How Do Administrative Data Compare with a Clinical Registry for Identifying 30-Day Postoperative Complications?

Lillian S Kao, MD, FACS, Justin B Dimick, MD, FACS, Geoff A Porter, MD, FACS, for Members of the Evidence-Based Reviews in Surgery Group

The term *evidence-based medicine* was first coined by Sackett and colleagues[1] as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." The key to practicing evidence-based medicine is applying the best current knowledge to decisions in individual patients. Medical knowledge is continually and rapidly expanding, and it is impossible for an individual clinician to read all the medical literature. For clinicians to practice evidence-based medicine, they must have the skills to read and interpret the medical literature so that they can determine the validity, reliability, credibility, and utility of individual articles. These skills are known as critical appraisal skills. Generally, critical appraisal requires that the clinician have some knowledge of biostatistics, clinical epidemiology, decision analysis, and economics as well as clinical knowledge.

The Canadian Association of General Surgeons (CAGS) and the American College of Surgeons (ACS) jointly sponsor a program titled, "Evidence-Based Reviews in Surgery" (EBRS). The primary objective of this initiative is to help practicing surgeons improve their critical appraisal skills. During the academic year, 8 clinical articles are chosen for review and discussion. They are selected not only for their clinical relevance to general surgeons, but also because they cover a spectrum of issues important to surgeons; for example, causation or risk factors for disease, natural history or prognosis of disease, how to quantify disease (measurement issues), diagnostic tests and the diagnosis of disease, and the effectiveness of treatment. Both methodologic and clinical reviews of the article are performed by experts in the relevant areas and posted on the EBRS website. As well, a listserv discussion is held where participants can discuss the monthly article. Fellows and candidates of the College can access Evidence-Based Reviews in Surgery through the American College of Surgeons website (www.facs.org/education/ebrs.html). All journal articles and reviews are available electronically through the website. Currently we have a library of more than 100 articles and reviews, which can be accessed at any time.

Beginning in October, a new set of articles will be available each month until May. Surgeons who participate in the current (modules) packages can receive CME credits by completing a series of multiple choice questions. For further information about EBRS the reader is directed to the ACS website or should email the administrator, Marg McKenzie at mmckenzie@mtsinai.on.ca.

In addition to making the reviews available through the ACS and CAGS websites, 4 of the reviews are published in condensed versions in the *Canadian Journal of Surgery*, 4 in the *Journal of the American College of Surgeons*, and 4 in *Diseases of Colon and Rectum* each year.

REFERENCE

1. Evidence-Based Medicine Working Group. Evidence-based medicine. JAMA 1992;268:2420–2425.

SELECTED ARTICLE

A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications

Lawson EH, Louie R, Zingmond DS, et al. Ann Surg 2012;256:973–981

Objective: 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.

Design: Retrospective analysis of prospectively collected data.

Data Source: Medicare (administrative claims) and American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP: clinical registry).

http://dx.doi.org/10.1016/j.jamcollsurg.2014.09.002
ISSN 1072-7515/14

Methods: Patient records (2005 to 2008) from ACS-NSQIP were linked to Medicare inpatient and outpatient claims datasets. The authors assessed the ability of Medicare inpatient claims and 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 (UTI), pneumonia, sepsis, deep venous thrombosis (DVT), pulmonary embolism, venous thromboembolism (DVT and/or pulmonary embolism), and myocardial infarction.

Results: 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: There are significant differences in the recording of 10 postoperative complications between administrative claims data (Medicare) and clinical registry (ACS NSQIP), even when outpatient claims data are included.

