three cases were eliminated from analysis, two because they contained rare pain procedures, and one because it was the only case comprised of a pathology procedure as CPT1. The remaining 10,737 surgeries were performed by 205 different surgeons and 136 different anesthesiologists. Of 10,737 surgeries, 5,269 cases (49%) were female and 943 cases (9%) were emergencies. General anesthesia was used in 7,653 cases (71%); 2,221 cases (21%) involved regional anesthesia; 461 cases (4%) had monitored sedation; and 402 cases (4%) involved only local anesthesia. The average age of patients was 48.9 ± 18.1 yr (mean $\pm$ SD).

Model Probability Distributions

Tables 1 and 2 display the results of fitting the lognormal and normal distributions, respectively, to 260 CPT1-2

Table 2. Comparisons of Sample Size and Shapiro–Wilk Goodness-of-fit *P* Values for Normal Distribution

Category	$P < 0.01$	$0.01 \leq P \leq 0.1$	$P \geq 0.1$	Totals
Surgical time				
Not fit, $n \leq 4$	NF	NF	NF	6052 (100)
Small, $4 < n \leq 30$	15 (6.15)	24 (9.84)	205 (84.0)	244 (100)
Medium, $n > 30$	5 (31.2)	4 (25.0)	7 (43.7)	16 (100)
Column totals	20 (7.69)	28 (10.8)	212 (81.5)	260 (100)
Total time				
Not fit, $n \leq 4$	NF	NF	NF	6052 (100)
Small, $4 < n \leq 30$	11 (4.51)	28 (11.5)	205 (84.0)	244 (100)
Medium, $n > 30$	7 (43.7)	2 (12.5)	7 (43.7)	16 (100)
Column totals	18 (6.92)	30 (11.5)	212 (81.5)	260 (100)

Tabular comparisons of sample size and Shapiro–Wilk goodness-of-fit *P* values for the normal distribution for TT and ST. 6,052 CPT1–2 combinations were not fit due to insufficient sample size ($n \leq 4$). Values shown are number of surgeries, and the percent of the total row appears in parentheses; n = 3,266 surgeries, 260 fit CPT1–2 combinations.

CPT = Common Procedural Terminology code; NF = not fitted.

Table 3. Summary Statistics for Dual Procedure Surgeries

	TT, min			ST, min		
	CPT1–2	MTE1	MTE2	CPT1–2	MTE1	MTE2
n	10,737	10,335	10,243	10,737	10,335	10,243
Minimum	15	20	30	2	5.0	2.0
Maximum	1,160	905	905	1,080	835	835
Mean	259	206	170	191	139	109
SD	162	123	111	145	104	91
SE	1.56	1.2	1.1	1.40	1.0	0.9
Median	220	178	137	155	113	79
CV, %	62	60	65	76	75	83
Summation	2,791,360	2,132,424	1,741,162	2,053,770	1,437,670	1,118,195

Summary statistics for 10,737 dual (CPT1–2) procedure surgeries for TT and ST and MTEs for their component procedures. The MTEs were not available for all surgeries.

CPT = Common Procedural Terminology code; CV = coefficient of variation; MTE = median time estimate; ST = surgical time; TT = total time.

combinations (3,266 surgeries) for TT and ST. Small samples (n ≤ 30) were fit better than moderate-sized samples for TT and ST. The decision to fit only surgeries with sample sizes n ≥ 5 reduced substantially the number of dual-procedure surgeries fit. In doing so, 6,052 infrequent CPT1-2 combinations were omitted from analysis.

A paired comparison of the lognormal and normal models was made using Friedman and Sign tests applied to the SW goodness-of-fit *P* value results for TT and ST in tables 1 and 2. Tests on 260 CPT1-2 combinations (3,266 surgeries) revealed that TT fit the lognormal and normal models no better (and no worse) than ST. The lognormal models fit TT and ST better than the corresponding normal model (Friedman tests, $P \geq 0.05$). The SW tests rejected the lognormal model for only 4–6% of dual CPTs tested.

Component Procedure Estimates

A cumulative frequency histogram indicated that CPT1 was not always the longest component procedure. MTE2 equaled or exceeded MTE1 for 3,538 surgeries or 35.8% of total CPT1-2 surgeries.

Table 3 summarizes basic statistics for dual-procedure surgeries for TT and ST and for the median time estimates (MTE1 and MTE2) for their corresponding component procedure times. MTEs were available for only 10,243 (95.3%) of CPT1s and 10,335 (96.2%) of CPT2s. There were 9,876 dual-procedure surgeries (92%) with MTEs available for *both* component codes (*i.e.*, MTE1 and MTE2 available simultaneously).

