UCLA

UCLA Previously Published Works

Title

Comparison between clinical registry and medicare claims data on the classification of hospital quality of surgical care.

Permalink

<https://escholarship.org/uc/item/7jm9475w>

Journal

Annals of surgery, 261(2)

ISSN

0003-4932

Authors

Lawson, Elise H
Zingmond, David S
Hall, Bruce Lee
et al.

Publication Date

2015-02-01

DOI

10.1097/sla.0000000000000707

Peer reviewed

Comparison Between Clinical Registry and Medicare Claims Data on the Classification of Hospital Quality of Surgical Care

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

Objective: To compare the classification of hospital statistical outlier status as better or worse performance than expected for postoperative complications using Medicare claims versus clinical registry data.

Background: Controversy remains as to the most favorable data source for measuring postoperative complications for pay-for-performance and public reporting policies.

Methods: Patient-level records (2005–2008) were linked between the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) and Medicare inpatient claims. Hospital statistical outlier status for better or worse performance than expected was assessed using each data source for superficial surgical site infection (SSI), deep/organ-space SSI, any SSI, urinary tract infection, pneumonia, sepsis, deep venous thrombosis, pulmonary embolism, venous thromboembolism, and myocardial infarction by developing hierarchical multivariable logistic regression models. Kappa statistics and correlation coefficients assessed agreement between the data sources.

Results: A total of 192 hospitals with 110,987 surgical patients were included. Agreement on hospital rank for complication rates between Medicare claims and ACS-NSQIP was poor-to-moderate (weighted κ : 0.18–0.48). Of hospitals identified as statistical outliers for better or worse performance by Medicare claims, 26% were also identified as outliers by ACS-NSQIP. Of outliers identified by ACS-NSQIP, 16% were also identified as outliers by Medicare claims. Agreement between the data sources on hospital outlier status classification was uniformly poor (weighted κ : –0.02–0.34).

Conclusions: Despite using the same statistical methodology with each data source, classification of hospital outlier status as better or worse performance than expected for postoperative complications differed substantially between ACS-NSQIP and Medicare claims.

Keywords: surgery, quality measurement, clinical registry, administrative claims

(*Ann Surg* 2015;261:290–296)

The Centers for Medicare and Medicaid Services (CMS) publicly report a number of hospital quality measures on their Hospital Compare Web site, including measures of structure, process, and outcome.¹ Recently passed legislation, including the Patient Protection and Affordable Care Act of 2010, suggests an increasing interest among policymakers to move toward outcomes-based assessments of quality. For example, outcome measures are included in CMS's Hospital Value-Based Purchasing program, which became effective in October 2012 and links a portion of hospital payment to better performance on quality measures for common, high-cost medical conditions and surgical procedures. This program is part of a broader governmental effort to improve the quality and efficiency of health care through new innovative patient care models, research on comparative effectiveness, and incentives and penalties linked to performance.^{2,3}

A valid and reliable measurement system is essential for quality comparisons such as those used for public reporting and pay-for-performance to ensure that high and low performing hospitals are accurately and reliably identified. Structural measures of quality (such as volume of procedures performed or staffing ratios) and process measures of quality (such as the timing of preoperative prophylactic antibiotic administration) are relatively easy to assess and typically do not require risk adjustment but are weakly linked to patient outcomes, undermining their validity and importance as quality measures.^{4–9} In contrast, outcomes, such as postoperative complications, may be difficult to identify and their related quality measures require risk adjustment. However, as quality measures, outcomes are more meaningful to patients and may have greater face validity among physicians.

In surgery, controversy remains as to the most favorable data source for the assessment of postoperative complications. Administrative data sources, such as Medicare claims, are attractive because they contain data for patients from virtually all hospitals nationwide. However, because these data are primarily submitted for obtaining hospital reimbursement, their use for quality measurement has been questioned. In contrast, clinical registry data are often considered a more valid and reliable source for quality measures. However, the rigorous standards that make clinical registries attractive also incur an added burden because of the resources necessary for data collection.^{10,11}

Previous work has shown differences in the patient-level coding of 30-day postoperative complications between clinical and administrative claims sources.^{12–16} The purpose of this study is to directly compare how these data sources classify hospital statistical outlier status as better, worse, or “as expected” performance for postoperative complications using a database that links records for surgical patients between a clinical registry [the American

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; ¶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; and Department of Surgery, John Cochran Veterans Affairs Medical Center, St Louis, MO; ||RAND Corporation, Santa Monica, CA; and **UCLA Jonathan and Karin Fielding School of Public Health, Los Angeles, CA.

