Predicting Anesthesia Times for Diagnostic and Interventional Radiological Procedures

Franklin Dexter, M.D. Ph.D. (Associate Professor and Director of the Division of Management Consulting, Department of Anesthesia, University of Iowa)

Jack C. Yue, Ph.D. (Associate Professor, Department of Statistics, National Chengchi University, Taipei, Taiwan)

Angella J. Dow, B.S. (Clerk IV, Department of Anesthesia, University of Iowa)

Address correspondence to Franklin Dexter, M.D., Ph.D., Division of Management Consulting, Department of Anesthesia, University of Iowa, Iowa City, Iowa 52242. Phone 319-621-6360, Fax 603-947-1304, Email Franklin-Dexter@UIowa.edu

Attribute work to Division of Management Consulting, Dept. of Anesthesia, University of Iowa

Abbreviated title. Predicting durations of radiological procedures

Abstracts describing this work have been submitted to the American Association of Clinical Directors meeting on October 23, 2005 in New Orleans, Louisiana, and to the 2005 American Society of Anesthesiologists' meeting.

Financial disclosures. Support was provided solely from institutional and/or departmental sources. FD is Director of the Division of Management Consulting, a Division of the Dept. of Anesthesia. He receives no funds personally other than his salary from the State of Iowa, including no travel expenses or honorarium, and has tenure with no incentive program.

Summary statement. We studied challenges and solutions to predicting anesthesia times for diagnostic radiological procedures (e.g., magnetic resonance imaging) and interventional procedures (e.g., neuroradiology).

Abstract

Introduction: A full-time non-operating room (OR) scheduler was hired by an anesthesia department's hospital. Non-OR sites (e.g., interventional radiology) were allocated time. Unexpectedly, challenges in implementation were inaccurate anesthesia times for radiology.

Methods: Anesthesia billing data and paper logbooks from non-OR sites were used.

Results: Case durations for computerized tomography (CT) and magnetic resonance imaging (MRI) classified by Current Procedural Terminology (CPT) were less accurate than for OR cases. There were many different CPT, almost all rare, and CPT reflected organs imaged, not scanning times. Anesthesia times estimated by expert judgment had better face validity, internal consistency, and accuracy. Methods were developed and validated to estimate upper and lower prediction bounds (e.g., for fasting "NPO") from the expert estimates. Scheduling interventional radiology was challenging because a few CPT accounted for most cases. Since interventional radiology scheduled cases into allocated time, and open access was planned within two-week increments, the relevant objective was to estimate the time to complete series of elective cases including turnover times. We describe and validate how to pick the time up to when interventional radiology schedules to be finished on a specified percentage of days by a specified time (e.g., 6 PM). Results are expected to be similar for International Classifications of Diseases Version 9.

Conclusions: We believe this paper to be the first investigation of the estimation of anesthesia times for cases performed outside of ORs. Case duration prediction based on CPT(s) performs poorly, unlike for OR cases.

Introduction

An anesthesia department aimed to improve case scheduling for anesthetics outside of operating rooms (ORs). Time was allocated to adult and pediatric cardiology and to interventional and neuroradiology (Table 1). A full-time non-OR scheduler was hired by the hospital to facilitate the transition. A Vice Chair became actively involved in non-OR scheduling. Together, they shepherded the process of working with non-OR sites as they choose days of the week for their allocations. During a transition period of a few months, three non-OR teams were used to eliminate a two-month backlog. Then, as forecasted based on expected workload, two teams were used each workday. With multiple meetings and support of the hospital and College of Medicine's leadership, implementation was relatively uneventful.

However, accurately estimating anesthesia times for diagnostic and interventional radiology was unexpectedly a problem. In retrospect, large inaccuracy had been present for a decade. Yet, without a focus on non-OR scheduling, the anesthesia department had not appreciated the magnitude of the case estimation problem. Starting and finishing computerized tomography (CT) and magnetic resonance imaging (MRI) anesthetics early or late had been attributed to other causes (e.g., how the time was planned and cases were scheduled).

The first attempt to improve estimates of anesthesia times for scheduling[1] radiology was to average[2,3] historical anesthesia times classified by the Current Procedural Terminology (CPT) codes used for billing. The assumption was made that such estimates of anesthesia times would be highly accurate for CT and MRI, because these procedures involve unchanging and unbiased physics, not different surgeons who may also deliberately underestimate times.

In this paper, we describe reasons identified for inaccuracies in using CPT for case scheduling of radiological procedures, both interventional and diagnostic. We describe some reasonable alternatives, and our experience in implementation. For application outside the USA, frequent reference is also made to the October 2004 edition of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) procedure codes.

Methods

Quantifying Inaccuracy in Predicting Anesthetic Time for Radiology

"Anesthetic time" referred to the time of continuous anesthesia presence. The first goal was to put the inaccuracy in predicting anesthetic time by radiology CPT into practical perspective.

The diagnostic radiological procedures studied were CT and MRI with anesthesia, almost all for children. Three years of billing data were available: September 1, 2001 through August 31, 2004. As described by Strum et al.,[4,5] cases studied were those of CPT[*] performed a moderate to large number of times ($N \geq 30$). The observed sample mean for each CPT was considered the scheduled duration. Uncertainty was quantified using the mean absolute error and the mean absolute percentage error.

The comparison group was the hospital's surgical cases during the same period. The cases were sorted in ascending order of duration. The fewest number of cases were used that provided the same overall mean duration as that of the studied MRI and CT cases. The cases' scheduled and actual OR times were obtained from the hospital's OR information system.

For quantifying uncertainty in anesthetic times for interventional radiology, cases were classified by combination of CPT and physician.[2,3,6] Data were available including the most common physician from July 1, 2003. The same final date of August 31, 2004 was used.

