Solving quality quandaries through statistics

Statistics Spotlight

ESTIMATION MODELS

On the Clock

by Julia E. Seaman and I. Elaine Allen

Optimizing estimation models to adjust duration of surgery times

Estimation models are widely used throughout different industries to maximize efficiency and make future plans. Unlike academic research, these models may emphasize utility and ease of use rather than accuracy and completeness.

For hospitals, running multiple operating rooms (OR) requires juggling numerous doctors, nurses and anesthesiologists, plus other staff for preparation and cleanup, and supply chain management personnel even before patients and payers get involved. Accurate models for surgery estimates are important for hospitals, just as similar models are used across disciplines to let companies run as efficiently as possible.

Setting the schedule for a hospital's ORs is a common and complex modeling problem for hospitals across the country. The main factor for timing any OR schedule is the specific procedure—ranging from quick 30-minute routine ones to day-long multisystem and complex operations. Over and underestimation of any OR operation can affect the hospital's efficiency, healthcare costs and a patient's individual healthcare. Surgeons' estimates have been shown to be poor predictors of procedure time, as well as duration estimates that excluded individual patient characteristics.[1-3]

Estimating surgery times within procedures and within specialties has been shown to be more accurate, but requiring complex models with many variables is not easy to implement across individual hospitals.[4, 5] Ultimately, predicting surgery duration must be relatively easy to implement with minimal complexity without sacrificing too much accuracy—a common tradeoff in the business world.

In this column, we will show how a simple update can greatly improve the accuracy of a surgery duration estimation model, even when the data set is quite complex and messy. We compared actual surgery duration with its scheduled surgery length over a one-year period at a single large metropolitan hospital. Scheduled surgery length was based on past surgery duration for an individual type of surgery without correction for important patient covariates.

On average, actual duration was consistently longer than scheduled duration for all surgeries, but this was not uniform over all types of surgeries and all patient characteristics. We focused on two specific patient characteristics—the American Society of Anesthesiologists (ASA) score and body mass index (BMI)—as simple ways to adjust the OR time for procedures.

Method

There were 15,186 surgeries performed from January through December 2016 at the hospital. After eliminating poor quality data due to recording and export error, 14,450 patient surgeries are included in the analyses. Log transformations of the predicted and actual surgery times were used for analyses given the skewed distributions (Figures 1-2, p. 52).

Comparisons between the scheduled and calculated surgery time were examined overall and by ASA class, BMI group and medical service using analysis of variance. A general linear model was fit to the data predicting the difference between scheduled and calculated surgery time to identify important covariates contributing to this difference and to develop a simple calibration method for the hospital.