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 through the Robert Wood Johnson Foundation Clinical Scholars Program. This study was funded by a contract from the Centers for Medicare and Medicaid Services (CMS). The remaining authors have nothing to declare.

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, David Geffen School of Medicine at UCLA, Department of Surgery, 10833 Le Conte Avenue, 72-215 CHS, Los Angeles, CA 90095. E-mail: elawson@mednet.ucla.edu.

Copyright © 2014 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/14/26102-0290

DOI: 10.1097/SLA.0000000000000707

College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP)] and Medicare inpatient claims. Our overarching goal is to provide information to guide current policy decisions regarding the measurement of hospital quality for surgical patients.

METHODS

Data Sources and Study Sample

This study used data from 2 sources, ACS-NSQIP and Medicare inpatient claims, which have been previously described.¹⁶ Briefly, ACS-NSQIP is a validated, institution-based, multispecialty, surgical registry of patient risk factors and 30-day postoperative outcomes. Hospital participation in ACS-NSQIP is voluntary but requires a dedicated data abstractor who is trained to use strict variable definitions and collection methods. The sampling strategy includes collecting data for the first 40 cases performed within consecutive 8-day cycles. Cases are accrued from several surgical specialties, including general, vascular, and specific subspecialties. Sources for data are medical records and the patient. Information in the database is de-identified. Hospitals are audited to ensure standardized data collection, with audit results demonstrating substantial or almost perfect agreement on the coding of most variables.^{17–19} The 100% Medicare Provider Analysis and Review (MedPAR) file contains inpatient hospital final action stay records for Medicare beneficiaries assembled from claims submitted by hospitals for reimbursement. Hospitals are identifiable and each Medicare beneficiary has a unique identification number allowing for linkage of subsequent hospital visits. The Medicare Denominator file was used to identify beneficiary mortality.²⁰

Eligible patient-level records from ACS-NSQIP, years 2005–2008, were linked to Medicare inpatient claim records in MedPAR using indirect patient identifiers and a deterministic linkage algorithm, as previously described. There was excellent agreement between ACS-NSQIP and MedPAR records on death during the primary hospitalization, supporting the validity of the linkage procedure.²¹ The study population was restricted to patients aged 65 years or older who underwent a surgical procedure during the years studied, were entered into the ACS-NSQIP database, and were successfully matched to Medicare claims data. Patients with procedures occurring in 2005 were excluded because we used 1 year of preoperative Medicare data for risk-adjustment of the Medicare outcomes. Similarly, patients with procedures occurring in December 2008 were excluded because we lacked a full 30 days of follow-up Medicare data to identify outcomes for this group. In addition, hospitals with fewer than 100 cases were excluded.

This work was supported by a contract from the CMS, which approved the use of Medicare claims data. The RAND Corporation institutional review board approved the study protocol. Data analysis was performed by the authors. All analyses were performed using SAS version 9.2 software (SAS Institute, Inc, Cary, NC).

Analyses

Hospital rates of the following 30-day postoperative complications were assessed using clinical data from ACS-NSQIP and inpatient claims data from MedPAR: superficial surgical site infection (SSI), deep/organ-space SSI, urinary tract infection, pneumonia, deep venous thrombosis (DVT), pulmonary embolism, and myocardial infarction. These complications represent the outcomes that the Surgical Care Improvement Project process measures were designed to improve.²² In addition, we assessed 2 combined variables to be consistent with postoperative complication definitions currently included as Hospital-Acquired Conditions²³ and/or Patient Safety Indicators:²⁴ any SSI (superficial and/or deep/organ-space) and venous thromboembolism (DVT and/or pulmonary embolism).