Estimating Anesthetic Times for Computerized Tomography

The chief CT technologist was provided a list of body regions scanned by CT (e.g., brain or thorax). She provided her expert judgment of the CT room time required when a case is performed with anesthesia. The next day, her estimates plus 15 min were converted into a presentation-format rule by the authors, where 15 min was the value used by anesthesia

---

[*] The CPT studied were 70553 MRI of brain without contrast followed by contrast and further sequences (N = 281); 70551 MRI of brain without contrast (N = 74); 70543 MRI orbit, face, and neck without contrast followed by contrast and further sequences (N = 51); 72141 MRI of spine without contrast (N = 43); 72156 MRI of spine without contrast followed by contrast and further sequences (N = 32); 70480 CT of orbit, sella, posterior fossa or outer, middle, or inner ear, without contrast (N = 35); and 74160 CT of abdomen with contrast (N = 51).

schedulers for years as the typical time to take the patient to the post-anesthesia care unit. The expert reviewed the rule, but made no changes. By this process the expert's estimates of anesthesia times were made *separate* from the data (i.e., *no data fitting was performed*).

To assess the accuracy of estimates, the scheduled procedure(s) needed to be known. For example, if the scheduled procedure was "CT of head" and yet "CT of brain" was performed, such poor scheduling may reduce the accuracy of estimates. Both anesthesia scheduling and CT had kept their paper logbooks listing scheduled procedure(s) back to June 1, 2003. The 1.7 yr of data provided 105 anesthetics for validation.

Residuals between actual anesthesia times and expert estimates were analyzed to determine if there was additional error that could be explained. Each body region scheduled for scanning was coded as 0 or 1 for each patient. Use of each CT machine was coded as 0 or 1. Mann-Whitney test was applied repeatedly, one body region or machine at a time, to test whether residuals differed. Because 20 comparisons were performed without correction for multiple comparison, by design the analysis was intentionally overly likely to detect a significant relationship even when detection may be spurious due to random error.

Residuals were compared to those of the original scheduled times to assure that estimation errors were no worse than that of the original method of scheduling. Better performance was preferred by the anesthesia department studied, and was achieved (see Results). Nonetheless, that was not of scientific importance, as the baseline method of scheduling lacked both face validity (e.g., a whole body CT scheduled for the same time as a sinuses only CT) and internal consistency (e.g., CT of the same sites scheduled for different times). The expert description, in contrast, will be shown to be systematic enough for implementation in enterprise wide scheduling software and generalized enough for implementation elsewhere. Thus, the scientific issue was to show that the estimates of anesthesia times by expert judgment were no worse than estimates by the schedulers, as that suggests that there was not important a priori knowledge that schedulers had about individual patients that the expert did not use.

Lower (5%) and upper (90%) prediction bounds were estimated. The next three paragraphs explain the logic. Since anesthetic times are uncertain, the expert's estimate alone could not be used for good decision-making.[7]

The average anesthetic time is relevant to scheduling a case into the time allocated for a service, or into time of another day[8,9]. The allocations for the services (Table 1) were calculated using the forecasted total hours of anesthesia including turnovers for each service.[8,10] When deciding whether to schedule a case into OR time that has been allocated based on the expected (mean) total hours of anesthesia time, the value used should be the unbiased estimator for the total time.[1,2,7] The same applies to moving a case, releasing allocated time, and assigning anesthesia providers.

The upper prediction bound[7,11,12] is the end-point that relates to inserting a case into a gap in the schedule (e.g., CT in the middle of the workday during time originally scheduled for a case that was subsequently cancelled). The scheduler needs to know the longest time that an anesthetic is likely to take. An upper prediction bound for the duration of a case is the value that will be exceeded by the next case of the same type at the specified rate. There is a 10% chance that the duration of a case will be longer than its 90% upper prediction bound.

The lower prediction bound is the end-point that relates to planning the availability of the subsequent patient to receive care by the same team.[7,13] The anesthesia scheduler and pediatric practitioners adjust children's fasting times using the calculated lower 5% prediction bounds for the shortest expected durations of preceding cases (below).

Previously described methods for calculating lower[7,13] and upper[7,11,12] prediction bounds for cases relied on the statistical distribution of OR times to follow two-parameter log-normal distributions, with different parameters for each combination of surgeon and scheduled CPT(s). However, in the Results we will show that this approach does not apply, because the estimated anesthetic time(s) were neither based on surgeon or scheduled CPT (s). Rather, the relevant issue was residuals from a rule based on expert judgment.

Suppose that there are N observed anesthesia times (AT_i, i = 1, 2, …, N), each with a corresponding (Z_i) expert estimate of that anesthesia time. Assume that the experts are providing the median anesthesia times for CT's of a specified region. This assumption was tested with the sign and the Wilcoxon signed-ranks tests.

Assume that the AT_i, i = 1, 2, …, N, follow two-parameter log normal distributions with individual medians but a common variance. Specifically, let μ_i refer to the expected value (i.e., mean) of the natural logarithm of AT_i and c refer to the ***c***ase effect specifying variation of the natural logarithm of AT_i around that mean.[14] By definition, c_i follows a normal distribution with mean of 0 and unknown variance of σ^2, and $AT_i = \exp(\mu + c_i) = \exp(\mu) \times \exp(c_i) = Z_i \times \exp(c_i)$. Therefore, we use Lilliefors' test as to whether the N observed $\ln(AT_i / Z_i)$ follow a normal distribution with mean of zero and common, unknown, variance σ^2. Simultaneously, we obtain s, the sample estimate of σ.