MTE1 and MTE2 estimates for component procedures CPT1 and CPT2 were examined using normal probability plots for ST and TT, and these values were lognormally distributed.

Table 4 cross-tabulates the component CPTs for dual-procedure surgeries by surgical specialty category. One thousand nine hundred sixty-nine different component CPT1s with case frequencies 344 to 1 (5 ± 15; mean ± SD) and 1,924 CPT2s with case frequencies 301 to 1 (6 ± 16; mean ± SD) were categorized. The diagonal of table 4 reveals that 7,489 cases (70%) of dual-procedure surgeries were comprised of component procedures both from the same surgical specialty.

Coding Permutations

Dual-procedure surgeries were examined for the presence of coding permutations. There were 5,978 different CPT1-2 combinations comprised of 10,737 surgical cases. Four thousand seven hundred twenty-seven cases (and an equal number of CPT1-2 combinations) were eliminated because they were singletons with sample size n = 1. There remained 1,249 different CPT1-2 combinations comprised of 6,010 surgeries, each with two or more cases (range, 2–196 cases per combination). Of 1,249 dual-procedure surgeries, each with two or more cases, 913 combinations had no permutations, and only 336 CPT1-2 combinations had two permutations.

There existed only 60 CPT1-2 combinations (1,862 surgeries) with two permutations and no fewer than 10 surgeries in each subset (10 patients in each subset was found by us as being minimal for statistical testing). Individual *t* tests grouped by permutation (pooled variance estimates) could be completed for only 52 of 60 dual procedures for lnTT and lnST because of insufficient numbers for some permutations. Only 5 of 52 dual procedures (9.6%) and 2 of 52 dual procedures (3.8%) differed significantly for the two permutations, with respect to lnST and lnTT, respectively. To put these results in perspective, if the null hypotheses were true and there were no differences among permutations, then 5% of *t* tests were expected to be positive by chance alone. Probability plots and Kolmogorov-Smirnov tests indicated that the uniform distribution fits the observed *P* values for both LnTT and LnST (Kolmogorov-Smirnov *P* values > 0.15 for both).

Aggregate Models

Table 5 displays analysis of variance (ANOVA) results (model 1) for the same 60 CPT1-2 combinations (1,862 surgeries), each with two permutations (and no fewer

Table 4. Frequency and Distribution of Component Common Procedural Terminology Codes for Dual Procedure Surgeries Cross-tabulated by Surgical Specialty Categories

Category Descriptor	CAT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Integument	1	576	76	2	14	15	0	6	3	1	0	0	8	0	1	5	3	2	1	2	1
Musculoskeletal	2	116	1670	22	12	10	0	62	4	1	0	0	2	0	0	74	6	3	13	3	0
Respiratory	3	55	54	605	22	9	7	136	4	0	0	0	1	0	5	4	4	7	1	0	0
Cardiovascular	4	22	35	16	1532	1	3	34	14	3	0	0	1	1	1	1	1	0	1	2	31
Hematology	5	91	23	5	30	26	0	29	5	1	0	2	8	0	0	2	0	0	2	5	1
Mediastinum	6	0	1	48	0	1	2	3	0	1	0	0	0	O	1	0	0	0	0	0	0
Digestive	7	86	40	85	36	35	0	808	263	9	0	30	181	2	1	8	0	3	14	3	2
Urinary	8	9	4	2	25	7	0	188	844	84	0	0	48	0	0	1	0	0	5	1	0
Male genital	9	2	1	0	1	3	0	13	73	72	0	0	0	0	0	0	0	0	0	10	0
Intersex	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Laparoscopic	11	6	1	1	2	1	0	21	10	17	0	48	40	3	2	0	0	0	0	0	0
Female genital	12	31	1	1	7	24	0	97	92	1	0	17	401	6	0	1	0	0	1	1	0
Maternity care	13	0	0	0	0	2	0	7	0	1	0	1	31	5	0	0	0	0	1	0	0
Endocrine	14	6	1	9	0	2	0	4	3	0	0	1	2	0	7	0	1	0	0	0	0
Nervous system	15	19	88	14	3	1	0	8	1	0	0	0	2	0	0	290	2	8	10	2	3
Eye and adnexa	16	38	4	7	1	0	0	0	0	0	0	0	1	0	1	2	545	0	1	0	45
Auditory	17	22	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	56	0	0	0
Radiology	18	4	4	0	1	1	0	0	16	0	0	0	6	0	0	0	0	0	0	0	0
Pathology	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Medicine	20	1	0	0	9	0	0	2	0	0	0	0	0	0	0	0	1	0	1	0	2

Frequency and distribution of component CPTs for dual procedure surgeries are cross-tabulated by surgical specialty categories (CPT1 in rows and CPT2 in columns). The diagonal of the table demonstrates that component CPTs were derived from the same surgical specialty in 70% (7,489) of surgeries; n = 10,737 dual procedure surgeries, 205 surgeons, 136 anesthesiologists.

