

Anesthesiology
2000; 92:1454-66
© 2000 American Society of Anesthesiologists, Inc.
Lippincott Williams & Wilkins, Inc.

Surgeon and Type of Anesthesia Predict Variability in Surgical Procedure Times

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


Background: Variability in surgical procedure times increases the cost of healthcare delivery by increasing both the underutilization and overutilization of expensive surgical resources. To reduce variability in surgical procedure times, we must identify and study its sources.

Methods: Our data set consisted of all surgeries performed over a 7-yr period at a large teaching hospital, resulting in 46,322 surgical cases. To study factors associated with variability in surgical procedure times, data mining techniques were used to segment and focus the data so that the analyses would be both technically and intellectually feasible. The data were subdivided into 40 representative segments of manageable size and variability based on headers adopted from the common procedural terminology classification. Each data segment was then analyzed using a main-effects linear model to identify and

quantify specific sources of variability in surgical procedure times.

Results: The single most important source of variability in surgical procedure times was surgeon effect. Type of anesthesia, age, gender, and American Society of Anesthesiologists risk class were additional sources of variability. Intrinsic case-specific variability, unexplained by any of the preceding factors, was found to be highest for shorter surgeries relative to longer procedures. Variability in procedure times among surgeons was a multiplicative function (proportionate to time) of surgical time and total procedure time, such that as procedure time increased, variability in surgeons' surgical time increased proportionately.

Conclusions: Surgeon-specific variability should be considered when building scheduling heuristics for longer surgeries. Results concerning variability in surgical procedure times due to factors such as type of anesthesia, age, gender, and American Society of Anesthesiologists risk class may be extrapolated to scheduling in other institutions, although specifics on individual surgeons may not. This research identifies factors associated with variability in surgical procedure times, knowledge of which may ultimately be used to improve surgical scheduling and operating room utilization. (Key words: Data mining; factors affecting surgery times; minimizing surgical costs; scheduling surgery; work rates; variability in surgical times.)

 Additional material related to this article can be found on the ANESTHESIOLOGY Web site. Go to the following address, click on Enhancements Index, and then scroll down to find the appropriate article and link. <http://www.anesthesiology.org>

* Associate Professor of Anesthesiology, Department of Anesthesiology, Queens University, Kingston General Hospital.

† Professor and Chairman, Department of Statistics, University of Pittsburgh.

‡ Professor of Business Administration and Intelligent Systems, The Joseph M. Katz Graduate School of Business, University of Pittsburgh.

§ Professor of Operations, Decision Sciences, and Artificial Intelligence, The Joseph M. Katz Graduate School of Business, University of Pittsburgh.

Received from the Department of Anesthesiology, Queen's University, Kingston, Ontario, Canada, and the Department of Statistics and The Joseph M. Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania. Submitted for publication October 12, 1998. Accepted for publication December 19, 1999. Supported by grant No. DMS-9704400 from the National Science Foundation (Dr. Sampson), Washington, DC, and in part by a grant from the Institute for Industrial Competitiveness of the Joseph M. Katz Graduate School of Business, Pittsburgh, Pennsylvania, and the Department of Anesthesiology, Queen's University, Kingston, Ontario, Canada.

Address reprint requests to Dr. Strum: Department of Anesthesiology, Queens University, KGH, 76 Stuart Street, Kingston, Ontario K7L 2V7, Canada. Address electronic mail to: strumd@post.queensu.ca

IN an era of cost-constrained health care, it is of utmost economic importance for medical institutions to effectively schedule and efficiently use expensive surgical resources. Unexplained variability in procedure times complicates surgical scheduling and reduces operational efficiency. If variability in procedure times could be controlled or better predicted, the cost of surgeries could be reduced through improved scheduling of surgical resources. To reduce the impact of variability in surgical procedure times, we need to model and thus predict surgical procedure times more accurately. The purpose of this article is to lay a foundation for developing statistical models that identify and estimate the effects of major sources of variability in surgical procedure times.

Modeling of variability in surgical procedure times has been of interest for at least 35 yr.¹ Various researchers have suggested statistical models for surgical procedure

times, including normal models,² lognormal models,^{3,4} and three-parameter lognormal models.⁵ However, it is known⁶ that when populations are heterogeneous mixtures, improvement occurs by identifying homogeneous subpopulations and modeling within these. A standard statistical procedure useful in this regard is linear modeling, which permits the identification of factors that cause nonhomogeneity in the population. By estimating the effects of the significant factors, improved models can be developed with better fits to the data, thereby permitting improved prediction of surgical procedure times.

We studied surgical variability and differences among surgeons at a large teaching hospital and also searched for demographic factors that might explain additional variability in the surgical procedure times. Our ultimate concern was describing variability and identifying the associated factors that might be predicted, controlled, or altered to improve scheduling and reduce the cost of surgical services.

Methods

We evaluated computerized records⁷ of all surgical cases performed at a large teaching hospital 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. Independent variables included in the record were surgical time (ST; defined as the time from incision to closure of the surgical wound), total procedure time (TT; defined as the time from entry into the operating suite until emergence from anesthesia), patient age and gender, surgeon, anesthesiologist, American Society of Anesthesiologists (ASA) risk class (a scale for grading systemic illness defined by the ASA), type of anesthesia, ICD-9 codes, and Current Procedural Terminology (CPT) codes.⁸ We classified the type of anesthesia administered into four categories: general, local, monitored, and regional.

Of 60,643 total records, 779 were omitted from analysis because of incomplete data on the ST or TT variables, leaving 59,864. Surgeries included between one and three CPTs. There were 46,322 patients with only one CPT code, 10,740 with exactly two different CPT codes, and 2,802 patients with three CPT codes. To reduce potential confounding factors, we chose to analyze those procedures with only a single CPT code.

Data Structures

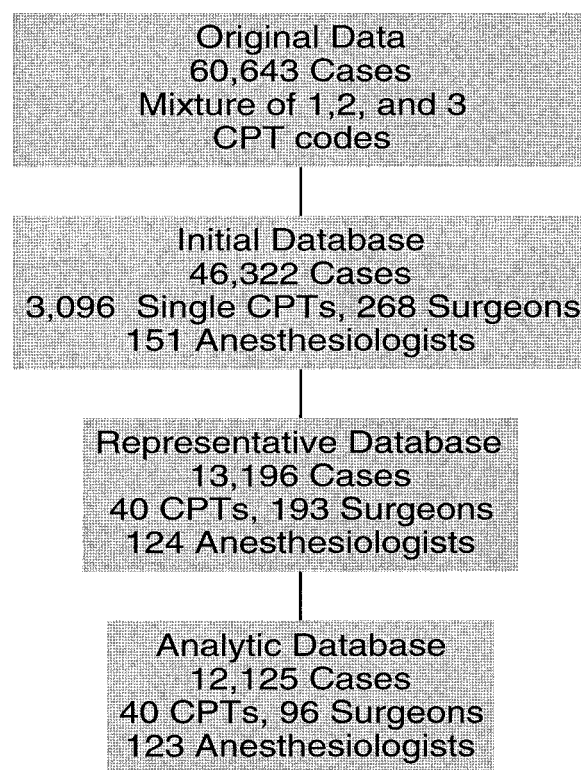


Fig. 1. Schematic illustration of the hierarchy of databases produced by data mining.

