A Comparison of Clinical Registry Versus Administrative Claims Data for Reporting of 30-Day Surgical Complications

Elise H. Lawson, MD, MSHS, Rachel Louie, MS,* David S. Zingmond, MD, PhD,* Robert H. Brook, MD, ScD,*†‡ Bruce L. Hall, MD, PhD, MBA,§|| Lein Han, PhD,¶ Michael Rapp, MD, JD,¶# and Clifford Y. Ko, MD, MSHS*§***

Objectives: To compare the recording of 30-day postoperative complications between a national clinical registry and Medicare inpatient claims data and to determine whether the addition of outpatient claims data improves concordance with the clinical registry.

Background: Policymakers are increasingly discussing use of postoperative complication rates for value-based purchasing. There is debate regarding the optimal data source for such measures.

Methods: Patient records (2005–2008) from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) were linked to Medicare inpatient and outpatient claims data sets. We assessed the ability of (1) Medicare inpatient claims and (2) Medicare inpatient and outpatient claims to detect a core set of ACS-NSQIP 30-day postoperative complications: superficial surgical site infection (SSI), deep/organ-space SSI, any SSI (superficial and/or deep/organ-space), urinary tract infection, pneumonia, sepsis, deep venous thrombosis (DVT), pulmonary embolism, venous thromboembolism (DVT and/or pulmonary embolism), and myocardial infarction. Agreement of patient-level complications by ACS-NSQIP versus Medicare was assessed by κ statistics.

Results: A total of 117,752 patients from more than 200 hospitals were studied. The sensitivity of inpatient claims data for detecting ACS-NSQIP complications ranged from 0.27 to 0.78; the percentage of false-positives ranged from 48% to 84%. Addition of outpatient claims data improved sensitivity slightly but also greatly increased the percentage of false-positives. Agreement was routinely poor between clinical and claims data for patient-level complications.

Conclusions: This analysis demonstrates important differences between ACS-NSQIP and Medicare claims data sets for measuring surgical complications. Poor accuracy potentially makes claims data suboptimal for evaluating surgical complications. These findings have meaningful implications for performance measures currently being considered.

From the *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; §Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL; ||School of Medicine and Barnes Jewish Hospital, Saint Louis Veterans Affairs Medical Center, and Center for Health Policy and Olin Business School, Washington University in Saint Louis, MO; ¶Centers for Medicare & Medicaid Services, Baltimore, MD; #Department of Emergency Medicine, George Washington University School of Medicine and Health Sciences, Washington, DC; and **VA Greater Los Angeles Healthcare System, 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 & Medicaid Services (CMS). None of the remaining authors had any conflicts of interests to declare. The views expressed in this article represent the authors' views and do not necessarily represent official policy or opinions of the Department of Health and Human Services, the Centers for Medicare & Medicaid Services.

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.

ISSN: 0003-4932/12/25606-0973
DOI: 10.1097/SLA.0b013e31826b4c4f

Keywords: Administrative claims, clinical registry, complications, postoperative, quality measurement, surgery

(*Ann Surg* 2012;256: 973–981)

Surgeons have a long history of tracking postoperative complications for their patients to understand treatment outcomes better and continuously improve the quality of surgical care. Dr Ernest Codman first described the concept of following patient outcomes with what he termed his "end result idea," which he explained as "The common sense notion that every hospital should follow every patient it treats, long enough to determine whether or not the treatment has been successful, and then to inquire, 'If not, why not?' with a view to preventing similar failures in the future."[1(p53)] Now, almost a century later, every hospital routinely holds Morbidity and Mortality conferences in which 30-day postoperative complications are presented and discussed with the aim of determining how care could be delivered differently to prevent future complications.

There are numerous data sources, including registries of clinical chart abstracted data and administrative claims databases, available for assessing postoperative complications. There are important differences between these 2 data sources. Proponents of clinical chart abstracted data assert that measures of quality using this source are more valid and reliable than measures using administrative claims data.[2,3] However, measures that require clinical data can incur an added burden for hospitals as these data are abstracted directly from patient records for the purpose of quality measurement. In contrast, administrative claims data are routinely collected for the purpose of submitting claims for payment and thus are available for patients nationwide at no additional burden to hospitals.

Studies comparing clinical registry data with administrative claims data for postoperative complications have typically found that administrative data sources have low sensitivity for detecting complications recorded in the clinical registry—meaning a substantial number of complications identified in the clinical registry are missed by the administrative sources.[4–7] One explanation for these observed differences is that the administrative sources studied focus only on inpatient events. When evaluating 30-day postoperative complications, it is important to follow a patient's course even after discharge from the primary hospitalization, or else a substantial number of complications and deaths may be missed. For example, previous work showed that 66% of surgical site infections (SSIs) and 42% of pulmonary embolisms occurred after discharge and that exclusion of postdischarge complications has a considerable effect on rankings of hospital quality.[8] Sources that rely on inpatient data alone for surgical quality measurement may thus be providing an incomplete assessment.

To perform an "apples-to-apples" comparison of the recording of 30-day postoperative complications, we compared a clinical registry, the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), with comprehensive inpatient and outpatient administrative claims data from Medicare. We focused on complications that are the focus of recent public reporting and pay-for-performance policies: superficial SSI, deep/organ-space

SSI, any SSI (superficial and/or deep/organ-space), urinary tract infection (UTI), pneumonia, sepsis, deep venous thrombosis (DVT), pulmonary embolism, venous thromboembolism (DVT and/or pulmonary embolism), and myocardial infarction. Our objectives were (1) to compare the recording of 30-day postoperative complications between ACS-NSQIP and Medicare inpatient claims and (2) to determine whether the addition of Medicare outpatient claims data improves concordance with the clinical registry.

