Original Paper

How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

Joan Porter[a]   Luke Mondor[a]   Moira K. Kapral[a–c]   Jiming Fang[a]
Ruth E. Hall[a, d, e]

[a]Institute for Clinical Evaluative Sciences, [b]Department of Medicine, University of Toronto, [c]Division of General Medicine and Toronto General Research Institute, University Health Network, [d]Institute for Health Policy, Management and Evaluation, University of Toronto, and [e]Ontario Stroke Network, Toronto, Ont., Canada

Abstract

Background/Aims: The reliability of diagnostic coding of acute stroke and transient ischemic attack (TIA) in administrative data is uncertain. The purpose of this study is to determine the agreement between administrative data sources and chart audit for the identification of stroke type, stroke risk factors, and the use of hospital-based diagnostic procedures in patients with stroke or TIA. ***Methods:*** Medical charts for a population-based sample of patients (n = 14,508) with ischemic stroke, intracerebral hemorrhage (ICH), or TIA discharged from inpatient and emergency departments (ED) in Ontario, Canada, between April 1, 2012 and March 31, 2013, were audited by trained abstractors. Audited data were linked and compared with hospital administrative data and physician billing data. The positive predictive value (PPV) of hospital administrative data and kappa agreement for the reporting of stroke type were calculated. Kappa agreement was also determined for stroke risk factors and for select stroke-related procedures. ***Results:*** The PPV for stroke type in inpatient administrative data ranged from 89.5% (95% CI 88.0–91.0) for TIA, 91.9% (95% CI 90.2–93.5) for ICH, and 97.3% (95% CI 96.9–97.7) for ischemic stroke. For ED administrative data, PPV varied from 78.8% (95% CI 76.3–81.2) for ischemic, 86.3% (95% CI 76.8–95.7) for ICH, and 95.3% (95% CI 94.6–96.0) for TIA. The chance-corrected agreement between the audited and administrative data was good for atrial fibrillation (k = 0.60) and very good for diabetes (k = 0.86). Hospital administrative data combined with physician billing data more than doubled the observed agreement for carotid imaging (k = 0.65) and echocardiography (k = 0.66) compared to hospital administra-

Ruth E. Hall
Institute for Clinical Evaluative Sciences
G1 06, 2075 Bayview Avenue
Toronto, ON M4N 3M5 (Canada)
E-Mail Ruth.hall @ ices.on.ca

KARGER

97

Cerebrovasc Dis Extra 2016;6:96–106

| DOI: 10.1159/000449288 | © 2016 The Author(s). Published by S. Karger AG, Basel www.karger.com/cee |

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

tive data alone. **Conclusions:** Inpatient and ED administrative data were found to be reliable in the reporting of the International Classification of Diagnosis, 10th revision, Canada (ICD-10-CA)-coded ischemic stroke, ICH and TIA, and for the recording of atrial fibrillation and diabetes. The combination of physician billing data with hospital administrative data greatly improved the capture of some diagnostic services provided to inpatients.

© 2016 The Author(s)
Published by S. Karger AG, Basel

Introduction

Administrative data are an important resource for the measurement of health system performance and have been used in the production of quality indicators, outcomes research, and trend analyses. A persistent concern with administrative data is the degree to which these data reflect reality because while primary data collected for a specific research purpose provide some assurance of the validity of information compiled, the same cannot be assumed with administrative data which are generally collected in pursuit of management rather than research objectives [1]. Validation studies are considered essential for determining the quality of administrative data for secondary use in health services research [2–4]. Studies undertaken to validate stroke and transient ischemic attack (TIA) diagnoses in administrative data have been comprehensively inventoried in systematic reviews [5, 6].

Stroke continues to be a leading cause of death in Canada and the US despite a secular decline in mortality rates in both countries [7, 8]. In 2009, approximately 1.1% of the community-dwelling population in Canada self-reported suffering the effects of a stroke, which is likely an underestimate of the overall population burden since those in institutional care were not included [8]. A more recent development has been an increasing incidence of stroke at younger ages, a trend that is likely to affect demand for health system resources in direct correspondence to the potential burden of disability suffered earlier in life [9, 10]. The epidemiology of stroke and evaluation of health system response to stroke will thus continue to feature prominently in health services research.

Since 2009, a number of initiatives designed to improve stroke/TIA administrative data coding have been implemented across Ontario, Canada, including stricter definition around the coding of ischemic stroke and mandatory collection of data elements specific to stroke management [11, 12]. In addition, a series of audits of the medical records of stroke and TIA patients seen at Ontario hospitals have been conducted, the most recent in 2012. In short, system-level improvements to coding, as well as access to audited stroke and TIA data provide the rationale and means to review administrative coding quality.

The purpose of this study is to determine the agreement between administrative data sources and chart audit for the identification of stroke type, stroke risk factors, and the use of hospital-based diagnostic procedures in patients with stroke or TIA. In a review of the quality of administrative data available in Canada, Roos et al. [3] defined 'reliability' as the agreement between data sources, a term that we adopt in our study to assess administrative data quality specific to stroke and TIA coding.