CAT = surgical specialty categories; CPT = Common Procedural Terminology code.

than 10 surgeries each), as above. Permutations of dual CPTs did not differ ($P < 0.05$) with respect to LnTT and LnST. CPTA-B and type of anesthesia were important determinants ($P < 0.05$) of time estimates for LnTT. The first-order interaction, CPTA-B * Anes, was not tested because too many CPTA-B combinations were associated with only a single type of anesthesia (general). All other first-order interactions were not significant. Results for lnST were similar to those for lnTT.

Primary Component Procedure

Table 6 displays details for an ANOVA of model 2 performed for dependent variable lnTT. Note CPT1-2 is the ordering given by the provider. All seven independent factors were retained as significant ($P < 0.05$), and together they explained 68.7% of the variability in lnTT. The independent factors in decreasing order of importance by F ratios were MTE1, MTE2, Anes, Emerg, Age, CAT1, and CAT2. The order for the independent factors was the same for a similar analysis of lnST, which is not reported in detail herein. Type III sums of squares were used in the ANOVA.

Longest Component Procedure

Table 7 displays details for an ANOVA of model 3 performed for dependent variable lnTT (n = 9,833 CPTL-S combinations). Note that CPTL-S are dual CPT codes ordered by decreasing MTE. All seven independent factors were retained as significant ($P < 0.05$), and together they explained 70.5% of the variability in lnTT. The independent factors in decreasing order of importance by F ratios were MTEL, Anes, MTES, Emerg, Age, CATL, and CATS. Type III sums of squares were used in the ANOVA, and a similar ordering of main effects was obtained for an ANOVA of lnST.

Table 5. Analysis of Variance Results for Model 1 on Dependent Variable lnTT

Independent Factor	SSE	df	MSE	F Ratio	P
CPT A-B	443.671	59	7.520	112.242	0.0000
Anes	17.657	3	5.882	87.798	0.0000
Permutations	0.036	1	0.036	0.537	0.4637
Permutations * Anes	0.073	3	0.024	0.362	0.7806
Permutations * CPTA-B	3.670	59	0.062	0.929	0.6304
Error	116.307	1736	0.067	—	—

Results of a three-factor main-effects analysis of variance on dependent variable lnTT (r^2 = 89.3%; df = 1,736; n = 1,862 surgeries). The interaction term CPT A-B * Anes could not be tested because a large number of CPT A-B surgeries were associated with only a single type of anesthesia (usually general).

Anes = type of anesthesia; CPT = Common Procedural Terminology code; lnTT = natural logarithm of total time; MSE = mean square error; SSE = sum squared errors.

Table 6. Main-effects Analysis of Variance on Model 2 for Dependent Variable lnTT

Independent Factor	SSE	df	MSE	F Ratio	P
MTE1	780.899	1	780.899	6064.183	0.0000
MTE2	110.599	1	110.599	858.876	0.0000
Anes	61.423	3	20.475	158.998	0.0000
Emergency status	2.341	1	2.341	18.177	0.0000
Age	1.475	1	1.475	11.457	0.0007
CAT1	15.357	17	0.903	7.015	0.0000
CAT2	11.543	18	0.641	4.980	0.0000
Error	1266.219	9833	0.129	—	—

Results of a seven-factor main-effects analysis of variance on dependent variable lnTT for CPT1–2 (r^2 = 68.7%; df = 9,833; n = 9,876 dual procedure surgeries). Anes= type of anesthesia; CAT = surgical subspecialty category for CPT1 or 2 as defined in Table 2; CPT = Common Procedural Terminology code; MSE = mean squared error; MTE = median time estimate for CPT 1 or 2; SSE = squared sum of errors.

In comparing the results of tables 7 and 8, we noted that the overall explanatory power of model 3 as measured by r^2 is slightly better than that of model 2. Furthermore, the relative importance of factors anesthesia and MTE2 were reversed between table 6 and table 7.

Simplified Models

Model 3 had greater explanatory power for lnTT and lnST regardless of the number of model factors included (table 8). The explanatory power of model 2 degraded noticeably when going from a three-factor to a two-factor model. The explanatory power of model 3 decreased noticeably only when going from a two-factor to a one-factor model. In fact, a one-factor model for lnTT using model 3 is superior to a three-factor model based on model 2. These observations further support the somewhat better explanatory power using ordered MTEs rather than provider-designated codes to estimate surgical times for dual-procedure surgeries.

