Predicting Anesthesia Times for Diagnostic and Interventional Radiological Procedures

Franklin Dexter, MD, PhD, Jack C. Yue, PhD, and Angella J. Dow, BS

Division of Management Consulting, Departments of Anesthesia and Health Management & Policy, University of Iowa, Iowa City, Iowa; Department of Statistics, National Chengchi University, Taipei, Taiwan

We studied anesthesia times for diagnostic and interventional radiology using anesthesia billing data and paper radiology logbooks. For computerized tomography and magnetic resonance imaging procedures, we tried to predict future anesthesia times by using historical anesthesia times classified by Current Procedural Terminology (CPT) codes. By this method, anesthesia times were estimated even less accurately than operating room cases. Computerized tomography and magnetic resonance imaging had many different CPT codes, most rare, and CPT codes reflected organs imaged, not scanning times. However, when, anesthesia times were estimated by expert judgment, face validity and accuracy were good. Lower and upper prediction bounds were also estimated from the expert estimates. For interventional radiology, predicting anesthesia times was challenging because few CPT codes accounted for most cases. Because interventional radiologists scheduled their elective cases into allocated time, the necessary goal was not to estimate the time to complete each case but rather the time to complete each day's entire series of elective cases including turnover times. We determined the time of day (e.g., 4 PM) up to when interventional radiology could schedule so that on 80% of days the anesthesia team finishes no later than a specified time (e.g., 6 PM). Both diagnostic and interventional radiology results were similarly less accurate when Version 9 of the International Classifications of Diseases' procedure codes was used instead of CPT.

(Anesth Analg 2006;102:1491–500)

The anesthesia department studied in this paper faced the common situation of increasing numbers of cases performed outside of operating rooms (ORs), including diagnostic and interventional radiology. An organizational decision was made to improve case scheduling for these anesthetics. Anesthesia time was allocated to interventional (neurological and peripheral) radiology. A full-time non-OR scheduler was hired by the hospital. A vice chair in the department became actively involved in non-OR scheduling. Together, they shepherded the process of working with non-OR sites as they chose days of the week for their allocations. During a transition period, three non-OR teams were used to eliminate a 2-mo backlog. Then, as forecast based on expected workload, two teams were used each workday. With multiple meetings and support of the hospital, implementation was relatively uneventful. However, there was limited evidence of improved efficiency of use (1,2) of the anesthesia providers.

In this paper, we describe how we then diagnosed an unrecognized problem: estimates of anesthesia times for diagnostic and interventional radiology were strikingly inaccurate. We developed, validated, and implemented solutions. In the Discussion, we explain under what circumstances our solutions can be useful. The solutions described are novel, as they are not the same as those reached in prior studies of OR case scheduling (2). Specifically, there is poor performance in averaging (3,4) historical anesthesia times classified by the Current Procedural Terminology (CPT) or International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) procedure codes.

Accepted for publication December 6, 2005.

An abstract describing this work was presented on October 26, 2005 at the American Society of Anesthesiologists' meeting, Atlanta, Georgia. An abstract was presented on November 13, 2005 at the INFORMS meeting, San Francisco, California.

Financial disclosure: Franklin Dexter is Director of the Division of Management Consulting, a Division of the Department of Anesthesia. He receives no funds personally other than his salary from the State of Iowa, including no travel expenses or honorarium, and has tenure with no incentive program.

Address correspondence to Franklin Dexter, MD, PhD, Division of Management Consulting, Department of Anesthesia, University of Iowa, Iowa City, Iowa 52242. Address e-mail to Franklin-Dexter@UIowa.edu.

DOI: 10.1213/01.ane.0000202397.90361.1b

0003-2999/06

Methods

Diagnostic and Interventional Radiology - Quantifying Inaccuracy in Anesthetic Times

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

For computerized tomography (CT) and magnetic resonance imaging (MRI), 3 yr of billing data were available: September 1, 2001 through August 31, 2004. Almost all cases were pediatric. The 566 cases studied were those of CPT code(s) performed a moderate to large number of times ($N \geq 30$) (5,6). The observed sample mean for each CPT(s) was considered the scheduled duration.

For comparison, the hospital's OR cases from the same period were considered. The cases' scheduled and actual OR times were obtained from the hospital's OR information system. Sorting the cases in ascending order of OR times, the first 41,770 cases provided the same overall mean duration (1.97 h) as that of the studied MRI and CT cases.

For interventional radiology, cases were classified by a combination of CPT and physician (3,4,7). Data were available, including that for the most active physician, from July 1, 2003. The same final date of August 31, 2004 was used. These 373 cases had a mean anesthesia time of 3.89 h. All 58,291 OR cases had a mean duration of 2.75 h. Therefore, the OR cases were sorted in descending sequence. The first 34,046 cases provided the same mean of 3.89 h as the 373 interventional radiology cases.

