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.

From the Department of Surgery, University of Illinois at Chicago.
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 weighting tables for individual disease status.9 The benefits of using a scoring system to accurately predict 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 Europe10 and the National Surgical Improvement Program (NSQIP) in the United States from the Department of Veterans Affairs.11

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.12 Copeland et al12 in 1991 addressed the problems with quality of outcome measurement, mainly some models reduced the number of variables with clear disadvantages 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 Surgeons13 and the Vascular Society of Great Britain and Ireland14 use this equation for audit and modify it accordingly, for example, C-POSSUM for colorectal scoring and V-POSSUM for vascular scoring. Moreover, many authors15-17 have applied POSSUM to specific areas of surgery with good results. Cagigas et al17 concluded that POSSUM might be appropriately used as a tool of surgical audit in bariatric surgery with vertical-banded gastroplasty. Midwinter et al18 showed that POSSUM satisfactorily predicted mortality and morbidity for patients undergoing a general vascular surgical procedure. Sagar et al19 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 colleagues4 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.18 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...
ample, cardiothoracic surgery. The Society of Thoracic variable measures exist more in some specialties, for ex-

The following steps can be adopted should funding 

ommendation should extend to both residents and attend-

come precisely of our general surgery operations. The rec-

sions that have been calibrated on a population of older 

soravely disadvantaged. This is in contrast to the POSSUM 

derived from a typical Western modern town (Liver-

pool, England). Most scoring systems have been de-

vised specifically for use in an intensive care unit where 

general surgery patients are in the minority. In compari-

son to the NSQIP, the resident or attending physician can 

can quickly fill a POSSUM score sheet before the operation 

with no extra costs incurred by trained nonclinical ad-

ministrative staff. These features have made the POSSUM 

score now the most appropriate of the available scores.59

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 qual-

ity 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 ob-

served ratios. Moreover, there is an increasing need to 

be able to provide relatives with details of predicted out-

come, 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 ad-

ressing the question of how the quality of surgery needs 

to be best measured.3 A large number of index predictor 

variable measures exist more in some specialties, for ex-

ample, 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 qual-

ity control. Should we general surgeons be performing 

this accurate prospective registry and audit data for col-

olon 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 in-

tercenter 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 gen-

eral surgeon.

Could surgery be so discrete? And if so, should math-

ematical models be the way to evaluate outcome?

It is still unclear to many surgeons that mathemati-

cal 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 out-

come precisely of our general surgery operations. The rec-

ommendation should extend to both residents and attend-

ings. 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 equa-

tions in each surgical department filled by a senior resi-

dent 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 stand-

dard score sheet; (3) start to compare outcome data among 

surgeons in each training level, for example, senior sur-

gical 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. There-

after, such academic centers can extend the exercise to pe-

ripheral community hospitals in association with col-

leagues in other countries willing 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 

come pathways to clinical excellence.

Corresponding author: Jeffrey H. Shuhaiber, MD, Depart-

ment 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


5-17.


3. Donabedian A. Quality assessment and assurance: unity of purpose, diversity of 


4. Lohr KN. Outcome measurement: concepts and questions [review]. Inquiry. 1988;


2001;181:76-80

6. Birksley JD, Hamby LS, Birksley CM, Deckers MV, Karon NM, Dow RW. Is un-

planned return to the operating room a useful quality indicator in general sur-


7. Aarsen MZ, Collogy BT. The risk of an unplanned return to the operating room in 


eight common operations: results from the VA National Surgical Quality Im-


9. Civetta JM, Hudson-Civetta JA, Nelson LD. Evaluation of APACHE II for cost con-


10. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical au-


11. Khuri SF, Daley J, Henderson W, et al. Risk adjustment of the postoperative mor-

bidity rate for the comparative assessment of the quality of surgical care: results 
of the National Veterans Affairs Surgical Risk Study. J Am Coll Surg. 1997;185:

315-327.


14. Vascular Surgical Society of Great Britain and Ireland at The Royal College of 


risk in vascular surgery using POSSUM and the Portsmouth predictor equation. 


17. Sagar PM, Hartley MN, Maney-Jones B, Sedman PC, Jay M, Macfie J. Com-

parative audit of colorectal resection with the POSSUM scoring system. Br J Surg.

1994;91:1492-1494.

18. Angell M, Kassirer JP. Quality and the medical marketplace—following el-


©2002 American Medical Association. All rights reserved.