Articles

Using Medicare Claims Data to Assess Provider Quality for CABG Surgery: Does It Work Well Enough?

Edward L. Hannan, Michael J. Racz, James G. Jollis, and Eric D. Peterson

Objectives. To assess the relative abilities of clinical and administrative data to predict mortality and to assess hospital quality of care for CABG surgery patients.
Data Sources/Study Setting. 1991–1992 data from New York's Cardiac Surgery Reporting System (clinical data) and HCFA's MEDPAR (administrative data).
Study Design/Setting/Sample. This is an observational study that identifies significant risk factors for in-hospital mortality and that risk-adjusts hospital mortality rates using these variables. Setting was all 31 hospitals in New York State in which CABG surgery was performed in 1991–1992. A total of 13,577 patients undergoing isolated CABG surgery who could be matched in the two databases made up the sample.
Main Outcome Measures. Hospital risk-adjusted mortality rates, identification of "outlier" hospitals, and discrimination and calibration of statistical models were the main outcome measures.
Principal Findings. Part of the discriminatory power of administrative statistical models resulted from the miscoding of postoperative complications as comorbidities. Removal of these complications led to deterioration in the model's C index (from C = .78 to C = .71 and C = .73). Also, provider performance assessments changed considerably when complications of care were distinguished from comorbidities. The addition of a couple of clinical data elements considerably improved the fit of administrative models. Further, a clinical model based on Medicare CABG patients yielded only three outliers, whereas eight were identified using a clinical model for all CABG patients.
Conclusions. If administrative databases are used in outcomes research, (1) efforts to distinguish complications of care from comorbidities should be undertaken, (2) much more accurate assessments may be obtained by appending a limited number of clinical data elements to administrative data before assessing outcomes, and (3) Medicare data may be misleading because they do not reflect outcomes for all patients.

Key Words. Administrative data, quality assessment, CABG surgery, Medicare data

In recent years, many efforts have been made to assess the quality of medical care using outcomes data. One of the most frequently explored areas for this research has been coronary artery bypass graft (CABG) surgery. In part this is because of the relative uniformity of the patients, the frequency with which the operation is performed, and the cost of the procedure. These studies have generally involved "isolated" CABG surgery; that is, no other major cardiac surgery (such as valve surgery) is performed during the course of the same hospital visit.

Studies that have assessed provider performance for CABG surgery generally have used clinical data either collected concurrently by cardiac surgery departments or abstracted from medical records (Tu, Jaglal, Naylor, et al. 1995; O'Connor, Plume, Olmstead, et al. 1992; Pennsylvania Health Care Cost Containment Council 1994; Hannan, Kilburn, Racz, et al. 1994). However, much of the information that is used for risk adjustment is typically available in administrative databases used for reimbursement or planning purposes. These databases usually contain demographics; the principal diagnosis and a limited number of secondary diagnoses; the primary procedure and a limited number of secondary procedures; provider identifiers; and admission, discharge, and surgery dates.

Administrative data are generally not used for risk adjustment in CABG surgery because, in the opinion of many experts, clinical risk factors that are not diagnoses (and therefore not in administrative databases) are essential to account fully for operative risk. These clinical risk factors include whether the patient has undergone previous open heart surgery, and ejection fraction, which is a measure of the left ventricle's ability to pump blood.

Other limitations of administrative data include the limited number of secondary diagnoses (and therefore risk factors) available in administrative databases, the inability of administrative databases to distinguish between comorbidities and complications, and the inflexibility of administrative databases in defining risk factors. It is important to distinguish between

---

This study was supported in part by grant HS06503 from the Agency for Health Care Policy and Research of the U.S. Department of Health and Human Services, Washington, DC.

Address correspondence and requests for reprints to Edward L. Hannan, Ph.D., Professor and Chair, Department of Health Policy and Management, School of Public Health, State University of New York at Albany, One University Place, Rensselaer, NY 1244-3456. Michael J. Racz, M.A. is Principal Research Support Specialist in the Department of Health Policy and Management, SUNY-Albany; James G. Jollis, M.D. and Eric D. Peterson, M.D., M.P.H. are Assistant Professors of Medicine at Duke University Medical Center. This article, submitted to *Health Services Research* on October 6, 1995, was revised and accepted for publication on May 28, 1996.

comorbidities (problems present at admission to the hospital) and complications (problems arising during the hospital stay) because comorbidities should be used in the risk-adjustment model, whereas complications are typically not used because they may reflect the quality of care provided. The inflexibility of administrative databases results from the rules that govern the coding of administrative data, the reliance on ICD-9-CM codes, and the inability to define subgroups of patients who may be at a higher risk. For example, there is an ICD-9-CM code for "acute myocardial infarction (heart attack)," but this code specifies only that the heart attack occurred in the previous eight weeks. Patients who have suffered a heart attack within a few hours prior to the procedure are at greatly increased risk compared to patients suffering a heart attack within a few days prior to the procedure, and administrative data are unable to separate these two groups of patients.

Despite these potential disadvantages of administrative data, there are compelling reasons to consider using them. They are far less expensive than clinical data in that no need exists to develop forms or data entry systems, and there is no ongoing expense in collecting data because they are already being collected for other reasons. With regard to available administrative databases, numerous states have hospital discharge databases that contain hospital discharge data for all patients discharged from acute care hospitals. However, the most well known and frequently used administrative database is the federal Health Care Financing Administration's (HCFA) Medicare Provider Analysis and Review File (MEDPAR), which contains hospital discharge information for all Medicare patients. This database is available for public use and has been employed by numerous researchers. Information contained in MEDPAR has been the basis of the HCFA mortality reports that were issued between 1987 and 1991 (Health Care Financing Administration 1991).