In ACS-NSQIP, postoperative complications are recorded as individual binary data fields (occurrence vs no occurrence of

specified complication). In contrast, MedPAR does not contain distinct data fields for complications. Instead, this information is identified from *International Classification of Diseases, 9th Edition (ICD-9)* codes entered into up to 10 diagnosis fields and up to 6 procedure fields. The study team created a crosswalk that matched variables in ACS-NSQIP with applicable ICD-9 codes in Medicare data, as previously described¹⁶ (see Table, Supplemental Digital Content 1, available at <http://links.lww.com/SLA/A556>). Because a “present-on-admission” qualifier is not currently available in Medicare data, codes that could represent a preoperative comorbidity rather than a postoperative complication were excluded from the crosswalk when possible.¹⁶ For MedPAR, codes in the index admission record and records for subsequent readmissions occurring within 30 days of surgery were examined. In addition, to further validate the linkage procedure, hospital rates of 30-day postoperative mortality were assessed as this represents an outcome that should have a high level of agreement between the 2 data sources.

Agreement between the data sources on unadjusted postoperative complication rates for each hospital was evaluated by the Spearman correlation coefficient. In addition, for each complication, hospitals were ranked by their rate and split into equal decile groups. Agreement on hospitals’ decile rank as determined by each data source was determined by the weighted kappa statistic. Interpretation of kappa follows Fleiss’s magnitude guidelines, which propose that kappa less than 0.4 indicates poor agreement, 0.4 to 0.75 indicates moderate agreement, and greater than 0.75 indicates excellent agreement.²⁵

The SAS routine PROC GLIMMIX was used to develop hierarchical multivariable logistic regression models for each postoperative complication using each data source (22 total models). Hospital was entered as a random effect. This approach accounts for clustering of patients within hospitals by allowing each hospital to have a different random intercept and incorporates the empirical Bayes method. The ACS-NSQIP models included 22 clinically relevant preoperative variables for case-mix adjustment (see Table 1). Missing values for body mass index and American Society of Anesthesiology class were imputed by hospital using the hot deck method. The Medicare claim models included the Elixhauser comorbidity variables for case-mix adjustment. These variables were created using previously validated and commonly used software that categorizes applicable ICD-9, Diagnosis-Related Group and Medicare Severity–Diagnosis-Related Group codes into meaningful comorbidity variables.²⁶ In addition, all clinical and claim models were adjusted for procedure-mix using a Current Procedural Terminology–based risk score, which is the standard for ACS-NSQIP quality measure modeling and has been included in quality measures approved by the National Quality Forum.²⁷ Statistical model diagnostics were calculated for each model, including relative goodness of fit (Akaike information criterion), discriminative power (C statistic), and calibration (Hosmer-Lemeshow statistic) (see Table, Supplemental Digital Content 2, available at <http://links.lww.com/SLA/A556>).

Individual hospital performance for a complication was estimated using the hospital intercept odds ratio derived from the applicable hierarchical multivariable logistic regression model. This odds ratio estimates the odds of having the complication at the specified hospital versus the odds of having the complication at a theoretical “average” hospital, adjusted for covariates in the model. Agreement between the data sources on risk-adjusted hospital performance for each complication was evaluated by the Spearman correlation coefficient. Hospital odds ratios for each complication were then ranked and split into equal decile groups. Agreement on hospitals’ decile rank as determined by each data source was determined by the weighted kappa statistic.

TABLE 1. Demographic and Preoperative Clinical Characteristics of Study Population From a Database Linking Clinical Surgical Registry Data (ACS-NSQIP) With Medicare Inpatient Claims Data (MedPAR).

