ONLINE FIRST

Predicting In-Hospital Mortality in Patients Undergoing Complex Gastrointestinal Surgery

Determining the Optimal Risk Adjustment Method

Jan Grendar, MD; Abdel A. Shaheen, MD, MPH; Robert P. Myers, MD, MSc; Robyn Parker, BSc; Charles M. Vollmer Jr, MD; Chad G. Ball, MD; May Lynn Quan, MD, MSc; Gilaad G. Kaplan, MD, MPH; Tariq Al-Manasra, MD; Elijah Dixon, MD, MSc

Objective: To compare the performance of Charlson/Deyo, Elixhauser, Disease Staging, and All Patient Refined Diagnosis-Related Groups (APR-DRGs) algorithms for predicting in-hospital mortality after 3 types of major abdominal surgeries: gastric, hepatic, and pancreatic resections.

Design: Cross-sectional nationwide sample.

Setting: Nationwide Inpatient Sample from 2002 to 2007.

Patients: Adult patients (≥18 years) hospitalized with a primary or secondary procedure of gastric, hepatic, or pancreatic resection between 2002 and 2007.

Main Outcome Measures: Predicting in-hospital mortality using the 4 comorbidity algorithms. Logistic regression analyses were used and C statistics were calculated to assess the performance of the indexes. Risk adjustment methods were then compared.

Results: In our study, we identified 46 395 gastric resections, 18 234 hepatic resections, and 15 443 pancreatic resections. Predicted in-hospital mortality rates according to the adjustment methods agreed for 43.8% to 74.6% of patients. In all types of resections, the APR-DRGs and Disease Staging algorithms predicted in-hospital mortality better than the Charlson/Deyo and Elixhauser indexes ($P<.001$). Compared with the Charlson/Deyo algorithm, the Elixhauser index was of higher accuracy in gastric resections (0.847 vs 0.792), hepatic resections (0.810 vs 0.757), and pancreatic resections (0.811 vs 0.741) ($P<.001$ for all comparisons). Higher accuracy of the Elixhauser algorithm compared with the Charlson/Deyo algorithm was not affected by diagnosis rank, multiple surgeries, or exclusion of transplant patients.

Conclusions: Different comorbidity algorithms were validated in the surgical setting. The Disease Staging and APR-DRGs algorithms were highly accurate. For commonly used algorithms such as Charlson/Deyo and Elixhauser, the latter showed higher accuracy.

Arch Surg. 2012;147(2):126-135. Published online October 17, 2011. doi:10.1001/archsurg.2011.296

Author Affiliations: Departments of Surgery (Drs Grendar, Shaheen, Ball, Quan, Al-Manasra, and Dixon and Ms Parker), Medicine (Dr Myers), and Medicine and Community Health Sciences (Dr Kaplan), University of Calgary, Calgary, Alberta, Canada; and School of Medicine, University of Pennsylvania, Philadelphia (Dr Vollmer).

ADMINISTRATIVE DATABASES are being increasingly used to evaluate outcomes following multiple surgical procedures.[1,2] These databases possess a comprehensive set of comparable information, including patient demographics, primary and secondary diagnoses, procedures, and details regarding hospital stay. These data are readily available in electronic form. The databases offer relatively easy access to a large sample size. Through comparison with clinical databases, some administrative databases have been validated for researching comorbidity-influenced outcomes in some patient populations. This was done for multiple diagnoses, codes, and databases.[3-18] However, administrative databases are not designed for research purposes and consequently their limitations must be considered. For example, studies of cardiac surgery using administrative databases have observed coding inaccuracies and a failure to record potentially important comorbidities.[19,20] Because outcomes in longitudinal studies can be significantly altered by comorbid conditions, a number of metrics to assess burden of disease and comorbid health conditions have been derived.[1,21-26]

See Invited Critique at end of article

To our knowledge, these comorbidity indexes have not yet been validated specifically for patients who have undergone major abdominal surgical resections. Consequently, it is not known how accurately these various algorithms can adjust in these populations.

The objective of our study was to assess and compare the 4 most common algorithms with regard to their ability to predict in-hospital mortality in patients undergoing major gastric, hepatic, and pancreatic resections. Secondarily, we compared the association between scores and total hospital charges as well as length of stay (LOS).

METHODS

DATA SOURCE

All data were extracted from the Healthcare Cost and Utilization Project Nationwide Inpatient Sample database for 2002 to 2007.[27] The Nationwide Inpatient Sample is the largest all-payer database of national hospital discharges (about 8 million patients per annum) maintained by the Agency for Healthcare Research and Quality. It represents a 20% stratified sample of nonfederal acute care hospitals in the United States including community, general, and academic centers, but not long-term care facilities. Each discharge abstract includes a patient identifier, demographic data, hospital transfer status, admission type (emergency, urgent, or elective), primary and secondary diagnoses (up to 15), procedures (up to 15), health insurance status, total hospital charges, and LOS. Since each record in the Nationwide Inpatient Sample is for a single hospitalization (ie, not a person), there could be multiple records for an individual if that individual had several hospitalizations. Nationwide Inpatient Sample data compare favorably with the National Hospital Discharge Survey, supporting the validity of this database.[28]

STUDY SAMPLE AND OUTCOMES

International Classification of Diseases, Ninth Version, Clinical Modification (*ICD-9-CM*) procedures codes (43.4-43.9, 50.2-50.5, 52.2, and 52.5-52.8) were used to identify adult patients (≥18 years) hospitalized with a primary or secondary procedure of gastric, hepatic, or pancreatic resection between 2002 and 2007. Because our primary outcome was to measure in-hospital mortality, we excluded patients with missing mortality data or those transferred to other institutions. Our secondary outcomes were LOS and total hospital charges adjusted for inflation to 2007 dollars using the US Consumer Price Index for medical care.[29]

RISK ADJUSTMENT MEASURES

In our study, we assessed 4 commonly used risk adjustment measures: the Deyo adaptation of the Charlson comorbidity algorithm,[25] the Elixhauser comorbidity algorithm,[26] the Disease Staging algorithm (Thomson Medstat Inc, Ann Arbor, Michigan),[30] and the All Patient Refined Diagnosis-Related Groups (APR-DRGs) algorithm (3M Health Information Systems, Wallingford, Connecticut).[31] These methods use information obtained in the hospital discharge abstract including demographics, diagnosis, and procedure codes. Both the Charlson/Deyo and Elixhauser algorithms are nonproprietary, can be routinely applied to administrative data using widely available computer algorithms, and identify 17 and 30 categories of comorbidities, respectively, using *ICD-9-CM* diagnosis codes.[25,26]

Both the Charlson/Deyo and Elixhauser algorithms contain a variable for liver disease as a comorbidity. Because we studied hepatic resections, we excluded these variables from our hepatic resection primary analysis.

