UCLA

UCLA Previously Published Works

Title

Using Both Clinical Registry and Administrative Claims Data to Measure Risk-adjusted Surgical Outcomes.

Permalink

<https://escholarship.org/uc/item/94h1d5nn>

Journal

Annals of surgery, 263(1)

ISSN

0003-4932

Authors

Lawson, Elise H
Louie, Rachel
Zingmond, David S
et al.

Publication Date

2016

DOI

10.1097/sla.0000000000001031

Peer reviewed

Using Both Clinical Registry and Administrative Claims Data to Measure Risk-adjusted Surgical Outcomes

Elise H. Lawson, MD, MSHS,*†‡ Rachel Louie, MS,§ David S. Zingmond, MD, PhD,§
Greg D. Sacks, MD, MPH,* Robert H. Brook, MD, ScD,§¶||
Bruce Lee Hall, MD, PhD, MBA,†**††‡‡§§ and Clifford Y. Ko, MD, MS, MSHS*†‡

Objective: To examine the validity of hybrid quality measures that use both clinical registry and administrative claims data, capitalizing on the strengths of each data source.

Background: Previous studies demonstrate substantial disagreement between clinical registry and administrative claims data on the occurrence of postoperative complications. Clinical data have greater validity than claims data for quality measurement but can be burdensome for hospitals to collect.

Methods: American College of Surgeons National Surgical Quality Improvement Program records were linked to Medicare inpatient claims (2005–2008). National Quality Forum–endorsed risk-adjusted measures of 30-day postoperative complications or death assessed hospital quality for patients undergoing colectomy, lower extremity bypass, or all surgical procedures. Measures use hierarchical multivariable logistic regression to identify statistical outliers. Measures were applied using clinical data, claims data, or a hybrid of both data sources. Kappa statistics assessed agreement on determinations of hospital quality.

Results: A total of 111,984 patients participated from 206 hospitals. Agreement on hospital quality between clinical and claims data was poor. Hybrid models using claims data to risk-adjust complications identified by clinical data had moderate agreement with all clinical data models, whereas hybrid models using clinical data to risk-adjust complications identified by claims data had routinely poor agreement with all clinical data models.

Conclusions: Assessments of hospital quality differ substantially when using clinical registry versus administrative claims data. A hybrid approach using claims data for risk adjustment and clinical data for complications may be a valid alternative with lower data collection burden. For quality measures

focused on postoperative complications to be meaningful, such policies should require, at a minimum, collection of clinical outcomes data.

Keywords: ACS-NSQIP, administrative claims, clinical registry, quality measure, surgery

(*Ann Surg* 2016;263:50–57)

Until recently, payment for health care services was determined by a fee-for-service model in which payment is linked to the volume of services provided.¹ By its very nature, this system encourages overuse of health care services and the fee-for-service model is often implicated in the ever-increasing rise in health care expenditures.² The Affordable Care Act of 2010 includes provisions designed to change these trends by tying hospital payment to performance on a series of quality measures, including measures of clinical processes of care and patient experience, thus fostering the delivery of high-quality, efficient care. This payment model, known as value-based purchasing, was implemented in fiscal year 2013 and is scheduled to expand to include outcome measures in 2014 and efficiency measures in 2015.³

Such quality measurement is not new to surgeons. Morbidity and mortality conference, a regular occurrence on most surgical services, is an example of quality measurement at the local level, whereas, nationally, the American College of Surgeons (ACS) has a long history of measuring and improving surgical quality through validated clinical registries such as the Commission on Cancer, the National Trauma Data Bank, and the National Surgical Quality Improvement Program (NSQIP). However, controversy remains regarding how to best measure surgical quality for the purposes of public reporting and pay-for-performance. Outcomes, such as postoperative complications, are generally considered to have greater clinical relevance and validity as indicators of quality than processes of care, such as the choice and timing of preoperative antibiotic administration. Current outcomes-based surgical quality measures use either administrative claims data (which are submitted by hospitals for billing purposes) or clinical registry data (which are collected by hospitals from medical records for quality improvement purposes). Each data source has its own advantages and disadvantages, with clinical registry data generally perceived as having greater validity for quality measurement than administrative claims data but at the price of substantially greater data collection burden for hospitals.^{4–6}

