The American Journal of Surgery 217 (2019) 447–451

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Physician derived versus administrative data in identifying surgical complications. Fact versus Fiction

Leonard R. Henry*, Michael J. Minarich, Rhonda Griffin, Urs W. von Holzen, Ashley N. Hardy, Hubert Fornalik, Roderich E. Schwarz

The Goshen Center for Cancer Care, Goshen, IN, 46526, USA

ARTICLE INFO

Article history:
Received 16 July 2018
Received in revised form
18 August 2018
Accepted 21 August 2018

ABSTRACT

Background: Administrative data are widely used as determinants of surgical quality. We compared surgical complications identified in a structured surgical review to coding and billing data of over a 19-month period.

Methods: A retrospective review of monthly morbidity and mortality conference reports was compared to a report over the same time period generated from hospital coding and billing data.

Results: 807 sequential operative procedures were included. Physician derived data compared to administrative data identified a complication of any severity in 205 (25.4%) versus 111 (13.8%) cases ($r = 0.39$), and major complications in 68 (8.4%) versus 46 (5.7%) cases ($r = 0.36$). Review of the administrative data regarding major complications identified 80 false negatives, 52 false positives, and 38 true positive designations. Overall sensitivity, specificity, positive and negative predictive values, and accuracy for administrative data in identifying major complications was 0.32, 0.99, 0.42, 0.99, and 0.99.

Conclusions: The correlation between physician determined and administrative data with regard to identifying surgical complications is poor. Administrative data are insensitive and lack positive predictive value.

Introduction

Recent efforts at healthcare reform have emphasized a transition of reimbursement for volume of healthcare delivery to one based on value of care delivery. Value in business terms is generally considered a derivation of quality of care dependent on the costs of care. Quality of care reporting has become an important component of several federal and private payer programs and is increasingly shared with the public for its consideration in health care choices.[1]

Much of this quality reporting is derived from administrative data, namely coding data submitted to payers for purposes of reimbursement. Coding and billing data are typically recorded by specialized non-physician personnel who translate clinical events into codes to be submitted for payment. The criteria for affixing codes in clinical circumstances is often based on criteria set forth by the payers, but in many instances relies upon the documentation of treating physicians. These data are readily available with little effort and are thus attractive for use in quality reporting. However, what is documented in the health record or interpreted by those conducting the coding and billing may not reflect the true severity of clinical events as they are perceived by treating physicians. This environment poses risks for health systems regarding quality reporting, as efforts to maximize reimbursement or to risk stratify for payment purposes may have negative consequences for value based payment programs.

Similar unintended consequences pose risks to individual surgeons or surgeon groups. Physician specific scorecards utilizing administrative data are already publicly available. Examples include the Propublica report released in 2015, detailing surgeon specific quality scorecards regarding 8 procedures[2]; surgeon ratings are also published by the checkbook organization[3] and also through the Centers for Medicare and Medicaid Services (CMS) physician compare website.[4] Many institutions also utilize administrative data to generate internal physician specific scorecards. It is widely recognized that there are limitations in administrative data and that the collection of reliable quality data would be preferred. However, high fidelity data acquisition can be time consuming and

* Corresponding author. Division of Surgical Oncology, Goshen Center for Cancer Care, Goshen, IN, 46526, USA.
E-mail address: LHenry@goshenhealth.com (L.R. Henry).

https://doi.org/10.1016/j.amjsurg.2018.08.015
0002-9610/

expensive, and there would be little reason for health care organizations to invest in this if coding and billing data could be demonstrated to be a reasonable surrogate.

The purpose of this study was to evaluate the quality of administrative data with regards to quality reporting in a group surgical oncology practice. The specific aim was to assess whether administrative data correlate with real outcomes from the perspectives of the treating surgeons, and if these data would be useful for internal quality monitoring of a surgical program. In addition, we sought to identify the performance characteristics of administrative data regarding major operative complications to identify which conditions were prone to over- or under-reporting in potentially publicly available data.

