Surgery has been trying to catch up to evidence-based medicine. Assessment of outcome in surgery is the child for quality assurance of patient care. We surgeons have our own set of mental variables that can predict good and poor outcomes. We value our experience and that of others, yet, are always inquisitive about which best predict morbidity and mortality. We all have our own functional equations for outcome that varies qualitatively and quantitatively. The main problem is lack of a uniform mathematical equation for individual patient risk factors that we refer to because of the limitations inherent in the equation or our understanding and awareness. Reviewing the literature in surgical outcome measurement, the impression is one of increasingly diverse messages with conclusions that are institution dependent. This can initiate confusion and controversy when comparing outcomes or extrapolating to one’s own practice while hindering training surgeons to contribute to the evolving evidence of objective quality measurement early in a career. Overall, we are falling behind in recognizing this evolving problem. In this article, I address this controversy and attempt to offer new avenues in achieving a consensus among us in patient risk-adjusted outcomes by adopting and modifying well-recognized risk scoring systems from either side of the Atlantic Ocean. The millennium should see the birth of a new generation of surgeons charged with evidence-based ideas in quality outcome measurement and ready to improve current mathematical models.

We surgeons have been involved in studying outcome—the end result of some action we perform. There is no absolute measure of quality of care in surgical practice. Mortality, morbidity, and length of hospital stay are far from perfect in evaluating the standard of quality. Moreover, these markers may not adequately reflect surgical and anesthetic variables.

Historically, Donabedian developed the classic triad for measuring quality in health care, namely, structure, process, and outcome. Outcome in surgery has been classically described by the “five Ds”—death, disability, dissatisfaction, disease, and discomfort. However, structure and process influence outcome. A shift in interest to include the effect of surgery on health status, functional status, and quality of life is evolving.

In general surgery, the rates of unplanned second laparotomies range from 0.6% to 10%. These rates are low but when comparing fixed intervals and tens of thousands of patients, the comparison between different hospitals can be beneficial. However, unplanned return to the op-
erating room will not have any value as an outcome for quality if patient populations are not adjusted for individual risk factors.

In the context of surgery, to measure outcome accurately, the variables should be easily quantifiable both preoperatively and intraoperatively—a challenging task. Moreover, when using, for example, unplanned return to the operating room to compare outcomes, correcting for patient variability mandates standardization. Such variables can form risk-scoring systems of which many already exist, for example, the Goldman cardiac risk index and prognostic nutritional index. Some scores are ideal for assessing the risk of mortality and morbidity in particular patients. Probably the best known and most widely used scoring system is the Acute Physiology and Chronic Health Evaluation [APACHE] II, which is ideal for the patient in the intensive care unit but requires 24 hours of observation and weighing tables for individual disease status. The benefits of using a scoring system are accurate prediction of outcome, which could influence treatment decisions and rationalize resources. Surgeons have been trained to provide the best quality of care and clinical judgments given the fairly structured surgical residency, so why should a difference in morbidity and mortality outcome exist among us?

A plausible answer is we rigorously fail to correct patients for their risk factors (ie, comorbidities). Nevertheless, there exists mathematical outcome predictor models for all comers in surgery—the 2 most validated models that correct for comorbidities and most critically discussed are the Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM) in Europe and the National Surgical Improvement Program (NSQIP) in the United States from the Department of Veterans Affairs.

POSSUM VS NSQIP

In England, the British government stated the absolute need for clinical governance to enable accurate assessment of the standard of care for individual patients with use for comparative analysis of national data. Copeland et al in 1991 addressed the problems with quality of outcome measurement, mainly some models reduced the number of variables with clear advantages using multifactorial analysis derived from complex mathematical equations. The need for a simple scoring system that can provide an efficient indicator of outcome risk became necessary.

The POSSUM score (Figure 1) has been devised specifically for prediction in surgical patients. The publication of this guideline was a landmark in British surgery. It uses statistically significant, readily available 12 physiological and 6 operative mortality and morbidity predictor variables (Table). A single scoring system applicable to all surgical patients would be an ideal means for facilitating a comparative surgical audit. Several scoring systems have various degrees of accuracy when predicting mortality, yet morbidity for more common complication in elective surgery, is almost universally ignored. In the POSSUM equation, morbidity and mortality is predicted and the individual score variables can be modified. These special features made the Royal College of Surgeons and the Vascular Society of Great Britain and Ireland use this equation for audit and modify it accordingly, for example, C-POSSUM for colorectal scoring and V-POSSUM for vascular scoring. Moreover, many authors have applied POSSUM to specific areas of surgery with good results. Cagigas et al concluded that POSSUM might be appropriately used as a tool of surgical audit in bariatric surgery with vertical-banded gastroplasty. Midwinter et al showed that POSSUM satisfactorily predicted mortality and morbidity for patients undergoing a general vascular surgical procedure. Sagar et al found POSSUM to be valuable for comparative audit in colorectal resection.