Data Structures

We reduced the data that we used so that the analysis would be both technically and intellectually feasible. In doing so, we attempted to maintain the breadth of surgical experience, all of the various surgical subsets and surgical subspecialties, and the physiologic diversity of our data, while at the same time minimizing the potential for biasing our results. As in any data mining (knowledge extraction from databases) project, one needs to decide which data to extract in light of the analytic approaches available and to document the process by which the extraction results.⁹ We documented our extraction process by describing the successively smaller interconnected databases that we used and summarized each in figure 1. The resultant data set was large and diverse.

We referred to the initial 46,322 cases, each described by only a single CPT, as the *initial database*. This database consisted of 3,096 different surgical procedures performed by 268 different surgeons and 151 different anesthesiologists in many combinations. The initial data-

Table 1. Initial Database: CPT Codes and Brief Definitions for the Three Most Numerous Procedures in Each Category, and the Range of Codes in Each Category

Cat	Descriptor	N	P	First Procedure Code	Second Procedure Code	Third Procedure Code	Range of CPT Codes
1	integumentary	3702	254	19120	19101	17108	10000 19999
2	musculoskeletal	9348	912	29881	27130	20680	20000 29999
3	respiratory	2195	219	31255	30520	31600	30000 32999
4	cardiovascular	8290	296	36489	33512	33513	33000 37999
5	hematology	647	27	38230	38500	38100	38000 38999
6	mediastinum	80	12	39400	39010	39000	39000 39999
7	digestive	4135	103	49000	47135	49505	40000 49999
8	urinary	4638	169	52000	52204	50360	50000 53999
9	male genital	1131	82	55400	55845	54520	54000 55899
10	intersex	0	0	NIL	NIL	NIL	55900 55999
11	laparoscopic	464	27	56301	56341	56300	56000 56399
12	female genital	2748	126	58150	58970	57520	56400 58999
13	maternity care	483	26	59841	59840	59510	59000 59999
14	endocrine	232	20	60500	60220	60540	60000 60999
15	nervous system	3448	243	61510	64721	61538	61000 64999
16	eye and adnexa	3174	182	66984	67107	67108	65000 68899
17	auditory	639	65	69631	69660	69632	69000 69999
18	radiology	292	14	77762	77761	77778	70000 79999
19	pathology	2	1	85102	NIL	NIL	80000 89999
20	medicine	674	15	93640	92018	93650	90000 99999

Cat = category number including descriptor; N = number of cases in each category; P = number of procedures in each category; 46,322 total cases and 3,096 different CPTs in the initial database.

VARIABILITY IN SURGICAL PROCEDURES

Table 2. Initial Database: Descriptive Statistics Summarizing ST for Each of the 20 Categories

Cat	N	P	Sum	%	Mean	SD	Median	Min	Max
1	3,702	254	313,234	5.13	85	81	60	5	805
2	9,348	912	1,105,155	18.1	118	84	102	3	960
3	2,195	219	240,327	3.93	109	91	85	2	650
4	8,290	296	1,510,173	24.7	182	132	175	2	1,055
5	647	27	81,544	1.33	126	94	108	8	595
6	80	12	8,079	0.13	101	84	73	23	555
7	4,135	103	663,491	10.9	160	158	105	4	1,130
8	4,638	169	352,742	5.77	76	114	33	1	980
9	1,131	82	149,810	2.45	132	102	115	2	740
10	0	0	0	0.00	0	0	0	0	0
11	464	27	50,715	0.83	109	69	96	5	490
12	2,748	126	316,468	5.18	115	93	89	5	750
13	483	26	21,130	0.35	44	32	35	5	210
14	232	20	40,350	0.66	174	90	154	40	690
15	3,448	243	781,436	12.8	227	165	195	1	1,030
16	3,174	182	326,381	5.34	103	63	90	2	460
17	639	65	77,109	1.26	121	74	107	5	730
18	292	14	17,632	0.29	60	29	55	5	300
19	2	1	96	0.00	48	31	48	26	70
20	674	15	53,832	0.88	80	60	70	1	480

Cat = category number; N = number of cases in the category; P = number of procedures in the category; Sum = total ST (min) for cases in the category; % = percent of ST in each category relative to the total; SD = standard deviation (min); Min = minimum ST (min); Max = maximum ST (min) (46,322 cases and 3,096 different procedures in the initial database).

base was segmented into 20 categories based on the 20 primary headings of the CPT classification. The definitions of these categories, a description of the three most common CPT codes in each category, and the range of codes for each category are listed in table 1.

There were imbalances in the 20 categories of the initial database with respect to the number of CPTs and total ST and TT. These imbalances were imposed by anatomy, physiology, and technology that dictate the number of discrete surgical procedures encompassed by each of the primary CPT categories. To illustrate these imbalances and to characterize the surgical experience at this institution by the type and quantity of surgeries performed, we described ST and TT segmented by each CPT category (tables 2 and 3).

In theory, to achieve our primary goal of understanding the various factors affecting ST and TT, one might like to use a large factorial statistical model using the initial database and taking into account all 3,096 CPT codes and 268 surgeons. This is beyond the limitations of current software and hardware,¹⁰ so that appropriate subsets of the initial database must be examined to use available techniques. In pursuing a data mining approach that would allow us to extract representative information, we chose to maintain diversity of physiology and surgical experience yet focus on CPT codes that con-

tained sufficient data so that standard statistical modeling would be appropriate.

We examined the three most numerous CPT codes occurring within each of the 20 categories, so that, at most, 60 CPT codes would be analyzed. To ensure that each surgeon had sufficient numbers of cases to estimate surgeon effect, we retained from these 60 possibilities only those CPT codes with two or more surgeons, each having 10 or more cases for that CPT code. Thus, if a CPT code had fewer than two surgeons with 10 or more surgeries each, we eliminated that CPT code from further consideration. The resultant database was called the *representative database* and contained 13,196 case records among 40 remaining CPT codes. We believed this database represented a cross-section of the surgical experience of the institution, and we therefore describe it in detail. For the representative database, tables 4 and Web tables 1-3 provide tabulations by CPT code of type and number of procedures, age, type of anesthesia, gender, ASA risk class, surgeon, and anesthesiologist. Table 5 describes ST and TT by CPT code for the representative database.

Because the analytic methods we used (as described in Statistical Analysis) require adequate sample sizes within the surgeon category and no missing values, we used a subset of the representative database in our analyses.