The purposes of this study are

1. To compare the ability of MEDPAR with the ability of a clinical database used in New York State–the Cardiac Surgery Reporting System (CSRS)–to predict in-hospital mortality for Medicare CABG patients;
2. To determine the extent to which individual hospital performance is assessed differently by the MEDPAR versus CSRS statistical models;
3. To determine how different the MEDPAR statistical model would be if CSRS information were used to distinguish between complications and comorbidities among secondary diagnoses reported in MEDPAR, and to eliminate all complications from the diagnoses

used in the MEDPAR model; also, to compare hospital assessments for the original and enhanced MEDPAR models;

4. To explore the extent to which MEDPAR could be improved by adding a small group of clinical data elements contained in CSRS to MEDPAR; also, to determine the extent to which the differences in hospital assessments mentioned in item 2 are decreased by the new model with appended clinical risk factors; and
5. To compare statistical models and hospital assessments for two groups of patients in New York: all Medicare CABG patients and all CABG patients.

Before describing the methods used in this study, it is necessary to explain these purposes further. With regard to the third objective, it is possible that a statistical model based on MEDPAR data will predict mortality accurately, but that part of the reason for the accuracy will be that complications of care are used in the prediction formula. Although complications are even more highly correlated with mortality than comorbidities, it is unfair to use them in the predictive instrument, which is created for the purpose of adjusting for preoperative severity of illness. Consequently, a predictive formula that unwittingly uses complications is problematic from the standpoint of providing an accurate view of preoperative risk.

The reason for the fifth objective is that although MEDPAR may actually predict mortality for Medicare CABG patients very well, either with or without the added clinical data, Medicare CABG patients only comprise a portion (about 45 percent in New York) of all CABG patients. Consequently, there is a danger that hospital assessments limited to Medicare patients, however accurate they may be for that group, are not accurate assessments of the care provided to all CABG patients in a hospital.

DATA AND METHODS

DATA SOURCES

The Cardiac Surgery Reporting System (CSRS) of New York State was initiated in 1989 by the New York State Health Department and its Cardiac Advisory Committee to improve the quality of cardiac surgical care in the state by virtue of risk-adjusting adverse outcomes and reporting these outcomes to providers.

Information in the system is collected under the supervision of the directors of cardiac surgery in the 31 New York hospitals with approval to perform open heart surgery. The system contains demographics; risk factors; complications; discharge status; admission, discharge, and procedure dates; procedures performed; and provider identifiers. A list of risk factors in the system is presented in Table 1.

MEDPAR is a derivative file created by HCFA by merging information on hospital characteristics and costs with part of the Hospital Stay Record File. In addition to many other data elements, it contains information concerning demographics; up to six procedures that were performed; up to ten diagnoses (principal and nine secondaries); hospital identifier; discharge status; and admission, discharge, and procedure dates. Risk factors (diagnoses) in MEDPAR that were used as potential predictors of in-hospital mortality are presented in Table 2 along with their associated ICD-9-CM ranges. These diagnoses and ICD-9-CM ranges were identified in an earlier study (Romano, Roos, Jollis, et al. 1994).

MATCHING DATA SETS

In order to compare the ability of the two databases in predicting in-hospital mortality, we first identified a group of patients that could be matched for the 1991–1992 calendar years. All 14,228 patients in MEDPAR reported as undergoing CABG surgery in New York (procedure codes 36.10 to 36.19 and no other procedure codes indicating major open heart surgery), who were discharged between January 1, 1991 and December 31, 1992, were first identified. These patients were then matched with all patients in CSRS, discharged in the same time interval and identified as having undergone CABG surgery and no other open heart surgery (a total of 30,972), whether or not the patients were New York residents. Note that the reason there were more than twice as many patients in CSRS is that MEDPAR contains only Medicare patients. Within hospitals, patients were matched using patient identifiers and discharge date. A total of 13,577 patients (95.4 percent of the patients identified in MEDPAR) were matched.

METHODS

The first step in the analysis consisted of developing backward stepwise logistic regression models for each of the two databases. The dependent variable in each model was binary, with "1" indicating an in-hospital death, and "0" coded as a live discharge. In-hospital mortality was used as an outcome measure rather than the 30-day post-surgery measure used in some

Table 1: CSRS Candidate Variables

	Definition
Body Surface Area	
Ejection Fraction (<20%, 20–39%)	
CCS Class IV	
One or More Previous Open Heart Operations	
Previous Myocardial Infarction	
Previous Stroke	
Carotid/Cerebrovascular Disease	>50% cerebral artery obstruction, history of non-embolic stroke, or surgery for such disease
Aortoiliac Disease	Significant vascular disease in the aorta or iliac arteries, or previous surgery for such disease
Femoral/Popliteal Disease	
Hemodynamically Unstable	Pharmacologic support required for BP or output
Shock	Low BP or cardiac output despite pharmacologic support
Hypertension	BP>140/90, history or current treatment for hypertension
IV Nitro within 24 hrs.	
ECG Evidence of Left Vent. Hypertrophy	
Congestive Heart Failure	Currently or recently treated for congestive heart failure with signs such as pulmonary edema, rales, pleural effusion, or the need for digitalis and diuretics
Major Acute Structural Defect	
Persistent Ventricular Arrythmia	Persistent despite specific therapy
Calcified Aorta	At risk for thromboembolism
Chronic Obstructive Pulmonary Disease	Functionally disabled, or require bronchodilator therapy, or have a forced expiratory volume in one second less than 75% of the predicted value or less than 1.25 litres, or have a room air PO2 < 60 or a PCO2 >50.
Diabetes	Diabetes requiring medication
Hepatic Failure	Bilirubin >2 mg/dl and serum albumin <3.5 grams/dl
Renal Failure	Creatinine >2.5 or require dialysis
Immunosuppression Therapy	Use of drugs that suppress immune system
Immunoincompetent Disease	Abnormal function of the immune system
Intra-aortic Balloon Pump Pre-Op	Arrive in OR with IABP or require its insertion
Cardiac Cath “Crash”	Require immediate surgery following cardiac cath
PTCA “Crash”	Require immediate surgery following angioplasty
Previous PTCA, this admission	
PTCA before this admission	
Thrombolytic Therapy within 7 days	
Smoking History, in past two weeks	