METHODS

Data Sources and Study Sample

We used 2 sources of data for this study: ACS-NSQIP and comprehensive Medicare inpatient and outpatient claims. The ACS-NSQIP is an 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. Data are collected across several surgical specialties, including general, vascular, and specific subspecialties. Patients who underwent surgical procedures during a hospital admission for trauma or transplantation, and patients who are classified as American Society of Anesthesiology (ASA) class 6 (brain-dead organ donor), are excluded from the sample frame. Sources for data are medical records and the patient. Data collected include demographics (sex, age); dates of admission, performance of the procedure, and discharge; preoperative risk factors and laboratory values; procedure performed by Current Procedural Terminology (CPT) code; indication for surgery by *International Classification of Diseases, 9th edition* (*ICD-9*) code; and postoperative complications occurring within 30 days of the index operation. Hospitals are identifiable, but patients in the database are deindentified. Hospitals are audited to ensure standardized data collection, with audit results demonstrating substantial or almost perfect agreement on the coding of most variables.[9] Participating hospitals receive semiannual reports with risk-adjusted outcomes from ACS-NSQIP that allow them to benchmark their performance with national averages.[9–11]

From Medicare, we used comprehensive claims data, including inpatient and outpatient data from institutions and physicians. Medicare is a health insurance program that enrolls people aged 65 years or older, some disabled people younger than 65 years, and all people with end-stage renal disease receiving dialysis. The 100% Medicare Provider Analysis and Review file (MedPAR) contains inpatient hospital and skilled nursing facility final action stay records for all Medicare beneficiaries receiving health care services at an inpatient facility in the United States. Each record in the MedPAR file represents an inpatient hospital stay for a beneficiary and summarizes all services rendered to that beneficiary from admission to discharge. As such, each record may represent one claim or multiple claims depending upon the extent of inpatient services used by the beneficiary. The Inpatient and Outpatient claim files contain final action claims data submitted for reimbursement by inpatient hospital providers and institutional outpatient providers, respectively. The Carrier claim file contains final action claims data submitted by non-institutional providers e.g., all office-based physicians. Each record in the Inpatient, Outpatient, and Carrier claim files is at the claim level. Data elements used from the earlier-mentioned files include demographics (sex, birth date); dates of admission, performance of procedures, and discharge; and diagnoses and procedures by *ICD-9* code. Hospitals are identifiable in these data sources and each Medicare beneficiary in the database has a unique identification number allowing for linkage of subsequent hospital visits without disclosing the patient's identity. The Medicare Denominator file was used to identify beneficiary mortality and date of death.[12]

Eligible patient-level records from ACS-NSQIP, years 2005 to 2008, were linked to Medicare inpatient claims data in MedPAR using indirect patient identifiers. Details of the linkage procedure are described in detail elsewhere.[13] Briefly, patient records were grouped by hospital and then matched by combinations of age, sex, dates of admission, performance of procedure, and discharge, procedure performed, and postoperative diagnosis. As previously reported, there was excellent agreement between ACS-NSQIP and MedPAR records on death during the primary hospitalization, supporting the validity of the linkage procedure. Our 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 inpatient claims data. We excluded patients with procedures occurring in December 2008 because we lacked a full 30 days of follow-up. Data from additional Medicare sources (Inpatient, Outpatient, and Carrier claim files and the Denominator file) were added to the linked database using the Medicare beneficiary identification number.

This work was supported by a contract from the Centers for Medicare & Medicaid Services (CMS), who approved the use of Medicare claims data. In addition, the RAND Corporation Institutional Review Board approved the study protocol. All analyses were conducted with SAS version 9.2.

Comparison of Patient-Level Postoperative Complications

Postoperative complications are recorded in ACS-NSQIP as individual binary data fields (occurrence vs no occurrence of specified complication) and have an associated data field for the date of occurrence. Each variable is strictly defined in the ACS-NSQIP operations manual and data abstractors undergo training and testing to ensure standardized data collection. In contrast, Medicare claims do not contain distinct data fields for complications. Instead, this information is identified from *ICD-9* codes entered into up to 10 diagnosis fields and up to 6 procedure fields in MedPAR. Whether a diagnosis code refers to a pre-existing condition (ie, a risk factor) or a postoperative complication (ie, an outcome) can sometimes be determined by the code description, but it is often unknown. Currently, condition present-on-admission qualifiers are not reported for MedPAR diagnosis codes.

To compare the postoperative complications entered into these databases, we created a crosswalk that matches variables in ACS-NSQIP with applicable *ICD-9* diagnosis codes in the Medicare data (Table 1). This crosswalk was created through careful review and classification of applicable *ICD-9* codes after consulting published literature and relevant measures from the Elixhauser Comorbidity Software,[14] Quality Indicators from the Agency for Healthcare Research and Quality,[15] and the list of hospital-acquired conditions from CMS.[16] We created and tested multiple variations of some Medicare variables to determine the optimal definition. Codes that could represent a preoperative comorbidity rather than a postoperative complication were excluded when possible. For example, we only included codes for *acute* myocardial infarction (*ICD-9* code prefix 410) and excluded codes for *old* myocardial infarction (*ICD-9* code prefix 412). In addition, the acute myocardial infarction *ICD-9* code prefix requires a fourth digit indicating the location of the infarction and a fifth digit indicating whether it is the first episode of care for a newly diagnosed myocardial infarction (fifth digit = 1), a subsequent follow-up episode of care (fifth digit = 2) or unspecified (fifth digit = 0). Codes with a fifth digit of 2 were excluded.

TABLE 1. *ICD-9* Diagnosis Codes Used to Identify 30-Day Postoperative Complications in Medicare Claims Data

ACS-NSQIP Defined Postoperative Complications	*ICD-9-CM* Diagnosis Codes
Superficial SSI	9985, 99851, 99859
Deep/organ-space SSI	99859
Any SSI	9985, 99851, 99859
UTI	1122, 5901*, 5903, 5908*, 5950, 5953, 5990, 99664
Pneumonia	0391, 1124, 1179, 1363, 46619, 480*, 481, 482*, 483*, 4841, 4846, 4847, 485, 486, 4870, 507*, 5130, 5168, 99731, 99739
Sepsis	038*, 78552, 99591, 99592, 9980, 99859, 99931
DVT	4511*, 4512, 45181, 4519, 4534*, 4538, 4539
Pulmonary embolism	4151*
Venous thromboembolism	4151*, 4511*, 4512, 45181, 4534*, 4538, 4539
Myocardial infarction	410*0, 410*1

We searched for codes recorded for episodes of care within 30 days of surgery. For brevity, the * represents all fourth or fifth digits that could designate an *ICD-9-CM* code. For example, 4151* = 41511, 41512, and 41519. Any SSI includes superficial and/or deep/organ-space SSI. Venous thromboembolism includes DVT and/or pulmonary embolism.

