Apples with apples, and the history of risk models

Paul Kurlansky, MD

Despite a plethora of recent literature, formal assessment of quality in health care dates back to the Hammurabi codes of Babylon some 4000 years ago. The modern era of quality assessment was pioneered by Florence Nightingale who published her experiences, as described by Racine,¹ with the British Army in 1858 (depicted in Figure 1). Following Flexner’s 1910 report on American medical education, Bostonian surgeon Ernest Codman earned the wrath of his colleagues by challenging them to study and report the “end results” of their efforts, as described by Shahian and colleagues.² The matter lay somewhat dormant until 1986, when the Health Care Financing Administration began to release a list of hospitals whose mortality rates for Medicare patients exceeded “predicted mortality rates.” Although the lists were originally designed for internal use, their release was demanded under the Freedom of Information Act, and in 1986, results began appearing in the lay press.³

Based on administrative claims rather than clinical data, this approach had numerous methodological flaws—especially the absence of appropriate risk adjustment—and the program was terminated in 1993.⁴,⁵ However, the impact was dramatic. Clinical registries were developed by the Department of Veterans Affairs,⁶ New York State,⁷ the Northern New England Cardiovascular Disease Study Group,⁸ and the Society of Thoracic Surgeons (STS).² The introduction of serious and increasingly sophisticated analytical methodology to the assessment of quality in cardiac surgery had begun, and appropriate assessment of risk was one of its cornerstones. Robust clinical registries—both voluntary and mandated—now populate many more states, and more than 90% of cardiac surgical programs in the country now participate in the STS, which now boasts more than 4.5 million records.⁹ Although originally designed to monitor and improve quality, clinical registries additionally serve to assess patient risk, facilitate surgical judgment, promote informed consent, provide a rich resource for clinical research, and most recently, support prospective randomized trials.

Assessment of surgical risk in coronary artery bypass grafting surgery had already been reported in the Collaborative Study in Coronary Artery Surgery study,¹⁰ and reports of well-validated risk models began to emerge from the STS, New York State, Virginia, and Northern New England. International efforts have followed suit with the well-publicized EuroSCORE (European System for Cardiac Operative Risk Evaluation)¹¹ and updated EuroSCORE II,¹² the JapanSCORE,¹³ the AUS-AVR score (Australian model for aortic valve replacement),¹⁴ as well as various institutionally based models. What is the rationale behind the proliferation of risk models? The 2 key elements that must be addressed with any formal assessment of risk are the internal validity of the model, and the external validity, or applicability to the patient or patient population in question. Standard assessments of internal validity are discrimination and calibration—the ability to discriminate one outcome from another and the ability to do so at varying levels of risk.

External validation refers to the ability of the model to perform well in patient populations different from those used to generate the model. Theoretically, a given model may have excellent predictive ability in the patient sample from which it was derived yet perform very poorly in a patient cohort with very different clinical or demographic characteristics. Differences in risk factors, disease severity, referral patterns, surgical technique, and patterns of care may all affect the ability of a given risk model to accurately predict operative mortality in a population distinct from the one in which it was generated. Moreover, even within a given patient population, changes in practice require that risk-model generation become an iterative process. For
example, the STS updates its model every 3 or 4 years, but in the interim, it annually publishes correction factors specific for the timeframe of a specific report.

This substrate of factors inspired Wang and colleagues\(^\text{15}\) to compare the performance of 4 contemporary risk models in this issue of the Journal. In an era of emerging and successful interventional approaches to aortic valve replacement, it has become increasingly important to accurately assess risk. Originally introduced as a technology focused on the high-risk patient, transcatheter aortic valve replacement (TAVR) may increasingly be applied to patients at more moderate risk.\(^\text{16}\) Being able to define criteria that determine which patients are in the high-risk group is essential, as they might be equally or even more appropriately treated via the percutaneous route. Moreover, since risk models for the TAVR procedure itself have begun to emerge, these calculations may become the determining factor in the selection of the most appropriate therapeutic approach for certain sets of patients.

With the limited data provided by a single institution, it is impossible to distinguish model performance from surgical performance. Indeed, the STS-predicted risk of mortality for the surgical arm of the PARTNER (Placement of Aortic Transcatheter Valve) trial was 11.7%, whereas the observed mortality was only 6.5%, likely owing to the high level of surgical expertise among the sites selected.\(^\text{17}\) Therefore, it is with caution that the results of Wang and colleagues’ study, somewhat unique in its use of the AUS-AVR risk model, should be added to the increasing body of clinical reports regarding the performance of various risk models in the elderly aortic stenosis patient group.

References