We also examined the Disease Staging and APR-DRGs algorithms, which are proprietary risk adjustment methods with logic unavailable for outside scrutiny. Results of these algorithms were provided at no cost by the Agency for Healthcare Research and Quality. In Disease Staging, severity is defined as the likelihood of death or organ failure resulting from disease progression and independent of the treatment process.[30] Disease progression is measured using 4 stages (with additional substages) of increasing complexity (stage 1 = no complications or problems of minimal severity; stage 2 = problems limited to a single organ or system, significantly increased risk of complications; stage 3 = multiple site involvement, generalized site involvement, poor prognosis; and stage 4 = death). Disease Staging uses age, sex, admission and discharge status, and diagnoses to generate a predictive scale for mortality.[30]

The APR-DRGs algorithm is a clinical model that expands on the basic DRG structure designed to group patients into approximately 500 categories with similar clinical features and resource use.[31] The APR-DRGs algorithm includes the addition of 4 subclasses to each DRG category to identify minor, moderate, major, or extreme risk of mortality, defined as the extent of physiologic decompensation or organ system loss of function. The process of classifying a patient consists of assessing the level of each secondary diagnosis; determining the base subclass for the patient based on all of his or her secondary diagnoses; and, finally, determining the final subclass of the patient by incorporating the impact of the principal diagnosis, age, procedures, and combinations of categories of secondary diagnoses.[31]

STATISTICAL ANALYSIS

Logistic regression analyses were used to assess the contributions of the individual Charlson/Deyo and Elixhauser comorbidities to predicted in-hospital mortality. We also categorized these comorbidities (as 0, 1, 2, or ≥3 present), as commonly performed in the literature,[32] to determine any impact on risk adjustment compared with the complete list of comorbidities. Similar models were generated for the Disease Staging Mortality Prediction Scale and the APR-DRGs risk of mortality subclass. Each regression model contained a set of independent variables adjusting for sociodemographic and clinical differences between patients including age, sex, health insurance (private, Medicaid, Medicare, self-pay, and other/unknown), race (white, African American, Hispanic, Asian/Pacific Islander, and other/unknown), emergency admission (emergency vs urgent and elective combined), and transfer status (transferred in vs not transferred in). We also adjusted for hospital characteristics including location and teaching status (rural, urban nonteaching, and urban teaching) and region of the United States (Northeast, Midwest, South, and West).

Linear regression models also determined the impact of the Charlson/Deyo and Elixhauser comorbidities on LOS and hospital charges, which were logarithmically transformed because of their skewed distributions. We present exponentiation of the coefficients from these models to reflect the percentage of change in resource use from having a particular comorbidity independent of other patient characteristics.

Performance of Risk Adjustment Methods

To assess the performance of the risk adjustment methods for predicting in-hospital mortality, we calculated the C statistic, an approximation of the area under the receiver operating characteristic curve and a measure of model discrimination.[33] The C statistic ranges from 0.5 to 1.0, with 1.0 indicating perfect discrimination and 0.5 indicating no ability to discriminate. The C statistic is commonly used to compare risk adjustment meth-

Table 1. Demographic, Clinical, and Admission Characteristics of Patients[a]

Characteristic	No. (%)		
	Gastric Resection (n = 40 746)	Hepatic Resection (n = 15 279)[b]	Pancreatic Resection (n = 13 236)[b]
Demographics			
Age, y, median (range)	63 (51-74)	58 (48-69)	60 (49-70)
Female	22 123 (54.3)	8025 (52.6)	6624 (50.1)
Health insurance			
Private	15 215 (37.3)	7878 (51.6)	6182 (46.7)
Medicare	19 255 (47.3)	5238 (34.3)	5041 (38.1)
Medicaid	3052 (27.5)	1098 (7.2)	974 (7.4)
Other	3224 (7.9)	1065 (7.0)	1039 (7.9)
Race			
White	21 161 (51.9)	8346 (54.6)	7401 (55.9)
African American	3818 (9.4)	974 (6.4)	922 (7.0)
Hispanic	3174 (7.8)	1099 (7.2)	910 (6.9)
Asian/Pacific Islander	1271 (3.1)	704 (4.6)	320 (2.4)
Other	11 322 (27.8)	4256 (27.2)	3683 (27.8)
Geographic region			
Northeast	7362 (18.1)	2978 (19.5)	2372 (17.9)
Midwest	8200 (20.1)	2591 (17.0)	2643 (20.0)
South	16 085 (39.5)	5881 (38.5)	5156 (39.0)
West	9099 (22.3)	3829 (25.1)	3065 (23.2)
Location/teaching			
Rural	3718 (9.1)	401 (2.6)	426 (3.2)
Urban	17 141 (42.1)	2686 (17.6)	2629 (19.9)
Teaching urban	19 871 (48.8)	12 189 (79.8)	10 180 (76.9)
Length of stay, d, median (IQR)	7 (4-11)	5 (3-8)	10 (7-17)
Charges, $, median (IQR)	37 770 (21 332-71 644)	40 823 (25 892-65 089)	64 522 (41 592-109 980)
In-hospital mortality	2409 (5.9)	667 (4.4)	784 (5.9)

Abbreviation: IQR, interquartile range.
[a] Patients with more than 1 resection type were excluded from this analysis (eg, patients with hepatic and pancreatic resections).
[b] Transplant patients were excluded (either liver or pancreas).

ods.[34-37] Ninety-five percent confidence intervals (CIs) and comparisons between C statistics were calculated using the method of DeLong et al.[38] Models were analyzed for each cohort of patients (gastric, hepatic, and pancreatic) separately.

Using sensitivity analyses, we analyzed subgroups of each cohort. Those subgroups were (1) limited to 1 type of surgical resection (excluding resections of more than 1 anatomic position, eg, having hepatic and pancreatic resection in the same admission, as those patients would have special characteristics); (2) admissions where the resection procedure of interest was the primary procedure; and (3) admissions where the resection procedure of interest was a secondary procedure.

For hepatic and pancreatic cohorts, we excluded transplant patients from our primary analysis (patients with *ICD-9-CM* procedure code 50.5 for liver transplant and 52.8 for pancreatic transplant). In supplemental analyses, we assessed the performance of the risk adjustment models in cohorts including those transplant patients.

These supplemental analyses allowed us to evaluate the utility of these methods in a broad spectrum of patients with the surgical resections of interest.

Ranking Patients by Predicted Probability of Death

Based on our multivariate models, we computed deciles of predicted probabilities of death for each risk adjustment method. As described by Iezzoni et al,[34] deciles were arranged in 10×10 tables for each of the 6 pairwise comparisons between methods. Our goal was to illustrate the magnitude of the differences in predicted probabilities of death between risk adjustment methods. Based on differences in the average probabilities of death between deciles, we considered a difference of 3 or more deciles between measures as clinically important (median 5.0% difference in mortality rates). For each comparison, we calculated the proportion of patients with "similar" and "different" predicted likelihoods of death (probabilities calculated by both methods $\leq$2 vs $\geq$3 deciles apart).[34]

All statistical analyses were performed using SAS version 9.1.3 software (SAS Institute Inc, Cary, North Carolina).

