Risk Adjustment in the American College of Surgeons National Surgical Quality Improvement Program: A Comparison of Logistic Versus Hierarchical Modeling

Mark E Cohen, PhD, Justin B Dimick, MD, MPH, Karl Y Bilimoria, MD, MS, Clifford Y Ko, MD, MS, MSHS, FACS, Karen Richards, Bruce Lee Hall, MD, PhD, MBA, FACS

BACKGROUND: Although logistic regression has commonly been used to adjust for risk differences in patient and case mix to permit quality comparisons across hospitals, hierarchical modeling has been advocated as the preferred methodology, because it accounts for clustering of patients within hospitals. It is unclear whether hierarchical models would yield important differences in quality assessments compared with logistic models when applied to American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) data. Our objective was to evaluate differences in logistic versus hierarchical modeling for identifying hospitals with outlying outcomes in the ACS-NSQIP.

STUDY DESIGN: Data from ACS-NSQIP patients who underwent colorectal operations in 2008 at hospitals that reported at least 100 operations were used to generate logistic and hierarchical prediction models for 30-day morbidity and mortality. Differences in risk-adjusted performance (ratio of observed-to-expected events) and outlier detections from the two models were compared.

RESULTS: Logistic and hierarchical models identified the same 25 hospitals as morbidity outliers (14 low and 11 high outliers), but the hierarchical model identified 2 additional high outliers. Both models identified the same eight hospitals as mortality outliers (five low and three high outliers). The values of observed-to-expected events ratios and p values from the two models were highly correlated. Results were similar when data were permitted from hospitals providing < 100 patients.

CONCLUSIONS: When applied to ACS-NSQIP data, logistic and hierarchical models provided nearly identical results with respect to identification of hospitals' observed-to-expected events ratio outliers. As hierarchical models are prone to implementation problems, logistic regression will remain an accurate and efficient method for performing risk adjustment of hospital quality comparisons. (J Am Coll Surg 2009;209:687–693. © 2009 by the American College of Surgeons)

The American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) collects robust patient-level data on preoperative risk factors and postoperative morbidity and mortality outcomes to assess surgical quality at $>$ 200 hospitals.[1] For purposes of comparing adverse outcomes across hospitals, event rates are adjusted by the surgical risk profile of patients treated at each hospital and with consideration of differences in case types performed. The primary statistic used to compare hospital quality in ACS-NSQIP is the ratio of observed events to expected events (O/E).[2,3] Observed events are the actual number of patients who had at least one complication or who died, and expected events are the expected

Disclosure Information: Nothing to disclose.

American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) Disclaimer: The ACS-NSQIP and the hospitals participating in the ACS-NSQIP are the source of the data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors. This article represents the personal viewpoint of the authors and cannot be construed as a statement of official ACS or ACS-NSQIP policy.

Received March 20, 2009; Revised August 12, 2009; Accepted August 19, 2009.

From the Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL (Cohen, Bilimoria, Ko, Richards); Michigan Surgical Collaborative for Outcomes Research and Evaluation (M-SCORE), Department of Surgery, University of Michigan, Ann Arbor, MI (Dimick); Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL (Bilimoria); Department of Surgery, University of California, Los Angeles, and Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA (Ko); and Department of Surgery, John Cochran Veterans Affairs Medical Center, Department of Surgery, School of Medicine, Washington University in St Louis, and Barnes Jewish Hospital, Center for Health Policy, and the Olin Business School at Washington University in St Louis, St Louis, MO (Hall).
Correspondence address: Mark E Cohen, PhD, American College of Surgeons, 633 N Saint Clair St, 22nd Fl, Chicago, IL 60611-3211. email: markcohen@facs.org

Published by Elsevier Inc.

ISSN 1072-7515/09/$36.00
doi:10.1016/j.jamcollsurg.2009.08.020

Abbreviations and Acronyms

ACS = American College of Surgeons
BLUP = best linear unbiased predictor
NSQIP = National Surgical Quality Improvement Program
O/E = observed events-to-expected events ratio

number of such events, statistically estimated based on risk profiles for patients and the procedures performed at each hospital. Values of O/E that are significantly different from 1.0 identify hospitals that are doing better (if O/E is significantly < 1.0; a low outlier) or worse (if O/E is significantly > 1.0; a high outlier) than expected based on their case mix. Depending on outcomes, ACS-NSQIP typically uses p value criteria of 0.10, 0.05, or 0.01. Statistically significant differences are those where the confidence interval for the O/E ratio does not overlap 1.0.

