Brief Tool to Measure Risk-Adjusted Surgical Outcomes in Resource-Limited Hospitals

Jamie E. Anderson, MPH; Randi Lassiter, BS; Stephen W. Bickler, MD; Mark A. Talamini, MD; David C. Chang, PhD, MPH, MBA

Objectives: To develop and validate a risk-adjusted tool with fewer than 10 variables to measure surgical outcomes in resource-limited hospitals.

Design: All National Surgical Quality Improvement Program (NSQIP) preoperative variables were used to develop models to predict inpatient mortality. The models were built by sequential addition of variables selected based on their area under the receiver operator characteristic curve (AUROC) and externally validated using data based on medical record reviews at 1 hospital outside the data set.

Setting: Model development was based on data from the NSQIP from 2005 to 2009. Validation was based on data from 1 nonurban hospital in the United States from 2009 to 2010.

Patients: A total of 631,449 patients in NSQIP and 239 patients from the validation hospital.

Main Outcome Measures: The AUROC value for each model.

Results: The AUROC values reached higher than 90% after only 3 variables (American Society of Anesthesiologists class, functional status at time of surgery, and age). The AUROC values increased to 91% with 4 variables but did not increase significantly with additional variables. On validation, the model with the highest AUROC was the same 3-variable model (0.9398).

Conclusions: Fewer than 6 variables may be necessary to develop a risk-adjusted tool to predict inpatient mortality, reducing the cost of collecting variables by 95%. These variables should be easily collectable in resource-poor settings, including low- and middle-income countries, thus creating the first standardized tool to measure surgical outcomes globally. Research is needed to determine which of these limited-variable models is most appropriate in a variety of clinical settings.


Many efforts have been made to define, measure, and evaluate quality surgical care, but these programs tend to focus on hospitals in urban areas, missing many suburban or rural hospitals and completely overlooking low- and middle-income countries (LMICs). In the United States, the most well known include the American College of Surgeons National Surgical Quality Improvement Program (NSQIP), and the Leapfrog Group’s surgical care standards. Many of these programs focus their research on data from urban and large suburban hospitals and target their programs toward these hospitals. For example, NSQIP collects data on more than 130 variables and includes a 30-day patient follow-up. The cost of participation in this quality improvement program is prohibitory for many small, rural medical centers. The NSQIP recently launched their small and rural program for hospitals that are designated rural by zip code or have fewer than 1680 “NSQIP eligible cases,” but this may miss many medium-sized hospitals in nonurban areas that may be too large for this program or too small to feasibly participate in the original NSQIP.

See Invited Critique at end of article

CME available online at www.jamaarchivescme.com and questions on page 794

In addition, surgical quality improvement programs have largely been isolated in developed countries. To improve global surgery, quality measurement tools must be developed to be broadly and internationally applicable. Allowing hosp
Take top 5 models

Perform multiple logistic regressions with 1 additional variable and rank according to AUROC

Figure 1. Stepwise methods for creating a 6-variable model based on area under the receiver operator characteristic curve (AUROC) values.

METHODS

Patient data from NSQIP from 2005 to 2009 were used to build a tool with a limited number of variables to predict inpatient mortality. This nationally validated program measures more than 130 variables on each patient and includes a 30-day patient follow-up. This data set was chosen for its breadth of variables available for each patient, both preoperatively and postoperatively.

A 6-variable tool was built using a list of all preoperative variables included in the NSQIP database, a total of 66 variables, to predict inpatient mortality. All continuous variables were kept as such except for age, which was grouped into 10-year categories.

We performed a 6-stage process to add each additional variable sequentially (Figure 1). For each stage, logistic regression was performed to predict inpatient death. After each regression, the area under the receiver operator characteristic curve (AUROC) for each model was calculated. The AUROC value is a discriminative measure to identify how well a model separates 2 groups (ie., survivors vs nonsurvivors). An AUROC value of 0.5 would indicate that the model separated the 2 groups no better than chance, whereas an AUROC value of 1.0 would indicate that the model completely separated the 2 groups. The AUROC statistic is actually the percentage of randomly selected pairs that are correctly predicted by the model. Thus, the AUROC value allows us to see which model can more accurately discriminate between the 2 groups of interest.