RESULTS

In total, 40 746 gastric resections, 15 279 hepatic resections, and 13 236 pancreatic resections met our inclusion criteria. The studied procedure was the primary diagnosis in 69% among gastric resections, 85% among hepatic resections, and 88% among pancreatic resections. The median age of the cohorts ranged from 58 to 63 years, 50% to 54% of patients were female, 52% to 56% were white, and 37% to 52% had private health insurance. Overall, in-hospital mortality was 5.9% in gastric and pancreatic resections and 4.4% in hepatic resections; median LOS ranged from 5 days (interquartile range [IQR], 3-8 days) for hepatic resections to 10 days (IQR, 7-17 days) for pancreatic resections, and median total hospital charges were lowest for gastric resections ($37 770 [IQR, $21 332-$71 644]) and highest for pancreatic resections ($64 522 [IQR, $41 592-$109 980]) (**Table 1**).

Table 2. Prevalence and Relationship of Charlson/Deyo Comorbidities to In-Hospital Mortality[a]

Comorbidity	Gastric Resection	Hepatic Resection[b]	Pancreatic Resection
Myocardial infarction			
Prevalence, %	4.4	2.6	3.3
In-hospital mortality OR (95% CI)	1.57 (1.35-1.84)	2.09 (1.44-3.05)	1.96 (1.46-2.65)
Congestive heart failure			
Prevalence, %	10.5	2.8	3.8
In-hospital mortality OR (95% CI)	2.10 (1.88-2.35)	2.14 (1.49-3.08)	3.13 (2.43-4.02)
Peripheral vascular disease			
Prevalence, %	3.2	1.1	1.9
In-hospital mortality OR (95% CI)	0.78 (0.62-0.97)	0.55 (0.25-1.24)	0.90 (0.55-1.47)
Cerebrovascular disease			
Prevalence, %	2.0	0.8	0.9
In-hospital mortality OR (95% CI)	1.69 (1.35-2.13)	10.00 (6.47-15.44)	2.21 (1.29-3.77)
Dementia			
Prevalence, %	0.6	0.1	0.1
In-hospital mortality OR (95% CI)	0.91 (0.58-1.41)	1.98 (0.52-7.53)	NA
Chronic pulmonary disease			
Prevalence, %	18.1	9.4	11.5
In-hospital mortality odds ratio (95% CI)	1.24 (1.12-1.38)	0.56 (0.40-0.77)	1.12 (0.90-1.39)
Rheumatologic disease			
Prevalence, %	2.0	1.0	1.4
In-hospital mortality OR (95% CI)	0.76 (0.54-1.07)	0.87 (0.35-2.17)	1.17 (0.64-2.16)
Peptic ulcer disease			
Prevalence, %	22.5	0.7	2.0
In-hospital mortality OR (95% CI)	3.87 (3.52-4.25)	1.69 (0.80-3.60)	1.71 (1.12-2.61)
Diabetes mellitus			
Prevalence, %	18.6	14.0	20.6
In-hospital mortality OR (95% CI)	0.51 (0.45-0.59)	0.41 (0.31-0.56)	0.47 (0.38-0.59)
Diabetes with chronic complications			
Prevalence, %	2.6	0.7	1.3
In-hospital mortality OR (95% CI)	0.51 (0.38-0.70)	0.87 (0.39-1.95)	0.50 (0.26-0.96)
Hemiplegia or paraplegia			
Prevalence, %	0.4	0.2	0.1
In-hospital mortality OR (95% CI)	4.99 (3.23-7.71)	6.56 (2.43-17.67)	0.92 (0.10-8.17)
Renal disease			
Prevalence, %	5.6	1.8	2.4
In-hospital mortality OR (95% CI)	2.89 (2.52-3.31)	2.03 (1.33-3.10)	3.39 (2.46-4.69)
Any malignancy			
Prevalence, %	29.7	40.1	51.9
In-hospital mortality OR (95% CI)	1.89 (1.69-2.13)	0.85 (0.71-1.01)	0.74 (0.62-0.89)
Moderate and severe liver disease			
Prevalence, %	3.1	3.3	1.0
In-hospital mortality OR (95% CI)	3.49 (2.85-4.28)	NA	3.38 (2.05-5.58)
Metastatic solid tumor			
Prevalence, %	13.1	48.8	25.8
In-hospital mortality OR (95% CI)	1.45 (1.27-1.66)	0.32 (0.26-0.39)	0.83 (0.67-1.02)
AIDS			
Prevalence, %	0.2	0.1	0.1
In-hospital mortality OR (95% CI)	2.14 (0.98-4.69)	0.44 (0.06-3.55)	5.28 (1.59-17.56)

Abbreviations: CI, confidence interval; NA, not applicable; OR, odds ratio.
[a] Coefficients for each comorbidity adjusted for age, sex, race, primary health insurer, emergency admission, hospital characteristics, and the remaining comorbidities.
[b] Moderate and severe liver disease comorbidity was removed from the adjusted model for hepatic resection.

CHARLSON/DEYO AND ELIXHAUSER COMORBIDITIES

Table 2 and **Table 3** provide the prevalence and independent effects of each comorbidity on in-hospital mortality, LOS, and hospital charges in each cohort.

In the gastric resection cohort, any malignancy (29.7%), peptic ulcer (22.5%), diabetes mellitus (18.6%), and chronic pulmonary disease (18.1%) were the most prevalent comorbidities. Neurological disorder (odds ratio [OR], 4.99; 95% CI, 3.23-7.71), peptic ulcer (OR, 3.87; 95% CI, 3.52-4.25), moderate and severe liver disease (OR, 3.49; 95% CI, 2.85-4.28), renal disease (OR, 2.89; 95% CI, 2.52-3.31), and congestive heart failure (OR, 2.10; 95% CI, 1.88-2.35) were associated with at least doubling of the adjusted odds of in-hospital mortality when using the Charlson/Deyo algorithm. In contrast, hypertension (45.5%), fluid and electrolyte disorders (20.7%), diabetes mellitus (17.8%), chronic pulmonary disorders (17.2%), and deficiency anemia (16.2%) were the more prevalent comorbidities in the Elixhauser algorithm. Coagulopathy (OR, 4.47; 95% CI, 3.89-5.13),

Table 3. Prevalence and Relationship of Elixhauser Comorbidities to In-Hospital Mortality[a]