We calculated patient-level rates of the following 30-day postoperative complications: superficial SSI, deep/organ-space SSI, any SSI (superficial and/or deep/organ-space), UTI, pneumonia, sepsis, DVT, pulmonary embolism, venous thromboembolism (DVT and/or pulmonary embolism), and myocardial infarction. These complications represent the outcomes that the Surgical Care Improvement Project process measures were designed to improve.[17] For Medicare inpatient claims, we searched for codes that would identify these complications in the record for the index admission or records for subsequent readmissions occurring within 30 days of surgery. For the outpatient data sources, we searched for codes associated with episodes of care occurring within 30 days of surgery. A patient was labeled as having a specified complication if any of the defined codes for that complication were identified in any of the Medicare data sources. We also tried stricter rules, such as (1) the presence of the code on 1 or more inpatient claims OR 2 or more outpatient claims and (2) the presence of the code on 1 or more physician outpatient claims OR 1 or more physician inpatient claims OR 2 or more claims from any source. However, these stricter rules did not substantially improve the results. In addition, to further validate the linkage procedure, we assessed occurrence of 30-day postoperative mortality as this represents an outcome that should have a high level of agreement between the 2 data sources.

Next, we calculated the sensitivity, specificity, positive predictive value, and negative predictive value of the Medicare variables for detecting ACS-NSQIP complications. The analyses were first performed with only inpatient Medicare claims data from MedPAR included and then repeated with the additional outpatient Medicare data sources included. Results of these 2 sets of analyses were compared. Patients with a complication recorded in Medicare but not recorded in ACS-NSQIP were considered false-positives, whereas patients *without* a complication recorded in Medicare and *with* a complication recorded in ACS-NSQIP were considered false-negatives. The κ statistic was calculated to determine agreement between the Medicare and ACS-NSQIP variables. The κ statistic takes into account agreement occurring by chance. Interpretation of the κ values follows Fleiss's magnitude guidelines, which propose that κ value of less than 0.4 indicates poor agreement, 0.4 to 0.75 indicates moderate agreement, and more than 0.75 indicates excellent agreement.[18]

RESULTS

The sample for the study included 117,752 patients from 214 hospitals. Table 2 lists demographic and preoperative clinical characteristics of the study population. The majority of patients underwent a procedure classified as general surgery or vascular surgery (57.1% and 29.7%, respectively).

Table 3 lists the percentages of patients with each 30-day postoperative complication as recorded in ACS-NSQIP, Medicare inpatient claims from MedPAR, and Medicare comprehensive claims (inpatient and outpatient). Most complications had higher rates in Medicare inpatient claims compared to ACS-NSQIP, with the exception of superficial SSI and any SSI. The superficial SSI rates were similar between ACS-NSQIP and Medicare inpatient claims (4.2% vs 4.0%, respectively), but any SSI rate was almost twofold higher in ACS-NSQIP (7.0% vs 4.0%, respectively). After inclusion of outpatient Medicare data, only any SSI had a higher rate in ACS-NSQIP and the difference was smaller (7.0% vs 6.1%). Myocardial infarction had the greatest difference in rate between the data sources, with an almost fivefold higher rate in Medicare inpatient claims compared to ACS-NSQIP (2.6% vs 0.5%, respectively) and a sevenfold higher rate in Medicare comprehensive claims (3.8%).

The sensitivity of Medicare inpatient claims for detecting complications in ACS-NSQIP ranged from poor to excellent (Table 4). For example, Medicare inpatient claims identified 27% of the superficial SSIs recorded in ACS-NSQIP (sensitivity = 0.27) and 78% of the myocardial infarctions (sensitivity = 0.78). Specificity was high for all complications examined, with Medicare inpatient claims data correctly not recording a specified complication for 94% to—more than 99% of the patients determined to not have the specified complication in ACS-NSQIP (specificity: 0.94–>0.99). Positive predictive value ranged from 0.16 to 0.52. Any SSI had the greatest positive predictive value (0.52), meaning of patients that Medicare inpatient claims identified as having an SSI, 52% had an SSI recorded in ACS-NSQIP as well. In contrast, among patients whom Medicare inpatient claims identified as having a myocardial infarction, 16% had a myocardial infarction recorded in ACS-NSQIP as well (positive predictive value = 0.16). In other words, 48% of the SSIs and 84% of the myocardial infarctions recorded in Medicare inpatient claims were false-positives, using ACS-NSQIP as the standard. Like specificity, negative predictive value was excellent for all complications studied, meaning that among patients identified as *not* having a specified complication in Medicare inpatient claims, 95% to more than 99% also did not have the specified complication recorded in ACS-NSQIP (negative predictive value: 0.95–>0.99). In other words, the percentage of false-negatives in Medicare inpatient claims ranged from less than 1% to 5%, using ACS-NSQIP as the standard. Agreement beyond chance between ACS-NSQIP and Medicare inpatient claims on whether or not a patient had a complication was poor to moderate (κ: 0.25–0.57).