Diagnostic Radiology - Anesthetic Times for CT

Radiology schedulers arranged for CT and MRI with anesthesia by scheduling the cases into anesthesia time allocated to many specialties. As these departments chose to share anesthesia time, the calculated weekly total (8) was large enough for the schedulers to offer patients the flexibility of many possible start times on multiple days of the week. As each CT or MRI case was scheduled separately, the goal in estimating the anesthesia times for diagnostic radiology was to estimate the times for individual cases.

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 typical CT scanner time required when such a case is performed with anesthesia. Her estimates plus 30 min were converted into a presentation-format rule by the authors (Table 1). The period of 30 min had been used by the anesthesia schedulers for years as their (expert) estimate of the time anesthesia providers take for anesthesia induction, emergence, and transport to the postanesthesia care unit. Using this process, the expert estimates were obtained independent of the actual times (i.e., no data fitting was performed).

Table 1. Estimates for Diagnostic Radiology Created by Expert Judgment

Computerized Tomography (CT)
Expert estimate of median time
- 30 min for anesthesia preparation, induction, emergence, and transport
- + 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

Lower 5% prediction bound = 55% of the expert's estimate
Upper 90% prediction bound = 160% of the expert's estimate

Magnetic Resonance Imaging (MRI)
Expert estimate of median time
- 45 min for anesthesia preparation induction, emergence, and transport
- + 45 min scanning time
- + 30 min scanning time if total spine or abdomen/pelvis extending into the thighs
- + 15 min scanning time if not using the Siemens Avanto MRI.

Lower 5% prediction bound = 63% of the expert's estimate
Upper 90% prediction bound = 143% of the expert's estimate

Expert estimates were created by chief technologists at the study hospital. The current study establishes the usefulness and validity of the *process* of using expert judgment, as well as the format of the prediction (e.g., fixed time plus increments). Each facility should follow the same process, replacing the numbers with values appropriate to its site. Then, to calculate lower and upper prediction bounds, review anesthesia times for historical cases, as obtained from anesthesia billing data. For each anesthetic, calculate the expert's estimate for the case. Compare the actual time to the expert's estimate. Apply equation (1) to obtain the lower and upper prediction bounds. An appropriate sample size is $N \cong 100$ previous anesthetics. As the prediction bounds are based on the expert's estimate, they apply provided the scanner to be used is known when the anesthetic is scheduled. Otherwise, the lower bounds should be reduced, and the upper bounds increased, by the differences in times among scanners.

To assess the accuracy of estimates when in practical use, 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 104 anesthetics for validation.

Residuals between actual anesthesia times and expert estimates were analyzed to determine if the expert estimates had any systematic bias. Each of the 10 body regions scheduled for scanning was coded as 0 or 1 for each patient. Use of each of the 4 CT machines was coded as 0 or 1. Mann-Whitney *U*-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 comparisons, intentionally (i.e., by design) the analysis was likely to detect a significant relationship even when detection may have been spurious due to random error. Essentially, this analysis used time-motion study data (i.e., the logbooks) to determine if the expert's estimates could be improved.

Residuals were compared to those of the original scheduled times to assure that the expert estimates were as accurate as the originally scheduled anesthesia times. If true, then that finding would suggest that schedulers did not have *a priori* knowledge about individual patients that improved prediction of anesthesia time but which expert technicians did not use. Although the expert technicians used only the body parts scanned, the schedulers could ask questions about the patient that may affect time in the suite (e.g., whether the patient is on a ventilator).

The allocation of the anesthesia time planned daily for all diagnostic radiology and other diagnostic procedures (i.e., anesthesia staffing and resource use) was calculated by forecasting the total hours of anesthesia time for such cases, including turnovers (9,10). Thus, when deciding whether to schedule a case into the allocated time, the estimate of the anesthesia time should be an unbiased estimator of the contribution of the case to the total time (2,3,11). This assumption that the bias of the expert estimate was negligibly different from 0 min was tested.

Upper (90%) and lower (5%) prediction bounds were also estimated (11). The upper prediction bound (12,13) is the end-point relevant to inserting a case into a gap in the schedule (e.g., CT is scheduled 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 will 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. For example, there is a 10% chance that the anesthetic will take more time than its 90% upper prediction bound. The lower prediction bound is the end-point relevant to planning the availability of the next patient to receive care by the same team (11,14). The anesthesia scheduler adjusts the fasting and arrival times of the subsequent child using the calculated lower 5% prediction bounds (14).