In a previous study, we demonstrated substantial disagreement between clinical registry and administrative claims data on the occurrence of risk-adjusted postoperative complications.⁷ The objective of this study was to examine the validity of surgical quality measures that use a hybrid of clinical registry and administrative claims data to capitalize on the strengths of each data source. We utilized 3 risk-adjusted measures of hospital performance for 30-day postoperative major complications or death after (1) colectomy, (2) lower extremity bypass (LEB), or (3) all surgical procedures. These 3

From the *Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA; †Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL; ‡VA Greater Los Angeles Healthcare System, Los Angeles, CA; §Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA; ¶RAND Corporation, Santa Monica, CA; ||UCLA Jonathan and Karin Fielding School of Public Health, Los Angeles, CA; **Department of Surgery, School of Medicine, Washington University in St Louis, and Barnes Jewish Hospital, St Louis, MO; ††Center for Health Policy and the Olin Business School at Washington University in St Louis, St Louis, MO; ‡‡Department of Surgery, John Cochran Veterans Affairs Medical Center, St Louis, MO; and §§BJC Healthcare, St Louis, MO.

Disclosure: E.H.L.'s time was supported by the VA Health Services Research and Development program (RWJ 65–020) and the American College of Surgeons (ACS) through the Robert Wood Johnson Foundation Clinical Scholars Program. R.L., D.S.Z., and B.L.H. are consultants for the ACS, and C.Y.K. is staff at the ACS. Supported by a contract from the Centers for Medicare & Medicaid Services. The remaining authors declare no conflicts of interest.

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

Reprints: Elise H. Lawson, MD, MSHS, Department of Surgery, David Geffen School of Medicine at UCLA, 10833 Le Conte Ave 72-215 CHS, Los Angeles, CA 90095. E-mail: elawson@mednet.ucla.edu.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/14/26105-0821

DOI: 10.1097/SLA.0000000000001031

measures were previously reviewed and endorsed by the National Quality Forum (NQF), and some hospitals are currently voluntarily publicly reporting their performance for these measures on Medicare's Hospital Compare Web site.⁸ Our overarching goal was to inform current policy efforts aimed at measuring the quality of surgical care for public reporting and value-based purchasing.

METHODS

Data Sources and Study Population

The primary data sources for this study were Medicare inpatient claims and clinical registry data from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), which have been previously described.⁶ Briefly, the 100% Medicare Provider Analysis and Review file (MedPAR) contains administrative claims data for all Medicare beneficiaries receiving health care services at inpatient facilities in the United States. Each record represents an inpatient hospital stay for a beneficiary. Relevant diagnoses, procedures, and services from hospital admission to discharge are included. The Medicare denominator file was used to assess mortality.⁹

The ACS-NSQIP clinical registry contains data on a systematic sample of patients undergoing surgical procedures in participating hospitals. To participate in the program, hospitals must employ a dedicated data abstractor who is trained to use strict variable definitions and collection methods. Cases are accrued across several surgical specialties, including general, vascular, and specific subspecialties. Patients admitted for major trauma or organ transplantation are excluded. Sources for data are medical records and the patient. Data collected include markers of the patient's clinical health status before surgery, aspects of the surgery performed, and the occurrence of rigorously specified complications or death within 30 days after surgery. Hospitals are periodically audited to ensure standardized data collection. Audit results in the past have demonstrated substantial or almost perfect agreement on the coding of most variables.^{10–12}

As previously described, a deterministic linkage algorithm was developed to match patient-level records from 2005 to 2008 between these 2 data sources using indirect identifiers. There was excellent agreement between matched records on death during the primary hospitalization, which supports the validity of the linkage procedure.¹³ Our study population was restricted to patients 65 years or older who underwent an inpatient surgical procedure between 2006 and 2008 and for whom we were able to identify a record of the hospital stay in both the Medicare claims and ACS-NSQIP databases. We excluded patients with procedures occurring in 2005 because we used Medicare data from one full year before the date of surgery to characterize patients' preoperative clinical health status. Similarly,

patients with procedures occurring in December 2008 were excluded because we lacked a full 30 days of follow-up of Medicare data to identify complications and deaths. Hospitals with fewer than 20 cases were excluded.

This work was supported by a contract from the Centers for Medicare & Medicaid Services, which approved the use of Medicare claims data. The RAND Corporation Institutional Review Board approved the study protocol. All analyses were performed using SAS version 9.3 software (SAS Institute, Inc, Cary, NC).