In stage 1, simple logistic regression was performed with each variable to predict inpatient death. The variable with the highest AUROC value to predict inpatient death was chosen from this first stage and used as the basis for stage 2. In stage 2, all other variables were added to the top variable chosen from stage 1. Multivariate logistic regression with inpatient death as the outcome was performed again for each variation of this 2-variable model, and AUROC values were found. The models with the top 5 AUROC values were chosen and used as the basis for stage 3. The method for stages 3 through 6 was the same as in stage 2: each additional variable was added to the 5 models chosen from the previous stage, multivariate logistic regression was performed to predict inpatient death, and the AUROC value was found. The 5 models with the highest AUROC value would become the basis for the next stage. This process was repeated until we created 6-variable models.

The models with the highest AUROC value at each stage were plotted to observe the diminishing returns of AUROC by each additional variable added (Figure 2). The models with the highest AUROC value were validated using patient data from a 110-bed hospital with a level IV trauma center that serves a community of approximately 25,000 people in California. A retrospective medical record review of 239 surgical patients from 2009 to 2010 was conducted to collect data on each variable of interest. Patients were chosen to represent a random sampling of common, low-mortality operations performed at this hospital (40 procedures on 153 patients) and less common, high-mortality procedures (18 procedures on 86 patients). Common procedures were found by ranking International Classification of Diseases, Ninth Revision (ICD-9) procedure codes. High-mortality procedures were found by ranking ICD-9 procedures among patients who died.
procedures were excluded. A random number of patients from each group were chosen to obtain a representative sample of both common and high-mortality operations performed at this hospital.

Patient data from this hospital were used to validate the models by rerunning the original multivariate logistic regressions and calculating AUROC values. Pseudo-$R^2$ values were also found for these models. Some variables, such as albumin, international normalized ratio, blood urea nitrogen, cancer status, ascites status, and surgical specialty of the surgeon, were not identified from medical record reviews; models with these variables were not available to include in the validation.

Statistical analysis was performed with Stata statistical software, version 11.0 (StataCorp). Statistical significance was defined as $P<.05$. This study received approval from the University of California, San Diego, Institutional Review Board.

**RESULTS**

Data from 631,449 patients from 2005 to 2009 were considered from the NSQIP database to create the limited risk-adjustment model, and data from 239 patients from 2009 to 2010 from the validation hospital were used to assess the risk-adjustment model (Table 1). Mean age and sex distribution are similar between the 2 study populations. By race, Hispanics constitute most cases at the validation hospital, whereas whites constitute most cases in the NSQIP data set.

The American Society of Anesthesiologists (ASA) physical status classification had the highest AUROC value (0.8479) in a single-variable model to predict inpatient mortality (Table 2). The top variables were ASA classification, albumin, functional status, age, sepsis status, and preoperative hematocrit. Combinations of these variables made up the 2- and 3-variable models. In the 4-variable model, emergency status and wound classification were added as significant variables. In the 5-variable model, cancer status, surgeon specialty, and ascites emerged as significant variables. In the 6-variable model, weight loss also emerged as a significant variable, but it is possible this is a surrogate for cancer status.

Using patient data from the validation hospital, the model with the highest AUROC value was a 3-variable model with age, ASA classification, and functional status (AUROC value of 0.9398) (Table 2). The model with the next highest AUROC value was a 2-variable model with ASA classification and functional status (AUROC value of 0.9290).

The AUROC values greater than 90% were achieved after only 3 variables (Figure 2). The AUROC values increased to 91% with 4-variable models and almost 92% with 6-variable models. There is little additional gain in AUROC for a 5- or 6-variable model compared with a 3- or 4-variable model. Including all 66 preoperative variables resulted in an AUROC value of 0.9104 (pseudo-$R^2=0.3342$), approximately the same AUROC value achieved with only 4 variables.

