Surgical Risk Preoperative Assessment System (SURPAS)

III. Accurate Preoperative Prediction of 8 Adverse Outcomes Using 8 Predictor Variables

Robert A. Meguid, MD, MPH,[*†] *Michael R. Bronsert, PhD, MS,*[*‡] *Elizabeth Juarez-Colunga, PhD,*[*‡§] *Karl E. Hammermeister, MD,*[*‡¶] *and William G. Henderson, PhD, MPH*[*‡§]

Objective: To develop accurate preoperative risk prediction models for multiple adverse postoperative outcomes applicable to a broad surgical population using a parsimonious common set of risk variables and outcomes.

Summary Background Data: Currently, preoperative assessment of surgical risk is largely based on subjective clinician experience. We propose a paradigm shift from the current postoperative risk adjustment for cross-hospital comparison to patient-centered quantitative risk assessment during the preoperative evaluation.

Methods: We identify the most common and important predictor variables of postoperative mortality, overall morbidity, and 6 complication clusters from previously published prediction analyses that used forward selection stepwise logistic regression. We then refit the prediction models using only the 8 most common and important predictor variables, and compare the discrimination and calibration of these models to the original full-variable models using the c-index, Hosmer-Lemeshow analysis, and Brier scores.

Results: Accurate risk models for 30-day outcomes of mortality, overall morbidity, and 6 clusters of complications were developed using a set of 8 preoperative risk variables. C-indexes of the 8 variable models are between 97.9% and 99.2% of those of the full models containing up to 28 variables, indicating excellent discrimination using fewer predictor variables. Hosmer-Lemeshow analyses showed observed to expected event rates to be nearly identical between parsimonious models and full models, both showing good calibration.

Conclusions: Accurate preoperative risk assessment of postoperative mortality, overall morbidity, and 6 complication clusters in a broad surgical population can be achieved with as few as 8 preoperative predictor variables, improving feasibility of routine preoperative risk assessment for surgical patients.

From the *Surgical Outcomes and Applied Research program, University of Colorado School of Medicine, Aurora, CO; †Department of Surgery, University of Colorado School of Medicine, Aurora, CO; ‡Adult and Child Center for Health Outcomes Research and Delivery Science, University of Colorado School of Medicine, Aurora, CO; §Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO; and ¶Division of Cardiology, Department of Medicine, University of Colorado School of Medicine, Aurora, CO.

Funding: The project was supported by funding from the Department of Surgery Surgical Outcomes and Applied Research program and by Dr Meguid's Academic Enrichment Fund from the Department of Surgery.

Disclosure: The ACS NSQIP and participating hospitals are the source of these data; the ACS has not verified and is not responsible for the statistical validity of the data analysis or the conclusions derived by the authors. The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Robert A. Meguid, MD, MPH, FACS, Division of Cardiothoracic Surgery, Department of Surgery, University of Colorado Denver | Anschutz Medical Campus, 12631 E. 17th Avenue, C-310, Aurora, CO 80045.
E-mail: ROBERT.MEGUID@UCDenver.edu.

ISSN: 0003-4932/14/26105-0821
DOI: 10.1097/SLA.0000000000001678

Keywords: overall surgical morbidity, parsimonious assessment, postoperative risk prediction models, surgical outcomes, surgical risk assessment

(*Ann Surg* 2016;xx:xxx–xxx)

Accurate preoperative quantitative assessment of postoperative adverse outcomes is likely to improve shared decision making, informed consent, and guide perioperative care.[1–5] However, formal risk assessment tools are not routinely used for preoperative assessment, education, and optimization of the surgical patient.[6,7] Estimation of patient risk is typically based on a combination of widely accepted or previously reported values, and subjective assessment of individual patient comorbidities by providers. Estimates can vary widely in accuracy.[8,9]

A significant proportion of these risk assessment tools are based on the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) Participant Use File (PUF).[10–15] These risk prediction models have predominately been applied postoperatively in comparing risk-adjusted surgical outcomes between hospitals for quality improvement, with information seldom reaching the individual patient and health care team in the preoperative setting.

Few published examples exist demonstrating consistent use of mathematical risk assessment preoperatively to inform the patient and surgical team about surgical risks. Two factors are likely to account for this: (1) the burden of data collection required by the risk prediction tools, and (2) the desire by many surgeons for greater operation-specificity in the risk model. The latter is a significant factor resulting in the proliferation of many risk models, which add to the data collection burden.

We argue that for a risk calculator to be accepted into widespread use, it must: (1) be easy to use; (2) be broadly encompassing of many different types of surgery and surgical complications; (3) provide reliable and meaningful risk estimates; (4) be based on readily available preoperative data; (5) be available to the public; and (6) be updated periodically. To achieve these goals, the tool must be based on a broadly representative sample of the current surgical population, encompassing most major surgical procedures. The model must be parsimonious in its data collection burden, facilitating rapid use and providing useful and meaningful data which are easy to interpret. It must be available in a user-friendly interface with efficient—ideally automatic—incorporation of data from the electronic health record (EHR) with minimal input of data by the surgical team, and subsequent electronic incorporation into the preoperative note.