Provided the assumptions hold, then lower[7,13] and upper[7,11,12] prediction bounds can be calculated for the next anesthetic (AT_{i+1}) based on the prior N anesthetics and the expert estimate of its duration (Z_{i+1}).[15] Let α refer to the desired quantile (e.g., $\alpha = 0.05$ for the 5% lower prediction bound and $\alpha = 0.9$ for the 90% upper prediction bound). The prediction bound equals $Z_{i+1} \cdot \exp\left(s \cdot \sqrt{1+1/N} \cdot T^{-1}\left[N-1, \alpha\right]\right)$, where $T^{-1}\left[N-1, \alpha\right]$ is the inverse of the Student t cumulative distribution function with (N–1) degrees of freedom.

All of the above steps for CT were repeated for MRI. Paper logbooks were available back to December 1, 2003. The 1.2 yr of data provided 154 anesthetics for review.

Diversity of Procedures in Diagnostic and Interventional Radiology

To understand why there were challenges in predicting anesthesia times for radiology, the distributions among all anesthetics of procedures involving radiology were studied. This methodology is presented after the above description of prediction bounds. However, the corresponding results will be presented first.

Sites studied were CT and MRI, interventional radiology, and pediatric cardiac catheterization laboratory. The latter control site was the only non-OR site other than radiology that was active for several years before initiation of the allocation of time for non-OR activities. All cases performed in 2004 were used.

A simple, but limited, method of quantifying the diversity of procedures is the percentage of anesthetics accounted for by the most common procedure(s). Standard errors for proportions were estimated by using the Clopper-Pearson confidence intervals.[16]

The diversity of the procedures performed at each of the non-OR sites was quantified using the internal Herfindahl index.[17,18] The internal Herfindahl index equaled the sum of the squares of the proportions of all anesthetics at a site that were accounted for by each procedure(s). That is, it equaled the probability that if two anesthetics were selected at random, both would be of the same procedure(s). Corresponding standard errors were calculated by Taplin's method.[19]

For example, suppose that 3 procedures were performed at a site in relative proportions of 50%, 40%, and 10%. Then, the internal Herfindahl index would equal 0.42, where $0.42 = (0.50)^2 + (0.40)^2 + (0.10)^2$. If a different site performed 3 procedures in relative proportions of 93%, 5%, and 2%, then the internal Herfindahl index would equal 0.87, where $0.87 = (0.93)^2 + (0.05)^2 + (0.02)^2$. Although both sites performed 3 procedures, the second site was less diverse because the proportions of each procedure were less balanced.

For example, suppose that a site performed only one procedure (e.g., therapeutic radiation treatment delivery). Then, the internal Herfindahl index would equal 1.0, where $1.0 = (1.00)^2$.

For example, suppose that a site performed 100 procedures, each with a relative proportion of 1%. Then, the internal Herfindahl index would equal 0.01, where $0.01 = 100 \times (0.01)^2$. The minimum value of the internal Herfindahl index equals one divided by the number of different procedures performed at the site.

To understand inaccuracy in predicting anesthesia times, we also studied the percentage of anesthetics at different sites that were of rare procedures.[20] A procedure(s) was considered rare if it accounted for 0.5% or less, or 0.1% or less, of anesthetics performed during a one-year period. Derivation and validation of this statistic and its standard error are in the Appendix.

When the number of previous anesthetics (N) is large, the difference between the lower and upper prediction bounds depends solely on the intrinsic variability (σ) and median expected duration (μ) of the procedure (above). In contrast, for rare procedures (e.g., $N = 2$), uncertainty in the sample estimates of these parameters contribute markedly to the width of the interval. This relationship is represented mathematically as having a large magnitude of the inverse of the cumulative distribution of the Student's *t* distribution (above).

For example, the uncertainty in the anesthetic time for the next insertion of myringotomy tube is less than the uncertainty in the anesthetic time for the next knee replacement. This is because myringotomy tube insertion takes less time (smaller μ) than knee replacement. However, there are plenty of historical data to review for both procedures. In contrast, consider disarticulation of the hip. Much of the uncertainty in the duration of an anesthetic is likely due to few or no prior recent cases of the same procedure on which to estimate the durations.

Predicting Duration of the Interventional Radiology Workday

By 2004, there were frequently series of scheduled (elective) interventional radiology cases. All series of cases with anesthesia scheduled to start before 9:00 AM and finish after 3:00 PM were reviewed from January 1, 2004 through January 18, 2005. Scheduled versus actual times to complete the series of cases were studied. The rational is explained in the Results.

Results

Quantifying Inaccuracy in Predicting Anesthetic Time for Radiology

Unexpectedly, the durations of CT and MRI anesthetics were less predictable than surgery, despite involving physics rather than surgeons.

For CT and MRI, the mean absolute error was 43 ± 1 min and mean absolute percentage error was 45% ± 1% (N = 566). In comparison, the briefest 41,770 of surgical cases had the same mean (1.97 hr) as that of the CT and MRI anesthetics. Among the surgical cases, the mean absolute error was 29 ± 1 min and mean absolute percentage error was 27% ± 1%. Thus, expected durations of CT and MRI anesthetics were less accurate than that of surgical cases.

For interventional radiology, using combination of CPT and radiologist, the mean absolute error was 50 ± 2 min and the mean absolute percentage error was 24% ± 1% (N = 373). The mean duration was 3.89 hr. All 58,291 surgical cases were used as a comparison (mean 2.75 hr). The mean absolute error was 42 ± 0 min and mean absolute percentage error was 28% ± 1%.