Table 3. Initial Database: Descriptive Statistics Summarizing TT for Each of the 20 Categories

Cat	N	P	Sum	%Total	Mean	SD	Median	Min	Max
1	3,702	254	471,169	5.33	127	93	100	12	850
2	9,348	912	1,689,938	19.1	181	97	165	14	1055
3	2,195	219	375,480	4.25	171	106	145	5	725
4	8,290	296	2,087,762	23.6	252	158	255	5	1185
5	647	27	117,346	1.33	181	110	160	30	675
6	80	12	13,144	0.15	164	95	135	85	685
7	4,135	103	934,943	10.6	226	181	165	20	1320
8	4,638	169	545,869	6.18	118	135	70	10	1065
9	1,131	82	200,066	2.26	177	121	150	18	895
10	0	0	0	0.00	0	0	0	0	0
11	464	27	79,241	0.90	171	79	160	60	575
12	2,748	126	472,573	5.35	172	105	145	25	885
13	483	26	41,558	0.47	86	50	70	30	625
14	232	20	57,168	0.65	246	104	225	75	875
15	3,448	243	1,035,385	11.7	300	177	275	25	1155
16	3,174	182	480,357	5.44	151	70	135	15	555
17	639	65	109,206	1.24	171	85	160	5	875
18	292	14	32,977	0.37	113	37	109	30	355
19	2	1	185	0.00	93	25	92.5	75	115
20	674	15	90,422	1.02	134	71	125	7	585

Cat = category number; N = number of cases in the category; P = number of procedures in the category; Sum = total ST (min) for cases in the category; % = percent of ST in each category relative to the total; SD = standard deviation (min); Min = minimum ST (min); Max = maximum ST (min) (46,322 cases and 3,096 different procedures in the initial database).

This subset, drawn from the representative database, retained only surgeons who performed 10 or more surgeries each; a total of 1,024 cases were thereby deleted.

A further reduction in cases was required due to analysis of variance (ANOVA) modeling because one factor, ASA risk class, was not recorded whenever surgery was performed with local anesthesia (*i.e.*, without an anesthesiologist present). Because all cases of local anesthesia had missing ASA values, eliminating all missing values would have eliminated local anesthesia from our analyses. From initial analyses, we knew that type of anesthesia was a more important determinant of surgical duration than ASA risk class, and we wanted to retain it. To decide when to keep or delete ASA missing values, we adopted an arbitrary "10% rule." When local anesthesia comprised less than 10% of a CPT's anesthesia types, we deleted local cases and retained ASA in our models. When local anesthesia comprised greater than 10% of cases, we retained all cases and omitted ASA as a factor in the corresponding ANOVA models. An additional 47 cases were deleted using the 10% rule, which we adopted to maintain the number and diversity of cases in our analyses. After omission of these 1,071 cases, we referred to the 12,125 cases remaining among 40 CPT codes as the *analytical database*. This database was compared with the representative database to ensure that they were similar.

Statistical Analysis

For the CPT codes in the representative database, we obtained various descriptive statistics and comparisons. Then, within each CPT code and based on the analytical database, we fit a five-factor main-effects linear model of the general form:

$$\text{Ln(ST)} = \text{surgeon} + \text{anesthesia} + \text{ASA} + \text{gender} + \text{age} + \text{error}$$

to each of ST and TT, where surgeon, anesthesia type, ASA risk class, and gender were treated as categorical variables, and age was treated as a continuous variable. In addition, we examined a six-factor ANOVA with anesthesiologist added to the preceding five factors. In certain cases, as when there was no variation in gender or in anesthesia for a CPT code, the nonvarying factor was not retained in the model. Because of the relatively large number of independent variables, it was not feasible to examine interaction effects. Because of the exploratory nature of our data analyses, no formal corrections for multiple comparisons were used. Because of the previous indication of lognormality of ST and TT,^{5,11,12} we conducted the ANOVAs using LnST and LnTT (Ln = the natural logarithm). Other investigators in related

VARIABILITY IN SURGICAL PROCEDURES

Table 4. Representative Database: Definitions of the 40 Procedures and Summary Statistics for Age (in years)

Cat	N	CPT	Procedure Description	Age (mean \pm SD)	Surgeons	Anesthesiologists
1	442	19101	Biopsy of breast	47.1 \pm 15.6	11	30
1	719	19120	Excision breast lesion	48.4 \pm 14.9	13	33
2	322	20680	Removal of support implant	40.0 \pm 15.9	41	71
2	297	27130	Total hip replacement	60.7 \pm 15.1	9	79
2	349	29881	Knee arthroscopy/surgery	40.9 \pm 13.9	9	71
3	138	30520	Repair of nasal septum	35.2 \pm 12.7	12	52
3	141	31255	Nasal endoscopy, ethmoidectomy	40.9 \pm 12.1	6	53
3	104	31600	Incision of windpipe	57.5 \pm 17.7	22	57
4	875	33512	CABG, vein, three	65.1 \pm 10.4	15	41
4	692	33513	CABG, vein, four	65.1 \pm 10.5	13	44
4	1602	36489	Insertion of catheter, vein	48.0 \pm 15.3	38	79
5	57	38100	Removal spleen, total	46.4 \pm 16.3	15	36
5	144	38230	Bone marrow collection	41.5 \pm 10.8	25	54
5	139	38500	Biopsy, removal, lymph node(s)	46.7 \pm 17.1	27	34
7	260	47135	Transplantation of liver	44.5 \pm 11.4	5	20
7	263	49000	Exploration of abdomen	49.9 \pm 16.2	40	75
7	231	49505	Repair inguinal hernia	47.7 \pm 18.8	17	63
8	185	50360	Transplantation of kidney	40.2 \pm 12.5	7	63
8	1521	52000	Cystoscopy, without biopsy	56.0 \pm 17.7	17	68
8	369	52204	Cystoscopy, with biopsy	53.2 \pm 16.7	14	72
9	54	54520	Removal of testis	62.3 \pm 19.7	10	32
9	291	55400	Repair of sperm duct	40.5 \pm 7.03	3	10
9	104	55845	Extensive prostate surgery	62.5 \pm 7.46	10	52
11	82	56301	Laparoscopy, tubal cauterly	30.0 \pm 5.81	8	23
11	70	56341	Laparoscopic cholecystectomy	42.8 \pm 15.2	5	30
12	198	57520	Conization of cervix	33.6 \pm 12.2	12	63
12	343	58150	Total hysterectomy	49.7 \pm 14.5	17	80
12	241	58970	Retrieval of oocyte	33.2 \pm 5.70	11	64
13	81	59840	Abortion and evacuation	27.9 \pm 7.81	14	35
13	204	59841	Abortion and curettage	28.9 \pm 7.28	17	42
15	318	61510	Removal of brain lesion	44.4 \pm 15.6	10	48
15	200	61538	Removal of brain tissue	32.9 \pm 12.4	4	39
15	264	64721	Carpal tunnel surgery	49.5 \pm 16.0	14	61
16	767	66984	Remove cataract, insert lens	66.1 \pm 16.3	9	74
16	281	67107	Scleral buckling, without vitrectomy	48.2 \pm 20.3	12	82
16	150	67108	Scleral buckling, with vitrectomy	39.4 \pm 18.6	8	67
17	99	69631	Repair eardrum structures	38.0 \pm 15.8	6	45
17	34	69632	Rebuild eardrum structures	44.0 \pm 13.1	4	23
17	74	69660	Revise middle ear bones	46.3 \pm 14.5	5	57
20	491	93640	Evaluation heart device	57.9 \pm 14.3	14	43