Methods

This study was approved by the Goshen Hospital Institutional Review Board. A list of all cases conducted by the Surgical Oncology division at the Goshen Center for Cancer Care is produced by the administrative assistant team and reviewed on a monthly basis during which complications related to therapeutic operations are discussed and graded for severity based on the classification of Clavien and Dindo[5]; accordingly, major complications were defined as grade 3 or higher events. The most severe complication is used to assign a (maximal) morbidity grade for each patient. To ensure the most accurate reporting possible, surgeons are encouraged to prospectively log their complications. Additionally, the conference is also attended by the surgical fellow and advanced care practitioners who provide a substantial portion of the inpatient care, and thus they provide an additional review to ensure complete reporting. Between Jan 1, 2016 and July 31, 2017 all operations, complication types and grades from this physician derived morbidity reporting were compared to a generated list of complications for the same time period derived from the hospital's coding and billing data repository. This data repository includes (among others) the submitted coding and billing data to payers, the reimbursement amounts provided back to the institution from the payers, and a report of all major or serious complications (sometimes multiple per patient) in alignment with the Centers for Medicare and Medicaid services (CMS) criteria.

The complications in the patient cohort were compared between the two data sets to determine their correlation statistic for classes of complications including any complication per patient, major complication per patient, or death (i.e. Clavien-Dindo grade 5 events). Given the wide scope of case types, correlations were also determined between the data sets for cases grouped as skin and soft tissue (skin, breast, endocrine), thoracic, abdomen and pelvis resections, and cases classified as other general surgical cases within the practice (for example, operations for wound care, small bowel obstructions, vascular access, empyema decortication, etc …, thus some cases classified as general surgical cases were also classified under thoracic or abdomen and pelvis). We also compared the data sets between those cases classified as inpatient category type cases versus those classified as outpatient (<2 night stay) type cases as a further surrogate measurement for case complexity as well as to test the assumption that complications identified in the outpatient setting may be readily missed because of the use of global post-operative codes.

To determine the performance characteristics of administrative data regarding major complications, the records of all listed major complications identified in the physician derived data set and the administrative data set were reviewed to determine if the complication actually reflected a major event. Major complications within 30 days of the operation in the cohort were identified in the physician derived data or administrative data due to any of 14 diagnoses (Table 1). To standardize the criteria of major complications we adopted the standards used by ACS NSQIP[6] where applicable (respiratory failure, renal failure, shock, sepsis, pneumonia), and the Framingham criteria for congestive heart failure.[7] In the absence of available standard definitions (for example peritonitis or bleeding that requires reoperation) we used medical documentation of the condition alone. Administrative codes for major complications were then classified by the first author as true negative, true positive, false negative and false positive for the entire patient cohort (N = 807) and sensitivity, specificity, positive and negative predictive values and overall accuracy were calculated.

Results

In the study time frame, 807 procedures were performed by 6 surgeons (range 41–269). The median age of the patients was 61 (range 16–97). The operations by organ or system were classified as: gynecologic (212), other general surgical (193), breast (73), endocrine (67), cutaneous (56), colorectal (39), lung (39), cytoreductive procedures with or without heated intraperitoneal chemotherapy (33), hepatobiliary (25), pancreatic (24), head and neck (19), gastroesophageal (14), and sarcoma or multi-visceral resections (13).

In the process of comparing physician derived data to administrative data for the operations (Table 2), a complication of any severity was identified in 25.4% versus 13.8% (r = 0.39) of patients; similarly, a major complication was identified in 8.4% versus 5.7% (r = 0.36), and a lethal event in 0.5% each (r = 0.75). Considering subtypes of operations, correlation for any complication ranged from 0.15 for outpatient type procedures to 0.43 for those of the abdomen and pelvis; those for major complications ranged from 0.00 for outpatient operations to 0.46 for those of the abdomen and pelvis. There were four lethal events attributed to operative complications, one of which occurred after a readmission and was not attributed to the operation by administrative data, and one which occurred in a patient presenting in extremis who underwent diagnostic laparoscopy for cancer diagnosis only (after which care was withdrawn), which was linked to the operation by administrative data. Therefore, the correlation for lethal events is not 1 (Table 2).