Our group has developed a series of surgical risk assessment models for each of the following 8 adverse outcomes: mortality, overall morbidity, and the following clusters of ACS NSQIP complications: infectious, cardiac/transfusion, pulmonary, venous thromboembolic, renal, and neurological complications.[16] These studies have demonstrated that these models can produce highly accurate risk estimations across a broad spectrum of operations and outcomes. These operations include 9 surgical specialties: general, vascular, orthopedic, otolaryngologic, urologic, thoracic, plastic, gynecologic, and neurosurgery. We have demonstrated that surgeon specialty-specific models and inclusion of preoperative laboratory values, which are often missing—particularly in healthier patients undergoing less complex operations—add minimally to the discrimination and calibration of the models. In the current article, we will explore whether limiting the independent variables in all 8 models to only the 8 independent variables that are common and important to all models will affect the accuracy of those models.

We hypothesize that a limited number of patient-level risk variables can be input into a Surgical Risk Preoperative Assessment System (SURPAS) to produce accurate preoperative risk estimates for a diverse array of adverse surgical outcomes in a broad surgical patient population.

METHODS

Data Source

We performed this study using the ACS NSQIP PUF, 2005 to 2012. In the second publication of this series, we demonstrated that inclusion of preoperative laboratory variables in models predicting postoperative mortality and overall morbidity contributes minimally to discrimination and calibration of the models.[16] Therefore, laboratory values were not included in the models, leaving a total of 28 preoperative variables from the ACS NSQIP ''Essentials'' data set as potential predictor variables of 30-day postoperative outcomes. These included 3 demographic variables (sex, age, and race/ethnicity), 17 preoperative comorbidities, 4 variables related to the operation [work relative value unit (RVU), inpatient/outpatient, surgical specialty, and emergency operation status], and 4 other more generic risk variables [American Society of Anesthesiology physical status classification (ASA class), functional health status before surgery, cigarette smoker, and transfer status]. The categories for the variable, race/ethnicity, are defined according to the Centers for Medicare and Medicaid Services standard.[17]

We excluded cases that were in surgical specialties with data no longer being collected by the ACS NSQIP (ophthalmology, podiatry, oral, and cardiac surgery) or due to missing key values. Two important preoperative variables (race/ethnicity and body mass index) had significant percentages of missing values (10.6% and 2.4%, respectively). As we did not want to exclude these cases, a category labeled ''missing'' was created for these 2 variables. The Colorado Multiple Institutional Review Board determined this study exempt from review as it uses publicly available de-identified data.

Primary Outcomes

The primary outcomes for our analyses were the occurrence of any of the following 8 adverse events within 30 days of surgery: mortality, one or more complications (overall morbidity), and clusters of infectious, cardiac/transfusion, pulmonary, venous thromboembolic, renal, and neurological complications. The development of these clusters using factor analysis is described in the first article of this series,[18] and includes the following 18 perioperative complications grouped into 6 clusters: (1) pulmonary cluster (on ventilator >48 h, intraoperative or postoperative unplanned intubation, pneumonia, and septic shock); (2) infectious cluster [sepsis, deep incisional surgical site infection (SSI), organ/space SSI, superficial SSI, urinary tract infection, and wound disruption]; (3) cardiac/transfusion cluster (intraoperative or postoperative myocardial infarction, intraoperative or postoperative cardiac arrest requiring cardiopulmonary resuscitation, and transfusion intra/postoperatively); (4) renal cluster (acute renal failure requiring dialysis and progressive renal insufficiency); (5) venous thromboembolic cluster (vein thrombosis requiring therapy and pulmonary embolism); and (6) neurological cluster (stroke/cerebrovascular accident).

Statistical Analyses

In the present article, we summarize and aggregate information from 8 previously described stepwise, forward selection, logistic regression analyses[16] to determine the minimum number of independent variables required for accurate predictions of all 8 adverse outcomes based on maintaining adequate discrimination and calibration. This minimum parsimonious variable set will comprise the data needed at the preoperative evaluation to accurately calculate the risks of perioperative mortality, overall morbidity, and each of the 6 complication clusters.[16] We first order the 28 preoperative predictor variables by the number of times that each variable enters the models for the 8 outcomes and by the average order of entry into the forward selection stepwise logistic regression models. We use these data to determine the minimum number of predictor variables that are common and important across all 8 outcome models. We then repeat the multiple logistic regression analysis using full models (not forward selection stepwise regression) with the smaller subset of predictor variables to predict the 8 adverse surgical outcomes. We compare the results of the parsimonious models to the full-variable models including all statistically significant predictor variables at the 0.05 level of significance, using discrimination (c-indexes),[19] calibration (Hosmer-Lemeshow analyses),[20,21] and a combined metric of both discrimination and calibration (Brier scores).[22]

We confirmed the results of the stepwise logistic regression analyses using bootstrap resampling. Two hundred bootstrap with replacement samples for each of the prediction models for the 8 adverse surgical outcomes were taken. Full-variable logistic regression models were performed on each sample, ordering the variables by magnitude of the Wald χ^2. The average rankings of the variables in the bootstrap samples were compared with the order of entry of the variables in the forward selection stepwise regression analysis.

All statistical analyses were performed using the program SAS, version 9.3 (SAS Institute, Inc, Cary, NC). The graphical displays of Hosmer-Lemeshow results were made using the statistical analysis software program, R.[23]

RESULTS

We excluded 2.0% (45,680/2,320,920) of cases (0.7% not in targeted surgical specialties, and 1.3% missing key data), resulting in the inclusion of 2,275,240/2,320,936 patients from the ACS NSQIP PUF for 2005 to 2012. The characteristics of the sample of 2,275,240 cases in the analytic data set for all preoperative variables are reported in the Supplemental Digital Content (SDC) Table 1, http://links.lww.com/SLA/A983, of an associated publication.[18] These include operations performed by 9 surgical specialties encompassing 3,045 distinct current procedural terminology (CPT) codes for the primary operation.