We found that 3 or 4 variables may be sufficient for adequate risk adjustment to measure surgical outcomes. We achieved AUROC values of greater than 90% with only 3 variables. On a scale of 0.5 to 1.0, with an AUROC value of 0.5 indicating that the model cannot distinguish between 2 groups any better than change and an AUROC value of 1.0 indicating that the model completely discriminates between the 2 groups, an AUROC value of greater than 90% is substantial.

Our data provide several examples of risk-adjustment models that may be appropriate for hospitals in resource-limited settings. In particular, a 3-variable model with ASA class, functional status, and age was found to have high discrimination within our nonurban validation hospital. However, the data presented allow for a wide range of possible risk-adjustment models, allowing surgical systems to choose the most appropriate model given their unique resources. For example, although it may be possible for hospital systems in one area to collect preoperative laboratory values, such as albumin or hematocrit, other hospital systems may find it easier to collect information on ASA classification or functional status.

Other studies found that a model based on only a few variables may provide enough discrimination to measure surgical outcomes. Rubinfeld et al found the AUROC value for mortality decreased only slightly from 0.907

**COMMENT**

**Table 1. Characteristics of the Patient Population**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>NSQIP Database</th>
<th>Validation Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All (N = 631 449)</td>
<td>Survived (n = 615 373)</td>
</tr>
<tr>
<td>Age, mean (SD), yr</td>
<td>57.4 (17.2)</td>
<td>57.1 (17.2)</td>
</tr>
<tr>
<td>Race, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>460 085 (78.8)</td>
<td>448 003 (78.8)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>41 165 (7.1)</td>
<td>40 493 (7.1)</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>13 221 (2.3)</td>
<td>12 946 (2.3)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>4291 (0.7)</td>
<td>4199 (0.7)</td>
</tr>
<tr>
<td>Female sex</td>
<td>356 475 (56.5)</td>
<td>348 808 (56.7)</td>
</tr>
<tr>
<td>Abbreviations: NA, not applicable; NSQIP, National Surgical Quality Improvement Program.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aAge was recorded within 5-year categories in the validation hospital and 10-year categories in the NSQIP database.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bRace was only reported for 583 833 of the 631 449 patients. Thus, the percentages are calculated from the total number of patients with a reported race (583 833).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ARCH SURG/VOL 147 (NO. 9), SEP 2012 WWW.ARCHSURG.COM

©2012 American Medical Association. All rights reserved.
using all variables to 0.902 using 10 variables and argue that only a few variables are required for predictive accuracy. Dimick et al\textsuperscript{10} found that limited models based on 5 or 12 variables had comparable discrimination to a 21-variable model using receiver operator characteristics. Birkmeyer et al\textsuperscript{11} also found high correlation between a 5-variable and a 20-variable morbidity risk model.

There is some concern that ASA class and functional status are not reliable measures because they are more subjective. Some data suggest that there is a lack of interrater reliability in assigning ASA class.\textsuperscript{12-14} Davenport et al\textsuperscript{15} found that although ASA class was the strongest single predictor of outcomes, combinations of other risk variables without ASA class were better predictors than ASA class alone. However, ASA class was significantly correlated with 57 of 59 NSQIP preoperative risk factors.\textsuperscript{15} In addition, Cohen et al\textsuperscript{16} did not find evidence that ASA class and functional status were inconsistently classified and argue that they improve model quality and should be used in surgical risk-adjusted assessments. Dimick et al\textsuperscript{10} also found that ASA class and functional status were the most important variables in all risk-adjustment models. Furthermore, ASA class and functional status were 2 of the most predictive preoperative risk variables of postoperative morbidity in the National Veterans Affairs Surgical Risk Study\textsuperscript{17} and have been shown to predict operative outcomes in specific procedures.\textsuperscript{18,19} Disagreement rates between ASA class and functional status, as well as other NSQIP variables, have also improved since implementation (functional status before operation: 11.38\% in 2005 to 3.4\% in 2008; ASA class: 2.65\% in 2005 to 1.82\% in 2008); the authors argue that this is possibly due to data collection training and ongoing support.\textsuperscript{20}

This study is strengthened by the fact that we developed our model using data from a large multicenter database from multiple years. Another strength of this study is that only a few variables are required for predictive accuracy. Dimick et al\textsuperscript{10} found that limited models based on 5 or 12 variables had comparable discrimination to a 21-variable model using receiver operator characteristics. Birkmeyer et al\textsuperscript{11} also found high correlation between a 5-variable and a 20-variable morbidity risk model and recommended that the new version of the NSQIP have no more than 5 to 10 core covariates.