	Patients in Linked Database (n = 110,987)
Surgery type, %	
General	56.8
Vascular	29.6
Cardiothoracic	5.2
Gynecology	0.9
Head and neck	0.2
Neurosurgery	0.2
Orthopedic	5.4
Plastic	0.2
Urology	1.4
Age, %, yrs	
65–74	47.9
75–84	39.8
≥85	12.3
Male sex, %	48.4
Body mass index, %, kg/m ²	
Underweight (<18.5)	3.1
Normal (18.5–24.9)	32.6
Overweight (25–29.9)	35.1
Class I obesity (30–34.9)	18.1
Class II obesity (35–39.9)	6.9
Class III obesity (≥40)	4.3
Functional status, %	
Independent	84.3
Partially dependent	11.3
Totally dependent	4.4
ASA class, %	
I	0.6
II	22.3
III	60.7
IV	15.7
V	0.8
Smoker, %	14.4
Ventilator dependent, %	1.8
Chronic obstructive pulmonary disease	11.4
Diabetes, %	
None	77.8
Oral medication	14.0
Insulin	8.3
Renal failure, %	1.1
Dialysis, %	1.1
Dyspnea, %	
None	80.6
With moderate exertion	16.4
At rest	3.1
Ascites, %	2.1
Congestive heart failure, %	2.6
Hypertension requiring medication, %	74.7
Disseminated cancer, %	3.4
Chronic steroid use, %	4.3
>10% weight loss in past 6 mo, %	4.3
Bleeding disorder, %	13.0
Sepsis, %	
None	87.9
Systemic inflammatory response	8.0
Sepsis	2.2
Septic shock	1.9
Wound class, %	
1–Clean	48.0
2–Clean/Contaminated	35.2
3–Contaminated	8.6
4–Dirty/Infected	8.3
Emergency procedure, %	15.2

Demographic and preoperative clinical characteristics identified from a clinical surgical registry (ACS-NSQIP) for patients operated on between the years 2006–2008.

A hospital was considered to be a high statistical outlier (worse than expected performance) if its odds ratio was greater 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 than 1 ($P \geq 0.05$) were labeled as “as expected” performance for their given patient population. Hospital outlier status classification was compared between ACS-NSQIP and Medicare claims, and agreement was evaluated by the weighted kappa statistic.

To test the sensitivity of our results, we repeated the analyses with (1) the inclusion of additional Medicare sources representing outpatient visits; and (2) case-mix adjustment using only age and sex for all clinical and claims models and no procedure-mix adjustment.

RESULTS

The study sample included 110,987 Medicare patients who underwent an inpatient surgical procedure from 192 hospitals. Table 1 lists demographic and preoperative clinical characteristics of the study population. The majority of patients underwent a procedure classified as general surgery or vascular surgery (57% and 30%, respectively).

Comparison of Hospital Unadjusted Performance Between the Data Sources

The median hospital percentage of patients with a 30-day postoperative complication was higher using Medicare inpatient claims data for most complications, with the exception of superficial SSI and any SSI, which were higher in ACS-NSQIP (Table 2). The median hospital percentage of patients with any SSI was almost twofold higher in ACS-NSQIP (6.5% vs 3.5%), whereas the median hospital percentages of patients with a urinary tract infection or myocardial infarction were more than twofold higher and almost fivefold higher in MedPAR, respectively. The interquartile range for hospital rates is reported in Table 2 for each postoperative complication using each data source. For example, using ACS-NSQIP data, half of the study hospitals had a percentage of patients with superficial SSI between 2.3% and 5.3%, with a quarter of hospitals having a lower percentage and a quarter having a higher percentage. Using Medicare data, half of the study hospitals had a percentage of patients with superficial SSI between 2.6% and 4.8%.

Correlation of hospitals' unadjusted complication rates between ACS-NSQIP and MedPAR was poor to moderate (correlation coefficients: 0.29–0.68). Of the complications studied, hospital rates of myocardial infarction were the least correlated between the data sources, whereas hospital rates of venous thromboembolism were the most correlated. Similarly, agreement beyond chance on hospital rank for complication rates was poor to moderate (weighted κ : 0.18–0.48). These findings indicate that hospitals in the study sample with relatively high rates of a complication according to ACS-NSQIP did not necessarily have relatively high rates in Medicare, and vice versa. In contrast, hospital rates of 30-day postoperative mortality were highly correlated between the data sources (correlation coefficient: 0.97) and agreement on hospital rank was excellent (weighted κ : 0.84), indicating that hospitals with a relatively high mortality rate in ACS-NSQIP also had a relatively high mortality rate in Medicare.