Comorbidity	Gastric Resection	Hepatic Resection[b]	Pancreatic Resection
Congestive heart failure			
Prevalence, %	8.5	2.4	3.3
In-hospital mortality OR (95% CI)	1.95 (1.72-2.21)	1.86 (1.25-2.76)	2.67 (2.01-3.55)
Cardiac arrhythmias			
Prevalence, %	12.8	7.8	9.8
In-hospital mortality OR (95% CI)	1.99 (1.78-2.22)	2.00 (1.55-2.58)	1.93 (1.56-2.39)
Valvular disease			
Prevalence, %	4.6	2.7	3.3
In-hospital mortality OR (95% CI)	0.61 (0.50-0.76)	0.64 (0.35-1.16)	0.92 (0.61-1.38)
Hypertension			
Prevalence, %	45.5	35.9	39.1
In-hospital mortality OR (95% CI)	0.37 (0.34-0.42)	0.38 (0.31-0.48)	0.49 (0.41-0.59)
Pulmonary circulation disorders			
Prevalence, %	0.7	0.4	0.3
In-hospital mortality OR (95% CI)	0.50 (0.29-0.85)	1.14 (0.41-3.16)	0.73 (0.19-2.75)
Peripheral vascular disorders			
Prevalence, %	3.1	1.2	2.1
In-hospital mortality OR (95% CI)	1.03 (0.81-1.31)	1.47 (0.72-3.02)	0.95 (0.56-1.61)
Paralysis			
Prevalence, %	0.8	0.3	0.2
In-hospital mortality OR (95% CI)	2.10 (1.43-3.09)	4.65 (1.99-10.84)	0.91 (0.19-4.39)
Other neurological disorders			
Prevalence, %	3.2	1.6	1.8
In-hospital mortality OR (95% CI)	1.76 (1.44-2.15)	2.59 (1.66-4.02)	0.77 (0.41-1.45)
Chronic pulmonary disease			
Prevalence, %	17.2	9.1	11.4
In-hospital mortality OR (95% CI)	1.24 (1.11-1.38)	0.59 (0.42-0.84)	1.20 (0.95-1.50)
Diabetes mellitus, uncomplicated			
Prevalence, %	17.8	13.8	19.8
In-hospital mortality OR (95% CI)	0.67 (0.58-0.77)	0.56 (0.41-0.76)	0.61 (0.49-0.77)
Diabetes, complicated			
Prevalence, %	2.7	0.9	2.0
In-hospital mortality OR (95% CI)	0.71 (0.53-0.95)	0.90 (0.41-1.96)	0.58 (0.32-1.06)
Hypothyroidism			
Prevalence, %	8.0	5.8	6.8
In-hospital mortality OR (95% CI)	0.47 (0.37-0.59)	0.34 (0.18-0.64)	0.57 (0.37-0.87)
Renal failure			
Prevalence, %	4.9	1.6	2.2
In-hospital mortality OR (95% CI)	3.38 (2.88-3.96)	2.86 (1.81-4.51)	4.38 (3.00-6.38)
Liver diseases			
Prevalence, %	6.8	15.3	3.3
In-hospital mortality OR (95% CI)	1.86 (1.56-2.22)	NA	1.30 (0.89-1.90)

(continued)

renal failure (OR, 3.38; 95% CI, 2.88-3.96), weight loss (OR, 2.39; 95% CI, 2.09-2.74), neurological disorders (paralysis) (OR, 2.10; 95% CI, 1.43-3.09), and cardiac arrhythmias (OR, 1.99; 95% CI, 1.78-2.22) were associated with the highest odds of in-hospital mortality when using the Elixhauser algorithm.

In the hepatic resection cohort, any malignancy (40.1%) or metastatic tumor (48.8%) and diabetes mellitus (14.0%) were the most prevalent comorbidities. Cerebrovascular disease (OR, 10.0; 95% CI, 6.47-15.44), neurological disorder (OR, 6.56; 95% CI, 2.43-17.67), congestive heart failure (OR, 2.14; 95% CI, 1.49-3.08), myocardial infarction (OR, 2.09; 95% CI, 1.44-3.05), and renal disease (OR, 2.03; 95% CI, 1.33-3.10) were each associated with more than a 2-fold increased risk of in-hospital mortality in the Charlson/Deyo algorithm. Conversely, hypertension (35.9%), metastatic cancer (14.8%), and diabetes (13.8%) were the more prevalent comorbidities in the Elixhauser algorithm. Paralysis (OR, 4.65; 95% CI, 1.99-10.84), coagulopathy (OR, 3.99; 95% CI, 3.18-5.00), renal failure (OR, 2.86; 95% CI, 1.81-4.51), fluid and electrolyte disorders (OR, 2.98; 95% CI, 2.46-3.60), neurological diseases other than paralysis (OR, 2.59; 95% CI, 1.66-4.02), and cardiac arrhythmias (OR, 2.00; 95% CI, 1.55-2.58) were associated with the highest odds of in-hospital mortality in the Elixhauser algorithm.

In the pancreatic resection cohort, any malignancy (51.9%) or metastatic tumor (25.8%) and diabetes mellitus (20.6%) were the most prevalent comorbidities. AIDS (OR, 5.28; 95% CI, 1.59-17.56), renal failure (OR, 3.39; 95% CI, 2.46-4.69), liver disease (OR, 3.38; 95% CI, 2.05-5.58), congestive heart failure (OR, 3.13; 95% CI, 2.43-4.02), and cerebrovascular disease (OR, 2.21; 95% CI, 1.29-3.77) had the highest association with in-hospital mortality in the Charlson/Deyo algorithm. In contrast, hypertension (39.1%), metastatic cancer (24.0%), fluid

Table 3. Prevalence and Relationship of Elixhauser Comorbidities to In-Hospital Mortality[a] (continued)