Table 2: MEDPAR Candidate Variables

	ICD-9 Codes
Acute Myocardial Infarction	410.x
Congestive Heart Failure	402.01, 402.11, 402.91, 402.92, 425.0, 428.x
Cerebrovascular Disease	362.34, 430–438, 781.4, 784.3, 997.0
Diabetes	250.x
Intra-aortic Balloon Pump	37.61
Liver Disease	456.0,465.1,456.2x
Chronic Obstructive Pulmonary Disease	490–494
Peripheral Vascular Disease	440.2, 441.x, 443.9, 444, 447.1, 785.4
Renal Disease	582.x, 583.0–583.7, 588.0–588.9
Renal Failure	585, V42.0, V45.1, V56
PTCA During Hospital Stay	36.01,36.02,36.05
Unstable Angina	411.1

studies, because the vast majority (94 percent) of patients who die within 30 days of CABG surgery are inpatient deaths, and because many patients die in the hospital more than 30 days after surgery of problems related to the surgery.

Candidate independent variables for CSRS and MEDPAR were age, sex, and the risk factors presented in Tables 1 and 2, respectively. In each model, the stepwise technique consisted of retaining only those variables that were significant for $p < .05$.

Both models were cross-validated by splitting the cases in half, developing a stepwise model on the first half of the data with $p < .10$, and then determining if the significant variables for the first half of the data were also significant for the other half (for $p < .10$). If not, they were eliminated. Variables that proved to be significant in both halves were then used in a stepwise model on the entire data set with $p < .05$. The discrimination of each model was assessed using the C statistic, and the calibration of each model was initially tested using the Hosmer-Lemeshow (H-L) statistic. However, particularly for the MEDPAR models, the H-L statistic was not found to be a reliable statistic. Frequently, the number of patients with the same predicted probability of death was high because of the small number of risk factors. Consequently, the decision of which patients to place in a given decile was arbitrary, and the value of the H-L statistic ranged from very good to very poor depending on this decision.

After comparing the discrimination of the two models, the next step consisted of calculating risk-adjusted mortality rates (RAMRs) for each of New York's 31 hospitals in which CABG surgery is performed, and comparing

RAMRs for the two models. The risk-adjusted mortality rate for each hospital was calculated by dividing its observed mortality rate by its expected mortality rate, and then multiplying this quotient by the overall mortality rate for all 13,577 patients. To obtain a hospital's expected mortality rate from one of the models, the predicted probabilities of death for all patients were summed and then divided by the number of patients. Hospitals were labeled as "outliers" if their RAMR was significantly ($p < .05$) higher or lower than the overall statewide mortality rate. Confidence intervals for RAMRs were also calculated (Breslow and Day 1991).

The next set of analyses consisted of examining the potential bias caused by the MEDPAR data set's inability to distinguish between comorbidities and complications for the two diagnoses (acute myocardial infarction and renal failure) that could be either a comorbidity or a complication, and for the one procedure (intra-aortic balloon pump) that could have been a risk factor (if inserted prior to the CABG surgery) or a complication (if necessitated by the CABG surgery). Note that although a date is present for the intra-aortic balloon pump (IABP) insertion, it is frequently the day of CABG surgery, and therefore is not indicative of whether the pump was inserted before the surgery or not.

Two different methods were employed using CSRS data to prevent complications from being unwittingly used as comorbidities in the MEDPAR model. First, if acute myocardial infarction (AMI) or renal failure appeared as a risk factor in MEDPAR but was not reported as a risk factor in CSRS, it was removed as a risk factor in MEDPAR. Similarly, if IABP was reported as a procedure in MEDPAR, but preoperative IABP was not reported as a risk factor in CSRS, it was removed as a risk factor in MEDPAR.

The second method for weeding out potential complications from what were initially treated as risk factors in the MEDPAR data consisted of eliminating a diagnosis (AMI or renal failure) as a risk factor in MEDPAR if it was reported as a complication in CSRS. Similarly, IABP was omitted as a risk factor in MEDPAR if it was reported as a complication and was not reported as a preoperative risk factor in CSRS.

After risk factors were removed from some of the MEDPAR cases using CSRS information, new MEDPAR models were developed. These models were compared with the original MEDPAR model to assess the extent to which the original MEDPAR model was influenced by inaccurate use of complications as risk factors. Also, these models were compared with the corresponding CSRS model in order to determine whether the difference between the predictive ability of CSRS and MEDPAR was considerably larger after complications were removed from the MEDPAR risk factors.

The next analyses compared RAMR correlations and outlier hospitals when CSRS data were applied to two different patient populations–all Medicare CABG patients in New York in 1991–1992, and all CABG patients in New York in 1991–1992. The purpose of this analysis was to determine if hospitals' overall performance for CABG surgery was consistent with their performance for Medicare CABG surgery patients, which is the only information available in the MEDPAR database.

The final investigation was aimed at determining whether the ability of MEDPAR to predict mortality for CABG surgery could be appreciably improved by adding a limited amount of clinical information to the administrative database. Thus, clinical data elements available in CSRS that were not available in MEDPAR (ejection fraction and previous open heart surgery) were added to the MEDPAR database. These two elements were chosen because they were among the most important predictors of CABG surgery mortality and because they are more reproducible (reliable) than many of the other measures.

As with the previous analyses, discrimination, correlations in RAMRs, and outliers were all used as criteria. The original MEDPAR, the appended MEDPAR, and CSRS for Medicare patients were all compared to determine the extent to which the appended MEDPAR bridged the gap between the original MEDPAR and CSRS in predictive ability.