The addition of outpatient data increased the sensitivity of Medicare data for detecting complications in ACS-NSQIP by 1.1- to 1.7-fold (Table 5). Sensitivity increased most for DVT (0.46–0.79) and least for myocardial infarction (0.78–0.87). Sensitivity was greatest for myocardial infarction and pulmonary embolism, with 87% of each of these events recorded in ACS-NSQIP also being recorded in Medicare comprehensive claims (sensitivity: 0.87 for both). Sensitivity remained lowest for superficial SSI, with Medicare comprehensive claims recording 39% of the superficial SSIs recorded in ACS-NSQIP (sensitivity: 0.39). The specificity of Medicare data was lowered with the addition of outpatient data, but was still excellent (0.91–0.99),

TABLE 2. Demographic and Preoperative Clinical Characteristics of Study Population From a Database Linking Clinical Surgical Registry Data (ACS-NSQIP) With Medicare Comprehensive Claims Data

	Patients in Linked Database (n = 117,752)
Surgery type, %	
General	57.1
Vascular	29.7
Cardiothoracic	5.1
Gynecology	0.9
Head and neck	0.2
Neurosurgery	0.2
Orthopedic	5.2
Plastic	0.2
Urology	1.4
Age (y), %	
65–74	48.0
75–84	39.8
≥85	12.2
Male sex, %	48.4
Body mass index (kg/m^2), %	
Underweight (<18.5)	3.5
Normal (18.5–24.9)	31.5
Overweight (25–29.9)	33.9
Class I obesity (30–34.9)	17.3
Class II obesity (35–39.9)	6.6
Class III obesity (≥40)	4.1
Unknown	3.2
Functional status, %	
Independent	84.3
Partially dependent	11.3
Totally dependent	4.4
ASA class, %	
I	0.6
II	22.6
III	60.4
IV	15.5
V	0.8
Smoker, %	14.4
2 drinks/d alcohol, %	2.9
Chronic obstructive pulmonary disease	11.3
Diabetes, %	
None	77.9
Oral medication	13.9
Insulin	8.2
Renal failure/dialysis, %	2.0
Dyspnea, %	
None	80.7
With moderate exertion	16.2
At rest	3.1
Ascites, %	2.1
Congestive heart failure, %	2.7
History of myocardial infarction, %	1.7
Hypertension requiring medication, %	74.6
Disseminated cancer, %	3.4
Chronic steroid use, %	4.3
>10% weight loss in past 6 mo, %	4.4
Bleeding disorder, %	12.9
Chemotherapy, %	1.4
Radiation, %	1.2
Sepsis, %	
None	88.0
Systemic inflammatory response	7.9
Sepsis	2.2
Septic shock	1.9
Emergency procedure, %	15.3

Demographic and preoperative clinical characteristics identified from a clinical surgical registry (ACS-NSQIP) for patients operated on between the years 2005–2008.

TABLE 3. Percentage of Patients With a 30-Day Postoperative Complication Recorded in a Clinical Surgical Registry (ACS-NSQIP), an Inpatient-Only Administrative Data Set (Medicare Inpatient Claims—MedPAR), and an Inpatient/Outpatient Administrative Data Set (Medicare Comprehensive Claims) for 117,752 Surgical Patients

	Percentage of Patients With Complication Recorded in ACS-NSQIP (95% CI)	Percentage of Patients With Complication Recorded in MedPAR (95% CI)	Percentage of Patients With Complication Recorded in Medicare Comprehensive Claims (95% CI)
Superficial SSI	4.23% (4.12–4.35)	3.96% (3.85–4.07)	6.13% (5.99–6.27)
Deep/organ-space SSI	2.93% (2.84–3.03)	3.86% (3.75–3.97)	5.91% (5.78–6.05)
Any SSI	6.95% (6.80–7.09)	3.96% (3.85–4.07)	6.13% (5.99–6.27)
UTI	3.48% (3.38–3.59)	7.39% (7.24–7.54)	10.87% (10.69–11.04)
Pneumonia	4.11% (3.99–4.22)	5.53% (5.40–5.66)	9.16% (8.99–9.32)
Sepsis	6.98% (6.83–7.12)	8.81% (8.65–8.97)	12.25% (12.07–12.44)
DVT	1.58% (1.51–1.65)	1.98% (1.91–2.06)	5.06% (4.93–5.18)
Pulmonary embolism	0.62% (0.58–0.67)	0.80% (0.74–0.85)	1.67% (1.60–1.75)
Venous thromboembolism	2.06% (1.98–2.14)	2.57% (2.48–2.66)	6.10% (5.96–6.23)
Myocardial infarction	0.54% (0.50–0.58)	2.59% (2.50–2.68)	3.79% (3.68–3.89)

All complications include any occurrence within 30 days of surgery. For the Medicare inpatient claims, we searched for codes included in MedPAR for the index admission or subsequent admissions within 30 days of surgery. Medicare comprehensive claims include inpatient and outpatient claims data from physicians and institutions. Any SSI includes superficial and/or deep/organ-space SSI. Venous thromboembolism includes DVT and/or pulmonary embolism.

CI indicates confidence interval.

TABLE 4. Comparison of Patient-Level 30-Day Postoperative Complications Recorded in a Clinical Surgical Registry (ACS-NSQIP) Versus an Inpatient Administrative Database (Medicare Inpatient Claims—MedPAR) for 117,752 Surgical Patients

	Percentage of Patients With Specified Complication Recorded in ACS-NSQIP Who Also Had the Complication Recorded in MedPAR (Sensitivity)	Percentage of Patients *Without* the Specified Complication Recorded in ACS-NSQIP Who Also *Did Not* Have the Complication Recorded in MedPAR (Specificity)	Percentage of Complications Recorded in MedPAR that Are False-Positives*	Percentage of Patients *Without* a Complication Recorded in MedPAR That Are False-Negatives*	Agreement Beyond Chance on the Recording of Complications Between ACS-NSQIP and MedPAR (κ)
Superficial SSI	27.0%	97.1%	71%	3%	Poor (0.25)
Deep/organ-space SSI	34.4%	97.1%	74%	2%	Poor (0.27)
Any SSI	29.9%	98.0%	48%	5%	Poor (0.35)
UTI	45.0%	94.0%	79%	2%	Poor (0.25)
Pneumonia	49.7%	96.4%	63%	2%	Poor (0.40)
Sepsis	46.3%	94.0%	63%	4%	Poor (0.36)
DVT	46.2%	98.7%	63%	<1%	Moderate (0.40)
Pulmonary embolism	64.8%	99.6%	49%	<1%	Moderate (0.57)
Venous thromboembolism	52.6%	98.5%	58%	1%	Moderate (0.46)
Myocardial infarction	78.4%	97.8%	84%	<1%	Poor (0.26)

* Using ACS-NSQIP as the standard.