Surgical Quality Measures

We utilized 3 NQF-endorsed risk-adjusted measures of surgical quality that reflect hospital performance for major complications and/or death occurring within 30 days after 3 groups of surgical procedures: colectomy, lower extremity vascular bypass (LEB), and all surgical procedures. These quality measures were originally developed using ACS-NSQIP data, and some ACS-NSQIP hospitals are currently voluntarily sharing their performance on 1 or more of these measures through a pilot public reporting program on Medicare's Hospital Compare Web site.⁸ The "all surgical procedures" measure was originally developed for patients 65 years or older, whereas the original colectomy and LEB measures only exclude patients younger than 18 years. Surgical procedures were identified by *Current Procedural Terminology (CPT)* code recorded in ACS-NSQIP.

The postoperative complications included in each binary composite outcome measure are reported in Table 2. The ACS-NSQIP program requires documentation of occurrence versus no occurrence of each specified complication for every patient, and each complication is specific and rigorously defined. In contrast, the Medicare database includes 10 diagnosis fields and 6 procedure fields, which are populated by *International Classification of Diseases, Ninth Revision (ICD-9)*, codes describing the patient's hospital stay. These codes thus represent medical conditions present before surgery and complications occurring after surgery. To compare these data sources, we created a crosswalk matching ACS-NSQIP surgical complication definitions with applicable *ICD-9* codes in Medicare data, as previously described.⁶ We searched for codes representing complications in the Medicare claims record for the index admission and for any subsequent readmissions occurring within 30 days of surgery.

Hierarchical multivariable logistic regression models were developed for each of the 3 quality measures (colectomy, LEB, and all-surgery). This method accounts for clustering of patients within hospitals by allowing each hospital to have a different random intercept and incorporates an empirical Bayes method.¹² Four modeling approaches were used for each quality measure, creating a total of 12 models. The 4 modeling approaches (described in Table 1) differ from each other by whether clinical registry or administrative

TABLE 1. Data Sources Used for Risk Adjustment and for the Composite Outcome 30-Day Postoperative Major Complications and/or Death in Each of Four Modeling Approaches for Surgical Quality Measures

Model Name	Data Source	
	Risk Adjustment	Outcome (30-d Postoperative Major Complications and/or Death)
Clinical registry model	ACS-NSQIP	ACS-NSQIP
Administrative claims model	Medicare Claims	Medicare claims
Clinical/claims hybrid 1 model	ACS-NSQIP	Medicare claims
Claims/clinical hybrid 2 model	Medicare Claims	ACS-NSQIP

ACS-NSQIP is a surgical clinical registry. Complications included in composite outcome are listed in Table 2.

TABLE 2. Variables Used for Risk Adjustment in Clinical Registry Models and 30-Day Postoperative Complications Included in Composite Outcome for Three Surgical Quality Measures

Surgical Quality Measure	Variables Used for Risk Adjustment in Clinical Registry Model	30-d Postoperative Complications Included in Composite Outcome
Colectomy	American Society of Anesthesiology class, functional status before surgery, indication for surgery, wound class, and emergency case	Death, unplanned reoperation, cardiac arrest requiring CPR, myocardial infarction, unplanned reintubation (without prior ventilator dependence), pneumonia (without pneumonia before surgery), pulmonary embolism, deep venous thrombosis requiring therapy, deep incisional surgical site infection, organ space surgical site infection, wound disruption, progressive renal insufficiency, acute renal failure (without preoperative renal failure/dialysis), sepsis, septic shock, and urinary tract infection
Lower extremity vascular bypass	Age, sex, American Society of Anesthesiology class, functional status before surgery, indication for surgery, emergency case, and preoperative laboratory values for albumin, aspartate aminotransferase, and creatinine	Death, cardiac arrest requiring CPR, myocardial infarction, prolonged ventilation (without prior ventilator dependence), organ space surgical site infection, acute renal failure (without preoperative renal failure or dialysis), sepsis, septic shock, and cerebrovascular accident/stroke
All-surgery	American Society of Anesthesiology class and functional status before surgery	Same as colectomy measure

Variables used for modeling are consistent with those specified in surgical quality measures approved by the NQF, as discussed in the text. Variables were originally selected for inclusion in these measures based on clinical and statistical significance and ease of data collection.

CPR indicates cardiopulmonary resuscitation.

claims data are used for risk adjustment and whether clinical registry or administrative claims data are used for the outcome (major complications and/or death).