Table 1 lists the 28 patient-level risk factors that entered into prediction models for the 8 adverse surgical outcomes of interest. These 28 predictor variables are ordered first in terms of their frequency of occurrence in the models that they entered and then by the average rank of order of entry. There is a natural breakpoint between the first 8 variables and the remaining 20 predictor variables in terms of average order of entry ($\leq$8.89 for the first 8 variables, and $\geq$12.22 for the remaining 20 variables). These top 8 variables [ASA class, work RVU, systemic sepsis (within 24 h before surgery), inpatient/outpatient operation, primary surgeon specialty, age

TABLE 1. Summary of Patient Risk Factor Rankings for the 8 Full Forward Stepwise Logistic Regression Models (n = 2,275,240)

Characteristics	Mortality	Overall Morbidity	Infectious	Cardiac/ Transfusion	Pulmonary	VTE	Renal	Neurological	Frequency of Occurrence	Average Rank
ASA class	1	1	4	1	2	1	1	1	8	2.22
Work relative value unit	10	3	1	2	3	2	3	6	8	4.22
Systemic sepsis (within 48 h)	2	4	2	6	1	3	2	11	8	4.33
Inpatient/outpatient	7	2	3	5	5	4	4	4	8	4.67
Primary surgeon specialty	15	9	5	3	7	9	7	2	8	7.22
Age, yrs	3	7	13	7	10	5	15	3	8	7.89
Functional health status before surgery	5	5	7	19	4	8	11	8	8	8.33
Emergency operation	6	8	8	13	6	14	10	7	8	8.89
Transfer status	14	11	16	8	13	10	14	16	8	12.22
Body mass index category	9	17	11	9	19	13	19	14	8	13.22
Ventilator dependent (within 48 h)	16	14	23	23	9	20	20	5	8	15.33
Open wound with or without infection	26	10	6	11	20	23	25	12	8	15.67
Diabetes mellitus	24	21	14	15	28	17	13	15	8	17.22
Cigarette smoker (within 1 yr)	25	20	12	27	11	19	26	10	8	17.56
Bleeding disorder requiring hospitalization	20	18	18	10	26	24	27	13	8	18.22
Disseminated cancer	4	15	15	12	25	6	23		7	14.29
Transfusion of >4 units PRBCs (within 72 h)	18	6	20	4	18	16	18		7	14.29
>10% loss of body weight (within 6 mo)	13	12	10	16	16	12	24		7	14.71
Steroid use for chronic condition	17	13	9	21	17	7	21		7	15.00
Ascites (within 30 d)	8	19	21	20	14	22	12		7	16.57
Sex	22		24	17	15	15	9	17	7	17.00
Race/ethnicity	27	22	19	14	22	11	16		7	18.71
Acute renal failure (rising creatinine to >3 mg/dL within 24 h)	23	23	22	22	21		5	18	7	19.14
Blood pressure >140/90 mm Hg or taking antihypertensive medicine		26		24	24	21	8	9	6	18.67
Severe chronic obstructive pulmonary disease	19	16	17	26	8		28		6	19.00
Dialysis or hemofiltration (within 2 wk)	11	27		18	27	25	6		6	19.00
Dyspnea (within 30 d)	12	24			12	18	17		5	16.60
Congestive heart failure (within 30 d)	21	25		25	23		22		5	23.20

ASA class indicates American Society of Anesthesiology physical status classification; PRBCs, packed red blood cells; VTE, venous thromboembolic.

[years], functional health status before surgery, and emergency operation) appeared in all 8 of the prediction models with average orders of entry ranging from 2.22 (ASA class) to 8.89 (emergency operation).

Because there was a ‘‘natural break’’ between the 8 most important predictor variables and the remaining 20 predictor variables, we recalculated the 8 prediction models using only these top 8 variables in standard full-variable multiple logistic regression analyses (SDC Tables 1 to 8, http://links.lww.com/SLA/A983). Figures 1 and 2 present Hosmer-Lemeshow graphs for the 8 variable models vs the full-variable models applied to all 2,275,240 cases in the database. Brier scores for these comparisons are presented in SDC Table 9, http://links.lww.com/SLA/A983, all differing less than 0.0028. Together, the c-indexes, Hosmer-Lemeshow graphs, and Brier scores show that the 8 variable models perform very similarly to the full-variable models.

Table 2 presents the odds ratios for the 8 variables for each of the 8 prediction models, allowing comparison of the effect of each risk factor across all 8 outcomes. Intercepts of each model are also presented. Comparison of the odds ratios for each of the predictor variables across the 8 models shows interesting differential effects of the different preoperative predictor variables on the 8 different postoperative adverse outcomes: age is a more important factor for 30-day mortality and stroke; ASA classes III to V are more important for mortality, pulmonary, and renal adverse outcomes; having an emergency operation is more important for the mortality and pulmonary adverse outcomes; partially or totally dependent functional status is most important for the mortality and pulmonary outcomes; having an inpatient vs outpatient operation is most important for the cardiac/transfusion adverse outcomes; systemic sepsis is most important for the mortality, overall morbidity, and pulmonary outcomes; and work RVU (a measure of complexity of the operation) is most important for the overall morbidity and cardiac/transfusion adverse outcomes.