All complications include any occurrence within 30 days of surgery. For the Medicare inpatient claims, we searched for codes included in the index admission or subsequent admissions within 30 days of surgery. Any SSI includes superficial and/or deep/organ-space SSI. Venous thromboembolism includes DVT and/or pulmonary embolism.

meaning that the vast majority of patients identified as not having a specified complication in ACS-NSQIP also were not recorded as having the specified complication in Medicare comprehensive claims.

The addition of outpatient Medicare data decreased the positive predicative value of Medicare data for detecting complications in ACS-NSQIP (Table 5). In other words, the percentage of patients with a false-positive complication recorded in Medicare comprehensive claims increased, using ACS-NSQIP as the standard. Pulmonary embolism had the greatest increase in false-positives (49%–68%), whereas superficial SSI, deep/organ-space SSI, and UTI had the smallest increases (71%–73%, 74%–76%, and 79%–81%, respectively). Myocardial infarction had the highest percentage of false-positives in Medicare comprehensive claims (88%), and any SSI had the lowest percentage (52%). Negative predictive value remained excellent with the addition of outpatient data. In other words, the percentage of false-negatives in Medicare comprehensive claims remained low (<1%–4%), using ACS-NSQIP as the standard. Overall, agreement beyond chance on the patient-level recording of postoperative complications between ACS-NSQIP and Medicare comprehensive claims was poor to moderate (κ: 0.21–0.47).

There was excellent agreement beyond chance between ACS-NSQIP and Medicare for 30-day postoperative mortality (κ: 0.95). The sensitivity of Medicare data for detecting mortality in ACS-NSQIP was 0.97, specificity more than 0.99, positive predictive value 0.93, and negative predictive value more than 0.99.

DISCUSSION

In this study, we directly compared the recording of postoperative complications for a cohort of surgical patients using a database containing clinical surgical registry data from ACS-NSQIP and an administrative claims data set. The unique aspect of this study is that the claims data set included both inpatient and outpatient data from Medicare linked to the clinical registry data at the individual patient level. The inclusion of outpatient claims data is important because

TABLE 5. Comparison of Patient-level 30-Day Postoperative Complications Recorded in a Clinical Surgical Registry (ACS-NSQIP) Versus an Inpatient/Outpatient Administrative Data Set (Medicare Comprehensive Claims) for 117,752 Surgical Patients

	Percentage of Patients With Specified Complication Recorded in ACS-NSQIP Who Also Had the Complication Recorded in Medicare Comprehensive Claims (Sensitivity)	Percentage of Patients *Without* the Specified Complication Recorded in ACS-NSQIP Who Also *Did Not* Have the Complication Recorded in Medicare Comprehensive Claims (Specificity)	Percentage of Complications Recorded in Medicare Comprehensive Claims That Are False-Positives*	Percentage of Patients *Without* a Complication Recorded in Medicare Comprehensive Claims That Are False-Negatives*	Agreement Beyond Chance on the Recording of Complications Between ACS-NSQIP and Medicare Comprehensive Claims (κ)
Superficial SSI	39.0%	95.3%	73%	3%	Poor (0.28)
Deep/organ-space SSI	48.0%	95.4%	76%	2%	Poor (0.29)
Any SSI	42.7%	96.6%	52%	4%	Moderate (0.42)
UTI	59.7%	90.9%	81%	2%	Poor (0.25)
Pneumonia	67.9%	93.4%	70%	1%	Poor (0.39)
Sepsis	58.8%	91.2%	67%	3%	Poor (0.37)
DVT	78.6%	96.1%	75%	<1%	Poor (0.36)
Pulmonary embolism	87.1%	98.9%	68%	<1%	Moderate (0.47)
Venous thromboembolism	81.7%	95.5%	72%	<1%	Poor (0.39)
Myocardial infarction	86.6%	96.7%	88%	<1%	Poor (0.21)

*Using ACS-NSQIP as the standard.

All complications include any occurrence within 30 days of surgery. Medicare comprehensive claims include inpatient and outpatient claims data from physicians and institutions. Any SSI includes superficial and/or deep/organ-space SSI. Venous thromboembolism includes DVT and/or pulmonary embolism.

ACS-NSQIP includes all complications occurring within 30 days of surgery, even if the complication occurred after discharge from the primary hospitalization.

The findings of this study demonstrate that although the sensitivity of inpatient Medicare data for detecting complications recorded in ACS-NSQIP was moderate to excellent for most complications (with the exception of the SSI complications) and the specificity was universally excellent, the rate of false-positives in the inpatient Medicare data was at best 48% and at worst 84%. Furthermore, agreement beyond chance on the patient-level coding of the 10 postoperative complications studied was poor to moderate. The addition of outpatient data improved the sensitivity of Medicare data for detecting complications in ACS-NSQIP, but at the cost of increased rates of false-positives as well.

Our findings are consistent with prior reports of single-institution studies comparing clinical and inpatient-only administrative data sets. One such study using data from a single institution found that the sensitivity of inpatient administrative data for detecting venous thromboembolism events recorded in ACS-NSQIP was 0.58 and the positive predictive value was 0.42—meaning 42% of venous thromboembolism events recorded in ACS-NSQIP for this single-institution were missed by the administrative claims data and 58% of the events recorded in the administrative claims database were false positives.[4] ACS-NSQIP has also been compared with claims data from the University HealthSystem Consortium (UHC) for a single institution, with the authors reporting that there were 2.5 times more complications recorded in ACS-NSQIP (28% vs 11%) than in UHC, and there was a 26% rate of discordance between the 2 sources.[5] A multi-institution study comparing postoperative complications reported in the Department of Veterans Affairs NSQIP (VA-NSQIP) versus the VA administrative claims data in the Patient Treatment File was performed and reported that sensitivity and positive predictive value of administrative claims data for detecting VA-NSQIP complications were poor or moderate at best.[6,7]

An important shortcoming to most studies that compare clinical and administrative data sets is that the administrative data source only reflects inpatient postoperative complications. Administrative data are derived from claims submitted for payment. Although claims are likely to be submitted for complications that occur after discharge in the outpatient setting, these claims are handled independent of the initial inpatient claims. This has implications for quality assessment, as a study using ACS-NSQIP data found that a considerable proportion of postoperative complications and deaths occurred after discharge from the index hospitalization. The authors further reported that hospital quality rankings and outlier status designations were substantially different between assessments including postdischarge events versus those using inpatient events alone.[8] Our study addresses this shortcoming in the current literature by providing an "apples-to-apples" comparison of ACS-NSQIP data with inpatient and outpatient administrative claims data from Medicare.

