Predicting Outcome from Past Performance

Since the healing arts began, people have sought three basic things from its practitioners: what is wrong with them (diagnosis), what to do for their problem (therapy), and what will happen to them in the future (prognosis). The article by Noordzij et al. entitled “Postoperative Mortality in The Netherlands. A Population-based Analysis of Surgery-specific Risk in Adults” is about prognosis or predicting the future for surgical patients. To be more precise, this article seeks information about the chance of dying within 30 days of surgery. The question is a good one and an important one because the answer applies to individuals seeking to make decisions and to those seeking to assess service performance in a more global sense. In contrast to diagnosis and therapy whose secrets and mechanisms seem readily uncovered by the scientific method, prognosis has been more difficult to unravel particularly when the therapy is procedural in nature and not part of a disease process whose mechanism and course is well known. In short, it is often easier to say what will happen when there is no procedure than what will happen as the result of the procedure.

For those trying to provide answers, the difficult questions are always about what information to apply as signs—dare I say omens—for the basis for the prediction and how well future predictions will perform. Most of our procedural outcome predictions are based on past performance. This work by Noordzij et al. is a good example of using information about past performance to find information that can be successfully applied for future predictions. The problem is that, similar to financial disclaimers or horse races, past performance is not always a good indicator of future performance.

Noordzij et al. examined a large national registry during a 15-yr period from January 1991 to December 2005 that preceded a major transformation in the Dutch healthcare system in 2006. For this observational study, they specifically analyzed a population of 3.7 million surgical procedures performed on adult patients having hospital-based nonambulatory open surgery in 102 hospitals. The authors used postoperative mortality, from any cause, within 30 days of surgery as an outcome. They identified 67,879 mortalities that fit this definition. Because surgical procedures were classified in the registry according to a standard national classification system, Classificatie van Verrichtingen, it was possible to calculate mortality rates for specific classes and subclasses of surgical procedures that ranged from a mortality incidence of 0.7–18.5% and seem to be largely consistent during the 15-yr observational period. The size of the registry, simplicity of the Classificatie van Verrichtingen classification system, and the clarity of the definition of mortality give confidence in the mortality incidences they report. Still, the question remains as to how valuable these surgical procedure groups are to prognostication?

The authors compared the mortality rates of their surgical groups with two other predictive systems, one system by Lee et al. and the other by Neary et al. (Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity system), and found a reasonable agreement among these three prognostication methods. All three these systems have in common that they are fundamentally empirical and intuitive in nature. That is, to say that they all rely on some known information beforehand to predict what will happen in the future based on what was observed in the past. In essence, when an omen is observed, there can be some confidence in guessing the outcome. However, the underlying theory as to why these predictive systems have prognostic value remains a mystery. In other words, why are these deaths occurring with such regularity?

Although not specifically seeking risk factors, implicit in the study by Noordzij et al. is the idea that there are factors associated with each surgical procedure group that can worsen risk, and if they can be identified, the predictive value will become more robust. Clearly, the idea of risk factors has a lot of face value in that we can readily think of comorbidities that influence therapy and ought to have an effect on outcome. Indeed, such risk factors are often associated with adverse outcomes through observational studies similar to the one by Noordzij et al. Therefore, it seems reasonable that the “severity of the surgical procedure” to use the authors’ words also will influence mortality but why should this be? Until we know why our procedural therapies fail, we must continue to rely on predictions based on past performance.

The National Surgical Quality Improvement Program used common surgical risk factors to validate the American Society of Anesthesiologists’ Physical Status classification

Accepted for publication November 17, 2009. The author is not supported by, nor maintains any financial interest in, any commercial activity that may be associated with the topic of this article.
levels (ASA PS) as a predictive of surgical mortality (30-day mortality). They found that the C-index for ASA PS in predicting 30-day mortality was 0.889 in contrast to 0.958 when using the 60 National Surgical Quality Improvement Program risk factors. Presumably, more (or more severe) risk factors are associated with higher ASA PS classifications that might explain this correlation. However, the interesting question is not why the correlation was so strong but why should the ASA PS be predictive at all because it has nothing to do with the surgical procedure by definition? How does this ASA PS correlation with mortality fit with the correlation between the severity of surgical procedures and mortality in the current report? Are certain surgical procedures more likely to be performed on one ASA group than another that might help explain the predictive value of the surgical procedure reported by Noordzij et al.? Is there a clue contained within these various prognostic omens that provide insight into mechanisms that might explain why sometimes our therapies end in morbidity and mortality?

Noordzij et al. sought to connect surgery-specific mortality risk as a prognostic tool that would be useful to patients, physicians, and administrators. The problem is that there is always a broader context to consider. How universal are these surgical procedure-based mortality rates? Caution should be exercised before widely applying surgical procedure-specific mortality rates as benchmarks or as fact beyond the context of the special and temporal boundaries of the study population. We need to keep in mind that therapy takes place in context of change and what might be true at one time or place may not be at another. Furthermore, until we understand why and how our therapeutic systems fail thereby allowing us to develop a structure of sound principles on which to base prediction, we must rely on the guideposts of past performance and hope that the future cooperates.

Albert J. Saubermann, M.D., Department of Anesthesiology, Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, New York. asauberm@mac.com

References