Comorbidity	Gastric Resection	Hepatic Resection[b]	Pancreatic Resection
Peptic ulcer disease (no bleeding)			
Prevalence, %	1.2	0	0.2
In-hospital mortality OR (95% CI)	0.96 (0.60-1.55)	NA	0.80 (0.11-6.14)
AIDS			
Prevalence, %	0.1	0.1	0.1
In-hospital mortality OR (95% CI)	1.09 (0.36-3.30)	0.57 (0.07-4.61)	1.03 (0.09-11.63)
Lymphoma			
Prevalence, %	0.7	0.4	0.4
In-hospital mortality OR (95% CI)	1.82 (1.25-2.65)	NA	1.86 (0.77-4.47)
Metastatic cancer			
Prevalence, %	12.5	14.8	24.0
In-hospital mortality OR (95% CI)	1.55 (1.37-1.76)	0.75 (0.58-0.96)	0.72 (0.59-0.88)
Solid tumor without metastasis			
Prevalence, %	1.9	9.6	2.5
In-hospital mortality OR (95% CI)	1.47 (1.13-1.91)	0.42 (0.29-0.62)	0.98 (0.61-1.56)
Rheumatoid arthritis/collagen vascular diseases			
Prevalence, %	2.0	1.1	1.4
In-hospital mortality OR (95% CI)	0.76 (0.52-1.09)	0.80 (0.29-2.26)	1.31 (0.67-2.57)
Coagulopathy			
Prevalence, %	4.2	5.1	4.2
In-hospital mortality OR (95% CI)	4.47 (3.89-5.13)	3.99 (3.18-5.00)	6.04 (4.82-7.57)
Obesity			
Prevalence, %	5.4	3.9	3.5
In-hospital mortality OR (95% CI)	0.64 (0.45-0.89)	0.43 (0.21-0.88)	0.54 (0.27-1.09)
Weight loss			
Prevalence, %	5.9	2.1	7.2
In-hospital mortality OR (95% CI)	2.39 (2.09-2.74)	1.83 (1.25-2.67)	1.41 (1.11-1.79)
Fluid and electrolyte disorders			
Prevalence, %	20.7	12.1	20.6
In-hospital mortality OR (95% CI)	2.87 (2.61-3.15)	2.98 (2.46-3.60)	2.34 (1.98-2.76)
Blood loss anemia			
Prevalence, %	10.2	0.7	1.4
In-hospital mortality OR (95% CI)	0.54 (0.47-0.63)	0.46 (0.18-1.20)	1.10 (0.62-1.97)
Deficiency anemia			
Prevalence, %	16.2	7.6	10.4
In-hospital mortality OR (95% CI)	0.39 (0.33-0.45)	0.61 (0.44-0.86)	0.44 (0.32-0.60)
Alcohol abuse			
Prevalence, %	3.9	2.6	4.4
In-hospital mortality OR (95% CI)	1.22 (0.98-1.52)	1.44 (0.96-2.16)	1.43 (1.02-2.02)
Drug abuse			
Prevalence, %	1.2	0.7	1.1
In-hospital mortality OR (95% CI)	1.08 (0.70-1.66)	0.48 (0.20-1.17)	0.70 (0.28-1.78)
Psychoses			
Prevalence, %	1.5	0.8	1.3
In-hospital mortality OR (95% CI)	1.02 (0.69-1.50)	0.24 (0.05-1.09)	0.75 (0.32-1.77)
Depression			
Prevalence, %	6.4	3.7	4.6
In-hospital mortality OR (95% CI)	0.46 (0.33-0.63)	0.45 (0.22-0.88)	0.49 (0.27-0.89)

Abbreviations: CI, confidence interval; NA, not applicable; OR, odds ratio.
[a] Coefficients for each comorbidity adjusted for age, sex, race, primary health insurer, emergency admission, hospital characteristics, and the remaining comorbidities.
[b] Moderate and severe liver disease comorbidity was removed from the adjusted model for hepatic resection.

and electrolyte disorders (20.6%), and diabetes (19.8%) were the more prevalent comorbidities in the Elixhauser algorithm.

Coagulopathy (OR, 6.04; 95% CI 4.82-7.57), renal failure (OR, 4.38; 95% CI 3.00-6.38), fluid and electrolyte disorders (OR, 2.34; 95% CI, 1.98-2.76), and congestive heart failure (OR, 2.67; 95% CI 2.01-3.55) were the most important independent predictors of the in-hospital mortality in the latter algorithm.

As shown in eTable 1 (http://www.archsurg.com), mortality, together with LOS and hospital charges, increased with the number of comorbidities according to both the Elixhauser and Charlson/Deyo algorithms in each of the 3 cohorts. The majority (62%-68%) of patients in the 3 cohorts also had 3 or more Elixhauser or Charlson/Deyo comorbidities.

For the gastric resection cohort, LOS and charges increased proportionally with the number of comorbidities. The mortality rate ranged from 1.6% to 2% for patients without comorbidities to 8.7% to 10.3% for patients with 3 or more comorbidities. Similar trends were observed in the pancreatic resection cohorts.

Interestingly, for hepatic resections, there was no significant clinical difference for LOS and charges. How-

Table 4. C Statistics (95% CI) for Predicting In-Hospital Mortality of the Risk Adjustment Models

Patient Group	C Statistic (95% CI)			
	Charlson/Deyo Algorithm	Elixhauser Algorithm	Disease Staging	APR-DRGs Risk of Mortality Subclass
Gastric resection				
Overall (n = 40 746)	0.792 (0.783-0.801)	0.847 (0.839-0.855)	0.903 (0.896-0.909)	0.941 (0.937-0.946)
Primary diagnosis (n = 12 575)	0.805 (0.784-0.826)	0.833 (0.813-0.853)	0.899 (0.882-0.917)	0.940 (0.929-0.952)
Secondary diagnosis (n = 28 171)	0.796 (0.786-0.806)	0.854 (0.845-0.863)	0.902 (0.895-0.910)	0.941 (0.936-0.946)
Hepatic resection				
Overall (n = 15 279)	0.757 (0.736-0.777)	0.810 (0.792-0.828)	0.857 (0.841-0.872)	0.937 (0.927-0.947)
Primary diagnosis (n = 13 003)	0.735 (0.710-0.759)	0.801 (0.778-0.824)	0.861 (0.843-0.879)	0.945 (0.933-0.956)
Secondary diagnosis (n = 2276)	0.809 (0.778-0.841)	0.847 (0.821-0.873)	0.854 (0.827-0.881)	0.903 (0.884-0.922)
Transplant patients (n = 20 056)	0.731 (0.715-0.747)	0.774 (0.759-0.789)	0.836 (0.823-0.850)	0.901 (0.891-0.912)
Pancreatic resection				
Overall (n = 13 236)	0.741 (0.722-0.759)	0.811 (0.794-0.828)	0.868 (0.855-0.881)	0.922 (0.911-0.932)
Primary diagnosis (n = 11 600)	0.755 (0.733-0.776)	0.825 (0.806-0.844)	0.849 (0.833-0.866)	0.923 (0.911-0.935)
Secondary diagnosis (n = 1636)	0.737 (0.702-0.771)	0.797 (0.764-0.830)	0.891 (0.868-0.915)	0.913 (0.893-0.934)
Transplant patients (n = 14 739)	0.752 (0.734-0.770)	0.811 (0.794-0.827)	0.873 (0.860-0.886)	0.923 (0.913-0.933)

Abbreviations: APR-DRGs, All Patient Refined Diagnosis-Related Groups; CI, confidence interval.

ever, mortality was inversely proportional to the number of comorbidities in the Charlson/Deyo algorithm.

DISEASE STAGING AND APR-DRGs

Additional analyses were performed to assess the relationships between Disease Staging (the Predicted Mortality Scale score quartile) and the 4 APR-DRGs risk of mortality subclasses with clinical outcomes (eTable 2). As observed in the nonproprietary algorithms, increased mortality risk or subclass score (Disease Staging and APR-DRGs, respectively) was associated with older age, longer LOS, higher hospital charges, and mortality. For the 3 cohorts, those outcomes were more pronounced among higher values of subclass scores. For example, mortality rates according to APR-DRGs ranged from 0.2% to 0.4% for the lowest quartile to 43.2% to 55.2% for the highest quartile. Using the Disease Staging scale, mortality rates ranged from 0.2% to 1.3% in the lowest quartile to 13.0% to 20.8% in the highest quartile.