The first modeling approach used ACS-NSQIP data for both risk adjustment and for the outcome (“clinical registry models”). The ACS-NSQIP variables used for risk adjustment in this approach are consistent with those used in the NQF-approved measures and are listed in Table 2.¹⁴ Of note, variables were originally selected for inclusion in these measures based on clinical and statistical significance and ease of data collection. Missing data for American Society of Anesthesiology class and laboratory values were imputed by hospital using the hot-deck method. ICD-9 codes representing the indication for surgery were grouped into 10 categories for the colon measure and 5 categories for LEB.

The second modeling approach used Medicare claims data for both risk adjustment and the outcome (“administrative claims models”). Risk adjustment was achieved by creating variables for coexisting diseases using secondary diagnostic codes and the methods of Elixhauser.¹⁵

The third and fourth models used a hybrid approach to the data sources. For the third model, ACS-NSQIP data were used for risk adjustment and Medicare claims data for the outcome (“clinical/claims hybrid 1 models”). Finally, for the fourth model, Medicare claims data were used for risk adjustment and ACS-NSQIP data for the outcome (“claims/clinical hybrid 2 models”).

In addition to risk adjustment for patient characteristics, all clinical, claims, and hybrid models were adjusted for “procedure mix” using a CPT-based linear risk score, as used in the original NQF measures and the standard for ACS-NSQIP modeling.¹² Statistical model diagnostics, including relative goodness of fit (Akaike information criterion), discriminative power (c-statistic), and calibration (Hosmer-Lemeshow statistic), are reported for each model (see Table 1, Supplemental Digital Content, available at <http://links.lww.com/SLA/A888>, for statistical model diagnostics).

Individual hospital performance for each quality measure was estimated using the hospital intercept odds ratio derived from the

applicable model. This odds ratio estimates the odds of a patient having a major complication or death at the specified hospital versus the odds of having this outcome at a theoretical “average” hospital, adjusted for the other variables included in the model. A hospital was considered to be a high statistical outlier (worse than expected performance) if its odds ratio was more than 1 ($P < 0.05$) and a low statistical outlier (better than expected performance) if its odds ratio was less than 1 ($P < 0.05$). Hospitals with odds ratios not significantly different from that of 1 ($P \geq 0.05$) were labeled as “as expected” performance for their given patient population.

The Spearman correlation coefficient evaluated agreement on hospital performance for each quality measure as determined by the clinical registry model versus the administrative claims or hybrid models. Hospital performance was also compared by ranking hospitals and then splitting them into equal decile groups. The weighted κ statistic evaluated agreement on hospitals’ decile rank as determined by the clinical registry model versus the administrative claims or hybrid models. Kappa values, which take into account agreement occurring by chance, were interpreted using guidelines that propose the following: $\kappa < 0.4$, poor agreement; $\kappa = 0.4–0.75$, moderate; and $\kappa > 0.75$, excellent agreement.¹⁶ Hospitals with risk-adjusted odds ratios in the highest decile were considered high percentile outliers for postoperative complications and/or death (worse performance), whereas hospitals in the lowest decile were considered low percentile outliers (better performance).

RESULTS

The samples for the study included 16,887 patients from 176 hospitals for the colectomy measure, 5378 patients from 115 hospitals for the LEB measure, and 112,040 patients from 208 hospitals for the all-surgery measure. For the colectomy measure, the most frequent indication for surgery was malignancy (47%), followed by diverticulitis (15%) and benign neoplasm (13%). The majority of LEB cases were classified as having elective (41%) or urgent (24%) indications. Of the remaining LEB cases, 10% were for embolism,

TABLE 3. Demographic and Preoperative Clinical Characteristics of Study Populations From a Database Linking Clinical Surgery Registry Data (ACS-NSQIP) With Medicare Inpatient Claims Data