### Table 2. Stepwise Process for Creating the Limited Model to Predict Inpatient Mortality\textsuperscript{a}

<table>
<thead>
<tr>
<th>Model</th>
<th>NSQIP Database</th>
<th>Validation Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC</td>
<td>Pseudo-$R^2$</td>
</tr>
<tr>
<td>1-Variable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA class</td>
<td>0.8479</td>
<td>0.2310</td>
</tr>
<tr>
<td>Albumin\textsuperscript{b}</td>
<td>0.8119</td>
<td>0.1512</td>
</tr>
<tr>
<td>Functional status</td>
<td>0.7676</td>
<td>0.1933</td>
</tr>
<tr>
<td>INR\textsuperscript{b}</td>
<td>0.7615</td>
<td>0.0366</td>
</tr>
<tr>
<td>BUN\textsuperscript{b}</td>
<td>0.7540</td>
<td>0.0909</td>
</tr>
<tr>
<td>2-Variable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA class, albumin\textsuperscript{b}</td>
<td>0.8870</td>
<td>0.2712</td>
</tr>
<tr>
<td>ASA class, functional status</td>
<td>0.8830</td>
<td>0.2668</td>
</tr>
<tr>
<td>ASA class, age (category)</td>
<td>0.8792</td>
<td>0.2465</td>
</tr>
<tr>
<td>ASA class, sepsis</td>
<td>0.8788</td>
<td>0.2841</td>
</tr>
<tr>
<td>ASA class, hematocrit</td>
<td>0.8744</td>
<td>0.2410</td>
</tr>
<tr>
<td>3-Variable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA class, age (category), sepsis</td>
<td>0.9019</td>
<td>0.3057</td>
</tr>
<tr>
<td>ASA class, functional status, age (category)</td>
<td>0.9015</td>
<td>0.3002</td>
</tr>
<tr>
<td>ASA class, albumin\textsuperscript{b}, age (category)</td>
<td>0.8982</td>
<td>0.2863</td>
</tr>
<tr>
<td>ASA class, albumin\textsuperscript{b}, functional status</td>
<td>0.8977</td>
<td>0.2964</td>
</tr>
<tr>
<td>ASA class, albumin\textsuperscript{b}, sepsis</td>
<td>0.8963</td>
<td>0.2951</td>
</tr>
<tr>
<td>4-Variable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status</td>
<td>0.9103</td>
<td>0.3253</td>
</tr>
<tr>
<td>ASA class, functional status, age (category), emergency</td>
<td>0.9085</td>
<td>0.3207</td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, albumin\textsuperscript{b}</td>
<td>0.9079</td>
<td>0.3168</td>
</tr>
<tr>
<td>ASA class, functional status, age (category), wound class</td>
<td>0.9073</td>
<td>0.3105</td>
</tr>
<tr>
<td>ASA class, functional status, age (category), albumin\textsuperscript{b}</td>
<td>0.9072</td>
<td>0.3120</td>
</tr>
<tr>
<td>5-Variable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, cancer\textsuperscript{b}</td>
<td>0.9133</td>
<td>0.3308</td>
</tr>
<tr>
<td>ASA class, functional status, age (category), emergency, sepsis</td>
<td>0.9131</td>
<td>0.3348</td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, wound class</td>
<td>0.9130</td>
<td>0.3281</td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, surgical specialty\textsuperscript{b}</td>
<td>0.9129</td>
<td>0.3294</td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, ascites\textsuperscript{b}</td>
<td>0.9126</td>
<td>0.3305</td>
</tr>
<tr>
<td>6-Variable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, cancer,\textsuperscript{b} emergency</td>
<td>0.9159</td>
<td>0.3403</td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, cancer,\textsuperscript{b} wound class</td>
<td>0.9154</td>
<td>0.3330</td>
</tr>
<tr>
<td>ASA class, age (category), sepsis, functional status, cancer,\textsuperscript{b} surgical specialty</td>
<td>0.9154</td>
<td>0.3344</td>
</tr>
<tr>
<td>ASA class, functional status, age (category), emergency, sepsis, wound class</td>
<td>0.9152</td>
<td>0.3370</td>
</tr>
<tr>
<td>ASA class, functional status, age (category), emergency, sepsis, weight loss\textsuperscript{b}</td>
<td>0.9151</td>
<td>0.3380</td>
</tr>
</tbody>
</table>