Table 3 shows a comparison of the c-indexes of the 8 variable models to the full models including all statistically significant predictor variables at the 0.05 level of significance. Across all 8 outcome models, the 8 predictor variable models retain at least 97.9% of the c-indexes of the full models.

The results of the bootstrap analysis are presented in SDC Table 10, http://links.lww.com/SLA/A983. The same 8 variables have the highest average ranks based upon the bootstrapped samples and Wald χ^2 as compared with the forward selection stepwise logistic regressions shown in Table 1. There are only slight differences in the orderings of the top 8 variables. Otherwise, the bootstrap analysis confirms the results of the forward selection stepwise regression analysis.

DISCUSSION

Summary of Results

We present a parsimonious set of accurate models predicting postoperative risk for 8 clinically relevant adverse outcomes for a broad surgical population using a limited set of independent preoperative variables. These 8 variable models are able to discriminate between an adverse outcome and no adverse outcome remarkably well (c-indexes 0.749–0.928, mean 0.835); these c-indexes are at least 97.9% of the full models with up to 28 independent variables. Furthermore, Hosmer-Lemeshow analyses demonstrate good calibration with similar model risk estimates by decile of risk with observed risk (Figs. 1 and 2), and differences in Brier scores were minimal (SDC Table 9, http://links.lww.com/SLA/A983).

Of the 8 independent variables that constituted those of the final models, 4 are operative characteristics (work RVU, inpatient/outpatient operation, primary surgeon specialty, and emergency operation status) and 4 are patient characteristics (ASA class, systemic sepsis, functional health status, and age), but all are measurable before surgery. Of the latter, only systemic sepsis is a specific comorbidity—the other 3 patient characteristics are generic risk factors that do not provide details about why they are risk factors. However, the ‘‘why’’ may be identified by as little as an examination of the patient’s problem list. Therefore, should a patient be identified as being at high risk for a specific adverse postoperative outcome, further work-up is warranted to identify specific risk factors and determine how to mitigate this risk.

Comparison With Other Studies

The ACS NSQIP, now in use in approximately 600 private sector and Department of Defense hospitals, has made major contributions to surgical quality improvement based on risk-adjusted outcomes.[10,24,25] However, a recent ACS NSQIP semiannual reports contain results for 384 risk models defined by combinations of outcomes, surgical specialties, individual operations, other surgical groupings, and available predictors accessed.[26] This plethora of risk models would require a major effort to translate to real-time prediction of risk estimates preoperatively and—more importantly—interpretation by surgical teams for their patients.

Of the previously published surgical risk assessment and predictive tools known to the authors, 2 studies are notable for attempting to develop predictive models for a broad surgical population.[7,11] The Universal ACS NSQIP Surgical Risk Calculator uses 24 independent variables in its online tool to predict risk for multiple adverse surgical outcomes (http://riskcalculator.facs.org),[11] whereas the Surgical Outcome Risk Tool (SORT) uses 6 independent variables.[7] These variables (ASA class, urgency of surgery, high-risk surgical specialty, surgical severity, presence of cancer, and age at least 65 yrs) are similar to the variables we have identified in our final models. SORT is based on data collected by anesthetists over a 7-day period in 2010 in 19,097 patients in 326 hospitals in England and Wales. Twelve percent (2,309/19,097) of cases were subsequently excluded mostly for missing data, including unknown mortality status. Thirty-day mortality was the only outcome assessed; the model discrimination was excellent (c-index = 0.91) and described as well calibrated using the Hosmer-Lemeshow test. Our work extends the work of Bilimoria and Protopapa in that we predict risk for 8 different important surgical adverse outcomes, and do it with only 8 preoperative predictor variables. However, aside from these 2 studies by Bilimoria et al and Protopapa et al, most of the previously proposed risk assessment tools are procedure- and disease-specific, and encompass many more patient variables. This increases the burden on the provider to collect and transcribe many data points for each patient, use multiple different risk algorithms throughout the course of a typical clinic, and explain predictions to patients. Instead of this approach of more and more specific surgical risk assessment algorithms, we have approached this challenge with the aim of parsimony and ease of use in the era of compressed time and increasing clinical demands. We are unaware of any published reports of the extent of use, effect on processes of care, or patient outcomes associated with the preoperative clinical implementation of these risk models. Although patient-specific reports can be printed from the ACS NSQIP risk calculator, we are not aware of any attempt at integration with the EHR, such that this information becomes part of the documented preoperative evaluation record.

Implementation of SURPAS

These parsimonious models are not intended as a comprehensive patient assessment, but a rapid and accurate risk assessment alerting care providers of patient risk. An increased risk of adverse