Commentary: The demand for reliable metrics of surgical quality is at an all-time high. Although morbidity and mortality conferences have traditionally been the primary venue for reviewing surgical complications, many departments are moving to a more systematic and rigorous approach to quality improvement.[1] Two widely used sources for outcomes data for measuring surgical quality are administrative databases and clinical registries. Administrative data consist of claims provided to health insurers; examples of administrative data sources include Medicare and Medicaid, state hospital discharge datasets, and Healthcare Cost and Utilization Project hospital databases.[2] Clinical registries consist of data abstracted from patient charts for quality improvement or research. The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) is a clinical registry that uses trained personnel to collect data, including 30-day surgical outcomes, which provides risk-adjusted outcomes.[3,4] Several single-center studies have compared the concordance between administrative data and clinical registries in identifying and predicting surgical complications; these have identified significant differences in both recording of risk factors and outcomes between the 2 data sources.[5,6] Using a larger patient population, Lawson and colleagues[7] addressed a similar question of how administrative data (Medicare inpatient and outpatient claims data) compared with a clinical registry (ACS NSQIP) for identifying 30-day surgical complications.[7] This study included more than 100,000 patients cared for at more than 200 hospitals between 2005 and 2008.

In addition to the sources from which the data are derived, Medicare and ACS NSQIP differ in inclusion and exclusion criteria, sampling strategy, data collected, data abstraction, follow-up, and resource requirements. Medicare data include only beneficiaries, defined as patients at least 65 years old, some disabled patients less than 65 years old, and all patients with end-stage renal disease receiving dialysis. On the other hand, ACS NSQIP includes all sampled patients who have undergone surgical procedures except for trauma patients, transplantation patients, and brain-dead organ donors. All eligible patients are included in the Medicare datasets, while ACS NSQIP uses a random sampling strategy based on consecutive 8-day cycles. Both Medicare and ACS NSQIP collect demographic data; dates of admission, procedures, and discharges; and diagnoses by *International Classification of Diseases, 9th edition (ICD-9)* code. However, Medicare identifies risk factors and complications using ICD-9 codes submitted with claims, while ACS NSQIP uses a full-time trained abstractor and strict definitions for risk factors and 30-day complications.

Medicare vs ACS NSQIP for reporting 30-day complications

Lawson and associates[8] used a linked dataset to perform this study. An algorithm using data common to both sources (ie, indirect patient identifiers including hospital, age, sex, diagnosis, procedure category, and dates of admission, procedure, and discharge) was used to link the Medicare and ACS NSQIP datasets.[8] The authors used multiple methods to ensure accurate linkage of the data, including an explicit and transparent algorithm, testing of alternative rules for combining data, manual inspection of matched records, and evaluation of agreement on unmatched variables such as mortality.[7,8]

Ideally, to determine which database is superior, both Medicare and ACS NSQIP data would have been compared with a reference gold standard or the "truth." In the case of surgical complications, the "truth" would be whether the patient suffered harm from a clinical syndrome consistent with the best definition of that complication. However, there is no current method for accurately measuring the "truth." Therefore, the authors calculated the diagnostic test properties (including sensitivity, specificity, false positives and false negatives, positive predictive value [PPV], and negative predictive value [NPV]) of the Medicare dataset for identifying 30-day surgical complications using the ACS NSQIP dataset as the gold standard.

Briefly, the frequency (prevalence) of identified complications was lowest for ACS NSQIP and highest for Medicare Comprehensive Claims (inpatient and outpatient administrative data), except for surgical site infections. There was as high as a 7-fold difference in specific complication rates (ie, for myocardial infarction) between Medicare Comprehensive Claims and ACS NSQIP. Sensitivity (Table 1), or the probability of having a complication reported by Medicare when one was reported by ACS NSQIP, ranged from poor to excellent for MedPAR (inpatient claims). Sensitivity was slightly improved when outpatient claims data were included; however, the percentage of false positives was greatly increased. Specificity, or the probability of not having a complication reported by Medicare when none was reported by ACS NSQIP, was high overall for MedPAR; specificity was lower but still high when outpatient claims data were included. The percentage of false negatives was low. The authors also described the positive and negative predictive values; positive predictive values were low to moderate, and negative predictive values were high. A detailed discussion of how these diagnostic test properties were calculated can be found on the EBRS website (http://www.facs.org/education/ebrs.html).