All major complications in either the physician derived data or administrative data were linked to one or more of 14 diagnoses. Of these the administrative data only reported major complications for respiratory failure, shock, sepsis, pneumonia, renal failure, myocardial infarction or cardiac arrest, congestive heart failure, deep venous thrombosis or pulmonary embolus (DVT/PE), or death. Major complications derived from physician data included these, but in addition included events diagnosed as post-operative peritonitis or empyema, neurologic complications (cerebrovascular accident, transient ischemic attack, or encephalopathy), hepatic failure, multisystem organ dysfunction, or post-operative events such as wound failure, post-operative bleeding or tissue ischemia. The performance characteristics of administrative data in regards to these operative complications are given in Table 2. Given the definitions used regarding major complications, there were 52 false positive reports, 80 false negative reports, and 38 true positive reports in the administrative data set. With respect to determinants for major postoperative complications, administrative data were particularly insensitive for diagnosis of sepsis, peritonitis or empyema, multisystem organ dysfunction, as well as for wound complications, bleeding, or tissue ischemia type complications (Table 1). In contrast, false positive coding results for administrative data coding major complications were most common for renal failure, respiratory failure, pneumonia and DVT/PE (Table 1).

Table 1
Diagnosis of major complications and performance characteristics of administrative data in identifying them.

	Respiratory failure	Renal failure	Shock	Sepsis	Pneumonia	MI or Cardiac arrest	Congestive Heart Failure	DVT PE	Peritonitis Empyema	Stroke, TIA, or Encephalopathy	Liver failure	MSOD	Wound failure, Bleeding, or Ischemia	Death	All
TN	790	772	787	786	797	802	803	799	767	801	803	801	793	797	11097
FN	5	2	9	13	2	1	1	2	27	0	2	4	11	1	80
TP	4	1	7	4	0	2	1	1	11	4	0	0	0	3	38
FP	6	30	2	3	6	0	0	4	0	0	0	0	0	1	52
Sensitivity	0.44	0.33	0.44	0.24	0.00	0.67	0.50	0.33	0.29	1.00	0.00	0.00	0.00	0.75	0.32
Specificity	0.99	0.96	1.0	1.0	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
PPV	0.40	0.03	0.78	0.57	0.00	1.00	1.00	0.20	1.00	1.00	.	.	.	0.75	0.42
NPV	0.99	1.00	0.99	0.98	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.99	1.00	0.99
Accuracy	0.99	0.96	0.99	0.98	0.99	1.00	1.00	0.99	0.97	1.00	1.00	1.00	0.99	1.00	0.99

MI: myocardial infarction; DVT PE: deep venous thrombosis, pulmonary embolus; TIA: transient ischemic attack; MSOD: multi-system organ dysfunction.

Table 2
a: Correlations of any complication, major complication, or death between physician derived data and administrative data for inpatient and outpatient operations.

	Number	Any Complication, n (% of patients)			Major Complication, n (% of patients)			Death, n (% of patients)			Grade 3–5 complications				
		Physician data	Administrative data	correlation (r)	Physician data	Administrative data	correlation (r)	Physician data	Administrative data	Correlation (r)	FN	TP	FP	Sensitivity	PPV
Inpatient	442	160 (36.3%)	108 (24.5%)	0.38	61 (13.8%)	45 (10.2%)	0.35	4 (0.9%)	3 (0.7%)	0.87	73	38	46	0.34	0.45
Outpatient	365	45 (12%)	3 (1.5%)	0.15	7 (2.0%)	1 (0%)	0.00	0 (0%)	1 (0%)	–	7	0	6	0.00	0.00
All	**807**	**205 (25.4%)**	**111 (13.8%)**	**0.39**	**68 (8.4%)**	**46 (5.7%)**	**0.36**	**4 (0.5%)**	**4 (0.5%)**	**0.75**	**80**	**38**	**52**	**0.32**	**0.42**
b: Representative comparisons between different selected operative types (not an exclusive list of operative types).															
Skin, Breast, Endocrine	215	44 (20.5%)	5 (2.3%)	0.23	10 (4.7%)	0 (0%)	–	0 (0%)	0 (0%)	–	9	1	0	0.10	1.00
Thoracic	65	25 (38.5%)	28 (43.1%)	0.38	12 (18.5%)	10 (15.4%)	0.13	1(1.5%)	1(1.5%)	1.00	9	9	10	0.50	0.47
Abdomen Pelvis	294	104 (35.4%)	60 (20.4%)	0.43	40 (13.6%)	28 (9.5%)	0.46	3(1%)	2(0.7%)	0.82	57	27	26	0.32	0.51
General	193	22 (11.4%)	24(12.4%)	0.35	7 (3.6%)	12 (6.2%)	0.29	1 (0.5%)	2 (1.0%)	0.71	7	4	20	0.36	0.17