Diversity of Procedures in Diagnostic and Interventional Radiology

Pediatric cardiac catheterization was used as a control, since there are many CPT and ICD-9-CM reflecting the many different congenital lesions. The most common CPT accounted for 18% of anesthetics, and the most common 3 CPT accounted for 39% of anesthetics (Table 2). The internal Herfinahl was 0.08, which is very low. Furthermore, most procedures were rare. Among anesthetics performed in the pediatric cardiac catheterization laboratory, 100% ± 0% were for procedures that each accounted for 0.5% or less of anesthetics. Thus, as expected, cardiac catheterization had a larger diversity of procedures, and a higher frequency of rare procedures, than does surgery.

For MRI and CT, the most common CPT (70553, MRI brain without contrast) accounted for 31% of anesthetics, and the most common 3 CPT accounted for 44% of anesthetics. The internal Herfindahl was 0.12. Among anesthetics for CT or MRI, 100% were for procedures accounting for 0.5% or less of anesthetics. The similarity of findings to pediatric cardiac catheterization laboratory partly explain why estimates of diagnostic radiology anesthesia times using CPT(s) were relatively inaccurate. Other reasons are below.

For interventional radiology, the most common CPT accounted for 63% of anesthetics, and the most common 3 CPT accounted for 77% of anesthetics. The internal Herfindahl was 0.42. Among cases with anesthesia, 37% were for procedures accounting for 0.5% or less of anesthetics. In fact, the 3 combinations of CPT and radiologist with $N \geq 30$ in the preceding section accounted for 69% of anesthetics. Thus, unlike for pediatric cardiac catheterization laboratory and diagnostic radiology, a challenge in predicting anesthetic times was that so many anesthetics were for a few CPT of long duration. The most common interventional radiology CPT (61624) was broad, not specifying the size of the lesion: “Transcatheter permanent occlusion or embolization (e.g., for tumor destruction, to achieve hemostasis, to occlude a vascular malformation), percutaneous, any method; central nervous system (intracranial, spinal cord).” The corresponding single ICD-9-CM was 39.72: “Endovascular repair or occlusion of head and neck vessels,” [including] “coil embolization or occlusion, …, endovascular graft(s),” [and/or] “… liquid tissue adhesive (glue) embolization or occlusion … for repair of aneurysm, arteriovenous malformation, or fistula.”

Estimating Anesthetic Times for Computerized Tomography

Expert judgment of anesthesia time for CT was:

30 min
+ (5 min scanning time for each of: head, facial bones, sinuses, or cervical spine)
+ (15 min scanning time for each of: brain, neck, thorax, abdomen, pelvis, or thoracic spine)
+ (5 min if not using the Siemens six slice spiral CT).

Analysis of residuals did not detect additional error that could be explained by body region(s) or the scanner used (all $P > 0.10$).

The bias of the estimate was negligible (2 ± 2 min).

The expert's estimate was for the median (P = 0.84 sign test, P = 0.50 Wilcoxon signed-ranks test). There were 50 under-estimates of the actual duration, 7 zero residuals, and 47 over-estimates of the actual duration.

Figure 1 shows a histogram of the natural logarithm of the actual anesthetic times to the expert estimate, with a superimposed normal curve (mean −0.01, standard deviation 0.36, N = 104). The distribution was close to that of a log-normal (Lilliefors P = 0.05). Strum et al. showed that and why goodness of fit tests tend to falsely reject null hypotheses of good fits for case duration data (i.e., P = 0.05 is acceptable).[21] The 5% lower prediction bound was 55% of the expert's estimate, where

$$0.55 = \exp\left(0.36 \cdot \sqrt{1+1/104} \cdot T^{-1}\left[104-1\,,\,0.05\,\right]\right).$$

The 90% upper prediction bound was 160% of the expert's estimate.

Rounding had a large effect on results, which was a reason why satisfying distributional assumptions was not overly important.[4,12] For example, depending on the machine used for CT of the brain, the expert's estimate above could be 40 min, which we could schedule for 45 min. Multiplying 0.55 by 40 min gives 22 min, which we practically would make 15 min as the lower bound. Multiplying by 1.60 by 40 min gives 64 min, which could be rounded to 1 hr or up to 1 hr 15 min being conservative.

The expert's estimate achieved a pairwise reduction in the mean absolute error of 11 ± 2 min versus that originally scheduled. The 5% lower bound calculated using the scheduled times was 48%, less than the above 55%. The 90% upper bound was 177%, more than the above 160%. Thus, the width of the uncertainty (5% to 90%) was reduced by 24%. The expert's estimate added face validity and internal consistency (see Methods) without evidence of losing a priori knowledge that the schedulers had about patients.

Estimating Anesthetic Times for Magnetic Resonance Imaging

Expert judgment of anesthesia time for MRI was:

30 min
+ (60 min scanning time)
+ (15 min scanning time if total spine or abdomen/pelvis extending into the thighs)
+ (15 min scanning time if not using the Siemens Avanto MRI).

Analysis of residuals did not detect additional error that could be explained by body region or the scanner used (all $P > 0.10$).

The bias of the estimate was negligible (2 ± 2 min).

The expert's MRI estimate did not differ from the hypothesis that the median was provided ($P = 0.68$ sign test, $P = 0.64$ Wilcoxon signed-ranks test, $N = 155$).

Figure 2 shows a histogram of the natural logarithm of the actual anesthetic times to the expert prediction, with a superimposed normal curve (mean -0.02, standard deviation 0.28). The distribution was close to that of a log-normal (Lilliefors $P = 0.69$). The 5% lower prediction bound was 63% of the expert's estimate, and the 90% upper bound was 143% of the estimate. For example, MRI of the brain would be scheduled for 90 min. The shortest time of 57 min could be considered 1 hr, and the longest of 128 min could be considered 2 hr 15 min.