Discussion

Choosing a highly appropriate probability model is an important first step in forecasting surgical procedure times with appropriately estimated probabilities. Our results indicated that the lognormal model was significantly better than the normal model for surgeries composed of exactly two procedures. These findings complement and affirm research conclusions based on previous analyses of single-procedure surgeries.

Cost structures have been used to determine the percentile point of time models used to allocate surgical specialty block times.[13] In an analogous procedure, minimal cost analyses may be used to allocate time to dual-procedure surgeries. Different point estimates may be chosen for varying cost structures, and fitting a statistical model to surgical procedure times is a good way to obtain these estimates.

Coding Permutations

We were unaware how permutations of component CPTs were determined for the dual-procedure surgeries in our data set. There were no known rules to determine which permutations of multiple procedure codes were "correct." We could not be certain if procedures in our database were ordered first done–first recorded, greatest medical impact first, surgeon's own specialty first, highest fee first, or simply arbitrarily. We were also unable to determine (from a nonscheduling perspective) if permutations represented different, similar, or identical surgeries. Coding permutations did not appear to matter from the perspective of scheduling surgical procedure times. However, it is possible that some functionally distinct permutations exist that (for legitimate reasons) should be modeled separately. This important question awaits further research.

Table 7. Main-effects Analysis of Variance on Model 3 for Dependent Variable lnTT

Independent Factor	SSE	df	MSE	F Ratio	P
MTEL	757.556	1	757.556	6249.135	0.0000
Anes	62.071	3	20.690	170.677	0.0000
MTES	10.2264	1	10.2264	84.359	0.0000
Emergency status	2.501	1	2.501	20.672	0.0000
Age	1.1534	1	1.154	9.518	0.0020
CATL	11.003	17	0.647	5.339	0.0000
CATS	7.548	18	0.419	3.459	0.0000
Error	1192.012	9833	0.121	—	—

Results of a seven-factor main-effects analysis of variance on dependent variable lnTT for CPTL-S (r^2 = 70.5%; df = 9,833; n = 9,876 dual procedure surgeries). Anes = type of anesthesia; CAT = surgical subspecialty category for CPTL or CPTS; CPT = Common Procedural Terminology code; MSE = mean squared error; MTE = median time estimate for CPTL or CPTS; SSE = squared sum of errors.

Table 8. Serial Main-effect Analysis of Variance with Independent Factors Subtracted from the Model

Main Effects	MTE1–2	MTE1–2	MTEL-S	MTEL-S
Dependent variable	lnST	lnTT	LnST	lnTT
7-Factor r^2	68.2%	68.7%	70.3%	70.5%
3-Factor r^2	67.1%	67.6%	69.4%	69.7%
2-Factor r^2	65.1%	66.3%	69.1%	69.5%
1-Factor r^2	61.4%	63.4%	66.9%	68.2%

Results of serial main effect analysis of variance with independent factors subtracted from the model one after another in order of increasing factor F ratio as determined from the initial seven-factor analysis of variance (tables 6, 7); n = 9,876 cases, df = 9,833.

lnST = natural logarithm of surgery time; lnTT = natural logarithm of total time; MTE = median time estimate; r^2 = % variation in the dependent variable explained by the independent variables.

Individual dual-procedure surgeries are difficult to model because they are performed infrequently. From this perspective, it is important to reduce coding permutations because (if modeled separately) these also reduce the number of cases modeled in each group. Reducing permutations is particularly important with multiple procedures, *e.g.*, six different permutations are possible from three component CPTs.

Based on our research, we propose that when modeling individual dual-procedure combinations, the surgeries be designated in order of decreasing MTEs for the component procedures. This approach would identify the longest of the two dual CPT surgeries and would reduce permutations, increase sample sizes, reduce the number of factors required for modeling, and encourage segmentation by type of anesthesia (and where practical by surgeon). This policy could be applied to surgeries with two or three CPT codes designated.

We compared two aggregate models, a provider-ordered model (CPT1-2) against another model that applied a uniform ordering based on the relative MTEs (CPTL-S). The r^2 measured for the model with ordered MTEs was numerically greater than for the provider-designated CPTs. Further research might indicate whether there are meaningful subsets of the data for which the model using the provider designation is superior or whether a mathematically more sophisticated model might result in a different conclusion than the one we found.

We recommend use of the lognormal model for estimating surgical times for surgeries with two component procedures. Our results help legitimize the use of log transforms to normalize surgical procedure times prior to hypothesis testing[14] using linear statistical models.