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

Assume that the AT_i, $i = 1, 2, \ldots, N$, 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 (log) of AT_i and let c_i refer to the ***case*** effect specifying variation of the natural logarithm of AT_i around that mean (15). By definition, c_i follows a normal distribution with mean of 0 and unknown variance of σ (2). Thus, $AT_i = \exp(\mu + c_i) = \exp(\mu) \times \exp(c_i) = Z_i \times \exp(c_i)$. Therefore, we use Lilliefors' test to evaluate 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 and upper 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}) (16). 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

$$Z_{i+1} \cdot \exp\left(s \cdot \sqrt{1 + 1/N} \cdot T^{-1}[N-1, \alpha] \right)$$

where T^{-1} $[N - 1, \alpha]$ is the inverse of the Student t cumulative distribution function with $(N - 1)$ degrees of freedom. For example, suppose that $s = 0.36$ and $n = 104$, as for CT. Then, the 5% lower prediction bound equals 55% of the expert estimate, where $55\% = 100 \cdot \exp(0.36 \cdot \sqrt{1 + 1/104} \cdot T^{-1}[104 - 1, 0.05])$.

Diagnostic Radiology - Anesthetic Times for MRI

All of the above steps for CT were repeated for MRI. There were 17 body regions and 3 scanners studied. Paper logbooks were available back to December 1, 2003. The 1.2 yr of data provided 154 anesthetics for review.

Interventional Radiology - Anesthetic Times

Interventional radiology scheduled its outpatient and same-day-admit cases into anesthesia time lasting the entire workday and allocated solely to interventional radiology (8). Open access was provided within a reasonable (2-wk) period (8,17). They were allocated 2 days a week, Tuesdays and Thursdays. If they had a new elective case to be scheduled and all allocated time was filled for the next 2 wk, then additional time was always made available on another day of the week for the extra case (8,9). That way, the new case could be performed within 2 wk, but not by extending the duration of the staffed workday. The corresponding goal in estimating anesthesia times for use in anesthesia scheduling was the determination of whether a series of cases including turnover times would fill the staffed workday (14). We calculated the latest time (e.g., 4 PM) up to when interventional radiology scheduled its elective cases so that the anesthesia providers would be done reliably no later than a specified time (6:00 PM). The latter is the time that the anesthesia department plans for the end of the workday.

By 2004, series of scheduled (elective) interventional radiology cases were frequently being performed. 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. From anesthesia billing data, the scheduled end of the last elective case of each workday was compared to the actual end of the anesthetic workday. All delays, turnovers, etc. were included. The difference was calculated between the actual versus scheduled ends of the anesthetic workday. A normal curve was fit to the differences, and tested for appropriateness of fit using Lilliefors' test.

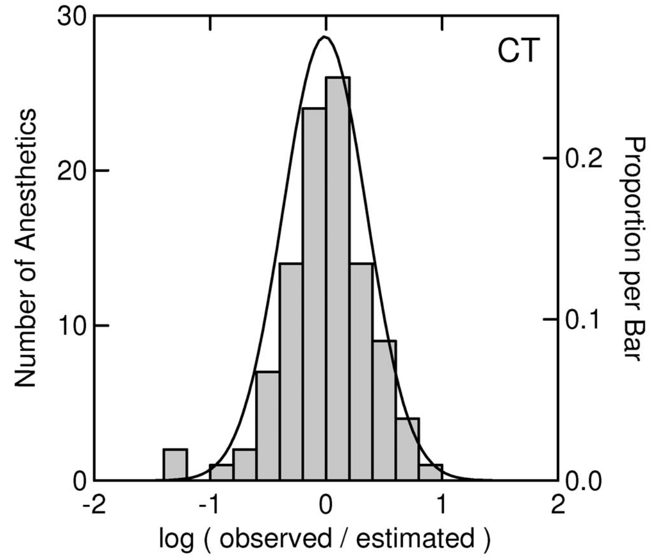

Figure 1. Histogram of the natural logarithm of the ratio of the actual anesthetic times for computerized tomography to the expert's estimate (n = 104). The superimposed normal curve has a mean of −0.01 and standard deviation 0.36. The distribution was close to that of a log-normal (P = 0.05).

Results

Diagnostic Radiology

For CT and MRI, the original mean absolute percentage error was 45% ± 1% (SE), less accurate than the 27% ± 1% for OR cases of comparable durations. The Appendix shows that this finding is a result of a fundamental characteristic of CPT and ICD-9-CM and thus is unlikely to be unique to the studied hospital. The inaccuracy resulted from CPT and ICD-9-CM being based on the organs imaged. The diversity of CPTs among cases was so large that there were few historical anesthetic times of the same CPT(s) to use for any one newly scheduled case. The ICD-9-CMs pooled organs with different imaging times. To avoid these problems, new estimates were created based on expert judgment, ignoring CPT and ICD-9-CM (Table 1). The mean absolute percentage error was 26% ± 2%, matching the above 27% for OR cases.

Analysis of residuals for CT did not detect additional error explainable by body region(s) or the scanner used (all $P > 0.10$). The bias of the estimate was negligible (2 ± 2 min). The estimate was for the median ($P = 0.84$ sign test, $P = 0.50$ Wilcoxon signed-ranks test). The pairwise reduction in the mean absolute error was 11 ± 2 min versus that originally scheduled.