Results

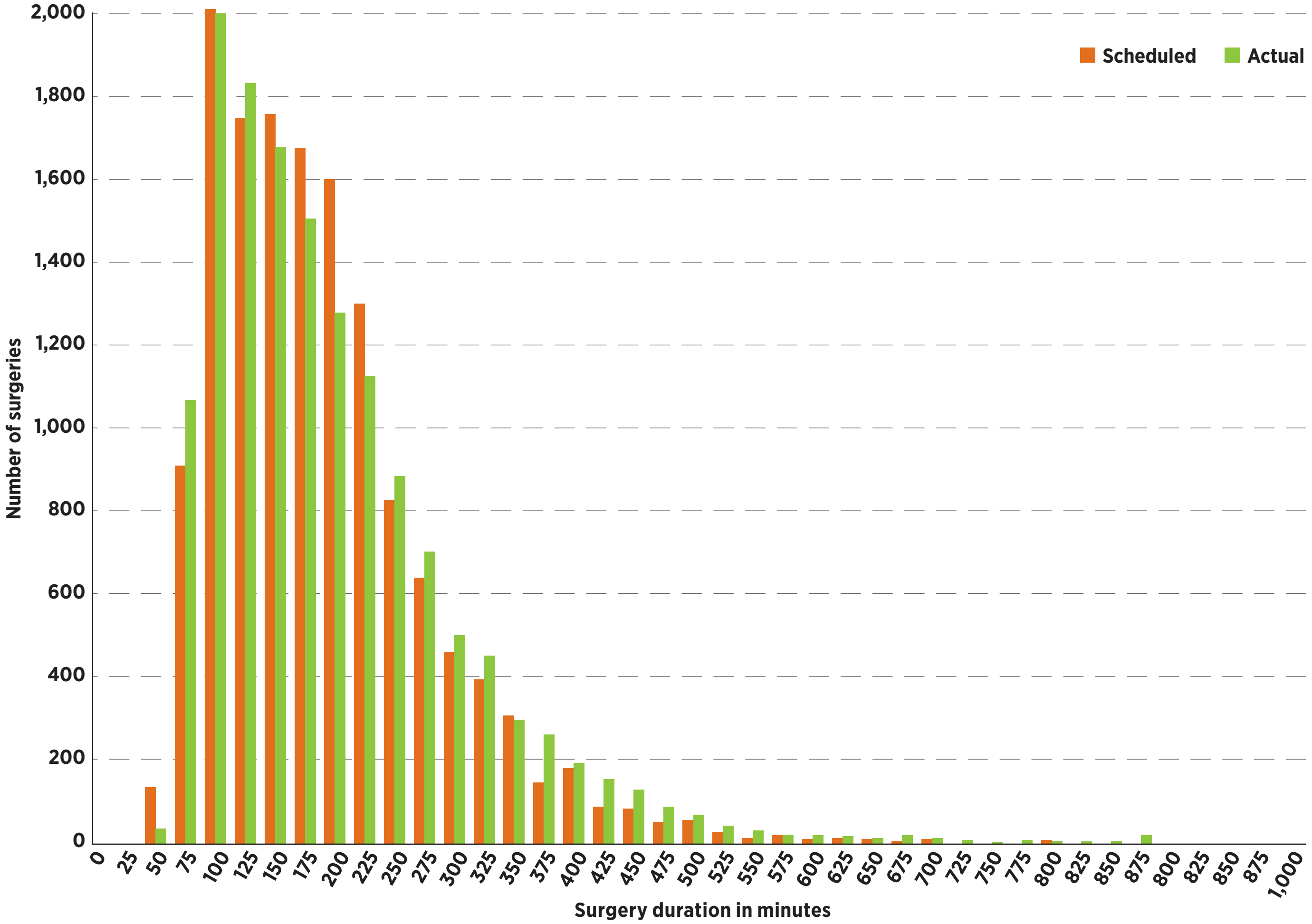
Overall, the mean difference between scheduled and calculated surgery times was an extra eight minutes (95% confidence interval [CI], six to 10 minutes), indicating the estimation model was relatively accurate. For those surgeries that were longer than scheduled, however, the mean increase in surgery time was 55 minutes (95% CI, 54 to 56 minutes). Over the entire year, surgeries lasted more than 2,300 hours longer than planned, which is a large, unexpected expense for work hours and patient health.

The subgroup analysis by ASA, BMI and type of medical

FIGURE 1

Surgery duration—scheduled vs. actual

The estimated and actual times have skewed distributions



service revealed a few interesting results because the general linear model showed the most important contributor to increased surgical time to be BMI ($p < 0.001$) and ASA class ($p = 0.025$).

First, it appeared that the current schedule estimation model already accounts for ASA because this variable had the highest correlation for scheduled duration. Second, the type of surgery did have a small impact on the difference between expected and actual duration, but it was highly variable. Further analysis by surgery service quickly became too complex. There were 20 total services (neurosurgery to orthopedics to pediatrics to general) with 12 to 3,000-plus surgeries per service, further divided into multiple sub-classes, and some patients received more than one procedure. Finally, a patient's BMI was not well correlated with the scheduled duration but was highly correlated with the actual duration.

Looking further into BMI, there was a distinct relationship on surgery time over scheduled duration (Figure 3). The normal range BMIs (< 25) averaged only five minutes over schedule, and patients with BMIs 25-40 were only slightly worse, around eight minutes, which is the overall group average.