RESULTS

Table 3 presents the prevalences and in-hospital mortality rates for each of the risk factors contained in both MEDPAR and CSRS, along with the kappa values for each risk factor. According to Landis and Koch (1977), kappa values of < .2, .21–.40, .41–.60, .61–.80, and .81–1.00 represent agreement levels of "slight," "fair," "moderate," "substantial," and "almost perfect," respectively. By these standards, there is a slight agreement between the databases for renal failure, fair agreement for acute myocardial infarction and cerebrovascular disease, moderate agreement for intra-aortic balloon pumps and congestive heart failure, substantial agreement for diabetes, and almost perfect agreement for in-hospital mortality and gender. For in-hospital mortality, for which perfect agreement is especially desirable, the differences were almost entirely due to undercoding in the CSRS at a time when some hospitals were

Table 3: In-Hospital Frequency of Risk Factor and Mortality Coding in CSRS and MEDPAR

	MEDPAR		*CSRS*		*MEDPAR and CSRS*		
	%	*Mortality Rate(%)*	*%*	*Mortality Rate(%)*	*%*	*Mortality Rate(%)*	*Kappa*
In-hospital Death	3.96		3.83		3.82		0.98
Female Gender	32.9	4.9	32.7	4.9	32.4	4.9	0.98
Intra-aortic Balloon Pump	4.6	25.2	3.3	17.9	2.1	16.8	0.51
Renal Failure	0.6	18.8	3.1	16.9	0.4	24.0	0.19
Acute MI*	23.2	6.4	60.2	5.0	21.0	6.5	0.25
Cerebrovascular Disease†	8.1	9.5	16.9	7.3	5.2	8.3	0.34
Diabetes	19.7	3.1	23.5	5.2	16.0	3.2	0.67
Congestive Heart Failure	14.3	9.0	17.5	9.1	8.7	10.3	0.46

* Acute Myocardial Infarction (Generally described in administrative databases as an MI during the hospital stay or up to eight weeks prior. In CSRS, this preoperative risk factor can be coded as MI within 6 hours prior to CABG procedure, MI within 6–23 hours prior to procedure, MI within 1–20 days prior to procedure, or MI 21 or more days prior to the procedure. The 60.2% reported above in CSRS for this risk factor includes any patient with one of these fields coded. In other words, CSRS captures all patients with an MI any time prior to the CABG procedure).

† Coded in CSRS as Previous Stroke and/or Carotid/Cerebrovascular Disease (see Table 1).

still not reporting in-hospital deaths that they concluded were not cardiac-related. These errors were corrected by matching with the Department's administrative database, SPARCS; and in the subsequent analyses, CSRS and MEDPAR match with regard to which patients died.

For various risk factors, lack of agreement was not necessarily an indication that coding in one of the databases was inaccurate. Instead, the differences may well have been primarily a function of different definitions. For example, renal failure had the least agreement. However, renal failure was defined as requiring dialysis or having high creatinine levels in CSRS, whereas it was defined in MEDPAR as kidney transplant, requiring dialysis, or chronic renal failure.

Table 4 presents odds ratios and p-values for significant risk factors for each of eight statistical models. Model M1 is based on all MEDPAR data defined in Table 2; M2 uses MEDPAR data with AMI, renal failure, and IABP suppressed as risk factors when they did not appear in CSRS as a risk factor; model M3 suppresses AMI, renal failure, and IABP if they appeared in CSRS as a complication. Models M4, M5, and M6 are enhancements of models M1, M2, and M3, respectively, in which information on ejection fraction (coded as less than 20 percent, 20–39 percent, and 40 percent and higher) and on whether the patient had previous open heart operations is

Table 4: Odds Ratios for CSRS and MEDPAR Models for Predicting Mortality of New York CABG Patients

		C1	*M1*	*M2*	*M3*	*M4*	*M5*	*M6*	*C2*
Risk Factors Statistically Significant in At Least One Administrative Database	Age	1.07	1.08	1.08	1.07	1.08	1.08	1.08	1.04
	Female	–	–	–	–	1.35#	1.37*	1.41*	1.52
	AMI†	–	1.46	1.60	–	1.40*	1.48	–	1.95*
	CHF	1.45*	2.07	2.35	2.46	1.73	1.84	1.90	1.35*
	Cerebro. Disease‡	1.78	2.83	2.83	2.85	2.90	2.90	2.94	1.64
	Renal Failure	2.87	3.68	–	3.58	3.70	–	3.68	2.71
	IABP	–	9.74	4.35	7.38	7.72	3.56	6.02	–
Risk Factors Statistically Significant ONLY in Clinical Database	EF<20%	1.94^	–	–	–	2.89	3.78	3.66	2.68
	20%<EF<39%	1.62	–	–	–	1.73	1.97	1.96	1.61
	Prev. Operations	2.81	–	–	–	2.53	3.18	3.05	3.33
	CCS Class IV	1.64	–	–	–	–	–	–	1.57
	Aortoiliac Disease	–	–	–	–	–	–	–	1.48*
	Hemo. Unstable	3.64	–	–	–	–	–	–	3.13
	Shock	13.44	–	–	–	–	–	–	8.10
	LV Hypertrophy	1.35#	–	–	–	–	–	–	1.33*
	Vent. Arrhythmia	–	–	–	–	–	–	–	1.70*
	Calcified Aorta	1.53#	–	–	–	–	–	–	1.65
	Diabetes	–	–	–	–	–	–	–	1.34*
	Hepatic Failure	7.41*	–	–	–	–	–	–	4.34#
	C Stat.	.789	.777	.709	.732	.796	.754	.773	.813

M1 = MEDPAR data.

M2 = M1 data in which AMI, Renal Failure, and IABP are not included as risk factors if they are not coded as risk factors in CSRS.

M3 = M1 data in which AMI, Renal Failure, and IABP are not included as risk factors if they are coded as complications in CSRS.