There are potential explanations for the observed differences in postoperative complication rates between ACS-NSQIP and Medicare claims data. An important issue encountered in this study was the difference in variable definitions between these 2 sources. An exact match in definitions was noted to be difficult because ACS-NSQIP utilizes detailed clinical definitions that are routinely more strict and specific than the *ICD-9* codes used for administrative claims. For example, it is often difficult to distinguish whether an *ICD-9* code represents a preoperative comorbidity or a postoperative complication, as a "present-on-admission" qualifier is not currently available in Medicare data. In addition, *ICD-9* codes include diagnoses labeled as "rule out" or "suspected" in the medical record for inpatients.[19] Finally, because claims data have a limited number of fields available for coding, the diagnosis codes that are recorded for the purpose of reimbursement may be different from those that would be recorded if the primary purpose was quality measurement.

Although studies demonstrate superior validity and reliability of clinical data, the burden associated with clinical data abstraction has remained as a barrier. Abstraction of clinical data requires substantial resources and is expensive and time consuming. Electronic health records (EHRs) are a potential solution to this burden, especially with the current national initiative encouraging implementation and meaningful use of such systems. It is paramount, however, that caution be exercised in the adoption of EHRs for quality measurement to avoid creating a new data source that simply recreates the deficiencies of administrative claims data. The chief strengths

of clinical data registries are the use of strict definitions with inclusion and exclusion criteria and tight control of data collection quality through continued training and testing of abstractors and auditing. These characteristic features of clinical data abstraction will need to be implemented within EHR data systems to ensure the production of robust and valid data. The use of provider templates have been touted as a potential solution; however, some have reported that the required use of electronic templates simply shifts the burden of data collection to clinicians while not producing documentation that is clinically useful. As the community advances toward quality measurement through EHRs, care should be taken in the design and implementation of this data collection system.

Our findings should be interpreted in light of several limitations. First, we linked records between these 2 databases using indirect identifiers, and thus we cannot know for certain that each record was matched correctly. However, we found excellent agreement between the 2 sources for both inpatient mortality and 30-day mortality, which supports the validity of the linkage procedure. Second, our crosswalk for postoperative complications may not be optimized despite the use of published literature and an extensive review of codes. Third, coding practice may 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. Finally, ACS-NSQIP hospitals in this data set are predominantly larger medical centers, which may limit the generalizability of our findings.

In conclusion, this analysis of more than 200 hospitals demonstrates important differences in the recording of 10 postoperative complications between a clinical surgical registry, ACS-NSQIP, and Medicare comprehensive inpatient and outpatient claims. These findings have meaningful implications for performance measures currently under consideration.

ACKNOWLEDGMENTS

This work was supported by a contract from the CMS. Dr Lawson'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.

REFERENCES

1. Mallon B. *Ernest Amory Codman: The End Result of a Life in Medicine*. Philadelphia, PA: WB Saunders; 2000.
2. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127:666–674.
3. 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.
4. 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.
5. Steinberg SM, Popa MR, Michalek JA, et al. Comparison of risk adjustment methodologies in surgical quality improvement. *Surgery*. 2008;144:662–667.
6. 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.
7. 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.
8. Bilimoria KY, Cohen ME, Ingraham AM, et al. Effect of postdischarge morbidity and mortality on comparisons of hospital surgical quality. *Ann Surg*. 2010;252:183–190.
9. 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.
10. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005;138:837–843.
11. 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.
12. Research Data Assistance Center (ResDAC). Available at: http://www.resdac.org/Medicare/Data_File_Descriptions_RIF.asp. Accessed November 18, 2011.
13. Lawson EH, Ko CY, Louie R, et al. Linkage of a clinical surgical registry with Medicare inpatient claims data using indirect identifiers. *Surgery*. In Press.
14. Elixhauser Comorbidity Software. Available at: http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp. Accessed November 18, 2011.
15. Agency for Healthcare Research and Quality (AHRQ) Quality Indicators. Available at: http://www.qualityindicators.ahrq.gov/Modules/iqi_overview.aspx. Accessed November 18, 2011.
16. Hospital-Acquired Conditions (HAC) in Acute Inpatient Prospective Payment System (IPPS) Hospitals. 2010. Available at: https://www.cms.gov/hospitalacqcond/downloads/hacfactsheet.pdf. Accessed November 18, 2011.
17. Surgical Care Improvement Project Core Measure Set. Available at: http://www.jointcommission.org/assets/1/6/Surgical%20Care%20Improvement%20Project.pdf. Accessed November 18, 2011.
18. Fleiss J, Levin B, Paik M. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ: Wiley; 2003.
19. ICD-9-CM Official Guidelines for Coding and Reporting. Available at: http://www.cdc.gov/nchs/data/icd9/icd9cm_guidelines_2011.pdf. Accessed November 18, 2011.

DISCUSSANT

DR. CRAIG KENT (Madison, WI): The emphasis on quality in today's healthcare environment is pervasive. Although few of us would argue against improving quality, this is akin to arguing against motherhood or apple pie, the Achilles heal in this effort is to sort out how to accurately and reproducibly measure outcomes.

For the time being, CMS has chosen its tool. They look to hospital administrative data. Unfortunately, administrative data consists of a set of codes that were not designed to measure quality, but rather designed for billing. Data input, for the most part, is not made by medical professionals but by individuals with only a computer coding background.

Moreover, we have endorsed this methodology, as clinical researchers, by using these same large administrative data sets for hundreds of studies on clinical outcomes. I am as guilty of this as anyone.

The data that Dr. Ko and colleagues have presented today is revealing and exposes the dangers of using administrative data to assess quality. At stake are literally millions if not billions of dollars in hospital reimbursements. These findings strongly suggest the need for CMS, and for us as clinicians, to identify a plan B for measuring quality.