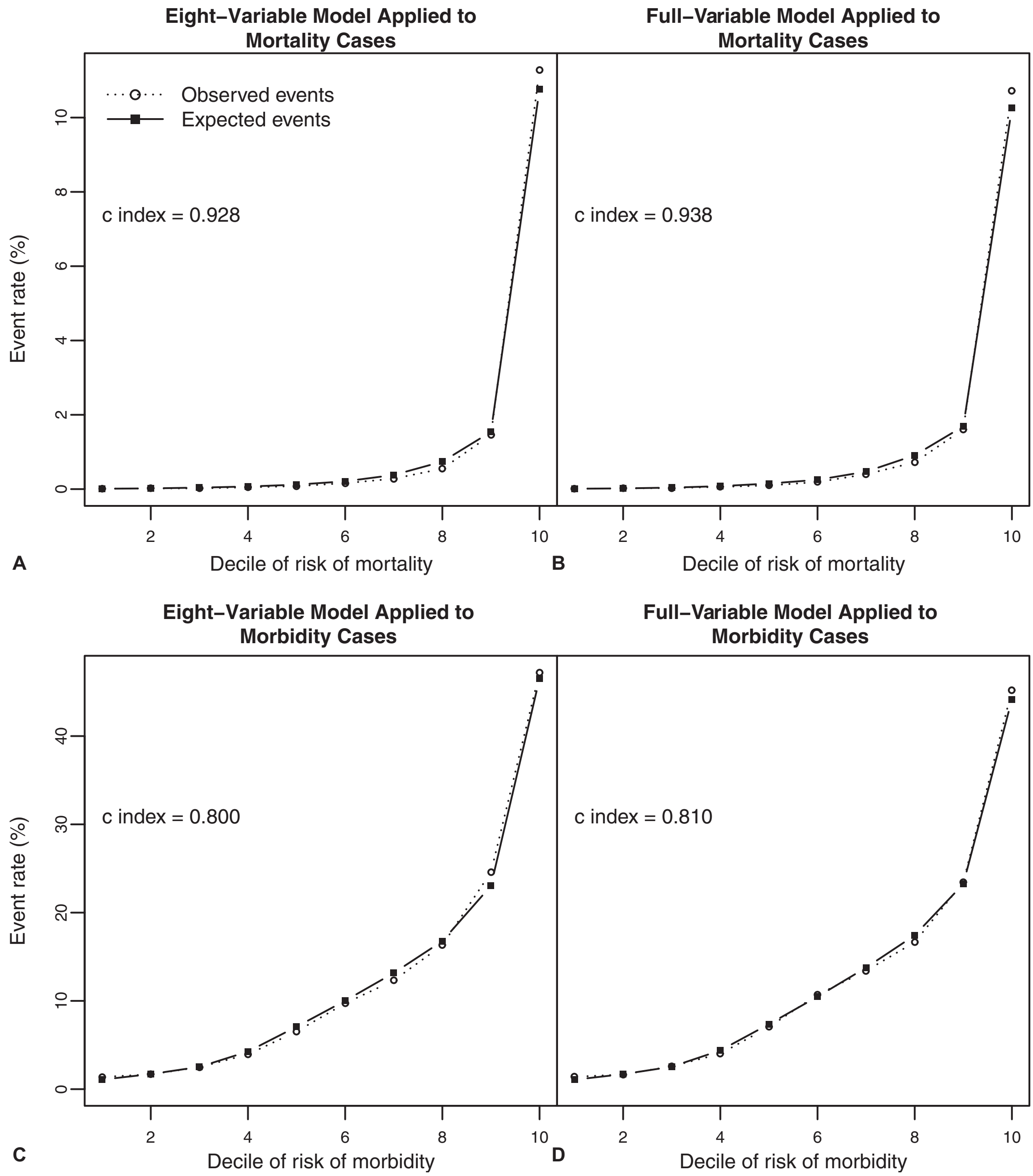

FIGURE 1. Hosmer-Lemeshow analyses showing observed to expected event rates in the groups defined by decile of risk for: (A) the model using the top 8 variables applied to all cases for the outcome of mortality, vs (B) the model including all significant variables applied to all cases for the outcome of mortality (the full-variable model); and (C) the model using the top 8 variables applied to all cases for the outcome of overall morbidity, vs (D) the full-variable model applied to all cases for the outcome of overall morbidity (n = 2,275,240 for all graphs).

postoperative events should prompt care providers to more closely evaluate specific problems and determine which can be intervened upon, or if surgical risk is prohibitive modification of the procedure or even the decision to operate. For example, the review of the problem list requires very little effort or cost, and should be recommended even for those at only moderately increased estimated risks of adverse outcome(s). This may result in identification of a modifiable risk factor, leading to further work-up and modification, to ultimately decrease patient risk of postoperative adverse outcome.

The next step in the overarching project is to incorporate a software application into the EHR to provide real-time risk assessment for surgical patients in the preoperative clinic. This information will serve to more accurately inform patients of their individual risk of adverse surgical outcomes, help with informed decision making,