The expert's MRI estimate resulted in a pairwise reduction of the mean absolute error of 12 ± 2 min. The corresponding 5% lower bound was 55%, less than the above 63%. The 90% upper bound was 158%, more than the above 143%. Thus, the width of the uncertainty (5% to 90%) was reduced by 23%. These findings suggest the schedulers did not have important, additional a priori knowledge lost in using the expert's estimate.

Quantifying Inaccuracy in Predicting Anesthetic Time for Radiology (Continued)

In the first section of the Results, we showed that CT and MRI had larger errors in predicting anesthesia times based on CPT than for ORs. Using the expert estimates, the mean absolute error was 19 ± 1 min and mean absolute percentage error was 26% ± 2%, less than for OR cases of comparable duration.

Radiology technicians recorded MRI scanning times in logbooks. Differences between anesthesia and scanning times were considered the anesthesia-controlled times. The expert estimate was 45 min for all patients. The mean absolute difference of the anesthesia-controlled times from 45 min equaled 19 ± 1 min, only slightly more than the 15 ± 1 min for the absolute difference of the actual and scheduled scanning times. Thus, further improvement in schedule estimates would require knowledge, at the time of case scheduling, of both patients' physiological condition for anesthesia and primary disease and machine for choosing the MRI scanning protocol. We failed at attempts to implement either with radiology scheduling.

Finally, findings that anesthesia times for CT and MRI were predicted poorly by CPT likely also applied to ICD-9-CM. The ICD-9-CM 88.38 included both CT of the sinuses and pelvis, differing by 10 min. The ICD-9-CM 88.97 included both MRI of the abdomen and orbit, differing by 15 min (above). Also, neither considered the substantive differences in scanning times among machines.

Predicting Duration of the Interventional Radiology Workday

Although CT and MRI were scheduled into open time, interventional radiology scheduled cases into allocated time (Table 1). Whereas individual cases were relevant to scheduling CT and MRI, the time to complete series of cases including turnover times applied to interventional radiology. Cases were scheduled sequentially. The scheduled end of the workday was compared to the actual end of the anesthetic workday. All delays, turnovers, etc. were included. Figure 3 shows a histogram of the differences in time between the actual versus scheduled ends of the anesthetic workday, with a superimposed normal curve (mean 0.75 hr, standard deviation 1.45 hr). As for series of OR cases, the distribution was close to normal (N = 57, P = 0.42).

The end of the workday was considered 6:00 PM. Relief was rarely available for anesthesia providers working in interventional radiology, even though they were not on call. Thus, the relative cost of an hour of over-utilized anesthesia time was considered costly, 4-fold more, than the cost of an hour of allocated time. For 4 out of 5 workdays (i.e., 80%),[22] the providers should finish early. The time up to when interventional radiology schedules cases was chosen so that the last anesthetic would be expected to end at 6:00 PM on at least 80% of workdays. Just as the 95% quantile of a normal distribution equals the mean + 1.65 $\times$ (standard deviation), the 80% quantile equals mean + 0.84 $\times$ (standard deviation). Substituting the mean and standard deviation above, the 80% quantile for the difference between actual and scheduled end of anesthesia equaled 2.0 hr, where 2.0 hr = 0.75 hr + 0.84 $\times$ 1.45 hr. In addition, the observed 80th percentile for the 57 workdays was 2.0 hr. Thus, interventional radiology scheduled cases with anesthesia sequentially to 4:00 PM.

Discussion

We believe this paper to be the first investigation of the estimation of anesthesia times for cases performed outside of ORs. If anesthetic times were known with relative certainty, then how the time was planned (i.e., allocated) into which the cases were scheduled could be separated from how anesthesia times were estimated. Since the anesthesia times were uncertain, estimation of anesthesia times could not be done rationally without considering simultaneously how the time was planned into which the cases were being scheduled.

Diagnostic radiology requested not to be allocated time. Partly, this was because CT, MRI, physiological imaging, echocardiography, etc., were performed in separate sites with different technicians. Partly, this was because these sites were scheduled like primary care outpatient clinics, with patients or secretaries calling, and a convenient time scheduled throughout the day. Scheduling such cases into open, first-come first-served time provided the most flexibility in start times when multiple days of the week were considered. These needs differed from those of interventional radiology at which the radiologists and technicians shared the same advantage as the anesthesia providers in scheduling cases sequentially.

Pediatric clinics were concerned about children fasting from the previous evening for diagnostic radiology in afternoons. For several years, anesthesiologists tried education to encourage the use of evidence-based practice. We perceive that lack of use reflected nurses' concerns in basing fasting instructions on scheduled start times. Implementation was achieved by our use of a dedicated non-OR scheduler who used the lower prediction bounds and told the nurse practitioners the recommended fasting period for each patient. The implication for this paper was that the expected (average) time was not the only important statistic for CT and MRI. The lower prediction bound was also needed.

Diagnostic radiology procedures were sometimes cancelled after scheduling, but several days before the date of the procedure. In ORs, pediatric cardiac catheterization laboratory, interventional radiology, etc., patients were routinely told their start times the day before surgery, so holes in the schedule could be filled by moving later cases to earlier times. However, we failed at encouraging such scheduling for diagnostic radiology. Many patients undergoing short procedures without anesthesia would have needed to have their times changed. The implication was that holes in the schedule were produced, and filled whenever possible by cases with anesthesia at other sites (e.g., cancelled MRI hole filled with a CT case). Upper prediction bounds show whether another case can fit within a hole in the schedule.