Lower and upper prediction bounds were calculated for CT (Table 1). Figure 1 shows the histogram of the natural logarithm of the ratio of the actual anesthetic times to the estimate, with a superimposed normal curve (mean −0.01, standard deviation 0.36, n = 104). The distribution was close to log-normal with a common proportional variance (Lilliefors $P = 0.05$). Strum et al. (5) showed that and why goodness of fit tests tend to falsely reject null hypotheses of good fits for anesthetic time data. $P = 0.05$ is acceptable when compared with the larger effect of rounding (5,13). For example, from Table 1, the upper bound for CT of the brain equaled 64 min, where 64 min equals 1.60 multiplied by the expert estimate of 40 min. The time of 64 min could be rounded down to 1 h or up to 1 h 15 min.

Analysis of residuals for MRI did not detect additional error explainable by body region or the scanner used (all $P > 0.10$). The bias of the estimate was negligible (2 ± 2 min). The estimate was for the median ($P = 0.68$ sign test, $P = 0.64$ Wilcoxon signed-ranks test). The pairwise reduction in the mean absolute error was 12 ± 2 min versus that originally scheduled.

Lower and upper prediction bounds were calculated for MRI (Table 1). Figure 2 shows a histogram of the natural logarithm of the ratio of the actual anesthetic times to the estimate, with a superimposed normal curve (mean −0.02, standard deviation 0.28, n = 154). The distribution was close to log-normal with a common proportional variance (Lilliefors $P = 0.69$).

Additional data were available for MRI regarding how to further improve the accuracy of estimates for anesthesia times. 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 (Table 1). The mean absolute difference of the actual anesthesia-controlled time from the expert estimate equaled 19 ± 1 min. Thus, accuracy may be improved by knowing and applying, at the time of case scheduling, each patient's physiological condition for anesthesia, the specific anesthesia providers, etc. Nevertheless, this inaccuracy was only slightly more than the 15 ± 1 min mean absolute

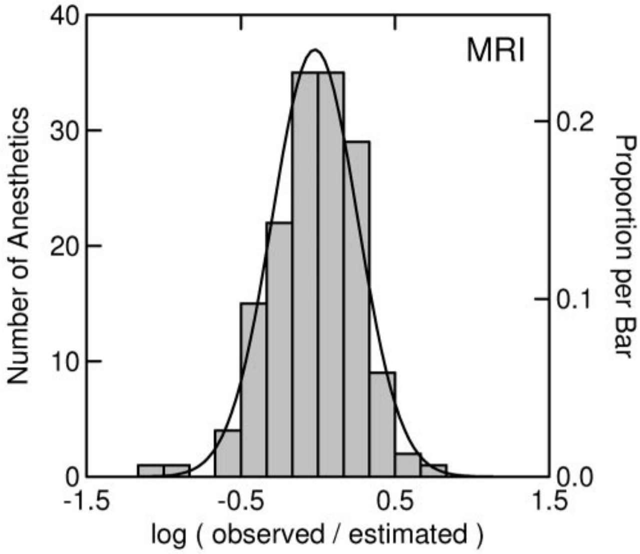

Figure 2. Histogram of the natural logarithm of the ratio of the actual anesthetic times for magnetic resonance imaging to the expert's estimate ($n = 154$). The superimposed normal curve has a mean of −0.02 and standard deviation 0.28. The distribution was close to that of a log-normal ($P = 0.69$).

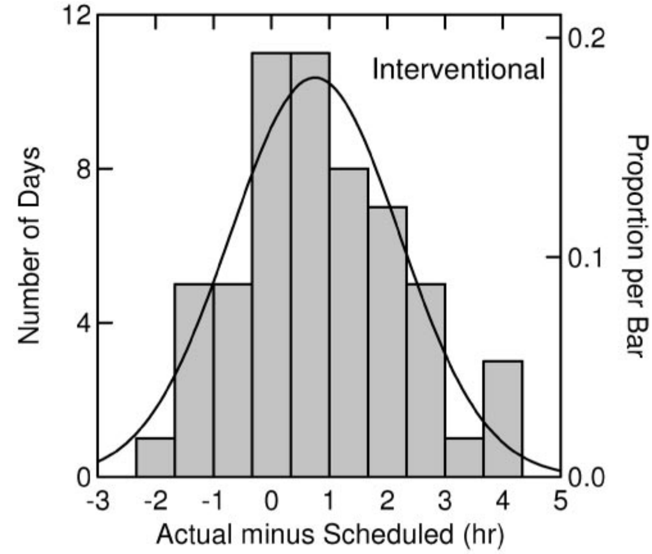

Figure 3. Histogram of the differences in time between the actual versus scheduled ends of the day for interventional radiology ($n = 57$ days). The superimposed normal curve has a mean of 0.75 h and standard deviation 1.45 h. The distribution was close to normal ($P = 0.42$).

difference of the actual scanning time and the expert estimate of the scanning time. Thus, accuracy may be improved by knowing and applying, at the time of case scheduling, each patient's primary disease for choosing the MRI scanning protocol. Neither the anesthetic factors nor primary disease for choosing the MRI scanning protocol was available at the study hospital when the anesthetic was scheduled.