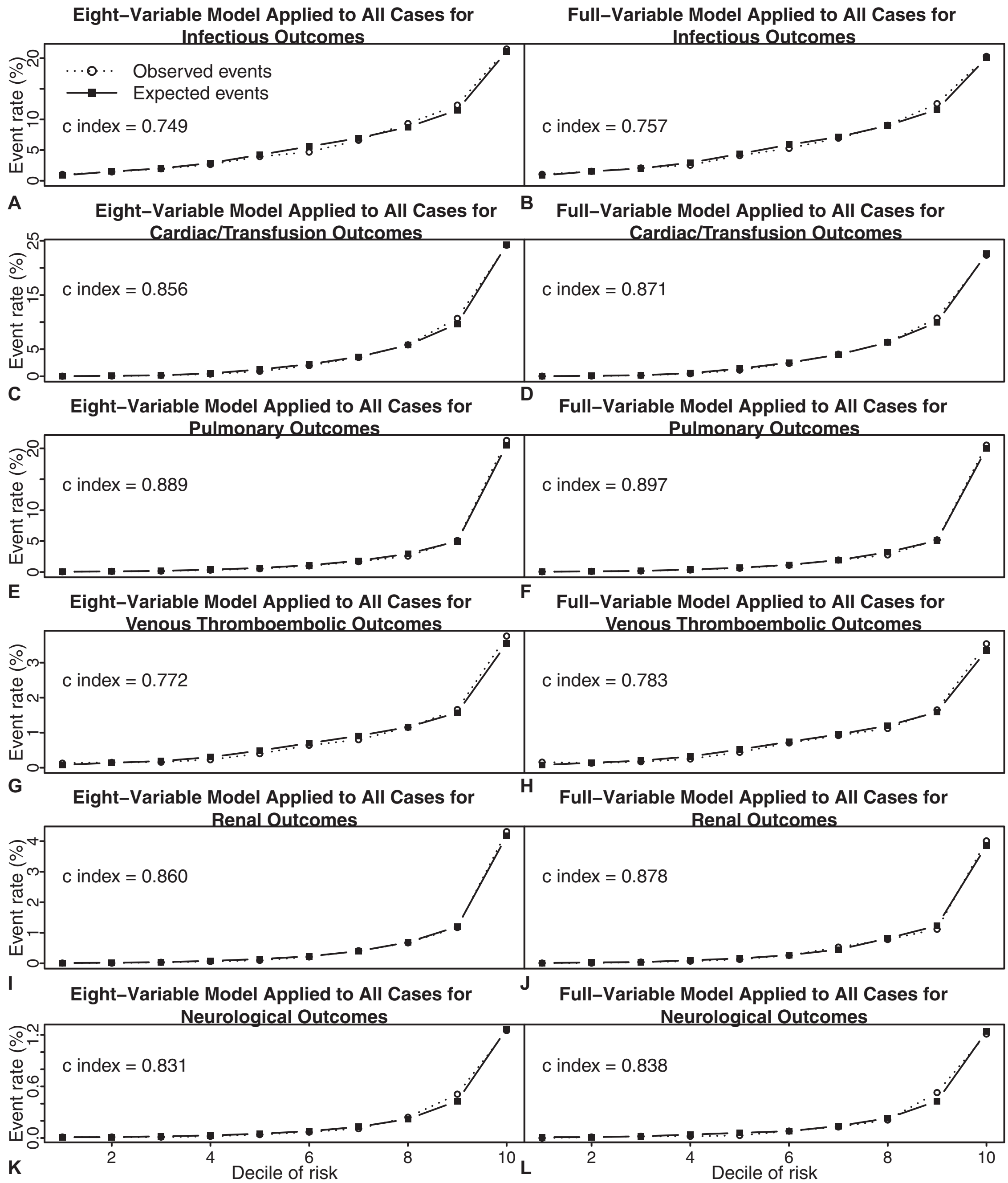

FIGURE 2. Hosmer-Lemeshow analyses comparing observed to expected event rates by decile of risk for the 6 complication clusters for the top 8 variables (A, C, E, G, I, and K), vs the full-variable models (B, D, F, H, J, and L).

and possibly guide perioperative care management with the aim of reduction of adverse outcomes. This new paradigm of preoperative patient risk assessment is based upon the ACS NSQIP PUF. However, this takes the utility of such data sets several steps beyond the current delayed retrospective risk adjustment for interhospital comparison, bringing the data to the clinic and bedside to guide patient management. This system, which we term the ‘‘Surgical Risk Preoperative Assessment System,’’ puts the EHR to work for the provider, the patient, and her/his family instead of simply being a data warehouse.

In the clinical model of SURPAS, we will emphasize the importance of use of the 8 variables in our final models, and request that providers enter those manually if not electronically available from the EHR. Age, emergency operation, inpatient/outpatient status, and primary surgeon specialty should be readily known to the surgical team or electronically extractable from the EHR. Procedure scheduling is commonly done using software-based look-up applications of the CPT name, which is directly linked to the CPT code and work RVU. Systemic sepsis could be identified by its ICD-9 or

TABLE 2. Odds Ratios From Multivariable Logistic Regressions of Top 8 Risk Factors on 30-day Mortality, Overall Morbidity, and 6 Complication Clusters (n = 2,275,240)