Interventional radiology performed its cases in individually allocated time. The anesthesia department studied received institutional financial support to provide open access within a reasonable (2 wk) period.[8,23] Together, there were fixed hours on any one day into which interventional radiology scheduled its cases. Each workday had insufficient time for the service to complete all its pending cases. Supporting this construct was the observation that the allocated time was never once released, other than many weeks in advance from vacations.[24] The issue then in estimating anesthesia times was to know the time up to when interventional radiology could schedule cases to be done reliably by the end of the workday (6:00 PM).[1,2] Whereas the uncertainty in the duration of a single case may be large, that of a series of cases was proportionally less.

Limitations

Although our findings are likely *valid* elsewhere, their *usefulness* depends on how anesthesia time is allocated for radiology. The anesthesia department studied did not provide open access on the requested day up to the limit of what could be done safely,[8,10,23] in contrast to the service provided in ORs[1,7,9,14]. At some hospitals, the expectation is that anesthesia providers should work late for any radiological procedure. Then, our approach for CT and MRI would apply, but not for interventional radiology. Nevertheless, we expect that the situation for the studied anesthesia department was commonplace, for two reasons.

First, radiology sites were not as interchangeable as ORs. When allocated time was full for the day and a service wanted to schedule another case, the allocated time of another service on the same day could not be released safely. For example, an interventional radiologist could not split and perform his other case inside an MRI machine. Consequently, if open access had been provided on the workday of choice of a service, then its new case would invariably have been performed in over-utilized anesthesia time rather than in the time originally allocated to another service that was likely not to use it.

Second, when OR staffing (i.e., short-term OR allocations) is based on departments with many surgeons, variations in OR workload from week to week are driven by variations in numbers of patients requesting to be scheduled for surgery.[25,26] In contrast, non-OR services (e.g., interventional radiology) were small, representing one or two physicians. Variation in workload was larger, reflecting vacations and meetings. Thus, providing open access on the workday of choice of the service would have resulted in substantial hours of under-utilized and over-utilized anesthesia time. The efficiency of use of allocated radiology time would have be substantially less than that of OR time.

Our results apply only to elective, scheduled radiology anesthetics. For example, our estimates are invalid for a CT at 2 AM in a combative trauma patient.

Analyses used data from one hospital. We studied the diversity of procedures performed with anesthesia in diagnostic and interventional radiology to understand the reason for our results. Based on those findings, we expect our results to be valid elsewhere.

Finally, we limited this paper to what we could implement. Substantial variability in anesthesia times remained. For MRI, we identified one cause (see Results), but lacked the ability to implement obtaining the needed data. In addition, for both CT and MRI, many children underwent other unrelated procedures while under general anesthesia (e.g., phlebotomy or bone marrow biopsy). Every such patient was included, even though these procedures were virtually never scheduled and thus contributed to inaccuracy in estimation of anesthesia times.

Appendix – Percentage of Cases that are of a Specified Type and are Rare

Each case is one of S procedure(s) $\{A_1, A_2, \ldots, A_S\}$ with relative frequencies $\{p_1, p_2, \ldots, p_S\}$. Some procedures, $C \leq S$, satisfy a dichotomous condition (e.g., diagnostic radiology). Consider the conditional probability of a case satisfying both the dichotomous condition and being rare, defined as $p_i \leq \lambda$, $i = 1, \ldots, C$. Without loss of generality, let procedures 1, 2, …, k be rare and satisfying the dichotomous condition ($k \leq C \leq S$). Also, without loss of generality, assume $p_1 \leq p_2 \leq \cdots \leq p_C$. Then, the statistic $\theta = \sum_1^k p_i \Big/ \sum_1^C p_i$, where $p_k \leq \lambda < p_{k+1}$. The corresponding nonparametric maximum likelihood estimator (NPMLE) $\hat{\theta} = \sum_1^{\hat{k}} X_i \Big/ \sum_1^C X_i$, where X_i is the observed number of cases of procedure i, $n = \sum_1^S X_i$. The $\hat{k}$ are estimated via $\sum_1^{\hat{k}} X_i \leq n\lambda < \sum_1^{\hat{k}+1} X_i$. Following Yue et al.,[27] $\hat{\theta}$ is asymptotically normally distributed with standard error $\sqrt{\hat{\theta}\left(1-\hat{\theta}\right) \Big/ \left(\sum_1^C X_i\right)}$.

For testing, the diagnostic and interventional radiology data were treated as population information in Monte-Carlo simulations. To also perform simulations with large sample sizes, three years of OR data were used for the percentage of cases that were rare and physiologically complex.[18,20,28,29] The NPMLE was within 1.96 standard errors of the true (simulated) mean other than when $n\lambda$ was just 5 cases per procedure for the OR data (Table 3). The bootstrap standard errors were accurate to within 1%, but not the asymptotic ones. Reasons were multi-factorial (not shown): $\theta \cong 0$, $\theta \cong 1$, $k \neq \hat{k}$, and non-smooth behavior near $p_i \cong \lambda$. In the Results, we report $\hat{\theta}$ with bootstrap standard errors.

Rarely procedures may be equivocal as to whether they satisfy the dichotomous category (e.g., a procedure may be performed in interventional radiology and in ORs). Even an unbelievably high 5% misclassification rate affected $\hat{\theta}$ by just 1% (Table 4).

References

1. Dexter F, Traub RD: How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesth Analg 2002; 94:933-942.

2. Dexter F, Traub RD, Qian F: Comparison of statistical methods to predict the time to complete a series of surgical cases. J Clin Monit 1999; 15:45-51.

3. Macario A, Dexter F: Estimating the duration of a case when the surgeon has not recently performed the procedure at the surgical suite. Anesth Analg 1999; 89:1241-1245.

4. Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times: comparison of the log-normal and normal models. ANESTHESIOLOGY 2000; 92:1160-1167.

5. Strum DP, May JH, Sampson AR, Vargas LG, Spangler WE: Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. ANESTHESIOLOGY 2003; 98:232-240.

6. Strum DP, Sampson AR, May JH, Vargas LG: Surgeon and type of anesthesia predict variability in surgical procedure times. ANESTHESIOLOGY 2000; 92:1454-1167.

7. Dexter F, Epstein RD, Traub RD, Xiao Y: Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. ANESTHESIOLOGY 2004; 101:1444-1453.

8. 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; 90:980-986.

9. Dexter F, Traub RD, Macario A: How to release allocated operating room time to increase efficiency. Predicting which surgical service will have the most under-utilized operating room time. Anesth Analg 2003; 96:507-512.

10. Dexter F, Macario A, Qian F, Traub RD: Forecasting surgical groups' total hours of elective cases for allocation of block time. Application of time series analysis to operating room management. ANESTHESIOLOGY 1999; 91:1501-1508.

11. Zhou J, Dexter F: Method to assist in the scheduling of add-on surgical cases – upper prediction bounds for surgical case durations based on the log normal distribution. ANESTHESIOLOGY 1998; 89:1228-1232.

12. Dexter F, Traub RD, Lebowitz P: Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations. Anesth Analg 2001; 92:943-946.

13. Dexter F, Traub RD: Statistical method for predicting when patients should be ready on the day of surgery. ANESTHESIOLOGY 2000; 93:1107-1114.

14. Dexter F, Abouleish AE, Epstein RH, Whitten CW, Lubarsky DA: Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. Anesth Analg 2003; 97:1119-1126.

15. Dahiya RC, Guttman I: Shortest confidence and prediction intervals for the log-normal. Can J Stat 1982; 10:277-291.

16. Newcombe RG: Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998; 17:857-872.

17. Dayhoff DA, Cromwell J: Measuring differences and similarities in hospital caseloads: a conceptual and empirical analysis. Health Serv Res 1993; 28:293-312.

18. Dexter F, Wachtel RE, Yue JC: Differentiating among pediatric hospitals' operative procedures: Surgery in infants and young children in the State of Iowa. ANESTHESIOLOGY 2003; 99:480-487.

19. Taplin RH: Harmony, statistical inference with the Herfindahl H index and C index. Abacus 2003; 39:82-94.

20. Wachtel RE, Dexter F: Differentiating among hospitals performing physiologically complex operative procedures in the elderly. ANESTHESIOLOGY 2004; 100:1552-1561.

21. Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times: comparison of the log-normal and normal models. ANESTHESIOLOGY 2000; 92:1160-1167.

22. Strum DP, Vargas LG, May JH: Surgical subspecialty block utilization and capacity planning. A minimal cost analysis model. ANESTHESIOLOGY 1999; 90: 1176-1185.

23. Dexter F, Macario A: Changing allocations of operating room time from a system based on historical utilization to one where the aim is to schedule as many surgical cases as possible. Anesth Analg 2002; 94:1272-1279.

24. Dexter F, Macario A, Traub RD, Hopwood M, Lubarsky DA: An operating room scheduling strategy to maximize the use of operating room block time. Computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. Anesth Analg 1999; 89:7-20.

25. 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; 91:262-274.

26. Dexter F, Traub RD, Macario A, Lubarsky DA: Operating room utilization alone is not an accurate metric for the allocation of operating room block time to individual surgeons with low caseloads. ANESTHESIOLOGY 2003; 98:1243-1249.

27. Yue CJ, Clayton MK, Lin FC: A nonparametric estimator of species overlap. Biometrics 2001; 57:743-774.

28. Dexter F, Thompson E: Relative value guide basic units in operating room scheduling to ensure compliance with anesthesia group policies for surgical procedures performed at each anesthetizing location. AANA J 2001; 69:120-123.

29. Dexter F, Macario A, Penning DH, Chung P: Development of an appropriate list of surgical procedures of a specified maximum anesthetic complexity to be performed at a new ambulatory surgery facility. Anesth Analg 2002; 95:78-82.

Table 1. Allocations of Anesthesia Time for non-Operating Room Locations

Week	Team	Monday	Tuesday	Wednesday	Thursday	Friday
1	A	Open	PCC	PCC	ACC/EP × 2	Open
	B	ECT	Int Rad	ECT	Int Rad	ECT
2	A	Open	Open	PCC	ACC/EP × 2	Open
	B	ECT	Int Rad	ECT	Int Rad	ECT

Open refers to first-requested first-scheduled time available for: (1) specialties that requested not to have allocated time (e.g., MRI and CT), (2) specialties with low workloads (e.g., dermatology clinic), and (3) specialties that have filled their allocated time for the two-week period and have another case to schedule. *PCC* refers to pediatric cardiac catheterization suite. *ACC/EP* × 2 refers to the adult cardiac catheterization and electrophysiology suite, staffed by one anesthesiologist medically directing two adjacent rooms with resident physicians and/or certified registered nurse anesthetists. *Int Rad* refers to neurological and non-neurological interventional radiology.

Table 2. Allocations of Anesthesia Time for non-Operating Room Locations

	Pediatric cardiac catheterization	Diagnostic Radiology (CT and MRI)	Interventional Radiology
Number	411	359	292
Most common CPT	18% ± 2%	31% ± 2%	63% ± 3%
Most common 3 CPT	39% ± 2%	44% ± 3%	77% ± 2%
Internal Herfindahl	0.08 ± 0.01	0.12 ± 0.01	0.42 ± 0.03
Percentage of anesthetics that were for procedures each accounting 0.5% or less of anesthetics in 2004	100% ± 0%	100% ± 0%	37% ± 3%
Percentage of anesthetics that were for procedures each accounting 0.1% or less of anesthetics in 2004	60% ± 7%	69% ± 5%	30% ± 4%

CPT stands for Current Procedural Terminology. CT stands for computerized tomography. MRI stands for magnetic resonance imaging.