Cat = category number; N = number of cases; CPT = current procedural terminology code; SD = standard deviation (13,196 total cases and 40 procedure procedures in the representative database).

research have also used log procedure duration as a response variable.¹³

Within each CPT, the significance level used for testing each factor was $P = 0.05$. For certain tabulations, we report the number of CPTs providing statistically significant results for certain factors. As an intuitive statistical yardstick in this regard, if there were no factor effects, we would have expected on the average only 2 of the 40 CPT analyses (5%) to falsely yield significant results (table 6).

To estimate the effect of surgeon within CPT, we used the adjusted means based on LnST for surgeons,¹⁴ some-

times exponentiating this value to obtain an estimate of the median surgeon time in actual minutes. A similar approach was used to obtain adjusted means for type of anesthesia. Various graphical representations of these adjusted values were developed.

Results

Detailed Description of Databases

The initial database was segmented into 20 categories by the primary headers of the CPT classification. These

Table 5. Representative Database: Statistics Summarizing ST and TT

CPT	Median ST	Mean ST	N	Median TT	Mean TT
19101	55	58.4 ± 25.3	442	85	88.9 ± 30.4
19120	50	54.4 ± 25.3	719	75	79.3 ± 29.5
20680	38.5	47.2 ± 40.3	322	80	91.5 ± 49.5
27130	200	207 ± 57.1	297	287	295 ± 60.8
29881	65	71.7 ± 28.9	349	120	125 ± 34.4
30520	85	92.1 ± 43.1	138	135	145 ± 50.1
31255	145	150 ± 64.2	141	200	204 ± 66.4
31600	46	52.7 ± 26.9	104	105	105 ± 34.4
33512	269	285 ± 85.7	875	360	379 ± 89.8
33513	295	302 ± 72.7	692	390	396 ± 77.0
36489	35	40.9 ± 26.7	1602	65	69.3 ± 35.7
38100	170	174 ± 60.8	57	240	246 ± 74.8
38230	120	130 ± 44.3	144	180	188 ± 48.0
38500	50	52.5 ± 24.8	139	80	86.6 ± 34.5
47135	540	560 ± 119	260	675	691 ± 126.4
49000	110	128 ± 81.3	263	175	194 ± 89.4
49505	90	93.6 ± 33.1	231	140	143 ± 38.5
50360	242	247 ± 49.4	185	328	328 ± 50.0
52000	15	18.7 ± 17.8	1521	35	42.7 ± 25.4
52204	30	37.2 ± 22.9	369	74	76.4 ± 32.9
54520	56	61.9 ± 26.7	54	105	107 ± 31.8
55400	130	135 ± 34.8	291	158	159 ± 35.9
55845	265	294 ± 114	104	355	388 ± 121.4
56301	50	55.9 ± 20.8	82	100	105 ± 27.4
56341	147	149 ± 44.6	70	215	219 ± 47.2
57520	32.5	34.9 ± 15.0	198	75	78.1 ± 21.0
58150	147	154 ± 57.4	343	210	223 ± 65.2
58970	55	61.0 ± 26.4	241	100	107 ± 31.6
59840	29	31.4 ± 15.2	81	60	66.1 ± 23.4
59841	30	31.9 ± 14.2	204	65	67.0 ± 20.6
61510	340	354 ± 116	318	445	453 ± 115.0
61538	470	467 ± 82.1	200	540	532 ± 75.4
64721	35	39.8 ± 24.1	264	80	86.6 ± 32.3
66984	67	71.2 ± 26.7	767	110	115 ± 30.5
67107	135	138 ± 41.5	281	185	192 ± 44.1
67108	240	241 ± 87.3	150	303	298 ± 93.8
69631	105	108 ± 39.6	99	150	158 ± 45.2
69632	110	115 ± 52.0	34	175	173 ± 63.0
69660	60	70.2 ± 29.3	74	85	104 ± 45.9
93640	75	86.2 ± 47.2	491	135	143 ± 53.2

Mean times (min) ± standard deviations; CPT = common procedural terminology code; N = number of cases for each CPT (13,196 total cases and 40 CPTs in the representative database).

20 categories, the number of cases in each, the descriptors for the three most common CPT codes in each, and the nominal range of CPT codes available to each category are reported in table 1.

Initial Database. There were 8,834,789 min of total operating-room usage (TT) in the initial database; 69% of this (or 6,109,704 min) were used for the surgical procedure (ST) alone, *i.e.*, with anesthesia and positioning times omitted. There were 1,424,197 min of initial anesthesia time, 639,240 min of preparation and surgical positioning time, and 661,648 min of anesthesia emer-

gence and transport time. Anesthesia induction and emergence time together consumed 24% of TT. For the categorized CPT codes, procedure durations for ST ranged from a minimum of 1 min to a maximum of 1,130 min (18.3 h), whereas median durations for each category ranged from 33 to 195 min. Procedure durations for TT ranged from a minimum of 5 min to a maximum of 1,320 min (22 h), whereas median durations for each category ranged from 70 to 275 min.

The relative importance of the 20 surgical subspecialty categories to the institution is illustrated by the detailed

VARIABILITY IN SURGICAL PROCEDURES

Table 6. Analytic Database: Tabular Comparisons of Main Effect ANOVAs by Level of Significance for ST and TT (12,125 Cases and 40 Procedures in the Analytic Database C)

P value	Age	Anesthesia	ASA	Gender	Surgeon
Surgical time (ST)					
$P \leq .05$	7 (17.5)	15 (44.1)	4 (14.3)	6 (19.4)	30 (75)
$P > .05$	33 (82.5)	19 (55.9)	24 (85.7)	25 (80.6)	10 (25)
Analyzed	40 (100)	34 (100)	28 (100)	31 (100)	40 (100)
Omitted	0	6	12	9	0
Total	40	40	40	40	40
Total time (TT)					
$P \leq .05$	5 (12.5)	18 (52.9)	5 (17.9)	6 (19.4)	30 (75)
$P > .05$	35 (87.5)	16 (47.1)	23 (82.1)	25 (80.6)	10 (25)
Analyzed	40 (100)	34 (100)	28 (100)	31 (100)	40 (100)
Omitted	0	6	12	9	0
Total	40	40	40	40	40

Procedure count, column % relative to number analyzed.

description of ST and TT in tables 2 and 3. Among categories, cardiovascular surgery stood apart, using the greatest quantities (25% and 24% of ST and TT, respectively) of operating-suite time. Four categories—cardiovascular, musculoskeletal, neurosurgery, and general surgery procedures—together accounted for approximately two thirds of total utilization (65% and 66% of ST and TT, respectively). Of 268 surgeons, the number of cases per individual surgeon ranged from 1 to 1,495 and for individual anesthesiologists ranged from 1 to 2,503 cases.