Abbreviations: ASA, American Society of Anesthesiologists; AUROC, area under the receiver operator characteristic curve; BUN, blood urea nitrogen; INR, international normalized ratio; NA, not applicable; NSQIP, National Surgical Quality Improvement Program.

\textsuperscript{a}The process used NSQIP data to create the model and data from a nonurban hospital for validation. Only the models with the 5 highest AUROC values for each stage are listed. The models are listed from high to low AUROC values at each stage.

\textsuperscript{b}Variable not collected from validation hospital.
that it was validated using patient data from a smaller non-
urban hospital, using data from both common pro-
dure types and less common, high-mortality procedures. Vali-
dating our study findings enabled us to judge the practicality
of collecting such variables in a resource-limited setting and,
in this case, in a setting that has not yet moved to elec-
tronic medical records. Our validation process also pro-
vided additional information as to which variables had the
highest discrimination among this population. This study is
also strengthened because we included data on all sur-
gical patients. Some quality improvement programs focus
specifically on certain surgical specialties. By using all pa-
tients in NSQIP and validating our models using a mix of
surgeons (including patients with the most common
procedures performed and those with less common but higher-mortality procedures), our findings can be widely
applicable to a variety of surgical fields.

One limitation of this study is that some of the top
variables from our models created by the NSQIP data were
unable to be collected from our validation hospital be-
cause they were not easily obtained through the paper
medical record review. However, the 2- and 3-variable
models using data from the validation hospital had very
high AUROC values, indicating that the additional miss-
ing variables would be unlikely to significantly affect
the results. Another limitation is that there are likely to be
coding errors, both in the NSQIP data and in data from
the validation hospital. However, these errors are likely
to be evenly and randomly distributed and thus should
not affect our conclusions. Furthermore, coding errors
will also be a reality when this model is used, so any cod-
ing errors present in our current data are likely to be simi-
lar to those encountered by this model in practice.

Our study has global implications. Although partici-
pation in programs such as the NSQIP offers adminis-
strative support and comparison of outcomes among par-
ticipating hospitals, the low-cost options reported can
expand the number of hospitals that participate in risk-
adjustment outcomes analysis and quality improve-
ment programs. Our work also allows the expansion of
risk-adjustment outcomes research to LMICs. With mini-
mal training, 3 or 4 variables can be easily and effi-
ciently collected by existing hospital personnel at small
or resource-limited hospitals in both developed and LMICs
with limited costs. From these variables, a hospital’s ob-
erved-to-expected ratio can be calculated to make compa-
rison about outcomes. By offering a simplified risk-
adjustment tool, we can compare surgical outcomes
among hospitals on a global scale, regardless of the spec-
trum of surgical procedures offered or hospital resources.

The area of global surgery has focused primarily on
issues of access, which are still problematic in many
LMICs. However, we should also begin to examine the
process and outcomes of a hospital’s surgical system to
develop more appropriate and cost-effective interven-
tions. Evaluating surgical outcomes requires risk adjust-
ment to take patient variability into account. Our study
suggests that simple but sufficient risk adjustment can
be achieved in these settings. Future validation in an LMIC
setting would be valuable.

Future risk-adjustment models should also consider
surgical complications and morbidity, in addition to mor-
tality. Although in-hospital mortality is simple to col-
lect and the ultimate outcome, other outcomes, such as
complications and morbidity, should not be over-
looked. Other important outcome indicators are disability-
adjusted life-years, which can be used to measure reduc-
tions in premature death and disability as a result of an
intervention. Disability-adjusted life-years are common-
ly used in LMICs, particularly in public health ef-
forts aimed at infectious diseases. By considering disability-
adjusted life-years as an outcome measurement, we can
begin to quantify surgical outcomes in terms of the amount
of reduction of death or disability and have a better under-
standing of the cost-effectiveness of surgical inter-
terventions, which is particularly crucial information in re-
source-limited settings.