Interventional Radiology

Using combinations of CPT code(s) and radiologist, the mean absolute percentage error between actual and scheduled anesthesia times was 24%± 1%, identical to the percentage error of 24%± 0% for OR times of cases of comparable durations. Results in the Appendix explain why durations of interventional radiology cases are not estimated more accurately, and that results are unlikely to be unique to the studied hospital. The most common single CPT and corresponding ICD-9-CM accounted for more than 2/3rd of anesthetics despite having widely different anesthetic times. Thus, precise estimates of anesthesia times are unlikely to be attainable by relying on CPTs or ICD-9-CMs.

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 (actual minus scheduled, mean 0.75 h, standard deviation 1.45 h). As for series of OR cases (3), the distribution was close to normal ($n = 57$, $P = 0.42$).

The anesthesia department considered the end of the workday to be 6:00 PM. Relief was rarely available for anesthesia providers working in interventional radiology, matching reports from other hospitals (18). Thus, the relative cost of an hour of overutilized anesthesia time was considered expensive: fourfold more than the cost of an hour of staffed time. Thus, for 4 of 5 workdays (i.e., 80%) (1), the providers should finish early. The latest time that interventional radiology should schedule cases was chosen so that the last anesthetic would 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 × (standard deviation), the 80% quantile equals the mean + 0.84 × (standard deviation). Substituting the mean and standard deviation from the preceding paragraph, the 80% quantile for the difference between actual and scheduled end of anesthesia equaled 2.0 h, where 2.0 h = 0.75 h + 0.84 × 1.45 h. In addition, the observed 80th percentile for the 57 workdays was 2.0 h. Thus, the studied interventional radiology department should schedule cases with anesthesia to end by 4:00 PM. This time creates a balance between ending early and ending late that is consistent with the expense of extending the workday beyond 6:00 PM.

Discussion

This paper validated a process for scheduling CT, MRI, and interventional radiology non-OR cases accurately. For anesthetics performed in CT or MRI suites, each facility should use its radiology technologists and anesthesia billing data to replace the numbers in Table 1 with its own values. For interventional radiology suites, each facility should compare its actual versus scheduled ending times for the last elective case of the workday, as performed in Results. For routine (nonscientific) work, each facility need not repeat the statistical development or validation. The use of CPT or ICD-9-CM need not be evaluated either, as our Appendix shows that the finding of a lack of benefit is likely to be valid elsewhere. The impact of our process on anesthesia staffing and resources was considered in the sixth paragraph of Methods, "Diagnostic Radiology - Anesthetic Times for CT" and first paragraph of the Methods, "Interventional Radiology - Anesthetic Times."

Diagnostic Radiology

We validated a process for choosing the anesthesia times of elective, scheduled CT and MRI. For example, the numbers in Table 1 would not apply to a CT at 2 AM in a combative trauma patient at the studied hospital. We did not test the validity of the process for such cases.

Our process for choosing anesthesia times for diagnostic radiology is useful for other facilities without allocated time exclusively for CT or MRI. At the studied hospital, CT, MRI, echocardiography, etc., did not each have allocated time, in part because the separate sites had different technicians. Partly, this was because these sites were scheduled like primary care clinics. Regardless of why, the implication was that the anesthesia time for each case needed to be estimated.

Our process would be useful for facilities at which elective studies on inpatients are performed in evening and night hours. In that circumstance, the anesthesia time for each CT or MRI would be scheduled individually.

Our process is useful for facilities at which anesthesia providers do not effectively have a fixed end of the staffed workday for diagnostic radiology. For example, this would apply if all of the week's elective anesthetics for MRI were performed sequentially on Mondays. Such scheduling would match how OR cases are scheduled at most surgical suites (1,2,11). Such scheduling can be practical for diagnostic radiology because how many diagnostic procedures are performed each workday can be insensitive to the schedules of a few physicians (as below). Such scheduling can also be practical if enough CT and/or MRI are scheduled sequentially to warrant an entire workday of an anesthesia provider but there are not enough consistently scheduled for concern of overutilized time. For example, there may invariably be more than 7 h of cases, but virtually never more than 10 h, with the end of the workday determined by a lack of patients needing an anesthesia provider for diagnostic radiology that week.

Our process would be valid, but less useful, if CT and MRI were scheduled like the interventional radiology cases.

Lower prediction bounds were useful for CT and MRI patients (11) because the pediatric clinics were concerned about children fasting for diagnostic procedures in afternoons. The anesthesia department used national recommendations for fasting (19). However, nurses worried about the scheduled start times changing (e.g., because the preceding case finished early). This matched a survey showing 64% of surgical nurses reported that the reason why fasting periods were longer than guidelines was uncertainty in start times (20).