Representative Database. Although the derivation of the representative database was chosen to foster analytical clarity by vastly reducing the number of CPT codes under consideration, it presumed both physiologic and surgical subspecialty representation and maintained a sizable fraction of patient procedures.

As indicated, only 40 of the 60 conceptually possible CPT codes in the initial database appear in the representative database. This database resulted in 13,196 case records, 193 surgeons, and 124 anesthesiologists. There were 2,316,472 min of total operating-room utilization (TT) in the representative database; 71% of this (or 1,633,817 min) was used for the surgical procedure (ST) alone, *i.e.*, with anesthesia and positioning times omitted. There were 368,220 min of initial anesthesia time, 142,967 min of preparation and surgical positioning time, and 171,468 min of anesthesia emergence and transport time. Roughly speaking, the representative database reflects the initial database in utilization characteristics, but only 26% of total minutes of operating suite

utilization. The representative database was segmented by CPT code and tabulated with respect to type of anesthesia, age, ASA risk class, gender, surgeon, and anesthesiologist (table 4, Web tables 1–3). The relative proportion of TT devoted to the various component times is essentially the same in both databases.

The mean age of all patients in the representative database was 50.6 ± 17.7 yr (mean \pm SD; $n = 13,196$ cases). The gender and ASA risk scores of the patients for each procedure are summarized in Web table 2. Of 13,196 patients in the representative database, 51% were female and 49% were male. Of all surgical procedures, 46% were associated with general anesthesia, 34% with local, and 10% each with monitored or regional anesthesia (Web table 3).

In the representative database, one third of cases (34%) was listed as "local." Of the 66% of remaining cases, *i.e.*, those who received anesthesia services and were therefore assigned ASA risk scores, 21% had an ASA score of 1, 33% had a score of 2, 36% had a score of 3, 9% had a score of 4, and less than 1% had a score of 5.

In summary, the characteristics of surgical cases in the representative database are, in our experience, typical of contemporary surgical and anesthetic academic practice. Moreover, within each of the various CPT codes in the representative database, there was a diversity of patient demographics and medical, anesthetic, and surgical characteristics.

Analytic Database. Descriptive comparisons of the representative and analytic databases indicated strong comparability. Thus, we believed that it was appropriate to assert that the ANOVA results we obtained in the next subsection can be expected to be inferences relevant to the representative database. In fact, the analytic database

|| In the initial database, there were 28 cases for which no surgeon was assigned or surgeon was missing; these cases subsequently were excluded in the analytic database.

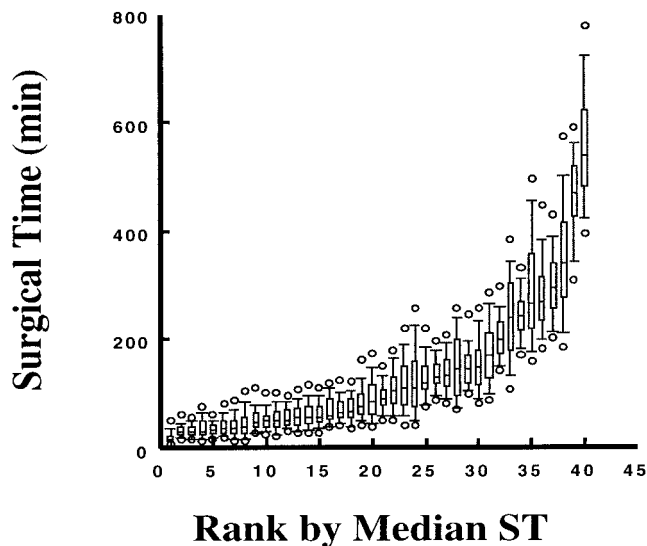


Fig. 2. Boxplot comparisons of ST ordered by increasing median surgical time (minutes). Each box indicates the range between the 25th and 75th percentile, capped bars indicate the 10th and 90th percentiles, and open circles indicate the 5th and 95th percentile of the extreme data points. A line in the middle of each box indicates the 50th percentile (13,196 cases, 40 procedures, and 193 surgeons in the representative database).

was technically necessary for certain analyses but was otherwise an undesired digression from the primary task, which was examination of the representative database. For the sake of brevity, tabulations of the analytic database are not reported.

Analytic Results

Summary statistics for ST and TT for the 40 procedures in the representative database are reported in table 5. Procedure duration data for TT and ST (fig. 2 and Web fig. 1, respectively) were reported for the representative database using multiple boxplots# in which procedures were ranked by median duration. The rank order for ST was similar to that for TT; however, among the shorter procedures, the addition of anesthesia time varied the order for some procedures. The range and variability of the data increased markedly with increased procedure duration. This relation is examined in further detail in the following section.

Main-Effect ANOVAs

To evaluate the association of independent variables (age, type of anesthesia, ASA risk class, gender, surgeon,

and anesthesiologist) with ST and TT, we used main-effects linear models based on the analytic database. For each CPT code, we tabulated the resultant *P* values corresponding to each independent variable. In order of importance among the 40 CPT codes, the main effects for ST were found to be statistically significant at the 0.05 level for surgeon (30 CPT codes; 75%), type of anesthesia (15 codes; 44%), gender (6 codes; 19%), age (7 codes; 17%), and ASA risk class (4 codes; 14%; table 6). The order and relative percentages were similar when the same analysis was performed on TT, except that ASA risk class moved up to fourth place in rank order of importance replacing age, which moved down to fifth place.

Adjusted Means

From our linear models, we obtained adjusted means for LnST and LnTT for each of the surgeons for each CPT code. These adjusted means are the estimated mean effects for LnST and LnTT for each surgeon, with the effects caused by other independent variables adjusted or controlled for. For instance, for CPT code 52000 and the dependent variable ST, the adjusted means for the 193 surgeons were the estimated mean LnST for each surgeon for cystoscopy adjusted for the effects caused by that surgeon's patients' ages, ASA class, etc. The adjusted mean LnST for surgeon was plotted against the weighted overall mean LnST for each procedure in figure 3. The variability of the logged adjusted means appeared homogeneous with respect to procedure duration.