Comparison of Hospital Risk-adjusted Performance Between the Data Sources

Correlation of hospitals' risk-adjusted odds ratios for postoperative complications as determined by ACS-NSQIP versus Medicare was poor to moderate (0.21–0.62) (Fig. 1). Hospital odds ratios for myocardial infarction and SSI were the least correlated between the data sources, whereas hospital odds ratios for DVT were the most

TABLE 2. Comparison of Unadjusted Hospital 30-day Postoperative Complication Rates Between a Clinical Surgical Registry (ACS-NSQIP) and Administrative Data From Medicare Inpatient Claims (MedPAR) for 192 Hospitals

Outcomes	Median Hospital Percentage of Patients With Complication Using ACS-NSQIP Data (Interquartile Range)	Median Hospital Percentage of Patients With Complication Using MedPAR Data (Interquartile Range)	Correlation of Hospitals' Postoperative Complication Rates Between ACS-NSQIP vs MedPAR	Agreement on Hospitals' Decile Rank for Postoperative Complication Rate Between ACS-NSQIP vs MedPAR (Weighted κ)
Superficial SSI	3.91% (2.25%–5.27%)	3.51% (2.60%–4.83%)	Moderate (0.46)	Poor (0.27)
Deep/organ-space SSI	2.78% (1.69%–3.88%)	3.43% (2.60%–4.69%)	Moderate (0.49)	Poor (0.34)
Any SSI	6.54% (4.86%–8.52%)	3.51% (2.60%–4.83%)	Moderate (0.55)	Poor (0.38)
Urinary tract infection	3.49% (2.19%–4.56%)	7.68% (5.93%–9.14%)	Poor (0.34)	Poor (0.23)
Pneumonia	3.56% (2.39%–5.30%)	5.58% (4.19%–6.92%)	Moderate (0.41)	Poor (0.28)
Sepsis	6.22% (4.38%–9.05%)	8.83% (7.02%–10.88%)	Moderate (0.51)	Poor (0.33)
DVT	1.39% (0.82%–1.94%)	1.74% (1.10%–2.48%)	Moderate (0.67)	Moderate (0.45)
Pulmonary embolism	0.54% (0.31%–0.87%)	0.75% (0.44%–1.06%)	Moderate (0.57)	Moderate (0.41)
Venous thromboembolism	1.79% (1.29%–2.57%)	2.40% (1.65%–3.24%)	Moderate (0.68)	Moderate (0.48)
Myocardial infarction	0.46% (0.20%–0.73%)	2.28% (1.62%–3.20%)	Poor (0.29)	Poor (0.18)

Any SSI includes superficial and/or deep and/or organ-space SSI. Venous thromboembolism includes DVT and/or pulmonary embolism.

FIGURE 1. Agreement on hospitals' risk-adjusted odds ratios for postoperative complications between a clinical surgical registry (ACS-NSQIP) and administrative data from Medicare inpatient claims (MedPAR), as determined by the correlation coefficient and weighted kappa for decile rank ($n = 192$ hospitals).

correlated. Agreement beyond chance on hospital decile rank for complication odds ratios was also poor to moderate (weighted κ : 0.13–0.42). As an example, these findings indicate that, after risk adjustment, hospitals with a high odds of patients getting a deep/organ-space SSI according to ACS-NSQIP do not necessarily have a high odds of deep/organ-space SSI according to Medicare, and vice

versa. Hospitals' risk-adjusted odds ratios for 30-day postoperative mortality were moderately correlated between the data sources (correlation coefficient: 0.70) and agreement on hospitals' decile rank for mortality odds ratios was moderate as well (weighted kappa 0.51).