For the 8% of patients with BMI ≥ 40, however, the difference between times jumped to 20 minutes overall (n = 1,174, 95% CI, 11 to 29 minutes) and 59 minutes (n = 697, 95% CI, 54 to 64 minutes), totaling an extra 386 hours (or 17% of the unexpected time) of OR time over one year for this group alone. Improved models for surgery scheduling, including BMI, can increase accuracy and reduce conflicts and extra costs associated with hospital overscheduling. Because it is not known a priori whether surgeries will be shorter or longer than the scheduled time, all actual surgery durations were included in the model.

Based on these results, a new prediction model was proposed to improve estimations by incorporating a simple correction factor. The difficulty with a single equation is that surgery case times are not normally distributed and are skewed. Therefore, for simplicity, the modifications focused to improve the times of patient groups that added the most time.

For patients with BMIs < 40, the original method is unchanged. For patients with BMIs ≥ 40, a correction factor of 0.4 * BMI was added to the scheduled time to add accuracy to the estimated surgery duration. While not optimized,

FIGURE 2

Difference in actual and estimated duration

Difference between the scheduled and estimated times is closer to a normally distributed data set

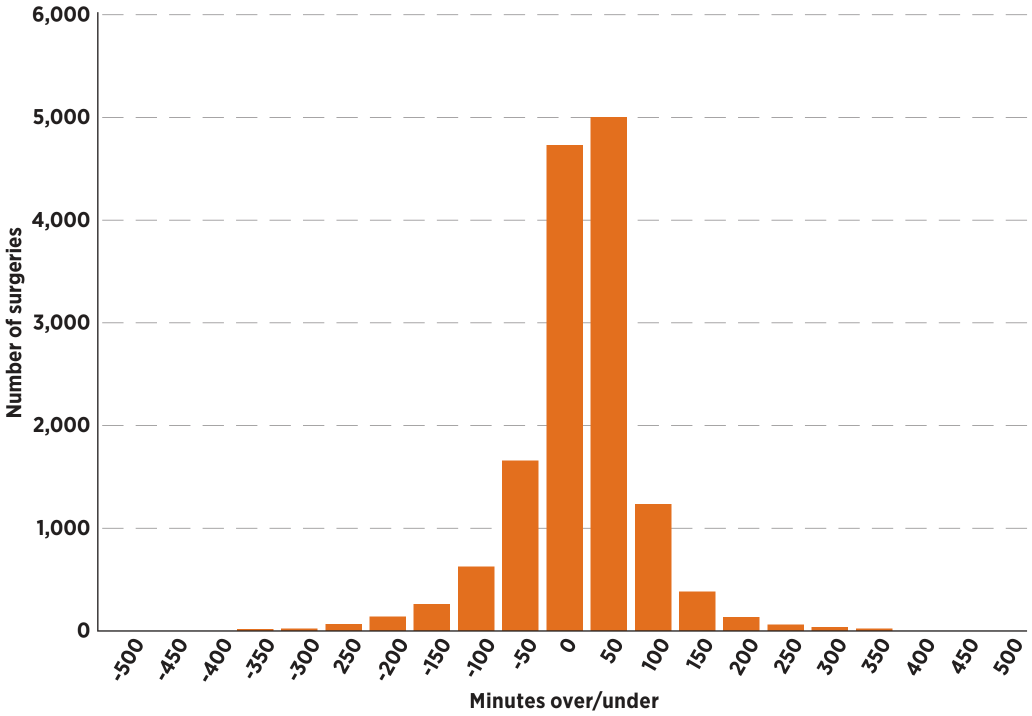
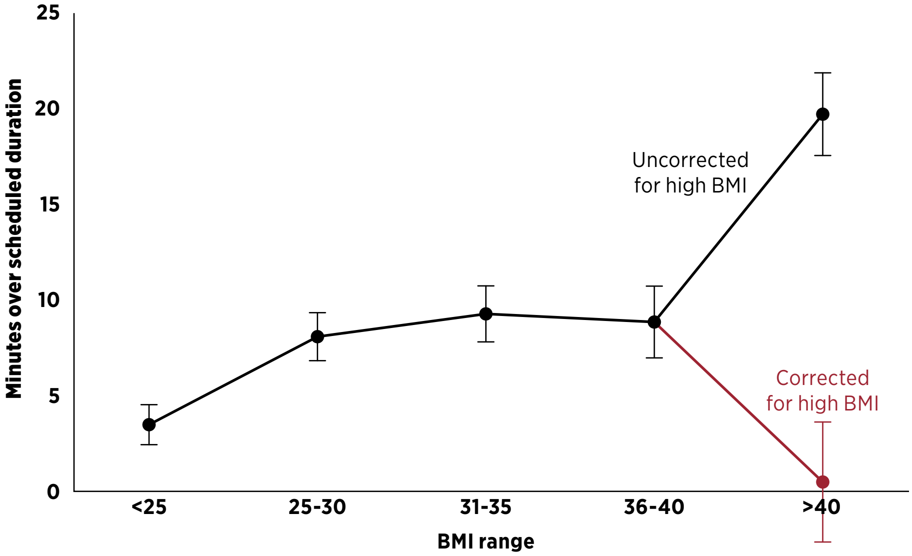