We subsequently plotted the unlogged adjusted mean

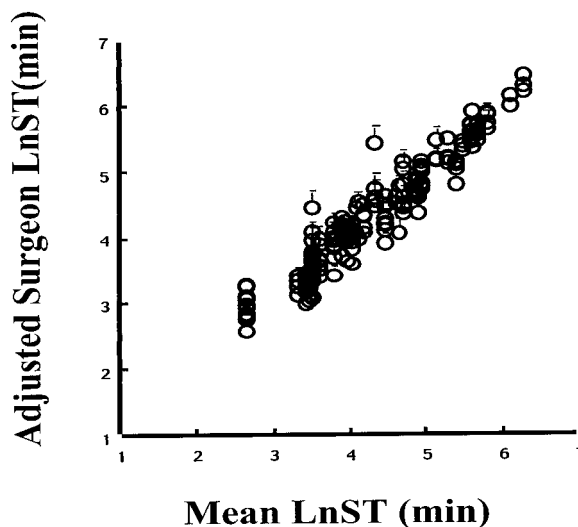


Fig. 3. Plot of the surgeon-adjusted means of the natural log of ST against the means of the natural log of ST (LnST) (12,125 cases, 40 procedures, and 96 surgeons in the analytic database).

Boxplots were ranked by median CPT duration to graphically separate CPT procedures of similar duration. In the event of a tie of medians, CPTs were ranked by means.

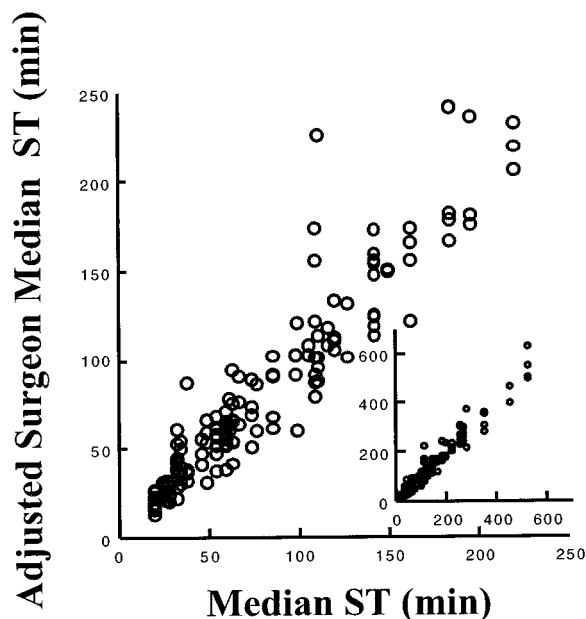


Fig. 4. Plot of the unlogged, adjusted, surgeon median ST plotted against the weighted median ST for each procedure, illustrating proportionately increased variability with longer procedure times ($n = 12,125$ cases, 40 procedures, and 96 surgeons in the analytic database).

(medians) for each surgeon against the overall median ST and TT for each procedure (fig. 4). This described, on the natural time scale, the linear model results, which used logged data. The variability among surgeons increased proportionately with respect to procedure time, implying that the variability among surgeons behaves as a multiplicative function of procedure time.¹⁵

The mean square error of the linear models decreased with the log of procedure duration (fig. 5), indicating that the case-specific pure error variability measured in the log scale decreased as surgical procedure time increased. On the other hand, the F ratio for surgeon effect for each CPT code, when plotted against average surgical duration (fig. 6), indicated a relatively homogeneous surgeon effect, no matter what the mean log surgical duration was. Thus, surgeon effect remained essentially the same, even for longer-duration CPTs, where case-specific error was smaller. This can be interpreted as meaning that logging of the ST effectively removed variability caused by the surgeon effect. If we accept this interpretation, then the majority of variability in surgical procedure times can be thought of as related to *surgeon work rate effect* (different surgeons work at different rates, and thus they differ from one another more and more as time passes).

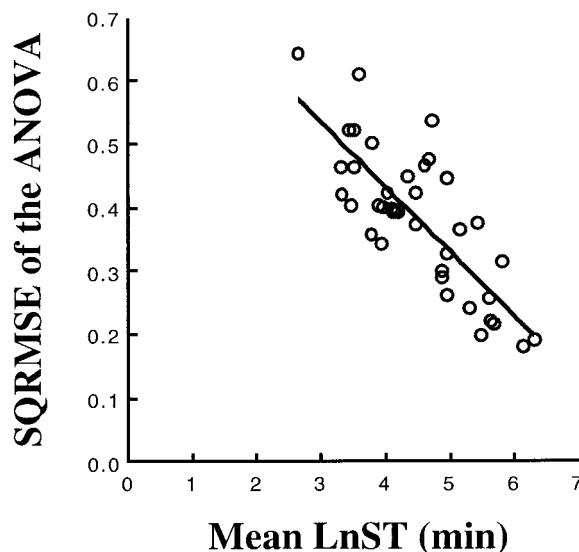


Fig. 5. Plot of the square root of the mean square error of the five-factor ANOVA model versus the mean LnST for each procedure (12,125 cases, 40 procedures, and 96 surgeons in the analytic database).

Unlogged adjusted means (medians) for anesthesia effect based on the linear model (which included surgeon effect) were plotted against median procedure time (fig. 7). General or regional anesthesia was often used for longer procedures, whereas local anesthesia was used often for shorter procedures. Compared with figure 4, there was no increased variability in anesthesia-adjusted median ST, which, importantly, indicated that the anes

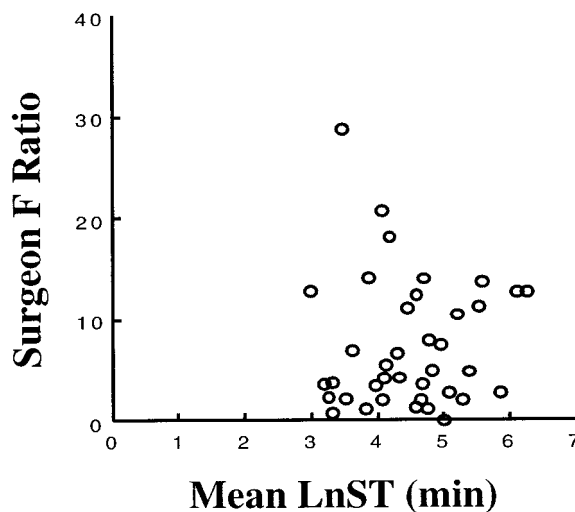


Fig. 6. Analytic database: plot of the F ratio (mean square error of the main-effect factor surgeon divided by mean square error of the total regression using the five-factor ANOVA model) with respect to weighted mean LnST (12,125 cases, 38 procedures, with 2 outliers omitted from the analytic database).