For most complications, ACS-NSQIP data classified a greater number of hospitals as statistical outliers for better or worse

performance than expected than did Medicare data (Table 3). For DVT, ACS-NSQIP, and MedPAR, both classified 20 hospitals as high outliers (worse performance than expected) and 5 hospitals as low outliers (better performance than expected). However, agreement on which hospitals were the high outliers was 50%, meaning 10 hospitals were classified as a high outlier by both data sources and 20 hospitals were classified as a high outlier by one data source and not by the other. There was 0% agreement on the 5 low outlier hospitals, meaning each data source identified a completely different group of hospitals as having better than expected performance.

Across all complications, on average 19% of hospitals were identified as outliers for better or worse performance using ACS-NSQIP (range 2–40%), compared with 12% using Medicare (range: 0–20%). Of the hospitals identified as outliers by Medicare, 26% were also identified as outliers by ACS-NSQIP, and of outliers identified by ACS-NSQIP, 16% were also identified as outliers by Medicare. Complete discordance, in which a hospital is labeled as a high outlier for performance by one data source and a low outlier by the other source, occurred infrequently. Agreement between the data sources on the classification of a hospital's outlier status as better, worse, or "as expected," performance was uniformly poor for the complications studied and for mortality.

The conclusions of this study were not altered by the inclusion of additional outpatient Medicare data sources in the analyses and agreement between ACS-NSQIP and Medicare comprehensive claims actually worsened (see Tables, Supplemental Digital Content 3, available at <http://links.lww.com/SLA/A556>). Excluding procedure-mix adjustment and including only age and sex for case-mix adjustment in the ACS-NSQIP and Medicare claims models for postoperative complications improved agreement between the data sources, but not enough to substantially change the conclusions of the study (see Tables, Supplemental Digital Content 3, available at <http://links.lww.com/SLA/A556>). For mortality, using only age and sex for case-mix adjustment resulted in near-perfect agreement between the data sources (correlation coefficient for hospitals' risk-adjusted odds ratios = 0.98).

DISCUSSION

Comparison of hospital quality is a cornerstone of policies aimed at improving care delivery and stemming the rapidly increasing cost of health care. In this study, 2 data sources under consideration for measuring hospital quality for postoperative complications were compared: clinical registry data from ACS-NSQIP and administrative data from Medicare inpatient claims. Agreement between these data sources was at best moderate for hospital-level unadjusted complication rates and worsened for risk-adjusted quality comparisons. Furthermore, agreement on hospital statistical outlier status for performance was universally poor for all postoperative complications studied. These findings have implications if these postoperative complications are chosen for public reporting and pay-for-performance policies, as it is possible that different hospitals would be rewarded or penalized depending upon the data source used.

Operationalizing the concept of hospital quality for surgical patients into measures that are understandable and meaningful has proved challenging. Public reporting and pay-for-performance policies initially focused on structural and process measures of quality because they are objective and thus relatively easy to assess. Some of these measures are falling out of favor. Surgical procedure volume has lost some level of validity as a structural measure of quality²⁸ due to a lack of evidence that selective referral to high-volume hospitals results in substantial benefit to patients.^{4,5} Similarly, process measures such as the timing of preoperative prophylactic antibiotic administration have been shown to be only weakly linked to outcomes, at best.^{9,29}

TABLE 3. Comparison of Hospital Risk-Adjusted Statistical Outlier Classification for 30-Day Postoperative Complications Between a Clinical Surgical Registry (ACS-NSQIP) and Administrative Data From Medicare Inpatient Claims (MedPAR) for 192 Hospitals