FIGURE 3

Surgeries with longer duration by BMI

Highlighting the impact of adding a correction factor to the highest BMI group



BMI = *body mass index*

this simple calculation reduced the average actual duration difference to 1.3 minutes (−2.8 to 5.5 minutes) from 20 minutes for these high BMI patients, and now accurately accounted for an additional 365 hours of previously unexpected OR time.

If further refinement is wanted, a flat addition of five minutes to all surgery estimates can increase the accuracy, but also may increase the amount of time an OR is unused, decreasing efficiency (see Figure 3).

The missing puzzle piece

Estimation models are important and widely used, but specific models may become outdated or new factors may become available over time. It is important to monitor the accuracy of the model and any changes in the underlying data sets.

For hospitals, there has been an increasing amount of accessible electronic data for patients that can easily pull characteristics to add for modeling. Additionally, in the last decades, the overall U.S. population's BMI has increased, which broke with the assumptions about patient demographics in the existing model.

Indeed, BMI was the ideal missing puzzle piece to add to the model because it was not currently modeled but highly correlated to the actual surgery duration. Finally, the best estimation models are ones that are actually used in the field. A less accurate, but easy-to-follow calculation is often more valuable than a complicated, highly accurate model. QP

REFERENCES

1. D.P. Strum, A.R. Sampson, J.G. May and L.G. Vargas, "Surgeon and Type of Anesthesia Predict Variability in Surgical Procedure Times," *Anesthesiology,* 2000, Vol. 92, No. 5, pp. 1,454-1,466.
2. Marinus J.C. Eijkemans, Mark Van Houdenhoven, Tien Nguyen, Eric Boersma, Ewout W. Steyerberg and Geert Kazemier, "Predicting the Unpredictable: A New Prediction Model for Operating Room Times Using Individual Characteristics and the Surgeon's Estimate," *Anesthesiology,* 2010, Vol. 112, pp. 41-49.
3. Elizabeth Travis, Sarah Woodhouse, Ruth Tan, Sandeep Patel and Jason Donovan, "Operating Theatre Time, Where Does It All Go? A Prospective Observational Study," *British Medical Journal*, Dec. 15, 2014, Vol. 349: 1-4.
4. Albert Wu, Ethan Y. Brovman, Edward E. Whang, Jesse M. Ehrenfeld and Richard D. Urman, "The Impact of Overestimations of Surgical Control Times Across Multiple Specialties on Medical Systems," *Journal of Medical Systems*, 2016, Vol. 40, No. 4, p. 95.
5. Albert Wu, Michael J. Weaver, Marilyn M. Heng and Richard D. Urman, "Predictive Model of Surgical Time for Revision Total Hip Arthroplasty," *Journal of Arthroplasty*, Feb. 7, 2017, pp. 1-5.

Julia E. Seaman is a research scientist at the University of California, San Francisco, focused on pharmaceutical chemistry. She also is a statistical consultant for the Babson Survey Research Group. She earned her doctorate in pharmaceutical chemistry and pharmacogenomics from the University of California, San Francisco.

I. Elaine Allen is professor of biostatistics at the University of California, San Francisco, and emeritus professor of statistics at Babson College in Wellesley, MA. She also is director of the Babson Survey Research Group at Babson College. She earned a doctorate in statistics from Cornell University in Ithaca, NY. Allen is a member of ASQ.