Limitations

Goodness-of-fit tests may inappropriately reject preferred models under a variety of circumstances if they are the only tools used for model selection. Causes of poor fits to a correct model include rounding of shorter procedure times, large sample sizes, untrimmed statistical outliers, and failure to properly segment sample mixtures. These and other causes of model rejection have been discussed in a previous publication.[4] We elected not to trim outliers from our data because we had no information to support doing so.

Uncertainty would arise with a modeling policy ordering the longest case first if the component CPTs possessed nearly equal time estimates. In that event, arrival of a new surgeon or additional experience with component CPTs could change the order of the component CPTs by altering the procedure-designated CPTL. Uncertainty about the longest procedure (thus the order of multiple procedures) is a limitation if this method of naming dual-procedure surgeries were to be adopted.

We showed previously[12] that surgeons differ in variability in surgical times involving a single CPT. We did not include the surgeon as a factor in building our models for dual CPTs. Our data set, although it was a complete census of all surgeries performed at a major hospital over a 7-yr period, was not large enough to permit a model to be estimated using the surgeon as a factor. Other factors known in practice to affect variability in surgical times were also omitted from our models.

In this manuscript, surgeon work rate effect was not investigated for dual-procedure surgeries. Research on single-CPT surgeries suggested that surgeon work rate effect[12] is an important factor whenever the lognormal is superior to the normal distribution for modeling surgeries. We did not examine surgeon effect in this study because too few samples of dual-procedure surgeries each contained enough surgeons with case numbers sufficient to support the analyses. Based on our previous research, however, we believe that surgeon work rate is second in importance after procedure code (and ahead of type of anesthesia) in explaining variability in dual-procedure surgeries.

Conclusions

We studied individual and aggregate models of dual-procedure surgeries. Individual models fit a mathematical distribution to surgeries segmented by dual CPT, type of anesthesia, or other factors in a subset of the data. Aggregate models used all the records from all dual CPTs to derive time estimates. Aggregate models make implicit assumptions that the variability in each of the data

subsets is similar. Dual CPT surgeries were better modeled by the lognormal distribution than by the normal distribution. Permutations of individual dual CPTs did not appear to represent statistically distinct procedures with respect to TT and ST. Our results suggested it might be practical to improve time estimates for infrequent dual CPT surgeries by simply considering the duration of the longest procedure and type of anesthesia.

References

1. 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

2. Dexter F, Macario A, O'Neill L: Scheduling surgical cases into overflow block time: Computer simulation of the effects of scheduling strategies on operating room labor costs. Anesth Analg 2000; 90:980–8

3. May JH, Strum DP, Vargas LG: Fitting the lognormal distribution to surgical procedure times. Decis Sci 2000; 31:129–48

4. 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

5. Hancock WM, Walter PF, More RA, Glick ND: Operating room scheduling data base analysis for scheduling. J Med Syst 1988; 12:397–409

6. Bashein G, Barna C: A comprehensive computer system for anesthetic record retrieval. Anesth Analg 1985; 64:425–31

7. Kirschner CG, Davis SJ, Jackson JA, Jacobson CA, O'Heron MR, O'Hara KE, Reyes D, Rozell D, Young RL: Physicians' Current Procedural Terminology 1997. Chicago, American Medical Association, 1997

8. Johnson NL, Kotz S, Balakrishnan N: Continuous Univariate Distributions, 2nd edition. New York, Wiley, 1994

9. Muralidhar K, Zanakis SH: A simple minimum-bias percentile estimator of the location parameter for the gamma, weibull, and log-normal distributions. Decis Sci 1992; 23:862–79

10. Shapiro SS, Wilk MB, Chen HJ: A comparative study of various tests for normality. J Am Statist Assoc 1968; 63:1343–72

11. D'Agostino RB: Tests for the normal distribution, Goodness-of-fit Techniques. Edited by D'Agostino RB, Stephens MA. New York, Marcel Dekker, 1986, pp 367–419

12. Strum DP, Sampson AR, May JH, Vargas LG: Surgeon and type of anesthesia predict variability in surgical procedure times. ANESTHESIOLOGY 2000; 92:1454–66

13. Strum DP, Vargas LG, May JH: Surgical subspecialty block utilization and capacity planning: A minimal cost analysis model. ANESTHESIOLOGY 1999; 90: 1176–85

14. Zhou XH, Gao S, Hui SL: Methods for comparing the means of two independent log-normal samples [published erratum appears in Biometrics 1997; 53:1566]. Biometrics 1997; 53:1129–35