	Colectomy Measure (n = 16,887)	Lower Extremity Vascular Bypass Measure (n = 5378)	All-Surgery Measure (n = 112,040)
Year of procedure, %			
2006	20.2	22.0	19.0
2007	38.2	38.1	36.3
2008	41.6	39.9	44.8
Age, %			
65–74 yr	43.4	48.9	47.9
75–84 yr	41.8	39.1	39.8
≥85 yr	14.8	12.1	12.3
Male sex, %	43.8	59.4	48.4
<i>Preoperative clinical characteristics from ACS-NSQIP data</i>			
Body mass index (kg/m ²), %			
Underweight (<18.5)	3.1	4.9	3.1
Normal (18.5–24.9)	35.2	38.5	32.6
Overweight (25–29.9)	34.9	34.8	35.1
Class I obesity (30–34.9)	17.1	15.5	18.0
Class II obesity (35–39.9)	6.4	4.7	6.9
Class III obesity (≥40)	3.3	1.7	4.3
Functional status, %			
Independent	83.7	81.7	84.2
Partially dependent	11.1	16.0	11.4
Totally dependent	5.2	2.3	4.4
ASA class, %			
I	0.8	0.2	0.6
II	30.3	5.0	22.3
III	54.1	73.8	60.6
IV	13.6	20.8	15.7
V	1.1	0.3	0.8
Smoker, %	10.7	31.6	14.4
>2 drinks/d alcohol, %	3.1	5.3	2.8
Chronic obstructive pulmonary disease, %	10.5	17.2	11.4
Diabetes, %			
None	81.0	64.4	77.8
Oral medication	13.1	18.8	14.0
Insulin	5.9	16.9	8.3
Renal failure/dialysis, %	1.95	2.3	2.0
Ascites, %	3.1	0.2	2.1
Congestive heart failure, %	2.7	3.2	2.7
Hypertension requiring medication, %	70.2	87.0	74.7
Disseminated cancer, %	4.8	0.5	3.4
Chronic steroid use, %	5.4	4.1	4.3
>10% weight loss in past 6 mo, %	5.8	2.0	4.3
Bleeding disorder, %	9.5	21.7	13.0
Sepsis, %			
None	84.1	94.1	87.9
Systemic inflammatory response	9.5	4.4	8.0
Sepsis	2.8	1.0	2.2
Septic shock	3.7	0.5	1.9
Emergency procedure, %	19.3	6.0	15.3
Wound class, %			
1—clean	0	89.0	47.9
2—clean/contaminated	76.1	2.4	35.2
3—contaminated	11.1	4.2	8.6
4—dirty/infected	12.8	4.4	8.3
<i>Preoperative clinical characteristics from Medicare inpatient claims data</i>			
Depression, %	5.2	5.0	5.6
Anemia, %	17.0	13.9	12.4
Fluid and electrolyte disorders, %	26.8	20.1	22.2
Weight loss, %	8.1	3.7	6.0
Obesity, %	5.0	3.8	5.7
Coagulopathy, %	4.0%	4.7	4.6
Solid tumor without metastasis, %	5.6	2.1	4.5
Metastatic cancer, %	17.4	1.1	7.9
Liver disease, %	1.6	0.8	1.7
Renal failure, %	8.0	17.7	9.9

TABLE 3. (Continued)

	Colectomy Measure (n = 16,887)	Lower Extremity Vascular Bypass Measure (n = 5378)	All-Surgery Measure (n = 112,040)
Diabetes, %			
With chronic complications	2.4	12.5	4.2
Without chronic complications	17.4	28.2	20.9
Chronic pulmonary disease, %	18.5	30.7	21.2
Congestive heart failure, %	11.7	9.2	10.1

ASA indicates American Society of Anesthesiologists.

10% were reoperations, and 14% were for an “other” indication. For the all-surgery measure, the majority of patients underwent a procedure classified as general surgery or vascular surgery (57% and 30%, respectively). Table 3 lists demographic and preoperative clinical characteristics of the study populations derived from ACS-NSQIP data and Medicare inpatient claims data.

For each measure, Medicare claims data identified more patients with a major postoperative complication and/or death than ACS-NSQIP data (Table 4). Agreement between the data sources on the patient-level occurrence of this composite outcome was moderate. For the colectomy measure, Medicare claims data identified 82% of patients who had a major complication and/or death recorded in ACS-NSQIP (sensitivity compared with ACS-NSQIP) and 81% of patients identified as *not* having this outcome in ACS-NSQIP (specificity compared with ACS-NSQIP). These findings were similar for the LEB and all-surgery measures (sensitivity 71% and 79%, respectively; specificity 87% and 84%, respectively). The LEB measure had the highest percentage of false-positives—defined as patients with a major complication and/or death recorded in Medicare claims data but not in ACS-NSQIP data—followed by the all-surgery and colectomy measures (false-positives 59%, 42%, and 38%, respectively). False-negatives—defined as patients identified as *not* having this outcome recorded in Medicare claims but having the outcome recorded in ACS-NSQIP—were relatively uncommon for all 3 measures (7% for colectomy, 4% for LEB, and 6% for all-surgery).