M4, M5, M6 = M1, M2, M3 data, respectively, with clinical data elements Previous Open Heart Surgery and Ejection Fraction added.

C1 = CSRS data on all NY Medicare CABG patients in 1991–1992.

C2 = CSRS data on all NY CABG patients in 1991–1992.

†Significant in clinical database as MI within 6 hours previous to CABG procedure.

‡Significant in clinical database as carotid/cerebrovascular disease.

* Denotes $.0001 < p \leq .001$; # denotes $.001 < p \leq .01$; ^denotes $.01<p<.05$; All other p-values are $<.0001$.

assumed to be available. Model C1 is based on CSRS for Medicare patients only and Model C2 is based on the CSRS database applied to all CABG surgery patients in New York, not just Medicare patients.

Table 5 presents hospital risk-adjusted mortality rates and their ranks based on each of the eight models (note that hospitals are numbered according to model C1's RAMR rank). The following is a series of analyses that compare various groups of these statistical models.

Table 5: Hospital Risk-Adjusted Mortality Rates (Ranks) for CSRS and MEDPAR Models

Hosp.	*Vol.*	*C1*	*M1*	*M2*	*M3*	*M4*	*M5*	*M6*	*C2*
1	130	1.59	0.89(1)	0.88(1)	0.94(1)	1.01(1)	1.02(1)	1.08(1)	2.65(11)
2	171	1.94	1.74(2)	2.47(2)	2.05(2)	1.97(2)	2.79(4)	2.34(2)	1.78(1)
3	612	2.55#	4.43(22)	4.05(16)	4.02(16)	4.08(15)	3.65(11)	3.68(13)	2.00(3)#
4	532	2.56	2.82(4)	2.56(3)	2.66(3)	2.89(3)	2.67(2)	2.80(3)	2.01(4)
5	405	2.60	3.15(9)	3.08(7)	3.19(8)	3.31(10)	3.33(10)	3.39(9)	1.91(2)#
6	166	2.73	2.92(5)	2.78(5)	2.82(5)	3.07(8)	3.00(5)	3.03(6)	2.06(5)
7	1314	2.82#	3.09(8)	2.90(6)	2.93(6)	3.03(6)	2.77(3)#	2.83(4)#	2.10(6)#
8	668	3.07	3.90(12)	3.58(10)	3.70(12)	3.39(12)	3.06(7)	3.13(7)	2.44(9)
9	352	3.29	2.98(6)	3.11(8)	3.38(10)	3.02(5)	3.20(9)	3.45(10)	2.20(8)
10	554	3.69	4.29(18)	3.69(12)	3.94(15)	4.22(16)	3.67(12)	3.89(15)	3.15(17)
11	243	3.75	2.79(3)	3.83(15)	3.32(9)	2.98(4)	4.04(16)	3.56(11)	2.56(10)
12	456	3.78	4.11(13)	3.71(13)	3.91(14)	4.26(17)	3.85(14)	4.10(18)	3.08(14)
13	473	3.82	3.19(10)	3.16(9)	3.00(7)	3.05(7)	3.04(6)	2.91(5)	3.28(20)
14	590	3.93	4.38(21)	4.34(20)	4.41(24)	3.90(13)	3.78(13)	3.83(14)	3.00(13)
15	686	4.26	3.28(11)	4.21(17)	3.87(13)	3.16(9)	3.92(15)	3.62(12)	2.11(7)#
16	335	4.27	4.60(24)	4.49(21)	4.22(18)	4.75(23)	4.65(21)	4.37(19)	3.43(22)
17	1002	4.30	4.12(14)	4.29(19)	4.25(19)	3.95(14)	4.17(17)	4.09(17)	2.94(12)
18	309	4.52	2.99(7)	2.67(4)	2.80(4)	3.39(11)	3.13(8)	3.26(8)	3.12(16)
19	480	4.68	4.26(17)	3.81(14)	4.39(22)	4.62(21)	4.44(19)	5.05(25)	3.28(19)
20	790	4.74	4.18(16)	4.28(18)	4.36(21)	4.32(18)	4.58(20)	4.58(22)	3.22(18)
21	448	4.74	4.34(19)	4.71(25)	4.35(20)	4.48(20)	4.79(24)	4.47(20)	3.10(15)
22	158	4.80	4.61(25)	4.73(26)	4.80(25)	5.19(26)	5.44(26)	5.55(26)	5.58(29)*
23	81	4.83	4.13(15)	3.64(11)	3.49(11)	5.04(24)	4.68(22)	4.49(21)	5.78(30)
24	445	5.05	4.67(26)	4.54(22)	4.41(23)	5.10(25)	5.11(25)	4.93(23)	3.43(21)
25	21	5.49	10.84(31)	8.43(31)	9.29(31)	8.80(31)	6.73(30)	7.33(30)	4.72(27)
26	234	5.56	6.91(30)	7.55(30)*	7.11(30)	6.99(29)	7.77(31)*	7.34(31)	4.64(26)
27	268	5.57	5.71(27)	6.21(29)	5.97(29)	5.60(27)	5.84(27)	5.74(27)	3.57(23)*
28	457	5.85	4.55(23)	4.71(24)	4.92(26)	4.64(22)	4.72(23)	4.96(24)	3.82(24)
29	345	6.40	5.96(28)	5.84(28)	5.80(27)	6.15(28)	5.84(28)	5.90(28)	4.00(25)
30	404	6.63	4.37(20)	4.55(23)	4.16(17)	4.33(19)	4.37(18)	4.03(16)	5.45(28)*
31	448	6.96*	6.63(29)*	5.39(27)	5.96(28)	7.25(30)*	6.19(29)*	6.79(29)*	6.10(31)*

Note: The ranks for hospitals using model C1 are the same as the hospital identifier in the first column. Hospital volumes for C1 and M1 – M6 total 13,577 cases with an overall mortality rate of 3.96 percent. For C2, the total number of cases is 30,972 with a mortality rate of 2.93 percent.