ACS-NSQIP uses logistic regression to estimate O/E ratios for each hospital. Logistic regression has well-documented advantages for this purpose, including being a robust methodology that can incorporate many predictors using efficient, well-documented, and well-maintained statistical software; being well-understood and widely accepted within the health research community; yielding results that are superior or equivalent to most alternative methods; being easily revised and associated with report-generating software so that analyses can be routinely reassessed and documents updated within tight production schedules; and producing information that is conceptually accessible to a wide range of different users.[4]

Hierarchical models have been advocated as a superior methodology for identifying hospital outliers and might be appropriate for ACS-NSQIP.[5-13] An important advantage of hierarchical models is their capacity to adjust for the fact that patients are not selected as independent observations but are clustered or nested within hospitals. The adjustment for this lack of independence theoretically results in more accurate performance estimates, and a frequent associated result is that fewer statistical outliers are identified. The magnitude of these effects will depend on features unique to each dataset, including homogeneity among hospitals (eg, ACS-NSQIP versus Veterans Affairs NSQIP versus all hospitals) and the level of clustering or lack of data independence (often investigated as "intraclass correlation"). The importance of these cluster effects will depend on programmatic goals. It is unknown whether hierarchical models will produce important differences in hospital quality assessments compared with logistic models in ACS-NSQIP. The objective of this study is to evaluate the magnitude and form of differences in estimated O/E ratios and p values using logistic versus one specific implementation of hierarchical modeling for morbidity and mortality after colorectal operations. Colorectal operations were selected for study because they are common procedures, with nontrivial rates of mortality and postoperative complications for which predictive models have been reasonably successful.

METHODS

Data acquisition and patient selection

The developmental history and current details of ACS-NSQIP, including sampling strategy, data abstraction procedures, variables collected, outcomes, and structure, are well described elsewhere.[1,14,15] In brief, the program collects detailed data on patient demographics, preoperative risk factors and laboratory values, operative variables, and postoperative events using standardized definitions. From the ACS-NSQIP database for January 1, 2008, through December 31, 2008, patients 16 years of age or older who underwent a major colorectal operation were identified using Current Procedural Terminology codes, and these data were used for model development. Preexisting blinded data were used. This work was not based on any data from the Veterans Affairs NSQIP.

Preoperative variables

A set of predictive variables, historically useful in modeling outcomes after colorectal operations in the ACS-NSQIP, was constructed from ACS-NSQIP data fields. For purposes of simplicity and consistency with routine ACS-NSQIP modeling, variables were made appropriately categorical and, to preclude empty cells that would adversely affect fitting algorithms, selected categories were collapsed. Patient demographic variables of age (younger than 65, 65 to 74, 75 to 84, 85+ years), gender, and lifestyle factors of smoking status (within 1 year of operation) and alcohol consumption (more than 2 drinks per day for 2 weeks before admission) were used. General factors considered were American Society of Anesthesiologists class (I/II, normal healthy/mild systemic disease; III, severe systemic disease; IV/V, severe systemic disease that is a constant threat to life/moribund), preoperative functional status (independent, partially dependent, totally dependent), dyspnea (none, moderate exertion, at rest), and body mass index (normal, < 18.5 to ≤ 25; underweight, ≤ 18.5; overweight, < 25 to ≤ 30; 3 levels of obese: < 30 to ≤ 35, < 35 to ≤ 40, and < 40). Comorbidities included ventilator dependence, sepsis (eg, systemic inflammatory response syndrome, sepsis, septic shock), a history of COPD, hypertension requiring medication, current pneumonia, ascites, coronary heart disease (includes angina, myocardial infarction within 30 days before operation, percutaneous cardiac intervention, or cardiac artery bypass operation), periph-