Correlation between the data sources on the hospital-level percentage of patients with a major postoperative complication or death was highest for the all-surgery measure, followed by colectomy and LEB (correlation coefficient = 0.75, 0.72, and 0.62, respectively). Agreement on hospitals’ decile rank for the percentage of patients with this composite outcome was moderate for colectomy and all-surgery and poor for LEB (Table 4).

When comparing with the clinical registry model, correlation of hospital quality determinations was lowest for the administrative claims model (correlation coefficient = 0.41 for colectomy, 0.42 for LEB, and 0.33 for all-surgery), slightly higher for the clinical/claims hybrid 1 model using administrative claims data for outcomes (correlation coefficient = 0.53 for colectomy, 0.48 for LEB, and 0.45 for all-surgery) and highest for the claims/clinical hybrid 2 model using clinical registry data for outcomes (correlation coefficient = 0.88 for colectomy, 0.93 for LEB, and 0.83 for all-surgery). Results were similar when comparing hospitals’ decile rank for quality between the models (see Table 2, Supplemental Digital Content, available at <http://links.lww.com/SLA/A888>, which reports agreement on hospital decile rank).

Table 5 demonstrates agreement between the clinical registry model and the administrative claims or hybrid models on hospital statistical outlier status for major complications and/or death. For LEB, the clinical registry model did not identify any high or low outlier hospitals. Complete discordance, in which a hospital is identified as a high statistical outlier by one model and a low statistical outlier by another, occurred only for the all-surgery measure. Disagreement, in which a hospital is identified as an outlier by one model and “as expected” by another, occurred more frequently.

Agreement with the clinical registry model on percentile outlier status (1st, 2nd–9th, or 10th decile) was highest for the claims/clinical hybrid 2 model (which uses claims data for risk adjustment and clinical data for outcomes) and lowest for the administrative claims model (Fig. 1). For LEB, 1 hospital was classified into the highest decile (worse performance) by the clinical registry model and the lowest decile (better performance) by the administrative claims model. For the all-surgery measure, this discordance was identified for 2 hospitals when comparing the clinical registry model with the claims and clinical/claims hybrid 1 models. No other instances of discordance were identified. For the colectomy and LEB measures, all hospitals that were classified as high or low

TABLE 4. Occurrence of Major Complications and/or Death After Specific Surgical Procedures as Recorded in a Clinical Surgical Registry (ACS-NSQIP) Versus an Inpatient Administrative Database (Medicare Claims)

Surgical Quality Measure	Percentage of Patients With a Major Complication and/or Death (95% CI)		Agreement Between Data Sources on Occurrence of a Major Complication and/or Death (κ)	Median Hospital Percentage of Patients With a Major Complication and/or Death in Each Data Source (Interquartile Range)		Agreement Between Data Sources on Hospitals’ Decile Rank for Major Complications and/or Death (Weighted κ)
	ACS-NSQIP	Medicare Claims		ACS-NSQIP	Medicare Claims	
Colectomy	27.1 (26.5–27.8)	36.0 (35.2–36.7)	Moderate (0.577)	27.7 (23.0–34.1)	37.6 (30.5–42.9)	Moderate (0.508)
Lower extremity vascular bypass	11.7 (10.8–12.6)	20.1 (19.0–21.2)	Moderate (0.441)	11.1 (7.3–15.4)	19.6 (13.9–25.0)	Poor (0.390)
All-surgery	21.8 (21.5–22.0)	30.0 (29.7–30.3)	Moderate (0.555)	22.2 (18.2–25.1)	30.5 (26.2–34.4)	Moderate (0.528)

CI indicates confidence interval.

TABLE 5. Comparison of Hospital Risk-adjusted Statistical Outlier Classification for 30-Day Postoperative Major Complications or Death Between Models Using Clinical Surgical Registry (ACS-NSQIP) Data Versus Models Using Administrative Data From Medicare Inpatient Claims (MedPAR) or a Hybrid of Data Sources for Three Quality Measures