*(#)Risk-adjusted Rate significantly higher(lower) than Statewide Rate.

MEDPAR Models: Separating Complications and Comorbidities

In contrasting models M1, M2, and M3, it should be noted that MEDPAR (M1) identified 3,144 patients (23.2 percent) as having an AMI. In M2 and M3, respectively, CSRS eliminated AMI as a risk factor for 300 and 130 patients.

Also, M1 identified 85 patients as having renal failure, and CSRS eliminated renal failure as a risk factor for 50 and 1 of these risk patients, respectively, in M2 and M3. For intra-aortic balloon pump, 624 patients were identified as having an IABP, and 285 and 218 of these patients had the risk factor suppressed in M2 and M3, respectively.

As indicated in Table 4, six variables in the original MEDPAR model M1 were significant independent predictors of in-hospital mortality: age, acute myocardial infarction, congestive heart failure, cerebrovascular disease, renal failure, and the presence of an intra-aortic balloon pump. The C statistic for this model was .777.

In M2, one of the three variables that were investigated (renal failure) was no longer significant, and a second one (IABP) was not as strong a predictor of mortality after CSRS was used to eliminate potential complications from the risk factor data. Furthermore, the C statistic demonstrated a worse model fit (C = .709).

When AMI, renal failure, and IABP were removed as risk factors in the MEDPAR model for cases in which they were reported as complications in CSRS (in M3), similar results occurred. AMI did not remain in the MEDPAR model, and IABP did not have as strong a relationship to mortality. The C statistic indicated a better fit than in the second model (.732) but considerably worse than in the original model.

Table 5 contrasts the three MEDPAR models with respect to hospital risk-adjusted mortality rates. As indicated, Hospital 11 had a rank in M1 (3) very different from its rank in M2 (15) and somewhat different from its rank in M3 (9). Also, Hospitals 3, 10, 11, 15, 16, and 21 had ranks in M1 that differed from either the M2 rank or the M3 rank by at least 6. In general, the correlations in risk-adjusted mortality among the models was high (.98 between models M1 and M3, and between models M2 and M3; and .94 between models M1 and M2). Also, the high outlier identified using M1 was not identified in either M2 or M3, and M2 identified a different hospital as a high outlier.

MEDPAR versus CSRS

The first model presented in Table 4 (C1) predicts in-hospital mortality for Medicare patients, using the CSRS clinical data. As shown in Table 4, this model includes several variables that are not available in MEDPAR (two ranges for ejection fraction, previous open heart surgery, Canadian Cardiovascular Society Class IV, hemodynamic instability, shock, electrocardiogram evidence of left ventricular hypertrophy, calcified ascending aorta, and hepatic failure. AMI and IABP, both of which were significant in the MEDPAR (M1) model, were not significant in the CSRS (C1) model. These differences should not be overemphasized because of intercorrelations among variables. The C statistic, .789, was only slightly better than that in M1, but was vastly superior to the corresponding values in M2 and M3, which are probably more accurate representations of the ability of MEDPAR to assess quality of care since they are more likely not to erroneously include complications.

With respect to hospital risk-adjusted mortality rates (see Table 5), the largest differences in hospital ranks between C1 and the three MEDPAR models were for Hospital 3 in C1 (recall that hospitals are named according to their ranking in C1), which had ranks of 22, 16, and 16 in the three MEDPAR models; for Hospital 18 in C1, which had ranks of 7, 4, and 4 in the MEDPAR models; for Hospital 23 in C1, which had ranks of 15, 11, and 11 in MEDPAR; and for Hospital 30 in C1, which had ranks of 20, 23, and 17 in MEDPAR. The respective correlations between the risk-adjusted mortality in C1 and the MEDPAR models were .69, .74, and .73. The CSRS model did identify the same high outlier as M1, but this outlier was not identified by either M2 or M3. Two low outliers identified in C1 (Hospitals 3 and 7) were not identified by the MEDPAR models.

Medicare Patients versus All Patients

The last model in Table 4 (C2) predicts the likelihood of mortality for all CABG surgery patients in New York in 1991–1992 using CSRS clinical data. In contrasting this model with CSRS (C1), which predicts mortality odds for all Medicare patients in New York using CSRS data, we find that several more variables prove to be significant predictors of mortality in the model for all patients. These include female gender, aortoiliac disease, diabetes, persistent ventricular arrhythmia, and a previous myocardial infarction within six hours of surgery.

For most of these risk factors, the reason they are significant in C2 but not in C1 appears to be insufficient statistical power in C1 because of the small numbers of patients with the risk factors. However, this is certainly not the case for female gender, which is a significant predictor of mortality for the entire New York population but not for the New York Medicare population. Another interesting finding is that the model for all patients has discrimination that is superior to the model for Medicare patients ($C = .813$ versus $C = .789$). However, we should caution that a comparison of C statistics based on different populations can be misleading.

In terms of hospital risk-adjusted mortality rates, six hospitals (1, 10, 13, 15, 22, and 23) had ranks in C1 and C2 that differed by at least 7. The correlation in RAMRs of the two models was .80. Also, whereas the CSRS model for Medicare patients identified only one high outlier and two low outliers, the CSRS model for all patients identified four high outliers (the one identified by the other model and three additional ones) and four low outliers (the two identified by the other model and two additional ones). In reviewing differences among outlier hospitals in the two models, it was found that of the three additional high outliers in C2, two were not identified by C1 because of lower statistical power (smaller sample sizes). Of the two additional low outliers in C2, one was not identified because of lower statistical power.

Adding Clinical Data to Administrative Data

Models M4, M5, and M6 in Table 4 are respective enhancements of the M1, M2, and M3 models in which the information added to the MEDPAR data consisted of information on ejection fraction and whether the patient had previous open heart operations.