FN: false negative; TP: true positive; FP: false positive; PPV: positive predictive value.

Discussion

Although independent physician review of operative cases and outcomes would likely represent the most trustworthy data upon which to judge surgical complications, we view this as impractical in most settings. Physician derived data obtained in a thorough and honest fashion therefore represents in our opinion, the practical gold standard upon which to judge operative complications. In this study we compared physician derived data to administrative data for identifying operative complications in a surgical oncology practice. Correlation between the two data sources for identifying any complication and major complications was poor, and administrative data were found to be generally insensitive and have low predictive value for identifying complications. The high reported specificity, negative predictive value and accuracy in this study are reflective of a very high number of true negative designations (over 11,000) in this dataset of 807 patients and considering 14 major complication diagnosis. Thus, we included the raw numbers of false positive, false negative, true positive and true negative events so the reader will not be misled. Our data corroborate the findings of others, who have demonstrated the limitations of administrative data when compared to more rigorous data capture methods.[8–11] In this study, the insensitivity was most apparent for cases classified as outpatient operations. In addition, we specifically identified certain diagnoses which were frequently over-reported as major complications (renal failure, respiratory failure, pneumonia, DVT/PE) and those which were often under-reported within administrative data (sepsis, peritonitis or empyema, multisystem organ dysfunction, as well as for wound complications, bleeding, or tissue ischemia type complications).

There are many reasons for the discrepancies observed within these results. A particularly important mechanism for differences between administrative and physician derived data in identifying and grading post-operative complications is based on imprecise medical documentation and the impact this may have on coders. This was most pronounced regarding the diagnosis of acute renal failure, for which the coding criteria may rely on guidance from payers but which is also heavily influenced by physician documentation. Renal failure was often documented by physicians for relatively trivial postoperative elevations in creatinine when no intervention was needed and the issue was of limited clinical consequence. When considering a definition of renal failure utilizing NSQIP criteria or translating the severity of the failure in terms of the Clavien Dindo scale, true events of renal failure were present in this cohort only once, yet acute renal failure was found to be represented in the administrative data 31 times, explaining the very low positive predictive value of the administrative data with regards to this particular code. Efforts to adopt standardized criteria between physicians and coders would be particularly useful regarding complications types that are open to some element of interpretation, such as post-operative renal failure, respiratory failure, shock, or sepsis.

In addition to the problems arising from the imprecision of medical documentation, it should be understood that the definitions of major complications between the data sets are different. The Clavien-Dindo scale grades severity of complications based on impact to the patient such that a pneumonia (for instance) that requires only the administration of antimicrobials without risk to life of the patient would not be classified as a major complication. However, administrative data rely only on the presence or the absence of the diagnostic code for the condition of interest (of which pneumonia was one) for determination regardless of the severity of the underlying condition or what interventions for treatment are required. Several of these reported events, however, would not be classified as major complications by clinicians based solely on their presence. Efforts to remedy this incongruence would need to be accomplished at the national level and would require either the payers or physician groups to adopt the criteria of the other or come to a common understanding on a new classification scheme. All of these potential events seem rather unlikely to occur, at least in the near term.

Other causes of incongruity in the data sets are that administrative data are not necessarily temporally linked with the operation. This had effects primarily in two ways. First, preexisting conditions may be inappropriately linked to outcomes in this setting. For example, a case of sepsis from an empyema requiring ventilation that requires a decortication may link sepsis and respiratory failure to the operation, though they were clearly present at the time (or before) of the operation. Secondly, patients discharged from the hospital and subsequently readmitted with major complications were frequently not linked to the index admission and as such the major complication was not attributed to the operation. Further, complications identified in the outpatient setting may be easily missed when physicians are entering global post-operative codes in the clinic. These latter two issues likely explain the very low sensitivity for administrative data in identifying major complications following outpatient procedures.