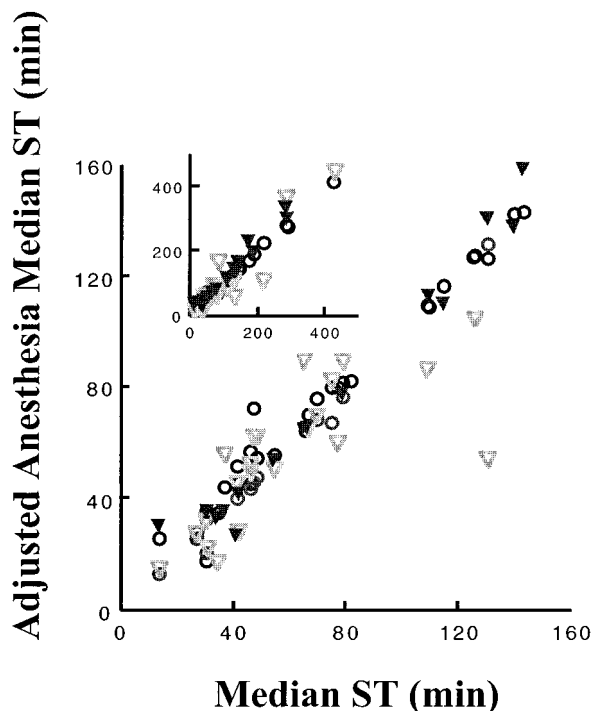


Fig. 7. Plot of the adjusted anesthesia median ST (minutes) versus median ST for each procedure. Closed circle = local anesthesia; open circle = general anesthesia; closed triangle = regional anesthesia; open triangle = monitored anesthesia (12,125 cases and 40 procedures in the analytic database).

thetia effect did not vary as a multiplicative function of surgical duration.

When anesthesiologist was added to the ANOVA model as a sixth independent variable, there were insufficient data to complete the ANOVA for 15 of the previous 40 procedures for both ST and TT. The main-effect anesthesiologist was significantly ($P < 0.05$) associated with ST for only 2 of 25 codes (8%) and with TT for only 3 of 25 codes (12%). Inspection of the main-effect surgeon with and without the addition of anesthesiologist revealed that the addition of anesthesiologist as a factor did not substantially affect the results for surgeon for either ST or TT. For this reason, we chose to report the results from the five-factor ANOVA (without anesthesiologists) with 40 procedures, rather than the six-factor ANOVA (with anesthesiologists) with 15 fewer procedures.

Discussion

Using a data mining approach, we examined a large and diverse surgical data set by segmenting it into 40 representative subsets based on headers from the CPT

classification. Each subset was analyzed using a main-effects linear model to identify and quantify specific sources of variability in surgical procedure times. The most important source of variability in surgical procedure times was surgeon work rate effect, which increased proportionate to procedure duration. Type of anesthesia, gender, age, and ASA risk class were, in order of importance, additional sources of variability in surgical procedure times. Case-specific relative variability was high for short cases, probably related to time penalties inherent in repeated turnovers of shorter surgeries.

Our analyses indicated that variability among surgeons may be a result of their working at relatively constant but differing rates, independent of procedure duration. The absolute impact of this work rate effect is pronounced absolute variability among surgeons for surgeries of long duration. As an example, if a surgeon's work rate is 10% apart from the average of his peers, then after 60 min the surgeon diverges by only 6 min. By contrast, if the surgeon works 10% apart from his peers for 10 h, the surgeon diverges by a full 60 min. We have demonstrated that this important work rate phenomenon exists at a single institution, but we believe it is generalizable to other institutions. Obviously, however, specifics for individual surgeons must be established independently at each institution.

Trends in surgical variability associated with type of anesthesia, age, gender, and ASA risk class are not proportionate to procedure duration. Trends in these factors may also be more easily extrapolated to surgical scheduling in other institutions. Anesthesiologist as a factor had little effect on the variability of ST and TT; however, we expect it would have much more effect on the variability of anesthesia time, a dependent variable not specifically addressed in this study.

It is interesting that type of anesthesia is a factor associated with variability in both ST and TT. It is not surprising that TT is affected by type of anesthesia because TT contains anesthesia induction and emergence times. However, it is surprising that type of anesthesia is strongly associated with ST, because ST (in theory) contains no component of anesthesia induction, anesthesia emergence, or positioning times within it. The authors believe the observed significant anesthesia factor effect is probably related to patient selection phenomenon, whereby healthy patients, with straightforward procedures, are favored for local or monitored anesthesia.¹⁶ Because this study is retrospective, however, we cannot establish causality and only note the association.

VARIABILITY IN SURGICAL PROCEDURES

Intrinsic case-specific variability is higher for short surgeries than for longer surgeries. The coefficient of variation is highest for short procedures, although the absolute variability is expected to be highest for longer procedures. We conjecture that case-specific absolute variability does not improve with repetition of shorter procedures because start-up delays for surgery, although not supposed to be “measured” by ST, may nonetheless affect it. Thus, when shorter procedures are repeated throughout the day, unmeasured variability may be added to the procedures by either changing patients (laboratory work, history, pathology may differ) or by changing personnel (scrub nurses, circulators, and house staff physicians may differ among cases).

One might presume that longer surgeries should have higher intrinsic case-specific variability. That this does not occur may be a result of the fact that with longer procedures, the surgical team may learn to adapt to one another and as a consequence may become more efficient for longer procedures. In addition, surgeons may be able to adjust ST for longer procedures in ways they are unable to accommodate for shorter procedures, saving time by abbreviating elements of the procedure. We believe that these trends in intrinsic case-specific variability should also extrapolate to other institutions.

The variability in surgical procedure times induced by surgeon work rate effect is of such absolute magnitude that it swamps the effects of other variables in surgeries of longer duration (fig. 2 and Web fig. 1). This has important implications in an era of managed care because increased variability in surgical schedules increases the costs of both underutilization and overutilization.^{17,18} It suggests two strategies that might be used to control costs. One clear strategy for an institution is to schedule surgeries by taking into account the specific surgeon, a particularly important factor for longer surgeries. An additional strategy would be to schedule the longest case first (which we have observed to be the most variable case),^{4,19,20} because doing so allows the surgical schedule to be dynamically altered by the addition or deletion of shorter subsequent surgeries in such a way as to minimize the costs to the institution of underutilization and overutilization.

A key factor determining the effectiveness of advanced scheduling and allocation systems is the method used to make accurate estimates of surgical procedure times.²¹ Although guidelines exist to estimate procedure times,^{3,22} surgeons work at different rates, and few procedures are of standard length. Surgeons' and schedulers' estimates and historical averages have all been tried,

but few attempts have been made to validate the accuracy of these estimates.^{23,24} Most schedules presently require adjustments made from memory by an experienced booking clerk. When the clerk is absent, variability increases, and the schedules are of inferior quality.^{20,25,26} Our results demonstrate that surgeon-specific databases that describe factors affecting the variability of surgeries are needed to improve scheduling²⁷; however, suitable databases are currently largely unavailable.