Table 3 (Appendix). Monte-Carlo Simulation to Test Estimator for Percentage of Cases (Anesthetics) that are Rare and Satisfy the Dichotomous Condition

λ	0.5%	0.2%	0.1%	0.08%	0.05%	0.03%	0.02%	0.01%
Diagnostic Radiology								
Cases ($n\lambda$)	127	51	25	20	13	8	5	2.5
θ – NPMLE $\hat{\theta}$	2.4%	-0.1%	4.8%	-1.2%	2.2%	1.4%	1.7%	0.8%
S.E. of NPMLE	8.8%	2.4%	5.2%	5.5%	4.6%	3.6%	2.8%	1.3%
True S.E. – Asymptotic S.E.	8.0%	-0.5%	2.1%	2.1%	0.5%	-1.0%	-2.1%	-3.9%
True S.E. – Bootstrap S.E.	0.1%	0.0%	-0.2%	-0.1%	0.1%	0.0%	0.0%	0.0%
Interventional Radiology								
Cases ($N\lambda$)	127	51	25	20	13	8	5	2.5
θ – NPMLE $\hat{\theta}$	-0.0%	0.0%	-4.4%	-6.5%	0.1%	2.4%	0.4%	0.7%
S.E. of NPMLE	2.8%	4.4%	4.4%	5.1%	2.8%	2.9%	2.6%	1.3%
True S.E. – Asymptotic S.E.	2.6%	2.6%	4.1%	4.8%	2.5%	2.7%	2.3%	1.0%
True S.E. – Bootstrap S.E.	0.0%	-0.1%	0.4%	0.8%	0.0%	0.1%	0.1%	0.0%
Physiologically Complex Procedures								
Cases ($N\lambda$)	255	102	51	41	25	15	10	5
θ – NPMLE $\hat{\theta}$	-0.9%	0.5%	-0.3%	0.9%	1.1%	0.9%	1.0%	2.0%
S.E. of NPMLE	1.4%	1.1%	0.9%	0.9%	0.8%	0.7%	0.6%	0.5%
True S.E. – Asymptotic S.E.	1.0%	0.7%	0.5%	0.4%	0.3%	0.3%	0.1%	0.1%
True S.E. – Bootstrap S.E.	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

“NPMLE” refers to nonparametric maximum likelihood estimator. “S.E.” refers to standard error. Monte-Carlo simulations were performed with empirical data treated as the population. At the top, the 1 yr of 25,389 anesthetics with 2,357 procedures included 359 anesthetics of 47 procedures involving computerized tomography or magnetic resonance imaging. In the middle, the 25,389 anesthetics of 2,357 procedures included 292 anesthetics of 30 procedures involving interventional radiology. At the bottom, the 3 yr of 50,982 operating room cases with 9,466 procedure(s) included 10,719 cases of 2,126 procedure(s) for which at least one of the procedures was physiologically complex[18,20,28,29] (i.e., 8 or more American Society of Anesthesiologists’ basic units). The “S.E. of NPMLE” is the standard deviation of the 1,000 bootstrap estimates of the NPMLE. The “Asymptotic S.E.” is the mean of the 1,000 calculations using the formula in the Appendix. For each of the 1,000 simulations, 1,000 bootstraps were taken. The “bootstrap S.E.” is the mean of the 1,000 standard deviations of the bootstrap estimates.

Table 4 (Appendix). Lack of Sensitivity of Results to Misclassification of the Dichotomous Condition

λ	0.5%	0.2%	0.1%	0.08%	0.05%	0.02%	0.01%
Cases ($n\lambda$)	255	102	51	41	25	15	5
θ − NPMLE $\hat{\theta}$ with 1% misclassification	-0.1%	-0.0%	-0.1%	-0.1%	-0.2%	-0.1%	-0.1%
θ − NPMLE $\hat{\theta}$ with 3% misclassification	-0.2%	-0.6%	-0.2%	-0.3%	-0.3%	-0.2%	-0.1%
θ − NPMLE $\hat{\theta}$ with 5% misclassification	-0.8%	-1.3%	-0.6%	-0.8%	-0.8%	-0.3%	-0.2%

NPMLE refers to nonparametric maximum likelihood estimator. First, 1000 Monte-Carlo simulations were performed using the large (operating room) dataset (Table 3) as the population. Then, additional 1000 simulations were performed with a binomial random variable (with probability 1%, 3%, or 5%) generated for each procedure. When true, the physiological complexity of the procedure was reversed. The bias is negative, because more than half of procedures, specifically 79%, were not physiologically complex.

Figure Legends

1. Histogram of the natural logarithm of the actual anesthetic times for computerized tomography to the expert's estimate (N = 104). The superimposed normal curve has a mean of –0.01 and standard deviation 0.36. The distribution was close to that of a log-normal (P = 0.05).
2. Histogram of the natural logarithm of the actual anesthetic times for magnetic resonance imaging to the expert's estimate (N = 155). The superimposed normal curve has a mean of -0.02 and standard deviation 0.28. The distribution was close to that of a log-normal (P = 0.69).
3. Histogram of the differences in time between the actual versus scheduled ends of the day for interventional radiology (N = 57 days). The superimposed normal curve has a mean of 0.75 hr and standard deviation 1.45 hr. The distribution was close to normal (P = 0.42).