The authors did not report accuracy, or the probability of a true result, which could have been misleading. For example, the accuracy of MedPAR for superficial surgical site infections was 94.1%; this resulted in 3,318 false positives and 3,636 false negatives (calculations not shown). So, the accuracy does not reflect the degree of discrepancy between datasets. In fact, the authors recorded the agreement beyond chance between the Medicare datasets and ACS NSQIP using κ. Most of the κ values indicated poor (<0.40) to moderate (0.40 to 0.75) agreement.

Based on these analyses, the authors concluded that there are significant differences in the recording of 10 postoperative complications between administrative claims data (Medicare) and clinical registry data (ACS NSQIP), even when outpatient claims data are included.

Sources of the differences in results

There are several potential reasons for the differences between the Medicare and ACS NSQIP datasets, which were also acknowledged and described by the authors in the discussion and the question-and-answer section. In the above example of superficial surgical site infections, there were equivalent numbers of both false negatives and false positives; there may have been multiple reasons for why the discrepancies between the datasets were present.

False negatives, or complications not recorded by Medicare, but recorded by ACS NSQIP, may be explained in part by the lack of postdischarge tracking of complications. A study of ACS NSQIP by Kazaure and coworkers[9] reported that 42% of postsurgical complications occurred after discharge.[9] MedPAR includes only data through discharge. Although Medicare claims that occurred within 30 days of a surgical procedure were searched for codes associated with the complications of interest, complications may have still been missed. Another reason for false negatives may have been coding limitations. Depending on the state, only a limited number of diagnoses (10) can be recorded by Medicare. So, if a patient has multiple comorbidities or has had multiple complications, other complications may have been intentionally or unintentionally left out.

False positives, or complications recorded by Medicare but not recorded by ACS NSQIP, may be explained by differences in definitions, coding, and present-on-admission diagnoses. First, ACS NSQIP has strict definitions for all variables, which may have resulted in fewer complications being recorded by ACS NSQIP than by Medicare. Second, coders for billing claims count "rule out" or "consider" as confirmatory for the complication unless otherwise clarified in the chart. Therefore, diagnoses that were considered but not borne out may still be counted as complications if documentation is unclear. Third, ACS NSQIP distinguishes between risk factors and postoperative complications, but Medicare does not record when a diagnosis was present on admission. Therefore, if a urinary tract infection was present on admission, it may have been counted as a complication. Nonetheless, Lawson and colleagues[7] attempted to exclude diagnoses present on admission when modifiers suggested a chronic condition (ie, old myocardial infarction).

False positives or negatives may be explained by other methodologic considerations. Although the authors exerted considerable effort to ensure that the data linkage algorithm performed accurately, mismatches may still have occurred. This could have resulted in either false positives or false negatives. However, given the multiple methods used for ensuring accurate linkage, the likelihood that this significantly affected the results of the study is low. Additionally, because no reference gold standard for postoperative complications exists, ACS NSQIP classification was used as the gold standard. Errors in classification of complications could have occurred in either direction; however, given the level of auditing performed, the percentage of errors in ACS NSQIP has been reported to be very low.[3]

Implications for quality improvement

Given the concerns about the validity and reliability of administrative data, why is there a remaining debate about data sources for surgical quality improvement?

Table 1. Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value

	Disease (30-day complication)		
	ACS NSQIP +	ACS NSQIP −	
Test (Medicare data)			
Medicare +	TP	FP	TP + FP
Medicare −	FN	TN	FN + TN
	TP + FN	FP + TN	

Sensitivity is the probability of having a complication reported by Medicare when one is reported by ACS NSQIP. **Specificity** is the probability of not having a complication reported by Medicare when none is reported by ACS NSQIP. The **positive predictive value (PPV)** is the probability of having a true complication (ie, in ACS NSQIP) if reported by Medicare. The **negative predictive value (NPV)** is the probability of truly not having a complication (ie, not in ACS NSQIP) if not reported by Medicare.