First, the issue of crosswalking the complications from NSQIP to administrative data is quite complex, as I am sure you know. NSQIP was well designed, it is relatively straightforward, and there are binary fields for each of the complications that were measured. However, for administrative data, the investigators had to ferret through numerous and often diverse sets of codes that represent one disease. Was this a difficult task?

Is there ambiguity in the comparison as a consequence of this?

Troubling is the fact that in MedPAR there is room for ten diagnostic codes, but it is not clear which of these codes represents pre, versus postoperative events. For example, a patient might have experienced a remote DVT, but the coder, trying to be complete, places this remote event into one of the ten coding positions. The investigators may then mistakenly interpret this code to represent a DVT that occurred during the index hospitalization. There are many ways to address this issue. One is to review codes from previous hospitalizations and eliminating those that represent chronic diagnoses. How have the investigators dealt with this issue?

The most important question is what is the solution? Coders extract data from medical records for administrative data. A nurse

extracts data from the same medical record for NSQIP. Presumably, a nurse does a better job of this than does a coder. Should all coders be nurses?

The nurse for NSQIP uses a well thought out system of coding that is specific for the diseases that he or she is designated to evaluate. Hospital coding, alternatively, is less precise, with too many and less specific codes. Is the solution for Medicare to change its method of coding? Does ICD 10 accomplish this goal?

Lastly, have you taken this data to CMS? What has been their response? If you have accomplished this, will you be successful in encouraging CMS to admit fallibility and acquiesce to a change?

DR. CLIFFORD KO:

First of all is the issue of ambiguity in the codes. Why do we have these differences in what we find if we look at a record clinically, versus what a coder would put down? There are a couple of reasons.

One reason is the definitions, which are very different form each other. What we think clinically, as providers, of an infection, or VTE, or whatnot, is very different than what the definition is in the codes. As we looked and worked with our coders and the coders across lots of the hospitals, the coders do a great job following their definitions. It is just that they are indeed different.

Another example that we delved into, is why do we have such a high false positive rate? Why would they find an infection when we did not find an infection at all, to the tune of almost 60% of the time that happening? The coders, for whatever reason, have these soft words, like "consider" or "rule out." Thus, if somebody writes in the chart, "rule out infection," the coder is trained to say that that is an infection, unless somebody, later on in the chart writes, "no infection," which, I guess, does not happen very much.

That is why these things, these words like "consider" and "rule out" or whatnot lead to things like the false positive.

It is absolutely the definition that is different and the application of the definitions that lead to the differences. It is kind of an ambiguous and inconsistent way of defining these terms.

You mentioned the issue of presence on admission. If somebody has an MI and they come into the hospital, and it is coded in the discharge list of diagnosis, MI, it is easy to say that an MI happened at discharge and that we account the MI to something happening postoperatively.

This has been addressed in a lot of the administrative codes by something called POA, present on admission. Some data sets have this, where if an MI is listed there is an asterisk or some kind of modifier that says it was present on admission. We would not get penalized for having an MI afterwards.

CMS data does not contain that, but they do have modifiers that say if it was there already or if it is a chronic condition versus an acute condition. A lot of these complications also are specific to the postoperative way of doing things.

I can tell you, in NSQIP, that we have come across this often. When we used to have present on admission, it still was not sufficient. We have developed, in the last couple of years, a modifier that is "present at the time of surgery." If it was at the time of surgery, not even at admission, but maybe between admission and the time of surgery, that has been a very helpful variable in distinguishing what is really happening with the patient.

As far as a solution, claims are probably good for certain things, such as mortality. When the patient dies, CMS knows it. We should probably use claims for those types of things.

Claims are good for readmission. There are a couple of papers in this meeting that looked at readmissions. Claims are good to signify those claims are good for length of stay.

If we are going to look at complications, like the things that Elise looked at, or like a lot of the things that we think are important, that we discuss at our M and Ms every week, claims are probably not the answer, and it is probably clinical data.

Then the question is how we get that. Do we hire an abstracter to collect all of this? Can we get it from the EHRs?

That kind leads to the fourth point of CMS. This actually was contracted work by CMS. That is how we got the 100% sample. The group that we work with at CMS is OCSQ, the office of clinical standards and quality. There are clearly a lot of offices in CMS, but this is the office that really tries, I believe, to do the right thing. They are trying to figure out how to improve the quality.

They come to the College of Surgeons, maybe you guys in the audience, to say, "If we are going to improve surgery, tell us how we should do it?"

This work came about because we went to them about five years ago, saying that SCIP is not the right thing; it is not correlating with outcomes. We did not publish at the time. We went to CMS because they were the ones who were putting this into play. They said, "All right, if process measures do not work, and structural measures like volume do not work, what would you suggest?"

At that time, we had NSQIP, and we knew that it improved care. We said, "How about outcomes?" Then, the big problem with outcomes is the burden. They said, "All right. Well, can we use administrative data?" So that largely came to this.

This work was originally supposed to and we are still doing it to try to see if we can merge clinical and administrative data so we could get the best of both worlds. If there are some things we can get administratively, and we do not need to have a data collector gather it, that would be great. If there are things that we have to do clinically, then we should do that. And maybe next year or the year after, we will have data to show that if we can merge it and then get the efficiencies of both.

CMS knows it. They are trying to work on it. They said that maybe the answer is the EHR. I said, "Well, if CMS says instead of the meaningful use that is going on currently, if they can really direct things more operationally for things like this that we are looking at, complications as an outcome, risk adjustments, etc., they said they would be very forward with that and would absolutely support that. I believe they are on board.

DISCUSSANT

DR. R. SCOTT JONES (Charlottesville, VA): This work substantiates and confirms the value of the NSQIP methodology for quality improvement. The risk adjusted outcome measures comprise the gold standard for our quality improvement efforts.

On a historical note, I would point out in the early days of NSQIP the VA did a study comparing VA administrative data and NSQIP data and reached exactly the same conclusions that Dr. Ko and his colleagues have reported here with the American College of Surgeons program and the Medicare database. This work provides another validation of the NSQIP methodology in the private sector.