			Complication Clusters					
	30-day Mortality	30-day Morbidity	Infectious	Cardiac/Transfusion	Pulmonary	VTE	Renal	Neurological
Characteristics	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**
Age, yrs	1.042	1.011	1.003	1.017	1.016	1.013	1.012	1.032
ASA physical status classification								
II vs I	2.777	1.569	1.547	1.868	2.618	1.685	3.631	2.878
III vs I	12.210	2.766	2.405	3.822	7.886	2.652	11.737	7.600
IV vs I	37.672	5.351	3.014	8.165*	19.743	3.724	24.920	11.481
V vs I	121.829	8.422	2.029	NA	20.853	2.747	38.243	9.167
Emergency operation	2.006	1.557	1.270	1.582	1.934	1.279	1.646	1.602
Functional health status before surgery								
Partially dependent vs independent	2.315	1.766	1.778	1.261	1.927	1.499	1.378	1.397
Totally dependent vs independent	3.648	2.405	1.802	1.138	3.289	1.696	1.577	1.856
Inpatient/outpatient								
Inpatient vs outpatient	4.190	3.757	3.063	12.211	6.012	3.358	5.832	3.097
Systemic sepsis within 48 h								
Other vs none	1.094†	1.135	1.008†	1.451	0.899†	0.953†	1.012†	1.226†
SIRS vs none	2.303	1.847	1.661	1.614	2.413	1.820	1.855	1.410
Sepsis vs none	2.443	2.667	1.957	2.316	3.257	2.206	2.300	1.173
Septic shock vs none	4.127	4.400	1.221	2.964	6.275	2.242	3.106	1.959
Primary surgeon specialty								
Gynecology vs general surgery	0.497	1.405	1.054	3.055	0.492	0.882	0.484	0.555
Neurosurgery vs general surgery	0.976†	0.701	0.422	1.288	0.704	1.332	0.264	3.385
Orthopedic surgery vs general surgery	0.579	1.290	0.389	4.550	0.332	1.073	0.426	0.806
Otolaryngology vs general surgery	0.575	0.790	0.558	1.475	1.139	0.571	0.330	1.212†
Plastic surgery vs general surgery	0.662	1.318	1.137	2.233	0.578	0.917†	0.625	0.962†
Thoracic surgery vs general surgery	1.495	0.912	0.436	1.395	1.620	1.224	0.918†	1.465
Urologic surgery vs general surgery	0.630	0.908	0.657	1.925	0.499	0.920	1.360	1.124†
Vascular surgery vs general surgery	0.853	0.863	0.587	1.641	0.703	0.702	1.102	2.742
Work relative value unit	1.020	1.047	1.035	1.055	1.041	1.035	1.034	1.026
Model intercept	−11.45	−5.56	−5.00	−9.10	−8.85	−8.07	−10.33	−11.77

All overall P values were <0.0001.
*For the cardiac/transfusion cluster ASA IV and V were combined.
†Individual odds ratio were not significant (ie, contained 1.0 within the 95% confidence interval).
ASA class indicates American Society of Anesthesiology physical status classification: I, a normal health patient; II, a patient with mild systemic disease; III, a patient with severe systemic disease; IV, a patient with severe systemic disease that is a constant threat to life; V, a moribund patient who is not expected to survive; NA, not applicable; SIRS, systemic inflammatory response syndrome; VTE, venous thromboembolic.

TABLE 3. Number and Percent of Events, Comparison of the C-indexes for the 8 Variable Models With the Full Models, and Percent of Full Model C-index Provided by the 8 Variable Model

Characteristics	Mortality	Overall Morbidity	Complication Clusters					
			Infectious	Cardiac/Transfusion	Pulmonary	VTE	Renal	Neurological
Events (%)	31,568 (1.4)	287,012 (12.6)	148,837 (6.5)	108,585 (4.8)	74,600 (3.3)	20,674 (0.9)	15,857 (0.7)	5,120 (0.2)
8 variable model c-index	0.928	0.800	0.749	0.856	0.889	0.772	0.860	0.831
Full model c-index	0.938	0.810	0.757	0.871	0.897	0.783	0.878	0.838
Number of variables in full model	27	27	24	27	28	25	28	18
Percent of the full model c-index provided by 8 variable model	98.9%	98.8%	98.9%	98.3%	99.1%	98.6%	97.9%	99.2%

VTE indicates venous thromboembolic.

ICD-10 codes, but should be confirmed by the surgical team. This leaves 2 very important predictor variables, ASA class and functional health status, which will require further work to enable input into SURPAS preoperatively. We plan to incorporate brief ACS NSQIP definitions into a portion of our software module for direct data entry by the surgical team.

Our aim is to minimize the data collection burden on the surgical team through a focus on parsimony in the number of independent variables required for individual patient risk estimation while maintaining accuracy of prediction of outcomes. This reduction in the burden of data collection on the provider should provide more time for interpreting and acting on the risk data in a patient-centered fashion.

Strengths

The strengths of this study include: (1) use of a very large sample size; (2) inclusion of risk, procedure, and outcomes data from a broad range of hospitals; (3) use of a broad range of operations and surgical specialties; (4) use of a systematic sampling method; and (5) collection of the ACS NSQIP data using a standardized protocol including central data auditing.

The use of forward selection stepwise regression analysis in small samples has been criticized in the past.[27,28] We believe that these criticisms do not hold in the presence of very large sample sizes. We compared the bootstrap analysis (SDC Table 10, http://links.lww.com/SLA/A983) with the forward selection stepwise logistic regressions (Table 1), and found that the same 8 variables have the highest average ranks with only slight differences in their orderings. This confirms the results of the forward selection stepwise regression analysis.

Limitations

The primary limitations are as follows. Within the ACS NSQIP, the sampling of surgical specialties has varied over time, resulting in the specialty proportions in the database not being strictly representative of the specialty proportions which exist at the participating hospitals. We do not include in our outcomes unplanned return to the operating room, prolonged length of stay, associated readmission, costs, and operative duration—these are largely a function of the adverse outcomes identified in postoperative follow-up. Outcomes studied and included in the risk models are those of the ACS NQSIP, which is intended to cover a broad spectrum of operations performed by 9 surgical specialties. Therefore, the predictor and outcome variables were necessarily chosen to be more generic (ie, potentially occurring across a large spectrum of operations) rather than more operation- or disease-specific. Thus, there may be important postoperative complications specific to a given operation not directly accounted for in the NSQIP. To the degree that these specific complications might be correlated with the occurrence of one of the 18 NSQIP complications (eg, gastrointestinal anastomotic leak encompassed by organ/space SSI, which is included in the infectious complication cluster), the specific complications are accounted for in the analysis of NSQIP data. The ACS NSQIP does not identify at which participating hospital each procedure occurred, thereby preventing hierarchical statistical analysis to cluster patients by hospital site. However, studies have found that results of prediction of surgical outcomes using either standard or hierarchical logistic regression analysis are similar.[29,30]

CONCLUSIONS

In conclusion, we have demonstrated that 8 independent variables can produce risk estimates of postoperative mortality, overall morbidity, and 6 complication clusters with discrimination and calibration equivalent to that of up to 28 independent variables.