On this side of the Atlantic Ocean, the NSQIP model (Figure 2) by Khuri and colleagues should be congratulated on taking differences in preoperative and intraoperative risk factors to be adjusted before comparing surgical outcomes. However, the process is fraught with methodological and policy challenges. Despite the fact that NSQIP is far more exhaustive than POSSUM (18 variables), it is finding difficulty as the model of choice in measuring outcome. Limitations to NSQIP include...
complex measurement of over 60 preoperative and 17 intraoperative patient risks into logistic regression equations that have been calibrated on a population of older preponderantly male veterans who are socioeconomically disadvantaged. This is in contrast to the POSSUM derived from a typical Western modern town (Liverpool, England). Most scoring systems have been devised specifically for use in an intensive care unit where general surgery patients are in the minority. In comparison to the NSQIP, the resident or attending physician can quickly fill a POSSUM score sheet before the operation with no extra costs incurred by trained nonclinical administrative staff. These features have made the POSSUM score the most appropriate of the available scores.

PROPOSAL

There is ongoing surgical research in structure, process, and outcome with no real consensus. The problem is similar to grading various degrees of coma prior to the universal use of the Glasgow Coma Scale. Volume rather than quality research is a concerning active issue in surgery today.

The millennium should see universal agreement with grading morbidity and mortality results of hospitals along a spectrum of scores that can compare predicted with observed ratios. Moreover, there is an increasing need to be able to provide relatives with details of predicted outcome, so there are no recriminations if the patient dies while identifying weakness and strength to improve the quality of the surgeon’s weekly morbidity and mortality conference. The surgical literature has many articles addressing the question of how the quality of surgery needs to be best measured. A large number of index predictor variable measures exist more in some specialties, for example, cardiothoracic surgery. The Society of Thoracic Surgeons requests their surgeons to fill in a data sheet on every case to add to the national data bank for quality control. Should we general surgeons be performing this accurate prospective registry and audit data for colon resection or pancreatectomy?

I address outcome risk assessment in surgery as a fundamental pillar to any surgeon. I hereby propose to surgeons in academic centers a plausible method to tackle our indifference of outcome risk measurement and intercenter variability. Moreover, I strongly anticipate that residents today will be the future leaders in surgery and adopting self-outcome risk assessment in residency will be key to a stronger and self-perpetuating objective general surgeon.

Could surgery be so discrete? And if so, should mathematical models be the way to evaluate outcome?

It is still unclear to many surgeons that mathematical and statistical analysis can be more efficacious than a surgeon’s intuition. However, the larger our score data sample (n = population sample), the greater our sample distribution curve will approximate reality.

The time has come to determine the predicted outcome precisely of our general surgery operations. The recommendation should extend to both residents and attendings. The following steps can be adopted should funding be available: (1) adopt and apply a simple recognized risk scoring systems, for example, POSSUM or NSQIP equations in each surgical department filled by a senior resident or attending about a specific operation; (2) create a departmental and central national score data registry from a timely organized collection exercise using the a standard score sheet; (3) start to compare outcome data among surgeons in each training level, for example, senior surgical residents across the nation to extend the exercise for attendings; and (4) modify the current predictor models for variables that are specialty and local practice specific. This application can be easily adopted, for example, by the large Veterans Affairs health system already attached to university academic departments across the nation. Therefore, such academic centers can extend the exercise to peripheral community hospitals in association with colleagues in other countries wishing to meet the challenge of outcome measurement in surgery in our era. The birth of an “International Society for Surgical Outcome” will be a “necessary evil” to track the quality of surgery and to discuss pathways to clinical excellence.

Corresponding author: Jeffrey H. Shuhaiber, MD, Department of Surgery, University of Illinois at Chicago, 840 S Wood St, Clinical Science Bldg Suite 518-E, Chicago, IL 60612 (e-mail: shuhaibr@uic.edu).

REFERENCES