Upper prediction bounds were useful for CT and MRI, showing when a case can fit into a hole in the schedule (11). Holes in the schedule result from cancellations and changes to start times. That differs from ORs and interventional radiology, where patients can routinely be told their start times the day before surgery, such that holes in the schedule are filled by moving later cases to earlier times.

The process that we validated for CT and MRI relied on expert judgment to provide the average anesthesia time and use of historical anesthesia times from billing data to estimate the proportional variation in the estimate (Table 1). Results for predicting OR times are similar (21). Provided experts (i.e., surgeons and schedulers) do not have an incentive to underestimate OR times (22), their estimate of the average OR time is nearly as accurate as that from historical data. However, historical data are still needed. For many OR management decisions, what matters is the uncertainty in the estimate of OR time (11). The historical data are used to estimate the proportional variation in anesthesia times and OR times (e.g., prediction bounds).

Interventional Radiology

The process we validated for interventional radiology is useful for a series of elective, scheduled cases. Our process is unlikely to be useful for single add-on cases. Our process is also unlikely to be useful for anesthesia departments that provide open access on the requested day for a nearly unlimited number of cases. At such facilities, the expectation is that anesthesia providers should work late for any radiological procedure. Nevertheless, we expect that our process will be useful for many, if not most, anesthesia departments, for three reasons.

First, consider a hospital with interventional radiology cases with anesthesia totaling less than one workday's worth every other week (e.g., 6 hours' total anesthesia time every other week). Then, our process would be valid but not useful. Nevertheless, such hospitals are uncommon, as many hospitals are experiencing increasing workloads of interventional radiology. Fewer than 5% of the members of the Neuroanesthesia Society of Great Britain and Ireland report having just 1 day per week of cases (18), and that is just for neurological radiology not including peripheral radiology procedures.

Second, radiology sites are not as interchangeable as ORs. When allocated time is full for the day and a service wants to schedule another case, the allocated time of another service on the same day cannot be released safely. For example, an interventional radiologist cannot split and perform his other case inside an MRI machine. Consequently, if open access were provided on the workday of choice of a service (unlike in our process), then a new case would invariably be performed in overutilized anesthesia time (e.g., after 6:00 PM) rather than in the time originally allocated to another service that is likely not to use it (i.e., before 6:00 PM). The anesthesia department studied did not find this to be practical nor did the hospital, and we suspect that this applies elsewhere as well.

Third, 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 (23,24). In contrast, non-OR services (e.g., interventional radiology) are often small, representing one or two physicians. Variation in workload on any one workday is larger, reflecting vacations and meetings (8). Thus, providing open access on the workday of choice of the non-OR service would result in substantial hours of underutilized and overutilized anesthesia time. The efficiency of use of allocated (1) radiology time would be substantially less than that of OR time. As such, anesthesia labor costs would be disproportionately high if our process were not followed.

Together, interventional radiology cases classified by CPT codes and radiologist were no more or less accurate than OR cases of corresponding durations. However, the consequence of that inaccuracy was to produce larger inefficiency of use (1) of anesthesia providers than for ORs, because of reduced flexibility in being able to move cases, assign different providers, etc. The result is that we expect that providing open access to anesthesia time on any workday (1,2,11) will be impractical for many other anesthesia departments. Instead, open access will be provided to assure that a case is performed within a few weeks (8,9,17). The usefulness of our process for anesthesia departments working in interventional radiology suites relies on this supposition.

Many factors affect the anesthesia times for each interventional radiology case (e.g., availability of equipment and the specific procedure). However, as explained in the first paragraph of Methods, "Interventional Radiology - Anesthetic Times," these issues were not considered because they have little to no effect on the anesthesia department's staffing, staff scheduling, or staff assignment. Our focus was the anesthesia department.

Conclusion

This is the first article to describe and validate a process for scheduling CT, MRI, and interventional radiology cases accurately. Anesthesia times for diagnostic radiological procedures (e.g., MRI) can be based on systematic rules from the facility's experts. On the other hand, a series of interventional radiological procedures performed on the same day can be scheduled consecutively up to a calculated time (e.g., 4 PM), assuring that the anesthesia team will usually be done by the end of the workday (e.g., 6 PM). The usefulness of the implemented methods is dependent on local factors (e.g., total hours of non-OR anesthesia cases per week and whether the duration of the anesthesia workday is limited when cases are scheduled).

Appendix – Quantifying Diversity of Radiological Procedures With Anesthesia

The Appendix shows that the above results are a consequence of fundamental characteristics of the CPT and ICD-9-CM coding systems. These results are important, as they suggest that our findings apply elsewhere.

Methods

One simple, but limited, method we used to quantify the diversity of procedures was the percentage of anesthetics accounted for by the most common procedure(s). Standard errors for the percentages were estimated by using the Clopper-Pearson confidence intervals (25).