MODEL PERFORMANCE

Table 4 provides C statistics for in-hospital mortality for each of the risk adjustment methods in the different resection cohorts. All risk-adjustment models were adjusted for age, sex, race, insurer, admission status, and hospital characteristics.

GASTRIC RESECTION COHORT

The C statistics were 0.792 (95% CI, 0.783-0.801), 0.847 (95% CI, 0.839-0.855), 0.903 (95% CI, 0.896-0.909), and 0.941 (95% CI, 0.937-0.946) for the Charlson/Deyo algorithm, Elixhauser algorithm, Disease Staging, and APR-DRGs risk of mortality subclasses, respectively. Similar findings were observed in patients whether gastric resection was their primary or secondary indication of admission. In the 3 groups of patients, the highest accuracy was observed in the APR-DRGs scale. The Elixhauser model was more predictive than the Charlson/Deyo algorithm ($P < .001$ for all comparisons).

HEPATIC RESECTION COHORT

The C statistics were 0.757 (95% CI, 0.736-0.777), 0.810 (95% CI, 0.792-0.828), 0.857 (95% CI, 0.841-0.872), and 0.937 (95% CI, 0.927-0.947) for the Charlson/Deyo algorithm, Elixhauser algorithm, Disease Staging, and APR-DRGs risk of mortality subclasses, respectively. The same trend of the C statistics was observed whether hepatic resection was the primary or secondary indication for admission or in transplant patients. The APR-DRGs algorithm was the most accurate scale and as compared with the Charlson/Deyo algorithm, the Elixhauser model was more accurate in all subgroups.

PANCREATIC RESECTION COHORT

For the whole cohort, the C statistics were 0.741 (95% CI, 0.722-0.759), 0.811 (95% CI, 0.794-0.828), 0.868 (95% CI, 0.855-0.881), and 0.922 (95% CI, 0.911-0.932) for the Charlson/Deyo algorithm, Elixhauser algorithm, Disease Staging, and APR-DRGs risk of mortality subclasses, respectively. Similar trends were observed in subgroup analyses (admissions with primary or secondary indications or in transplant patients). As observed in the other 2 types of resections, the APR-DRGs scale was the most accurate scale, and the Elixhauser index had higher accuracy than the Charlson/Deyo algorithm.

AGREEMENT BETWEEN RISK ADJUSTMENT METHODS

Table 5 shows the percentage of patients with similar, and very different, predicted probabilities of death for pairs of risk adjustment methods, as well as the percentage of patients who died within each group. Agreement between methods relied on the type of measures (propri-

Table 5. Agreement Between Relative Predicted Probabilities of Death by Pairs of Risk Adjustment Measures

Cohort	Risk Adjustment Method A	Risk Adjustment Method B	Patients, % (Patients Who Died, %): A's Prediction Similar to B's	A's Prediction Much Higher Than B's Prediction	B's Prediction Much Higher Than A's Prediction
Gastric	Charlson/Deyo	Elixhauser	69.6 (6.5)	15.0 (2.9)	15.5 (6.2)
	Charlson/Deyo	Disease Staging	72.9 (6.1)	13.0 (1.7)	14.1 (8.9)
	Charlson/Deyo	APR-DRGs	70.3 (6.2)	13.7 (0.6)	15.9 (9.4)
	Elixhauser	Disease Staging	66.5 (7.1)	15.8 (2.1)	17.7 (4.9)
	Elixhauser	APR-DRGs	66.4 (7.2)	15.9 (0.8)	17.7 (5.6)
	Disease Staging	APR-DRGs	77.0 (6.8)	9.4 (1.3)	13.6 (4.0)
Hepatic	Charlson/Deyo	Elixhauser	63.9 (5.3)	16.6 (1.9)	19.5 (3.4)
	Charlson/Deyo	Disease Staging	59.6 (5.3)	18.9 (1.7)	21.5 (4.3)
	Charlson/Deyo	APR-DRGs	43.8 (7.0)	27.5 (0.5)	28.7 (4.1)
	Elixhauser	Disease Staging	59.9 (5.8)	19.1 (1.2)	21.0 (3.1)
	Elixhauser	APR-DRGs	49.3 (6.8)	26.8 (0.4)	24.0 (3.9)
	Disease Staging	APR-DRGs	66.1 (5.6)	15.4 (0.5)	18.5 (3.2)
Pancreatic	Charlson/Deyo	Elixhauser	77.1 (5.9)	12.1 (2.9)	10.8 (9.2)
	Charlson/Deyo	Disease Staging	69.5 (6.0)	13.7 (1.4)	16.8 (9.2)
	Charlson/Deyo	APR-DRGs	57.7 (6.9)	20.4 (0.4)	21.9 (8.6)
	Elixhauser	Disease Staging	66.5 (6.6)	15.3 (3.2)	18.2 (5.8)
	Elixhauser	APR-DRGs	60.2 (7.5)	19.2 (0.7)	20.6 (6.1)
	Disease Staging	APR-DRGs	74.6 (6.6)	10.8 (1.5)	14.6 (5.7)

Abbreviation: APR-DRGs, All Patient Refined Diagnosis-Related Groups.

etary vs nonproprietary). In gastric resections, agreement between the Charlson/Deyo algorithm and both Disease Staging and APR-DRGs was 72.9% and 70.3%, while for the Elixhauser algorithm, agreement with the latter 2 measures was 66.5% and 66.4%, respectively. Agreement between the Elixhauser and Charlson/Deyo algorithms was 69.6% while the highest agreement was between Disease Staging and APR-DRGs (77.0%). The agreement between the Charlson/Deyo and Elixhauser algorithms was higher than the agreement between either of them and proprietary measures in hepatic and pancreatic resections (63.9% vs 43.8%-59.9% and 77.1% vs 57.7%-69.5%, respectively).

Patients viewed as more sickly by the nonproprietary algorithms had a lower death rate than those using the proprietary measures. For example, among gastric resections, 15.9% of patients were viewed as "sicker" by the Elixhauser algorithm (compared with the APR-DRGs model); only 0.8% died. Similarly, of the 17.7% of patients viewed as more sickly by APR-DRGs, 5.6% died. Similar trends were observed in the other 2 cohorts.

COMMENT

It is important to assess prognosis for each individual patient prior to major surgical procedures, particularly if outcomes differ depending on patient comorbidities. In our study, we described the ability of the 4 most commonly used risk adjustment algorithms to predict in-hospital mortality in patients after gastric (40 000), hepatic (15 000), and pancreatic (13 000) resections. After assessment of the performance of each algorithm, we were able to validate each of the 4 methods for this patient population, as their overall C statistics for all 3 types of resections were more than 0.73. A C statistic approximating 0.75 is considered acceptable for discrimination and validation of methods for ongoing use.