In our center, the administrative data report came with major complications already assigned (in alignment with CMS criteria presumably) but only pertained to a narrow set of certain conditions. These included post-operative acute respiratory failure, acute renal failure, shock, sepsis, pneumonia MI or cardiac arrest, congestive heart failure, DVT/PE and death. Interestingly, we recognize that these are conditions of importance for federal programs such as the PSI 90 patient safety and adverse events program (post-operative acute renal failure, post-operative respiratory failure, postoperative sepsis, perioperative DVT/PE[12]) or represent targeted conditions (myocardial infarction, congestive heart failure, pneumonia) for the readmissions reduction program.[4] This strongly suggests to us that administrative reports on surgical complications are (at least in part) a non-specific application of data gathered for other purposes. And it also suggests that metric vendors providing summarized administrative data for institutions are to some degree aligned and influenced by the payers. On the contrary, many major surgical complications such as those requiring interventions or reoperations such as wound failure, bleeding, peritonitis, or other major complications with significant organ functional impact and of importance for patients such as neurologic events, or multisystem organ dysfunction were not classified as major complications by administrative data, and in some instances do not appear to have any codes routinely affixed to them at all. All of these help to explain the multidirectional low correlation between administrative data and physician derived data.

The findings of this study have important implications for organizations and surgeons. For organizations these data should make evident that an inherent conflict is present in this mixed payment climate (fee for service and value based payments) in that maximizing coding for purposes of reimbursement or risk stratification poses potential risks to quality measures used for value based payments. This likely will serve as an impediment in the future as any over-reporting of untoward events will negatively impact on hospital rating and consumer choices. In addition, the data suggest that valid and reliable grading of surgeon performance or program performance with regard to operative complications cannot rely on administrative data. We hope these data prompt surgeons toward a more attentive posture to what is documented in the medical record on their post-operative patients by themselves, their assistants, and especially consulting physicians. These data also highlight that the added time and resources required for surgeons and programs to track their own outcomes, and report

them, are wise investments in this era emphasizing transparency of outcomes. To do anything less results in quality reporting that no surgeon can and no patient should believe.

Conflict of interest statement

The authors have no relevant financial conflicts of interest to disclose regarding this manuscript.

References

1. Henry LR, von Holzen UW, Minarich MJ, et al. Quality measurement affecting surgical practice: utility versus utopia. *Am J Surg*. 2017;215(3):357–366.
2. Propublica newsroom. Available at:: https://projects.propublica.org accessed March 14, 2018.
3. The Consumers' Checkbook. Available at:: www.checkbook.org/surgeonratings accessed March 14, 2018.
4. The Centers for Medicare and Medicaid Services. Available at:: www.cms.gov accessed March 14, 2018.
5. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240(2):205–213. https://doi.org/10.1097/01.sla.0000133083.54934.ae.
6. ACS NSQIP Operations manual. Available (for membership fee) at: https://www.facs.org/quality-programs/acs-nsqip accessed March 21, 2018.
7. Ho KKL, Anderson KM, Kannel WB, Grossman W, Levy D. Survival after the onset of congestive heart failure in the Framingham Heart Study subjects. *Circulation*. 1993;88:107–115.
8. 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.
9. 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.
10. Bordelanou L, Cauley CE, Antonelli D, et al. Truth in reporting: how data capture methods obfuscate actual surgical site infection rates within a healthcare network system. *Dis Colon Rectum*. 2017;60:96–106.
11. Awad MI, Shuman AG, Montero PH, Palmer FL, Shah JP, Patel SG. Accuracy of administrative and clinical registry data in reporting postoperative complications after surgery for oral cavity squamous cell carcinoma. *Head Neck*. 2015;37(6):851–861.
12. Agency for healthcare research and quality. Available at: www.AHRQ.gov. Accessed March 14, 2018.