A second method of quantifying the diversity of the procedures performed at each non-OR site was use of the internal Herfindahl index (26,27). 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 (28).

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

Table 2 (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%
SE of NPMLE	8.8%	2.4%	5.2%	5.5%	4.6%	3.6%	2.8%	1.3%
True SE – Asymptotic SE	8.0%	−0.5%	2.1%	2.1%	0.5%	−1.0%	−2.1%	−3.9%
True SE – Bootstrap SE	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%
SE of NPMLE	2.8%	4.4%	4.4%	5.1%	2.8%	2.9%	2.6%	1.3%
True SE – Asymptotic SE	2.6%	2.6%	4.1%	4.8%	2.5%	2.7%	2.3%	1.0%
True se – Bootstrap SE	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%
SE of NPMLE	1.4%	1.1%	0.9%	0.9%	0.8%	0.7%	0.6%	0.5%
True SE – Asymptotic SE	1.0%	0.7%	0.5%	0.4%	0.3%	0.3%	0.1%	0.1%
True SE – Bootstrap SE	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

"NPMLE" refers to nonparametric maximum likelihood estimator. SE 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 diagnostic radiology (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 (27,29,31,32) (i.e., 8 or more American Society of Anesthesiologists' basic units). The SE of NPMLE is the standard deviation of the 1000 bootstrap estimates of the NPMLE. The Asymptotic SE is the mean of the 1000 calculations using the formula in the Appendix. For each of the 1000 simulations, 1000 bootstraps were taken. The bootstrap SE is the mean of the 1000 standard deviations of the bootstrap estimates.

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's anesthetics were 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.

The third method of quantifying diversity was use of the percentage of anesthetics at different sites that were of rare procedures (29). A procedure(s) was considered rare if it accounted for 0.5% or less, or 0.1% or less, of anesthetics performed anywhere by the department during a 1-yr period. Let each case be of 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 $1, 2, \ldots, k$ be the ones that are rare and that satisfy the dichotomous condition ($k \leq C \leq S$). Also, without loss of generality, assume $p_1 \leq p_2 \leq \ldots \leq p_C$. Then, the statistic $\theta = \Sigma_1^k p_i / \Sigma_1^C p_i$, where $p_k \leq \lambda < p_{k+1}$. The corresponding nonparametric maximum likelihood estimator (NPMLE) $\hat{\theta} = \Sigma_1^{\hat{k}} x_i / \Sigma_1^C x_i$, where x_i is the observed number of cases of procedure i, $n = \Sigma_1^s x_i$. The k are estimated via $\Sigma_1^{\hat{k}} x_i \leq n\lambda < \Sigma_1^{\hat{k}+1} x_i$. Following the proof in Yue et al. (30), $\hat{\theta}$ is asymptotically normally distributed with standard error $\sqrt{\hat{\theta}(1-\hat{\theta})/(\Sigma_1^C x_i)}$.

To test these equations, the diagnostic and interventional radiology data were treated as population information in Monte Carlo simulations. To examine sensitivity to sample size, a larger sample was used: 3 yr of OR data, with the dichotomous condition being the physiological complexity of the procedure(s) of each case (27,29,31,32).

The NPMLE $\hat{\theta}$ 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 2). The bootstrap standard errors, but not the asymptotic ones, were accurate to within 1%. Reasons were multifactorial (not shown): $\hat{\theta} \cong 0$, $\hat{\theta} \cong 1$, $k \neq \hat{k}$, and non-smooth behavior near $p_i \cong \lambda$. Thus, below we report $\hat{\theta}$ with bootstrap standard errors. There were a few procedures for which it was equivocal as to whether it satisfied the dichotomous category (e.g., a procedure performed once in interventional radiology but often in ORs). This was unimportant, as even an

Table 3 (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 2) 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.

unbelievably high 5% misclassification rate affected $\hat{\theta}$ by just 1% (Table 3).

Results

All anesthetics performed in the anesthesia department in 2004 were used.

Pediatric cardiac catheterization was used as a control (n = 411). Before considering radiology, we wanted to study a subspecialty that we knew would have many different CPT and ICD-9-CM procedure codes because there are many different congenital cardiac lesions. The most common CPT codes accounted for 18% ± 2% of anesthetics, and the most common 3 CPTs accounted for 39% ± 2% of anesthetics. The internal Herfinahl was 0.08 ± 0.01, which is very low (27), as expected from Spangler et al. (33). Furthermore, most procedures were rare. Among anesthetics performed in the pediatric cardiac catheterization laboratory, 100% ± 0% (60% ± 7%) were for procedures that each accounted for 0.5% (0.1%) or less of all anesthetics performed by the department in 2004 at any location. Thus, as expected, pediatric cardiac catheterization had a larger diversity of procedures, and a higher frequency of rare procedures, than we previously reported (27,29) for surgery.