We observed in the early days as the NSQIP program evolved in the private sector that the electronic health records were not quite as good as they are now. The manual data collection was often difficult. With continuing improvements in the electronic health records there may be opportunities to collect NSQIP data as we collect clinical care data. Perhaps that could improve the efficiency and decrease the cost of the NSQIP.

Do you have any interest in comparing the NSQIP database with other databases such as the National Inpatient Database or the University Hospital Consortium database? Would that be worthwhile in the future?

I think, from what we have seen in this year's program at the American Surgical and programs for the last several years, this NSQIP

program is certainly here to stay as a quality improvement tool as well as an important research tool improving health services.

DR. CLIFFORD KO: Number one, with the EHR, I think everyone in the room agrees that is probably the answer. As we have delved into it, there are many complexities of the EHR. Talking to many vendors, we can very quickly, in the EHR, get the codes. We can get the administrative codes. We can get just as bad data, but really quickly. That is the first step.

The second step is, a lot of EHRs have made templates. Thus you have a template for everything, and you just enter it in, you check off a box or whatnot. What a lot of folks have found is that increases the work. It does not help clinically. People are bringing charts home, after their day of work, to fill out these templates. That is perhaps not the answer.

Perhaps a third way of doing it is how an abstracter looks at a data set or a list of variables, and then how they go through the chart to do that. That is much more complex. It is reading text. Part of it is the ASA funded project, of doing natural language processing or whatnot. It is much more difficult, but that is probably the correct answer.

Take for example, pneumonia. You are going look at something on the CT, the chest x-ray, the auscultation, the labs, the vitals or whatnot. How do you put all of that together? That is what a data abstracter does. That is probably the right way to do it. It is not easy, but that is probably the answer.

The second question was, would we look at other claims or administrative data sets? Those other data sets use discharge codes, the same ICD-9s that CMS used. I suspect that we would find the same thing if we did that.

DISCUSSANT

DR. STEVEN STEINBERG (Columbus, OH): We did a similar study, just at our own medical center a few years ago, comparing outcomes from NSQIP versus administrator data that was submitted to UHC, very similar to the data that is submitted to CMS. We found exactly the same outcomes at, obviously, a smaller level.

It is very encouraging to hear that CMS is working with you to improve the data that they are putting out, that we are being judged by. Hopefully, that will come to fruition in the next few years.

What would you suggest for us here and now? Right now, all of us in the room are being judged in our local medical centers, by our boards, by our administrators, on our outcomes, much less considering how we are being judged by payers and CMS. Most of what we are being judged on are outcomes based on administrative data, whether it is from UHC or CMS or wherever.

How do we deal with that situation now? I think it would not take much convincing for everybody in the room to buy into the idea that clinically derived outcome data from NSQIP is much more accurate than any of those other programs because they are all using the same administrative data. How do we deal with that now?

DR. CLIFFORD KO: How would we talk to our C suite in doing this? I can tell you from our experience at the college, and talking to a lot of hospitals and a lot of C suite people, there is variability in terms of enlightenment of the C suite of the hospitals. Some C suite folks get it. They say, "All right. We know that this data is not perfect. If you have something better, let us use it."

Others just say, "We have what we have, and we are using it and it is good enough, and we are going to do it." I think that we just have to constantly persevere and just show them over and over again.

There are many examples where surgeons have said, "Look, this is wrong. We have gone through the chart. There is not a DVT there when they said there was." Which slowly changed their minds. Although for some folks, it is not going to be so easy.

DISCUSSANT

DR. ED LIVINGSTON (Dallas, TX): You do not have to convince anybody in this room of the value of clinically derived data. You really must present this at an equivalent meeting of hospital CEOs. That is where we face the limitation, because NSQIP is expensive. There is resistance to add expenses to a hospital's bottom line in the modern era.

How can this be done in a more cost effective way, so that there is more penetration of NSQIP into the hospital environment?

Within your data, was there enough concordance on any particular complication, where you could use administratively derived data instead of clinically derived data to enter that information into your database, and then free up the nurses to collect other kinds of information?

Secondly, the coders are certified, and they are tested, and they are highly trained. As you pointed out, Dr. Ko, they do exactly what we tell them to do in terms of the definitions of the codes.

Is there room for us to refine the definitions within ICD-9 or ICD-10 so that they can do a better job of capturing this information and approximate what the nurses do?

DR. CLIFFORD KO: Clinical data are expensive to obtain. Whatever the registry, NSQIP, STS, the vascular registry, any time you are going to get clinical data, it is expensive. Again, EHR looks like it is the best answer, at least right now.

When you ask if there is something that we could use with claims instead of clinical, the things I listed, mortality, the crude readmission rate, and maybe things like length of stay. If you get to these other things, then the question is going to become, how willing are we to accept something that we know is not good?

If we are 90% there, perhaps, if we are 60% there, would we do that?

There is also a difference in terms of quality improvement and in payment. If it is quality improvement and it is kind of blurry and it is kind of in that direction, you kind of know you have the right direction but you do not have the right quantity, maybe which is okay. You have to get better and you are not going to get penalized, but it is really trying to accomplish quality improvement.

If it is for payment, and if it is 5% or 3% or even 1% of our income, then, at least the feedback that we get, is that it better be as good as it can get. If you know that it is not as good as it can get, and then we will all hear it. I think it also depends on the use for the accuracy.

DISCUSSANT

DR. JOHN ROBERTS (San Francisco, CA): There are different incentives to entering data in NSQIP and the CMS claims data based on the hospital financial performance and those kinds of things.

I was wondering, if you assumed that NSQIP data was the gold standard, did you see differences in the accuracy of the CMS claims data by hospital?

DR. CLIFFORD KO: We have not looked at that, but I know that in the literature and we have looked at it with other data sets that there is variability in coding by hospitals. There are hospitals that code a lot. As Dr. Kent was saying, there are ten slots for codes for an ICD-9 in some states. Like California, there are 25 slots. There are huge differences in how many of those slots are filled up routinely by hospitals. Some put just two. Some routinely use all 10. There are differences in terms of coding across hospitals.