A comparison of M2 and M5 shows that the clinical variables (ejection fraction and previous operations) added to M2 were significant predictors of mortality, and that female gender also proved to be significant in M5. Also, the C statistic improved considerably given the limited possible range of this statistic (from .709 to .754). Very similar results occurred when comparing M3 and M6. The three clinical risk variables were all significant in M6, female gender was significant, and the C statistic improved from .732 in M3 to .773 in M6.

With respect to correlations in hospital risk-adjusted mortality rates, the enhanced models correlated well with their respective administrative counterparts (each pair had a correlation of .94). However, it is noteworthy that the enhanced models correlated with the clinical models much better than the administrative models did. Both enhanced models had correlations of .82 with C1, which was similar to the correlation between the two clinical models (.80). The respective correlations of M5 and M6 with C2 were .69 and .71. In contrast, the respective correlations of M2 and M3 with C1 were .74 and .73, and their respective correlations with C2 were only .57 and .58.

The enhanced models did identify some different outliers than C1, although they tended to be more similar in outlier identification than the MEDPAR models that were altered to remove complications.

DISCUSSION

Administrative data have the advantage of being inexpensive to collect, easily accessible, and capable of identifying regional differences in outcomes. Many previous studies have used administrative data to screen for quality of care problems (Dubois, Brook, and Rogers 1987; Hannan et al. 1989; Roos et al. 1985; Iezzoni, Daley, Heeren, et al. 1994), but numerous potential difficulties with these data have been identified, particularly when the data are used to assess quality of care rather than just to screen for potential quality problems (Fisher, Whaley, Krushat, et al. 1992; Iezzoni 1990; Iezzoni, Foley, Daley, et al. 1992; Jollis, Ancukiewicz, DeLong, et al. 1993; Hsia, Krushat, Fagan, et al. 1988).

Clinical data, which do not have most of these problems (Pryor, Califf, Harrell, et al. 1985), have been used successfully, in particular, to predict mortality for CABG surgery. Two previous studies have compared the ability of administrative and clinical data to predict mortality for CABG surgery. One study compared New York's administrative hospital discharge data system (SPARCS) with its clinical data set (CSRS or the Cardiac Surgery Reporting System) in predicting mortality (Hannan et al. 1992). They concluded that the statistical model developed using clinical data had substantially more predictive ability than its administrative counterpart and that identification of outlier hospitals differed substantially. However, it was also found that adding only three clinical data elements to the administrative data eliminated much of the difference in effectiveness of the two systems.

Krakauer, Bailey, Skellan, et al. (1992) compared a statistical model based on HCFA data equivalent to current MEDPAR data to a statistical model based on clinical data abstracted from medical records. They concluded that the administrative data could be used to characterize variations in mortality rates and for further epidemiological analyses of factors related to patient mortality, but that they did not positively identify outlier hospitals.

The current study differs from the studies just discussed in that it also attempted (1) to assess changes that occurred when MEDPAR data were made more valid by limiting its risk factors to those that were verifiable by clinical data relating to the same patients, and (2) to assess the differences in hospital quality assessments based solely on Medicare CABG surgery patients compared with those based on all CABG patients.

Several important findings resulted from this study. First, when the clinical (CSRS) data were used to eliminate secondary diagnoses that were suspected to be complications rather than risk factors (comorbidities), the

MEDPAR models had substantially lower predictive ability than they previously had had. Also, the hospital risk-adjusted mortality rates from the original MEDPAR model had a correlation of only .69 with the CSRS risk-adjusted rates, but this correlation rose to .73 and .74 when attempts were made to eliminate complications using the other two MEDPAR models (M2 and M3).

We conclude that this study supports the contention that efforts be made to distinguish between comorbidities and complications among secondary diagnoses in administrative data systems. Probably the most effective way to do this is to introduce a binary code associated with each secondary diagnosis in administrative data that indicates whether or not the condition was present at admission. This has already been done in New York's administrative data system (SPARCS) and is also being implemented in California. We recommend that it be considered for Medicare data and for other administrative data systems that are used or are being considered for outcomes analyses. However, it is also important that medical records personnel be educated to code this information accurately.

A second finding was that the administrative models that were altered to eliminate potential complications yielded predictive ability appreciably inferior to that of the CSRS model for Medicare patients. However, consistent with earlier findings that compared clinical models with administrative models enhanced with a few clinical variables, much of the gap in predictive ability was recaptured when ejection fraction and previous open heart operations were added as risk factors to the MEDPAR data. The discrimination of the two enhanced models approached that of the clinical model and were considerably better than in the original MEDPAR model.

We should add that the only additions to administrative data that were explored in this study were clinical risk factors. However, another possibility would be to use administrative data to redefine current ICD-9-CM codes in order to identify a higher-risk population than is currently defined. For example, although AMI is available in CSRS for all possible time periods prior to surgery, the analyses conducted here and past analyses indicate that patients with recent MIs are at much higher risk (the CSRS model for all patients identified the group with MIs within six hours of surgery to be at increased risk, but not any other group of recent MI patients). Thus, it appears that if patients could be defined with this level of precision, the predictability of administrative models would improve. The reason these potential changes were not explored in this study was our desire to limit ourselves to those changes that appeared to be the easiest to implement.

In addition, an administrative database linking admissions over time could be used to add "prior CABG surgery" to the list of potential risk factors. However, it should be noted that surgery occurring before the patient became a Medicare beneficiary would not have been detected.

As a result of these findings, we recommend that agencies in charge of administrative databases consider adding limited clinical data elements to their administrative data. The implementation of this change would require a system that would alert coders when a certain kind of case was being coded, and would require a change in data entry that would present the new field to coders for this specific type of case. The types of cases that could be enhanced in this manner could increase over time, and eventually investigators could use the administrative database for outcomes research on a variety of patients.