For MRI and CT (n = 359), the most common CPT (70553, MRI brain without contrast) accounted for 31% ± 2% of anesthetics for MRI and CT, and the most common 3 CPT accounted for 44% ± 3% of anesthetics for MRI and CT. The internal Herfindahl was 0.12 ± 0.01. Among anesthetics for MRI or CT, 100% ± 0% (69% ± 5%) were for procedures accounting for 0.5% (0.1%) or less of anesthetics of all types. The diversity of procedures was similar to that of pediatric cardiac catheterization, explaining why estimates of diagnostic radiology anesthesia times using CPT(s) were relatively inaccurate.

Another reason for inaccurate anesthesia times was that the codes pooled organs with different imaging times. For example, the ICD-9-CM 88.38 included both CT of the sinuses and pelvis, differing by 10 min in the expert estimates (Table 1). The ICD-9-CM 88.97 included both MRI of the abdomen and orbit, differing by 15 min (Table 1). Neither considered the substantive differences in scanning times among machines (Table 1).

For interventional radiology (n = 292), the most common CPT accounted for 63% ± 3% of anesthetics, and the most common 3 CPT accounted for 77% ± 2% of anesthetics. The internal Herfindahl was 0.42 ± 0.03. Among cases with anesthesia, 37% ± 3% (30% ± 4%) were for procedures accounting for 0.5% (0.1%) or less of anesthetics. In fact, the 3 combinations of CPT and radiologist with N ≥ 30 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."

References

1. Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. Anesthesiology 1999;90:1176–85.
2. Dexter F, Traub RD. How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesth Analg 2002;94:933–42
3. 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
4. Macario A, Dexter F. Estimating the duration of a case when the surgeon has not recently performed the procedure at the surgical suite. Anesth Analg 1999;89:1241–5.
5. Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: comparison of the log-normal and normal models. Anesthesiology 2000;92:1160–7.
6. Strum DP, May JH, Sampson AR, et al. Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. Anesthesiology 2003;98:232–40.

7. Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. Anesthesiology 2000;92:1454–67.
8. Dexter F, Macario A, Traub RD, et al. 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.
9. 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;90:980–6.
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;91:1501–8
11. Dexter F, Epstein RD, Traub RD, Xiao Y. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. Anesthesiology 2004;101:1444–53.
12. Zhou J, Dexter F. Method to assist in the scheduling of add-on surgical cases: upper prediction bounds for surgical case durations based on the log normal distribution. Anesthesiology 1998;89:1228–32.
13. Dexter F, Traub RD, Lebowitz P. Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations. Anesth Analg 2001;92:943–6.
14. Dexter F, Traub RD. Statistical method for predicting when patients should be ready on the day of surgery. Anesthesiology 2000;93:1107–14.
15. Dexter F, Abouleish AE, Epstein RH, et al. Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. Anesth Analg 2003;97:1119–26.
16. Dahiya RC, Guttman I. Shortest confidence and prediction intervals for the log-normal. Can J Stat 1982;10:277–91.
17. 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;94:1272–9.
18. Willoughby L, Morgan R. Neuroanaesthetists' experience of workload-related issues and long-duration cases. Anaesthesia 2005;60:151–4.
19. Ferrari LR, Rooney FM, Rockoff MA. Preoperative fasting practices in pediatrics. Anesthesiology 1999;90:978–80.
20. O'Callaghan N. Pre-operative fasting. Nurs Stand 2002;16(36):33–7.
21. Dexter F, Ledolter J. Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historical data. Anesthesiology 2005;103:1259–67.
22. Dexter F, Macario A, Epstein RH, Ledolter J. Validity and usefulness of a method to monitor surgical services' average bias in scheduled case durations. Can J Anaesth 2005;52:935–9.
23. Dexter F, Macario A, Lubarsky DA, Burns DD. Statistical method to evaluate management strategies to decrease variability in operating room utilization. Application of linear statistical modeling and Monte-Carlo simulation to operating room management. Anesthesiology 1999;91:262–74.
24. 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;98:1243–9.
25. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998;17:857–72.
26. Dayhoff DA, Cromwell J. Measuring differences and similarities in hospital caseloads: a conceptual and empirical analysis. Health Serv Res 1993;28:293–312.
27. 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;99:480–7.
28. Taplin RH. Harmony, statistical inference with the Herfindahl H index and C index. Abacus 2003;39:82–94.
29. Wachtel RE, Dexter F. Differentiating among hospitals performing physiologically complex operative procedures in the elderly. Anesthesiology 2004;100:1552–61.
30. Yue CJ, Clayton MK, Lin FC. A nonparametric estimator of species overlap. Biometrics 2001;57:743–74.
31. 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;69:120–3.
32. 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.
33. Spangler WE, May JH, Strum DP, Vargas LG. A data mining approach to characterizing medical code usage patterns. J Med Syst 2002;26:255–75.