Limitations

The 20 CPT categories from which we chose data represent a full spectrum of adult surgical physiology and subspecialties. CPT codes are important because they are the most widely accepted nomenclature for reporting of physician procedures and services under government and private health insurance programs.²⁸ In our experience, surgeons typically specialize in surgeries contained within a single CPT primary category, and the uniform language is useful for administrative management, claims processing, and a basis for local, regional, and national utilization comparisons. The CPT classification was a useful means to subdivide our data without sacrificing diversity. Nonetheless, our data were not representative of the incidence of particular surgeries among the population because of the necessity of arriving at our analytic database.

To better estimate surgeon effects, we reduced the diversity of our data (both in number of different procedures and surgeons) by eliminating surgeons with minimum experience from our analysis. However, we believe that the variability among surgeon work rates that we detected is probably similar to the population. For none of the other main effect variables did we insist that there be a minimum number of cases in each categorization of that variable. This may have reduced slightly our ability to detect statistically the effects of these other variables. As a result, the percentages of the significant variables identified in table 6 may be somewhat higher if we had insisted on minimum sample sizes for these variables as well.

Our data are the experience of a single academic healthcare institution and therefore contain biases inherent in the hospital's own subspecialization, as well as in the teaching of residents. In addition, procedures with distinct CPT codes are sometimes similar, for example, three- and four-vessel coronary artery bypass grafts, cystoscopy with and without biopsy, etc. Thus, our insistence on choosing the three most frequent procedures in each category also reduces the diversity of our sample

subsets to some extent. However, this compromise is necessary to obtain sufficient cases for us to explore differences among surgeons for each CPT chosen.

Conclusions

Although the surgeon work rate phenomenon we observed may be generalizable to other institutions, surgeon specifics will not be. Type of anesthesia, age, gender, and ASA risk class are additional factors that affect variability in surgical procedure times, and this knowledge may be extrapolated to other institutions. Case-specific variability unexplained by the preceding factors is high for short procedures relative to longer procedures and may be an unmeasured time penalty inherent in the repeated turnovers of shorter surgeries. The evidence suggests that surgeon-specific procedure times and knowledge of the sources of surgical variability are needed to improve modeling and thus surgical scheduling. Poor scheduling leads to suboptimal use of the surgical suites. Surgical suites are costly functional areas within hospitals and must be scheduled efficiently for the financial health of the institution as a whole.

The authors thank Dr. Gerard Bashein for assistance with the manuscript.

References

- Rossiter CE, Reynolds JA: Automatic monitoring of the time waited in out-patient departments. *Med Care* 1963; 1:218-25
- Barnoon S, Wolfe H: Scheduling a multiple operating room system: A simulation approach. *Health Serv Res* 1968; 3:272-85
- Hancock WM, Walter PF, More RA, Glick ND: Operating room scheduling data base analysis for scheduling. *J Med Syst* 1988; 12:397-409
- Robb DJ, Silver EA: Scheduling in a management context: Uncertain processing times and non-regular performance measures. *Decision Sci* 1996; 6:1085-106
- Strum DP, May JH, Vargas LG: Surgical procedure times are well modeled by the lognormal distribution. *Anesth Analg* 1998; 86:S47
- Bliss CI: Provisionally normal distributions, *Statistics in Biology*. New York, McGraw-Hill, 1967, pp 152-85
- Bashein G, Barna C: A comprehensive computer system for anesthetic record retrieval. *Anesth Analg* 1985; 64:425-31
- Kirschner CG, Burkett RC, Marcinowski D, Kotowicz GM, Leoni G, Malone Y, O'Heron M, O'Hara KE, Scholten KR, Willard DM: *Physicians Current Procedural Terminology 1995*, 4th Edition. Chicago, American Medical Association, 1995
- John GH: Enhancements to the data mining process, Department of Computer Sciences. Palo Alto, Stanford University, 1997
- Fayyad U, Piatetsky-Shapiro G, Smyth P: From data mining to knowledge discovery in databases. *Artif Intell* 1996; Fall:37-53
- Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times. Pittsburgh, Katz Graduate School of Business, University of Pittsburgh, 1996
- Strum DP, May JH, Vargas LG: Lognormal distributions are a best fit model for scheduling surgical procedure times. *Anesth Analg* 1997; 84:S56
- Every NR, Spertus J, Fihn SD, Hlatky M, Martin JS, Weaver WD: Length of hospital stay after acute myocardial infarction in the myocardial infarction triage and intervention (MITI) project registry. *J Am Coll Cardiol* 1996; 28:287-93
- Wilkinson L: SYSTAT: The System for Statistics, Version 5.2. Evanston, IL, SYSTAT, Inc., 1992
- Strum DP, Vargas LG, Sampson AR, May JH: Individual surgeon variability is a multiplicative function of surgical time. *Anesth Analg* 1999; 88:S48
- Strum DP, Vargas LG, May JH: Should anesthetic type, surgeon, and location be used to schedule surgical procedures? *Anesth Analg* 1996; 82:S433
- Goldman J, Knappenberger HA, Moore EW: An evaluation of operating room scheduling policies. *Hosp Manage* 1969; 107:40-51
- Strum DP, Vargas LG, May JH: Surgical subspecialty block utilization and capacity planning: A minimal cost analysis model. *ANESTHESIOLOGY* 1999; 90:1176-85
- Kennedy MH: Bin-packing, knapsack, and change-constrained approaches to operating room scheduling, Department of Decision Sciences and Engineering Systems, Ann Arbor, Rensselaer Polytechnic Institute, 1994, pp 205-6
- McQuarrie DG: Limits to efficient operating room scheduling. Lessons from computer use-models. *Arch Surg* 1981; 116:1065-71
- Ozkarahan I: Allocation of surgical procedures to operating rooms. *J Med Syst* 1995; 19:333-52
- Dexter F: Application of prediction levels to OR scheduling. *AORN J* 1996; 63:1-8
- Wright IH, Kooperberg C, Bonar BA, Bashein G: Statistical modeling to predict elective surgery time: Comparison with a computer scheduling system and surgeon-provided estimates [see comments]. *ANESTHESIOLOGY* 1996; 85:1235-45
- Magerlain JM, Martin JB: Surgical demand scheduling: A review. *Health Serv Res* 1978; 13:418-33
- Goldman J, Knappenberger HA, Shearon WT: A study of the variability of surgical estimates. *Hosp Manage* 1970; 110:46,A,D
- Strum DP, Vargas LG, May JH, Bashein G: Surgical suite utilization and capacity planning: A minimal cost analysis model. *J Med Syst* 1997; 21:309-22
- Bruwer AM: Monitoring the cost-effective use of operating theatres. *Nursing RSA* 1994; 9:21-3
- 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