Outcome	No. Statistical High Outlier Hospitals (Worse Than Expected Performance) Identified by Each Data Source		No. Statistical Low Outlier Hospitals (Better Than Expected Performance) Identified by Each Data Source		No. Hospitals Classified as High Outliers by Both Data Sources	No. Hospitals Classified as Low Outliers by Both Data Sources	No. Hospitals Labeled as a "High Outlier" by ACS-NSQIP and a "Low Outlier" by MedPAR, or Vice Versa		No. Hospitals Labeled as an "Outlier" by ACS-NSQIP Model and "as Expected" by MedPAR, or Vice Versa	Agreement on Outlier Status Between ACS-NSQIP and MedPAR (Weighted κ)
	ACS-NSQIP	MedPAR	ACS-NSQIP	MedPAR			ACS-NSQIP	MedPAR		
Superficial SSI	30	16	20	2	2	0	1	1	62	Poor (−0.02)
Deep/organ-space SSI	20	16	13	3	2	1	1	1	44	Poor (0.05)
Any SSI	30	16	18	2	4	0	1	1	56	Poor (0.04)
Urinary tract infection	25	19	14	17	4	3	2	2	57	Poor (0.10)
Pneumonia	33	20	17	19	6	3	5	5	61	Poor (0.10)
Sepsis	48	17	29	14	6	5	2	2	82	Poor (0.10)
DVT	20	20	5	5	10	0	0	0	30	Poor (0.34)
Pulmonary embolism	4	0	0	0	0	0	0	0	4	Poor (0.30)
Venous thromboembolism	20	17	5	4	9	0	0	0	28	Poor (0.34)
Myocardial infarction	14	20	2	3	4	1	0	0	29	Poor (0.20)

Outlier status was determined using the hospital intercept odds ratio from the applicable hierarchical multivariable logistic regression, as described in the text. A hospital was considered to be a high outlier (worse than expected performance) if its odds ratio was > 1 ($P < 0.05$) and a low outlier (better than expected performance) if its odds ratio was < 1 ($P < 0.05$). Hospitals with odds ratios not significantly different than 1 ($P \geq 0.05$) were labeled "as expected" for their given patient population.

As a result, policymakers are making a move toward the development and implementation of outcomes-based measures, such as hospital performance for postoperative complications, because they more directly represent a potentially preventable source of patient harm and excess cost. However, there are methodological concerns with outcomes-based measures that still need to be considered. For example, controversy remains as to the most favorable data source for such quality measures. Although previous studies have evaluated the effect of data source at the patient-level,^{12–16} this study is the first large multi-institutional comparison of hospital quality measurements for postoperative surgical complications between a clinical registry and an administrative claims data source.

In this study, we observed substantial disagreement between ACS-NSQIP and Medicare claims on hospital complication rates, which is likely due to differences in data collection methodology.³⁰ The data in ACS-NSQIP are collected from the medical record (or from the patient when the medical record is not sufficient) by abstractors who are continually trained and tested on the use of detailed clinical variable definitions with strict inclusion and exclusion criteria. In contrast, ICD-9 codes tend to be less explicit and specific. For example, diagnoses labeled as “rule out” or “suspected” in the medical record would not be recorded in ACS-NSQIP but are routinely recorded as “present” in inpatient claims data.³¹ In addition, the diagnosis codes that are recorded in claims data for the purpose of reimbursement may be different from those that would be recorded if the primary objective was quality measurement, especially given the limited number of fields available for coding.

Certainly not all outcomes are better recorded in a clinical registry than in an administrative data set. For example, metrics such as mortality, readmission, and length of hospital stay are likely recorded accurately in administrative data—perhaps even more accurately than in a clinical registry. There may thus be an advantage to using administrative data to measure such outcomes.

Many hospitals lack the resources necessary to participate in registries, as clinical data abstraction remains a time-intensive and costly process. Automated abstraction of data from electronic health records (EHRs) may ease this burden and is likely to proliferate as electronic submission of clinical quality measures is now included in the criteria for demonstrating meaningful use of EHRs.³² It is worth noting, however, that care should be taken in the development of the EHR data source for quality measurement. Automated data entry that obviates the need for an independent data abstractor instead relies heavily on physicians to be comprehensive and reliable in documenting relevant patient risk factors and complications. Without an incentive for detailed, accurate documentation, risk factors may inadvertently be up-coded and complications omitted, for example, due to the ease of checking boxes in an electronic template. We certainly do not want to develop a system that quickly and easily obtains data of poor quality.

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 thus there is less certainty regarding the accuracy of 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 might not be perfect, despite the use of published literature and an extensive review of codes. Third, our data are from 2005 to 2008. It is possible that coding practices have changed since that time period, 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, we

have no way of confirming the 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

Classification of hospital statistical outlier status as better, worse, or “as expected” performance for the postoperative complications studied differed substantially between ACS-NSQIP and Medicare claims. On the basis of these results, public reporting and pay-for-performance policies focused on measuring the quality of surgical care using the postoperative complications included in this study should use data sources that employ methods to ensure the accuracy of abstracted data.