Quality Measure (No. Hospitals Included in Measure)	Model Name	Hospitals With Worse Than Expected Performance		Hospitals With Better Than Expected Performance		No. Hospitals Labeled as a “High Outlier” by ACS-NSQIP and a “Low Outlier” by Specified Model, or Vice Versa	No. Hospitals Labeled as an “Outlier” by ACS-NSQIP Model and “As Expected” by Specified Model, or Vice Versa
		Number Identified by Each Model	Number Identified by Clinical Model and Specified Model	Number Identified by Each Model	Number Identified by Clinical Model and Specified Model		
Colectomy (176 hospitals)	Clinical registry	10	Ref	5	Ref	Ref	Ref
	Administrative claims	3	0	3	1	0	19
	Clinical/claims hybrid 1	6	2	7	1	0	22
	Claims/clinical hybrid 2	9	5	6	3	0	14
	Clinical registry	0	Ref	0	Ref	Ref	Ref
Lower extremity vascular bypass (115 hospitals)	Administrative claims	4	0	2	0	0	6
	Clinical/claims hybrid 1	2	0	1	0	0	3
	Claims/clinical hybrid 2	0	0	0	0	0	0
	Clinical registry	23	Ref	20	Ref	Ref	Ref
	Administrative claims	22	4	23	9	2	58
All-surgery (208 hospitals)	Clinical/claims hybrid 1	26	10	26	10	2	51
	Claims/clinical hybrid 2	26	16	28	14	0	37

See Table 1 for description of data sources used for each modeling approach. “Clinical registry” models use data from ACS-NSQIP for both risk adjustment and the composite outcome. “Administrative claims” models use administrative claims data from Medicare for both risk adjustment and the composite outcomes. “Clinical/claims hybrid 1” models use ACS-NSQIP data for risk adjustment and administrative claims data from Medicare for the composite outcome. “Claims/clinical hybrid 2” models use administrative claims data from Medicare for risk adjustment and ACS-NSQIP data for the composite outcome.

FIGURE 1. Agreement on hospital quality for 30-day postoperative major complications or death between a model using data from a clinical surgical registry (ACS-NSQIP) and models using administrative data from Medicare inpatient claims (MedPAR) or a hybrid of data sources (hybrid 1—clinical registry data for risk adjustment and administrative claims data for outcomes; hybrid 2—administrative claims data for risk adjustment and clinical registry data for outcomes). See Table 1 for description of modeling approaches. Hospital quality assessed by outlier classification for risk-adjusted odds ratio for the composite outcome, with outliers identified by percentile method, as described in the text.

statistical outliers were also classified as high or low percentile outliers, respectively.

DISCUSSION

There are numerous ways to measure quality in surgery, each with advantages and disadvantages. In this study, we explored the use of a hybrid approach to quality measurement using both clinical registry data from ACS-NSQIP and administrative claims data from Medicare. We tested the validity of this hybrid approach using 3 NQF-endorsed quality measures that assess risk-adjusted hospital performance for major postoperative complications and/or death after colectomy, LEB, or all surgical procedures. We found that agreement on hospital quality was poor between clinical registry and administrative claims data for all 3 measures. Hybrid models using clinical registry outcomes data (claims/clinical hybrid 2 models) had moderate agreement with clinical registry models on hospital quality determinations, whereas hybrid models using administrative outcomes data (clinical/claims hybrid 1 models) had routinely poor agreement with clinical registry models.

There are drawbacks to using either clinical registry or administrative claims data for national policies aimed at measuring surgical quality for public reporting or pay-for-performance. Clinical registries are generally considered to be a more valid and reliable data source for quality measurement than administrative claims because the data are collected specifically for that purpose.^{4–6} In ACS-NSQIP, trained and tested clinical personnel abstract data and hospitals are audited to ensure standardized data collection.^{10–12} In contrast, administrative claims data are collected for the purpose of reimbursement. In previous work, we demonstrated poor agreement between ACS-NSQIP and Medicare claims on the patient-level coding of postoperative complications, and other researchers have published similar results.^{6,17–20} This study demonstrates that these patient-level differences in coding result in substantial differences in determinations of hospital risk-adjusted surgical quality when using a clinical registry source versus an administrative claims source.

Unfortunately, participation in a clinical registry can be burdensome to hospitals, especially smaller hospitals with limited resources, because of the need to employ a data abstractor and the labor intensity of abstracting data from medical records. Some of this data collection burden may soon be mitigated by the increasing use of electronic health records; however, this is not currently an option for most hospitals. Here, administrative claims data offer an advantage because the data are routinely available in large aggregated databases and can be used for quality measurement without incurring any extra data collection burden to hospitals.

The quality measures used in this study were previously evaluated and endorsed by the NQF, which is an independent organization that reviews, endorses, and recommends performance measures for health care. Most performance measures that are used by Medicare for public reporting or pay-for-performance have been previously recommended by the NQF, although this is not a strict requirement. To be recommended by the NQF, a measure must be reviewed by a multidisciplinary panel of experts and meet strict criteria regarding relevance, reliability, validity, and feasibility.