Furthermore, surgical quality assessments must in-
clude considerations of structure, process, and out-
comes to evaluate and improve the entire system of sur-
gical care. We encourage the World Health Organization
to expand their Tool for Situational Analysis to Assess
Emergency and Essential Surgical Care to include data
collection on preoperative variables to perform ade-
quately risk-adjustment analyses. With these addi-
tional data, the situational analysis tool can help record
and compare risk-adjusted surgical outcomes within and
among hospitals in LMICs. In conclusion, we propose
that future risk-adjustment tools be based on 6 or fewer
variables to allow for surgical outcomes to be measured
and compared within and among hospitals in resource-
limited settings.

Accepted for Publication: March 9, 2012. Published
Online: May 21, 2012. doi:10.1001
/archsurg.2012.699

Correspondence: Jamie E. Anderson, MPH, Depart-
manship of Surgery, University of California, San Diego, 200
W Arbor Dr, Ste 8400, San Diego, CA 92103 (jaa002
@ucsd.edu).

Author Contributions: Study concept and design:
Anderson, Lassiter, Bickler, Talamini, and Chang. Acquisi-
tion of data: Anderson and Chang. Analysis and inter-
pretation of data: Anderson, Lassiter, Bickler, and Chang. Draft-
ing of the manuscript: Anderson, Bickler, and Chang. Critical
revision of the manuscript for important intellectual
content: Anderson, Lassiter, Bickler, Talamini, and Chang.
Statistical analysis: Anderson, Lassiter, and Chang. Ob-
tained funding: Anderson, Bickler, Chang, and Talamini.
Administrative, technical, and material support: Bickler and
Chang. Study supervision: Bickler and Chang.

Financial Disclosure: None reported.

Funding/Support: Ms Anderson was supported by award
T35HL007491 from the National Heart, Lung, and Blood
Institute. Dr Chang is partially supported by SCANNER
grant R01 HS19913-01 awarded by the Agency for Health-
care Research and Quality. Drs Chang and Bickler are also
partially supported by the Medical Education Partner-
ship Initiative grant 1R24TW008910-01 awarded by the
National Institutes of Health Fogarty International
Center.

Previous Presentation: This paper was presented as a po-
dium presentation at the 83rd Annual Meeting of the Pac
coast Surgical Association; February 20, 2012, Napa

©2012 American Medical Association. All rights reserved.
Valley, California, and is published after peer review and revision.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

Additional Contributions: The NSQIP and its participating hospitals are the source of data used in this research; they have not verified and are not responsible for the statistical validity of the data analysis or conclusions of the authors.

REFERENCES


INVITED CRITIQUE

NSQIP Lite

A Potential Tool for Global Comparative Effectiveness Evaluations

The need to compare outcomes across hospitals is of paramount importance to our patients, physicians, and payers. Administrative databases are inherently limited in scope as has been described in several recent articles in this and other journals. To date, the National Surgical Quality Improvement Program (NSQIP) remains the most robust risk-adjusted and reliable tool available and, most important, the only tool that is readily accepted by most surgeons. A significant problem with NSQIP is that its expense limits the number of participating hospitals and excludes most of our smaller and rural hospitals—hospitals about which one might legitimately wish to ask certain quality and safety questions.

Anderson et al present a compelling pilot model that suggests that as few as 3 simple NSQIP data points, obtainable at significantly lower cost, are all that may be needed to predict inpatient mortality in a risk-adjusted manner across a wide variety of clinical settings. Although the statistical methods are dense, one should not overlook the importance of this article. The development of a simple and inexpensive tool that could be used in the most resource-poor settings in this country and around the world is of enormous importance. For the first time, a tool would exist that would give surgeons the ability to measure the effect of changes to the health care provision systems as they are being implemented in widely diverse settings. One would have a tool that gives teeth to the surgical compar-