REFERENCES

1. US Department of Health and Human Services. Hospital compare. Available at: <http://www.hospitalcompare.hhs.gov/>. Accessed October 17, 2012.
2. Centers for Medicare and Medicaid Services (CMS), HHS. 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.
3. Hospital Readmissions Reduction Program—The Patient Protection and Affordable Care Act, Public law 111–148, section 3025, March 23, 2010. 124 STAT. 120. Available at: <http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>. Accessed November 18, 2011.
4. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002;137:511–520.
5. Livingston EH, Cao J. Procedure volume as a predictor of surgical outcomes. *JAMA*. 2010;304:95–97.
6. Bradley EH, Herrin J, Elbel B, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA*. 2006;296:72–78.
7. Werner RM, Bradlow ET. Relationship between Medicare’s hospital compare performance measures and mortality rates. *JAMA*. 2006;296:2694–2702.
8. Finfer S, Chittock DR, Su SY, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360:1283–1297.
9. Stulberg JJ, Delaney CP, Neuhauser DV, et al. Adherence to surgical care improvement project measures and the association with postoperative infections. *JAMA*. 2010;303:2479–2485.
10. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127:666–674.
11. Romano PS, Chan BK, Schembri ME, et al. Can administrative data be used to compare postoperative complication rates across hospitals? *Med Care*. 2002;40:856–867.
12. 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.
13. 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.
14. Steinberg SM, Popa MR, Michalek JA, et al. Comparison of risk adjustment methodologies in surgical quality improvement. *Surgery*. 2008;144:662–667.
15. 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.
16. 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.
17. 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.
18. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005;138:837–843.

19. Ingraham AM, Richards KE, Hall BL, et al. Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. *Adv Surg*. 2010;44:251–267.
20. Research Data Assistance Center (ResDAC). Available at: http://www.resdac.org/Medicare/Data_File_Descriptions_RIF.asp. Accessed November 18, 2011.
21. 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.
22. Surgical Care Improvement Project Core Measure Set. Available at: http://www.jointcommission.org/assets/1/6/Surgical_Care_Improvement_Project.pdf. Accessed November 18, 2011.
23. Centers for Medicare and Medicaid Services. Hospital-acquired conditions (HAC) in acute inpatient prospective payment system (IPPS) hospitals. Available at: <https://www.cms.gov/hospitalacqcond/downloads/hacfactsheet.pdf>. Accessed November 18, 2011.
24. Agency for Healthcare Research and Quality (AHRQ) Quality Indicators. Inpatient quality indicators overview. Available at: <http://www.qualityindicators.ahrq.gov/Modules/iqi-overview.aspx>. Accessed November 18, 2011.
25. Fleiss J, Levin B, Paik M. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ: Wiley; 2003.
26. Elixhauser Comorbidity Software. Available at: <http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp>. Accessed November 18, 2011.
27. 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.
28. The Leapfrog Group. Fact sheet: evidence-based hospital referral. Available at: <http://www.leapfroggroup.org/media/file/Leapfrog-Evidence-based-Hospital-Referral-Fact-Sheet.pdf>. Accessed September 25, 2012.
29. Ingraham AM, Cohen ME, Bilimoria KY, et al. Association of surgical care improvement project infection-related process measure compliance with risk-adjusted outcomes: implications for quality measurement. *J Am Coll Surg*. 2010;211:705–714.
30. Rosof B. The importance of accurate data in quality-of-care measurement. *Ann Intern Med*. 2012;157:379–380.
31. ICD-9-CM Official Guidelines for Coding and Reporting. Available at: http://www.cdc.gov/nchs/data/icd/icd9cm_guidelines_2011.pdf. Accessed November 18, 2011.
32. Medicare and Medicaid Programs; Electronic Health Record Incentive Program—Stage 2; Final Rule. *Fed Regist*;77:53967–54162.