We anticipate that incorporating these risk estimates into the EHR can further reduce the data collection burden and present this information in a portion of the preoperative note, which will be available to all of the patient's care providers during the surgical episode. However, much remains to be learned about how patients understand and use this information, and whether it will reduce the incidence of adverse operative outcomes.

REFERENCES

1. Trevena LJ, Zikmund-Fisher BJ, Edwards A, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med Inform Decis Mak*. 2013;13(suppl 2):S7.
2. Arterburn D. Health plan leadership promotes shared decision making through use of decision aids by surgeons, contributing to lower joint replacement rates and overall health care costs. https://innovations.ahrq.gov/profiles/health-plan-leadership-promotes-shared-decisionmaking-through-use-decision-aids-surgeons. 8-28-2013. Agency for Healthcare Research and Quality. Innovations Exchange: Innovations and Tools to Improve Quality and Reduce Disparities. 7-13-2014. Ref Type: Report.
3. Arterburn D, Wellman R, Westbrook E, et al. Introducing decision aids at Group Health was linked to sharply lower hip and knee surgery rates and costs. *Health Aff (Millwood)*. 2012;31:2094–2104.
4. Salzburg statement on shared decision making. *BMJ*. 2011;342:d1745.
5. Barnett S, Moonesinghe SR. Clinical risk scores to guide perioperative management. *Postgrad Med J*. 2011;87:535–541.
6. Moonesinghe SR. Individualised surgical outcomes: please look the other way. *Postgrad Med J*. 2013;89:677–678.
7. Protopapa KL, Simpson JC, Smith NC, et al. Development and validation of the Surgical Outcome Risk Tool (SORT). *Br J Surg*. 2014;101:1774–1783.
8. Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? *J Am Coll Cardiol*. 2003;42:851–853.
9. Ivanov J, Borger MA, David TE, et al. Predictive accuracy study: comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery. *Ann Thorac Surg*. 2000;70:162–168.
10. Cohen ME, Bilimoria KY, Ko CY, et al. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg*. 2009;208: 1009–1016.
11. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217:833–842.
12. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg*. 2013;217:336–346.
13. Paruch JL, Ko CY, Bilimoria KY. An opportunity to improve informed consent and shared decision making: the role of the ACS NSQIP Surgical Risk Calculator in oncology. *Ann Surg Oncol*. 2014;21:5–7.
14. Raval MV, Cohen ME, Ingraham AM, et al. Improving American College of Surgeons National Surgical Quality Improvement Program risk adjustment: incorporation of a novel procedure risk score. *J Am Coll Surg*. 2010;211: 715–723.
15. Vaid S, Bell T, Grim R, et al. Predicting risk of death in general surgery patients on the basis of preoperative variables using American College of Surgeons National Surgical Quality Improvement Program data. *Perm J*. 2012;16:10–17.
16. Meguid RA, Bronsert MR, Juarez-Colunga E, et al. Surgical Risk Preoperative Assessment System (SURPAS): II. Parsimonious risk models for postoperative adverse outcomes addressing need for laboratory variables and surgeon specialty-specific models. *Ann Surg*. 2016. In press.
17. User Guide for the 2012 ACS NSQIP Participant Use Data File. 2013. American College of Surgeons National Surgical Quality Improvement Program.
18. Meguid RA, Bronsert MR, Juarez-Colunga E, et al. Surgical Risk Preoperative Assessment System (SURPAS): I. Parsimonious, clinically meaningful groups of postoperative complications by factor analysis. *Ann Surg*. 2016. In press.
19. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
20. Hosmer DW Jr, Lemeshow S. *Applied Logistic Regression*. 2nd ed, New York, NY: Wiley; 2000.
21. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982;115:92–106.
22. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–138.
23. R Core Team. R: a language and environment for statistical computing. Available at: http://www.r-project.org/2014. Accessed March 16, 2015.
24. Ramanan B, Gupta PK, Gupta H, et al. Development and validation of a bariatric surgery mortality risk calculator. *J Am Coll Surg*. 2012;214:892–900.
25. Parikh P, Shiloach M, Cohen ME, et al. Pancreatectomy risk calculator: an ACS-NSQIP resource. *HPB (Oxford)*. 2010;12:488–497.
26. American College of Surgeons National Surgical Quality Improvement Program. ACS NSQIP® Semiannual Report, July 10, 2015. Chicago American College of Surgeons; 2015.
27. Harrell FE Jr. Multivariable modeling strategies. In: Harrell Jr FE, editor. *Regression Modeling Strategies*. New York, Philadelphia: Springer; 2001:53–85.
28. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol*. 1999;52:935–942.
29. Hannan EL, Wu C, DeLong ER, et al. Predicting risk-adjusted mortality for CABG surgery: logistic versus hierarchical logistic models. *Med Care*. 2005;43:726–735.
30. Cohen ME, Dimick JB, Bilimoria KY, et al. Risk adjustment in the American College of Surgeons National Surgical Quality Improvement Program: a comparison of logistic versus hierarchical modeling. *J Am Coll Surg*. 2009;209:687–693.