Another important (and perhaps surprising) finding was that the clinical model for all New York CABG patients differed considerably from the clinical model developed using only Medicare patients. Even though the former model was noticeably superior with regard to both discrimination and calibration, this was not too troubling because both models had excellent predictive ability. What was troubling was that the correlations in hospital risk-adjusted mortality rates between the two models were only moderately good. Even more disconcerting was that the model for all patients yielded four high-outlier and four low-outlier hospitals, whereas the Medicare model yielded only one high-outlier hospital and no low-outlier hospitals. In some cases, these differences appeared to be related to statistical power, but for other hospitals, the two models yielded quite different RAMRs.

These results call into question the use of Medicare models to assess overall hospital quality. Although part of the reason for the differences might have been statistical power problems, these same problems are likely to arise for other procedures and medical conditions. As a consequence, we recommend that, in assessing overall hospital quality using administrative data, it is far preferable to use an administrative database that includes all patients rather than one limited to Medicare patients.

REFERENCES

Breslow, N. E., and N. E. Day. 1991. *Statistical Methods in Cancer Research.* New York: Oxford University Press.

Dubois, R. W., R. H. Brook, and W. H. Rogers. 1987. "Adjusted Hospital Death Rates: A Potential Screen for Quality of Care." *American Journal of Public Health* 77 (9): 1162–66.

Fisher, E. S., F. S. Whaley, W. M. Krushat, D. J. Malenka, C. Fleming, J. A. Baron, and D. C. Hsia. 1992. "The Accuracy of Medicare's Hospital Claims Data: Progress Has Been Made, but Problems Remain." *American Journal of Public Health* 82 (2): 243–48.

Hannan, E. L., H. R. Bernard, J. F. O'Donnell, and H. Kilburn. 1989. "A Methodology for Targeting Hospital Cases for Quality of Care Record Reviews." *American Journal of Public Health* 79 (4): 430–36.

Hannan, E. L., H. Kilburn, Jr., M. L. Lindsey, and R. Lewis. 1992. "Clinical versus Administrative Data Bases for CABG Surgery: Does It Matter?" *Medical Care* 30 (10): 892–907.

Hannan, E. L., H. Kilburn, Jr., M. Racz, E. Shields, and M. R. Chassin. 1994. "Improving the Outcomes of Coronary Artery Bypass Surgery in New York State." *Journal of the American Medical Association* 271 (10): 761–66.

Health Care Financing Administration. 1991. "Medicare Hospital Mortality Information, 1987–1988–1989." HCFA Publication No. 00744. Washington, DC: Government Printing Office.

Hsia, D. C., W. M. Krushat, A. B. Fagan, J. A. Tebbutt, and R. P. Kusserow. 1988. "Accuracy of Diagnostic Coding for Medicare Patients under the Prospective Payment System." *The New England Journal of Medicine* 318 (6): 352–55.

Iezzoni, L. I. 1990. "Using Administrative Diagnostic Data to Assess the Quality of Diagnostic Care: The Pitfalls and Potential of ICD-9-CM." *International Journal of Technology Assessment in Health Care* 6 (2): 272–81.

Iezzoni, L. I., S. M. Foley, J. Daley, J. Hughes, E. S. Fisher, and T. Heeren. 1992. "Comorbidities, Complications, and Coding Bias." *Journal of the American Medical Association* 267 (16): 2197–203.

Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, J. S. Hughes, E. S. Fisher, C. Duncan, and G. A. Coffman. 1994. "Using Administrative Data to Screen Hospitals for High Complication Rates." *Inquiry* 30, no. 1 (spring): 40–55.

Jollis, J. G., M. Ancukiewicz, E. DeLong, D. B. Pryor, L. H. Muhlbaier, and D. B. Mark. 1993. "Discordance of Databases Designed for Claims Payment versus Clinical Information Systems: Implications for Outcomes Research." *Annals of Internal Medicine* 119 (6): 844–50.

Krakauer, H., R. C. Bailey, K. J. Skellan, J. D. Stewart, A. J. Hartz, E. M. Kuhn, and A. A. Rimm. 1992. "Evaluation of the HCFA Model for the Analysis of Mortality Following Hospitalization." *Health Services Research* 27 (3): 317–35.

Landis, J. B., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–74.

O'Connor, G. T., S. K. Plume, E. M. Olmstead, L. H. Coffin, J. R. Morton, C. T. Maloney, E. R. Nowicki, D. G. Levy, J. F. Tryzelaar, and F. Hernandez. 1992. "Multivariate Prediction of In-hospital Mortality Associated with Coronary Artery Bypass Graft Surgery." *Circulation* 85 (6): 2110–118.

The Pennsylvania Health Care Cost Containment Council. 1994. Coronary Artery Bypass Graft Surgery, Technical Report, Vol. II: 1991. The Pennsylvania HCCCC, Harrisburg, February.

Pryor, D. B., R. M. Califf, F. E. Harrell, M. A. Hlatky, K. L. Lee, D. B. Mark, and

R. A. Rosati. 1985. "Clinical Data Bases: Accomplishments and Unrealized Potential." *Medical Care* 23 (5): 623–47.

Romano, P. S., L. L. Roos, J. G. Jollis, K. Doliszny, and the Ischemic Heart Disease Patient Outcomes Research Team. 1994. "A Comparison of Administrative versus Clinical Data: Coronary Artery Bypass Surgery as an Example." *Journal of Clinical Epidemiology* 47 (3): 249–60.

Roos, L. L., S. M. Cageorge, E. Austen, and K. N. Lohr. 1985. "Using Computers to Identify Complications after Surgery." *American Journal of Public Health* 75 (11): 1288–95.

Tu, J. V., S. B. Jaglal, C. D. Naylor, and the Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. 1995. "Multicenter Validation of a Risk Index for Mortality, Intensive Care Unit Stay, and Overall Hospital Length of Stay after Cardiac Surgery." *Circulation* 91 (3): 677–84.