Sensitivity = TP/(TP + FN)

Specificity = TN/(FP + TN)

PPV = TP/(TP + FP)

NPV = TN/(FN + TN)

Accuracy = (TP + TN)/(TP + FP + FN + TN)

FN, false negative; TP, true positive.

The primary reason is that the expense, resources, and time necessary to maintain clinical registries for all diagnoses and procedures are prohibitive. On the other hand, administrative databases are readily available, inexpensive, and increasingly linkable.

Rather than dismissing administrative data altogether, strategies for optimizing valid and reliable data collection are necessary. First, surgeons should recognize that there is a hierarchy of accuracy of outcomes in administrative data. Mortality and length of stay are accurate. Complications that lead to billable procedures and have a related billing code (ie, tracheostomy, feeding tubes, and reoperation) are accurate. Other valid and reliable outcomes that can be derived from administrative data have also been identified.[10,11] Second, surgeons should determine the outcomes of interest and use methods that capitalize on the most reliable elements of administrative data (ie, length of stay). For example, an extended length of stay might be used as a reliable proxy for nonfatal complications in administrative data. By linking extended length of stay with complication codes, hospitals could identify patients with serious complications. Although this would reduce sensitivity (by excluding patients with less severe complications), such data would allow hospitals to identify a patient population on which they should focus their quality improvement efforts. Third, a hybrid approach that incorporates the advantages of both data sources should be developed. Demographics and basic outcomes measures (ie, mortality, length of stay, reoperation) can be garnered from administrative data. These data can be supplemented by a small number of clinical data elements (ie, comorbid diseases, processes of care, and complications) that are collected as part of clinical work flow in an ideal world. Electronic health records may facilitate such data collection, but caution must be used in order to ensure the validity and reliability of such extracted data. Fourth, data from both sources may be used to complement each other. Although exact rates of complications may be inaccurate, trends over time identified from administrative data may instigate collection of more granular data in a clinical registry to guide quality improvement. Administrative data may then be used to evaluate sustainability of outcomes after improvement efforts. Lastly, the contribution of administrative data analyses to the understanding of population-level variations in care and outcomes should not be diminished.[12-14] When the limitations are carefully addressed and the data are properly analyzed,[15] administrative datasets can provide insight into quality of care.

REFERENCES

1. Hutter MM, Rowell KS, Devaney LA, et al. Identification of surgical complications and deaths: an assessment of the traditional surgical morbidity and mortality conference compared with the American College of Surgeons-National Surgical Quality Improvement Program. J Am Coll Surg 2006;203:618–624.
2. Riley GF. Administrative and claims records as sources of health care cost data. Med Care 2009;47:S51–55.
3. 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.
4. Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? Am J Surg 2009;198:S19–27.
5. Davenport DL, Holsapple CW, Conigliaro J. Assessing surgical quality using administrative and clinical data sets: a direct comparison of the University HealthSystem Consortium Clinical Database and the National Surgical Quality Improvement Program data set. Am J Med Qual 2009;24:395–402.
6. Steinberg SM, Popa MR, Michalek JA, et al. Comparison of risk adjustment methodologies in surgical quality improvement. Surgery 2008;144:662–667; discussion 662–667.
7. 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.
8. 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.
9. Kazaure HS, Roman SA, Sosa JA. Association of postdischarge complications with reoperation and mortality in general surgery. Arch Surg 2012;147:1000–1007.
10. Hudson M, Avina-Zubieta A, Lacaille D, et al. The validity of administrative data to identify hip fractures is high—a systematic review. J Clin Epidemiol 2013;66:278–285.
11. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. BMJ 2007;334:1044.
12. Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. Ann Int Med 2003;138:288–298.
13. Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. Ann Int Med 2003;138:273–287.
14. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. N Engl J Med 2002;346:1128–1137.
15. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol 2012;65:126–131.