When we compared the performance of each method, the proprietary comorbidity indexes (Disease Staging and APR-DRGs) outperformed the nonproprietary ones (Charlson/Deyo and Elixhauser). These results were generally consistent across subgroups. These proprietary methods may also include some complications of care, in contrast to the other methods that include only comorbidities. This issue likely biases any comparison in favor of the proprietary methods. Additionally, although these 2 scores have consistently outperformed other algorithms in multiple studies evaluating various primary diagnoses,[1,3,4,20] their cost makes them less accessible to researchers with limited budgets. It is therefore important to consider the 2 widely available nonproprietary methods that are based on simple computer algorithms separately from the proprietary methods. We primarily wanted to scrutinize the nonproprietary scores, and all comparisons between these 2 groups were added for informative purposes.

Comparison of the 2 proprietary methods showed that APR-DRGs was more accurate than Disease Staging. Both methods had remarkably high C statistics (0.836-0.945) in all subgroups. In the comparison of nonproprietary methods, we demonstrate that even though both scores have been validated for use with these patient populations, the Elixhauser algorithm showed significantly higher accuracy among all subgroups.

In our study, we did not perform any clinical data abstraction from medical records and so we did not validate the quality of the administrative database comparing it with the "gold standard" for this patient popu-

lation. In this article, we are therefore only able to compare the 4 comorbidity indexes in their relative performance.

In addition to the comparison of risk adjustment methods, we also described how single comorbidities influence outcome in this patient population. For all 3 types of resections, renal disease, congestive heart failure, coagulopathy, and fluid and electrolyte disorders were associated with at least a 2-fold increase in odds of in-hospital mortality, as well as moderate to severe liver disease in gastric and pancreatic resection cohorts (this comorbidity was removed from the adjusted model for hepatic resection). The fluid and electrolyte disturbances were also among the most prevalent diseases, occurring in 12% to 20% of patients. Another important predictor of increased mortality risk, LOS, and hospital charges was cerebrovascular disease, which displayed a large variance in ORs among the cohorts. In gastric and pancreatic resections, cerebrovascular disease was associated with ORs of 1.69 and 2.21, respectively, but patients with cerebrovascular disease undergoing hepatic resections experienced a 10-fold increase in their risk of dying. When we compared the Charlson/Deyo and Elixhauser indexes, the biggest discrepancy was peptic ulcer disease in patients who underwent gastric resection. This was identified as a comorbidity in 22.5% of patients in the Charlson/Deyo method (OR, 3.87) but only 1.2% of patients with the Elixhauser method (OR, 0.96). This difference is likely due to the exclusion of gastric bleeding in the Elixhauser method.

Other conditions were also associated with prolonged hospital stay yet decreased in-hospital mortality rates. As described by others,[33,39,40] these findings likely represent an inverse effect of patient severity on coding of certain common but relatively unthreatening diseases. For example, a seriously ill patient with multiple serious diseases concurrent to a diagnosis of obesity is less likely to have it listed among his or her comorbidities in the discharge abstract compared with an otherwise healthy obese patient.

On comparison of the relative predicted probabilities of death, methods agreed in only 43.8% to 74.6% of scenarios. In all cohorts, patients viewed as sicker by the nonproprietary algorithms had lower rates of death compared with the evaluation of sick patients by the proprietary algorithms. A succinct rationale for this observation is currently unknown.

Our secondary end points, total hospital charges and length of hospital stay, also increased proportionally with the burden of comorbidities. This is an expected outcome as sicker patients could reasonably be expected to spend a longer time in care and incur more expenses while in the hospital.

In conclusion, we found that the selection of a particular method for comorbidity risk adjustment after gastric, hepatic, and pancreatic resections has a significant impact on analysis of in-hospital mortality. This is explained by the reality that different methods classify the same patients into different risk categories. As a result, provider-specific outcomes may be evaluated differently depending on the method selected. Finally, the importance of a particular comorbidity can also be misinterpreted because of inadequate risk adjustment.

Accepted for Publication: July 28, 2011.
Published Online: October 17, 2011. doi:10.1001/archsurg.2011.296
Correspondence: Elijah Dixon, MD, MSc, Division of General Surgery and Surgical Oncology, Faculty of Medicine, University of Calgary, North Tower, 10th Floor, Foothills Medical Centre, Department of Surgery, 1403 29 St NW, Calgary, AB T2N 2T9, Canada (elijah.dixon@albertahealthservices.ca).
Author Contributions: Drs Dixon and Shaheen had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Grendar, Shaheen, Myers, Ball, Kaplan, and Dixon. *Acquisition of data:* Myers, Ball, and Al-Manasra. *Analysis and interpretation of data:* Grendar, Shaheen, Myers, Parker, Vollmer, Ball, Quan, Kaplan, and Dixon. *Drafting of the manuscript:* Grendar, Shaheen, Myers, Ball, Al-Manasra, and Dixon. *Critical revision of the manuscript for important intellectual content:* Grendar, Myers, Parker, Vollmer, Ball, Quan, Kaplan, and Dixon. *Statistical analysis:* Shaheen, Ball, and Dixon. *Administrative, technical, and material support:* Grendar, Myers, Parker, Ball, Al-Manasra, Kaplan, and Dixon. *Study supervision:* Myers, Ball, Quan, and Dixon.
Financial Disclosure: None reported.
Online-Only Material: The eTables are available at http://www.archsurg.com.

REFERENCES

1. Myers RP, Quan H, Hubbard JN, Shaheen AA, Kaplan GG. Predicting in-hospital mortality in patients with cirrhosis: results differ across risk adjustment methods. *Hepatology*. 2009;49(2):568-577.
2. Davila JA, El-Serag HBGI. GI epidemiology: databases for epidemiological studies. *Aliment Pharmacol Ther*. 2007;25(2):169-176.
3. Hall BL, Hirbe M, Waterman B, Boslaugh S, Dunagan WC. Comparison of mortality risk adjustment using a clinical data algorithm (American College of Surgeons National Surgical Quality Improvement Program) and an administrative data algorithm (Solucient) at the case level within a single institution. *J Am Coll Surg*. 2007;205(6):767-777.
4. Gordon HS, Johnson ML, Wray NP, et al. Mortality after noncardiac surgery: prediction from administrative versus clinical data. *Med Care*. 2005;43(2):159-167.
5. Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM, Spinelli JJ. Comorbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? *J Clin Epidemiol*. 2000;53(4):343-349.
6. Ma C, Crespin M, Proulx MC, et al. Accuracy of administrative data in identifying ulcerative colitis patients presenting with acute flare and undergoing colectomy [abstract]. *Am J Gastroenterol*. 2010;105:S467 Abstract 1270.
7. Quan H, Khan N, Hemmelgarn BR, et al; Hypertension Outcome and Surveillance Team of the Canadian Hypertension Education Programs. Validation of a case definition to define hypertension using administrative data. *Hypertension*. 2009;54(6):1423-1428.
8. Quan H, Li B, Saunders LD, et al; IMECCHI Investigators. Assessing validity of *ICD-9-CM* and *ICD-10* administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 2008;43(4):1424-1441.
9. Luthi JC, Troillet N, Eisenring MC, et al. Administrative data outperformed single-day chart review for comorbidity measure. *Int J Qual Health Care*. 2007;19(4):225-231.
10. Quan H, Parsons GA, Ghali WA. Validity of procedure codes in *International Classification of Diseases, 9th revision, Clinical Modification* administrative data. *Med Care*. 2004;42(8):801-809.
11. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived from *ICD-9-CCM* administrative data. *Med Care*. 2002;40(8):675-685.

12. Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract*. 2010;89(2):189-195.
13. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. 2009;9:5.
14. Clement FM, James MT, Chin R, et al; Alberta Kidney Disease Network. Validation of a case definition to define chronic dialysis using outpatient administrative data. *BMC Med Res Methodol*. 2011;11:25.
15. So L, Beck CA, Brien S, et al. Chart documentation quality and its relationship to the validity of administrative data discharge records. *Health Informatics J*. 2010; 16(2):101-113.
16. Bozic KJ, Chiu VW, Takemoto SK, et al. The validity of using administrative claims data in total joint arthroplasty outcomes research. *J Arthroplasty*. 2010;25 (6)(suppl):58-61.
17. Southern DA, Roberts B, Edwards A, et al. Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. *Can J Public Health*. 2010;101(1):61-64.
18. Jetté N, Reid AY, Quan H, Hill MD, Wiebe S. How accurate is ICD coding for epilepsy? *Epilepsia*. 2010;51(1):62-69.
19. Welke KF, Diggs BS, Karamlou T, Ungerleider RM. Comparison of pediatric cardiac surgical mortality rates from national administrative data to contemporary clinical standards. *Ann Thorac Surg*. 2009;87(1):216-223.
20. Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation*. 2007;115(12):1518-1527.
21. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
22. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care*. 2004;42(4):355-360.
23. Farley JF, Harley CR, Devine JW. A comparison of comorbidity measurements to predict healthcare expenditures. *Am J Manag Care*. 2006;12(2):110-119.
24. Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001;154(9):854-864.
25. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with *ICD-9-CM* administrative databases. *J Clin Epidemiol*. 1992;45(6):613-619.
26. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
27. HCUP Databases. Healthcare Cost and Utilization Project (HCUP) Web site. www.hcup-us.ahrq.gov/databases.jsp. Accessed February 2010.
28. Whalen D, Houchens R, Elixhauser A. 2004 HCUP Nationwide Inpatient Sample (NIS) comparison report. HCUP Method Series Report 2007-03. Healthcare Cost and Utilization Project (HCUP) Web site. http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp. Published December 2, 2006. Accessed February 2010.
29. Consumer Price Index: all urban consumers (US medical care 1982-1984=100). US Dept of Labor Bureau of Labor Statistics Web site. http://www.bls.gov/cpi/home.htm. Accessed February 2010.
30. Thomson Medstat Inc. Medstat Disease Staging Software version 5.24: reference guide. http://www.hcup-us.ahrq.gov/db/nation/nis/Disease%20Staging%20V5.24%20Reference%20Guide.pdf. Published November 2006. Accessed February 2010.
31. Averill RF, Goldfield N, Hughes JS, et al. All Patient Refined Diagnosis-Related Groups (APR-DRGs) version 20.0: methodology overview. http://www.hcup-us.ahrq.gov/db/nation/nis/APR-DRGsV20MethodologyOverviewandBibliography.pdf. Published July 2003. Accessed February 2010.
32. Thombs BD, Singh VA, Halonen J, Diallo A, Milner SM. The effects of preexisting medical comorbidities on mortality and length of hospital stay in acute burn injury: evidence from a national sample of 31,338 adult patients. *Ann Surg*. 2007; 245(4):629-634.
33. Shwartz M, Ash AS. Evaluating risk-adjustment models empirically. In: Iezzoni LI, ed. *Risk Adjustment for Measuring Health Outcomes*. 3rd ed. Chicago, IL: Health Administration Press; 2003:231-273.
34. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. *Ann Intern Med*. 1995;123(10):763-770.
35. Iezzoni LI, Ash AS, Shwartz M, Landon BE, Mackiernan YD. Predicting in-hospital deaths from coronary artery bypass graft surgery: do different severity measures give different predictions? *Med Care*. 1998;36(1):28-39.
36. Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. *Med Decis Making*. 1996;16(4):348-356.
37. Shwartz M, Iezzoni LI, Ash AS, Mackiernan YD. Do severity measures explain differences in length of hospital stay? the case of hip fracture. *Health Serv Res*. 1996;31(4):365-385.
38. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.
39. Hughes JS, Iezzoni LI, Daley J, Greenberg L. How severity measures rate hospitalized patients. *J Gen Intern Med*. 1996;11(5):303-311.
40. Jencks SF, Williams DK, Kay TL. Assessing hospital-associated deaths from discharge data: the role of length of stay and comorbidities. *JAMA*. 1988;260 (15):2240-2246.

INVITED CRITIQUE

How Should We Risk-Adjust Hospital Outcome Comparisons?

The need for reliable measures of surgical quality is at an all-time high. Most experts agree the best way to assess surgical quality is to directly measure the outcomes of care. However, to compare outcomes across hospitals, differences in patient risk must be adequately accounted for. Selecting appropriate risk adjustment models is of crucial importance. Otherwise, hospitals and surgeons will not be fairly judged. In this study, Grendar and colleagues[1] compare several different risk adjustment models for use with administrative data. These models rely on using secondary billing codes to identify factors (eg, coexisting diseases) associated with adverse outcomes.

Although Grendar and colleagues use sophisticated statistical analyses of 4 different algorithms, their study does not directly tell us how good these models would be if applied in the real world.

First, they did not have a "gold standard" risk adjustment method in these comparisons. To assess whether these models perform well, they should be compared with models developed using data abstracted directly from the medical record. For example, the American College of Surgeons National Surgical Quality Improvement Program uses trained nurse abstractors to record information on patient risk factors. Based on the analyses in the Grendar et al article, it is unclear how these 3 models compare with this approach. In other words, we know which model is "better," but how do we know if it is "good enough"?

Second, Grendar and colleagues did not perform a hospital-level analysis. Their analysis focused on patient-level risk prediction. These analyses focus on the C statistic, which assesses discrimination, and predicted