The findings in this study should be interpreted in light of several limitations. First, patient-level records between the 2 databases were linked using indirect identifiers and there may be some small level of inaccuracy in the matches. However, we found excellent agreement between the 2 sources for 30-day postoperative mortality, which supports the validity of the linkage procedure. Second, the variable crosswalk developed to compare postoperative

complications between the 2 data sources is not perfect despite the use of published literature and an extensive review of codes. Third, coding practice could change over time, especially as payment reforms are initiated, disease diagnoses are refined, new procedures are introduced, and specific codes are chosen as performance metrics. Fourth, although previous audits have demonstrated substantial or almost perfect agreement on the coding of most variables in ACS-NSQIP, there is no way to confirm total accuracy of data recorded by hospitals. Finally, ACS-NSQIP hospitals in this data set are predominantly larger medical centers, which may limit the generalizability of the findings.

CONCLUSIONS

Assessments of hospital surgical quality differ substantially depending upon whether clinical registry or administrative claims data are used for the analysis. A hybrid approach using administrative claims data for risk adjustment and clinical registry data for postoperative complications may be a valid alternative to relying exclusively on clinical registry data. This approach would lower the data collection burden for hospitals by capitalizing on more easily accessible administrative claims data from Medicare while maintaining the validity of the quality measures by using rigorously specified clinical data for the outcomes. These findings have implications for policies focused on evaluating surgical quality for public reporting and pay-for-performance. Administrative claims data may be valid for some surgical quality measures, such as processes of care or outcomes such as mortality and readmission. However, for quality measures focused on postoperative complications to be meaningful, such policies should require, at a minimum, collection of clinical outcomes data.

REFERENCES

1. Acute Care Hospital Inpatient Prospective Payment System. Payment System Fact Sheet Series. 2013. Available at: <http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/AcutePaymtSysfshst.pdf>. Accessed September 19, 2013.
2. Emanuel EJ, Fuchs VR. The perfect storm of overutilization. *JAMA*. 2008; 299:2789–2791.
3. Medicare and Medicaid programs: hospital outpatient prospective payment; ambulatory surgical center payment; hospital value-based purchasing program; physician self-referral; and patient notification requirements in provider agreements. *Fed Regist*. 2011;76:74122–74584.
4. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127:666–674.
5. Romano PS, Chan BK, Schembri ME, Rainwater JA. Can administrative data be used to compare postoperative complication rates across hospitals? *Med Care*. 2002;40:856–867.
6. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg*. 2012;256:973–981.
7. Lawson EH, Zingmond DS, Hall BL, et al. Comparison between clinical registry and Medicare claims data on the classification of hospital quality of surgical care. *Ann Surg*. 2015;261:290–296.
8. Hospital Compare. American College of Surgeons National Surgical Quality Improvement Program surgical outcome measures. Available at: <http://www.medicare.gov/hospitalcompare/acs-surgical-measures.html>. Accessed December 27, 2013.
9. Research Data Assistance Center (ResDAC). Available at: http://www.resdac.org/Medicare/Data_File_Descriptions_RIF.asp. Accessed November 18, 2011.
10. Ingraham AM, Richards KE, Hall BL, Ko CY. Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. *Adv Surg*. 2010;44:251–267.
11. Shiloach M, Frencher SK Jr, Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg*. 2010;210:6–16.
12. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix

- adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg*. 2013;217:336–346.
13. Lawson EH, Ko CY, Louie R, et al. Linkage of a clinical surgical registry with Medicare inpatient claims data using indirect identifiers. *Surgery*. 2013;153:423–430.
 14. Merkow RP, Hall BL, Cohen ME, et al. Validity and feasibility of the American College of Surgeons colectomy composite outcome quality measure. *Ann Surg*. 2013;257:483–489.
 15. Elixhauser Comorbidity Software. Available at: <http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp>. Accessed November 18, 2011.
 16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
 17. Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg*. 2002;194:257–266.
 18. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health Serv Res*. 2009;44:182–204.
 19. Steinberg SM, Popa MR, Michalek JA, et al. Comparison of risk adjustment methodologies in surgical quality improvement. *Surgery*. 2008;144:662–667.
 20. Cima RR, Lackore KA, Nehring SA, et al. How best to measure surgical quality? Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicators (AHRQ-PSI) and the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) postoperative adverse events at a single institution. *Surgery*. 2011;150:943–949.