Methods

Study Setting

By population, Ontario is the largest province in Canada with 13.6 million residents (2014), 86% of whom live in an urban area (2011). Under a publicly funded and administered health care system, provincial residents are insured for all medically necessary health care

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

services delivered in hospitals, emergency departments (ED), and physician offices. In this study, three data sources are used: hospital administrative data (inpatient department and ED), the Ontario Stroke Audit (OSA) data, and physician billing data.

Data Sources

Hospital Administrative Data

The hospital administrative data consist of the Discharge Abstract Database (DAD), a database that includes demographic, diagnostic, and treatment information associated with each inpatient hospitalization in Ontario, and the National Ambulatory Care Reporting System (NACRS), a database documenting similar information related to visits to ED of Ontario acute hospital. Up to 25 International Classification of Diagnosis, 10th revision, Canada (ICD-10-CA) diagnoses may be recorded on the inpatient record and up to 10 on the ED. In the inpatient data, each diagnosis is 'typed' according to whether the patient was admitted with a pre-existing condition (type 1) or the patient developed a condition following admission (type 2). Up to 20 diagnostic or therapeutic procedures in the inpatient record and 10 in the ED may be recorded based on the Canadian Classification of Interventions (CCI) coding standard.

Both the DAD and NACRS administrative data are collected and maintained by the Canadian Institute for Health Information (CIHI).

Ontario Stroke Audit

The audit is a population-based sample of patients with acute stroke or TIA presenting to acute care hospitals across the province. The sample was identified from the DAD and NACRS administrative data based on a most responsible diagnosis (in inpatient data, the diagnosis that accounted for the greatest length of stay) or main problem (the diagnosis determined at the end of the ED visit to be the clinically significant reason for the visit) of intracerebral hemorrhage (ICH, I61.x), ischemic stroke (I63.x, I64.x, H34.1), or TIA (G45.x excluding G45.4) and with a discharge date between April 1, 2012 and March 31, 2013. Stroke or TIA diagnoses identified as suspected or questionable were not included. Only the first stroke or TIA event for a patient in the period under study was kept in the sample. Audited ED cases were restricted to those visits that did not result in transfer to inpatient care.

A District Stroke Center (DSC) is an acute hospital with written stroke protocols, clinicians with stroke expertise, and linkages to rehabilitation and secondary prevention service providers. A Regional Stroke Center (RSC) equates to an American comprehensive stroke center and meets all requirements of a DSC with the addition of neurosurgical facilities and interventional radiology. A nondesignated center is an acute hospital that does not meet the definition of either a RSC or DSC. A nondesignated Telestroke center is an acute hospital with a computed tomography (CT) scanner and video-conferencing capability enabling communication between an ED physician and a stroke expert (stationed elsewhere in the province) related to the diagnosis and treatment of patients presenting with stroke or TIA symptoms. All 11 RSCs and 16 of 17 DSCs are located in urban areas while 24% of nondesignated centers and 38% of Telestroke hospitals are located in small towns, defined as a place outside the commuting zone of centers with a population of 10,000 or more [13].

The final diagnosis recorded in the audited data was based on meticulous chart review by abstractors who had received comprehensive training by research staff and physician specializing in stroke. The training focused on stroke and TIA care management, specifically key aspects of diagnosis and treatment. As part of their training, abstractor candidates were required to complete test charts and achieve a score of at least 85% for key variables on 8 of the 10 test charts to qualify for the role. Once they were in the field, abstractors were instructed to call the research office whenever there was uncertainty around information in the chart which was then passed on to the research physician for review and resolution. Data collected

EXTRA
Cerebrovascular Diseases

Cerebrovasc Dis Extra 2016;6:96–106

| DOI: 10.1159/000449288 | © 2016 The Author(s). Published by S. Karger AG, Basel
www.karger.com/cee |

99

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

from the patient chart included presenting symptoms and stroke severity, past medical history, diagnostic tests administered, treatment provided, in-hospital complications, and discharge destination. Abstractors coded TIA if the patient's definitive or probable diagnosis at the time of discharge from hospital met several criteria: symptoms/neurological deficits persisted for less than 24 h of onset, normal neuroimaging, and thrombolysis was not administered. Abstractors were not blinded to the discharge summary sheet in the medical chart.

A total of 15,802 cases were audited and following exclusions, 14,520 cases remained in the final sample used in this study. Description of the audit sampling strategy and a complete breakdown of exclusions is reported elsewhere [14].

Physician Billing Data

The Ontario Health Insurance Plan (OHIP) is a database containing the records of fee-for-service (FFS) billings by physicians, as well as shadow billing among physicians paid through alternative payment plans. As of March 31, 2013, approximately 64% of physicians received compensation through FFS billing [15]. A physician may participate in both FFS and alternative payment plans; however, physicians compensated under alternative funding arrangements are expected to shadow bill; a shadow bill includes identical information to a FFS claim but where the amount claimed is set to CD 0.

The OHIP database includes encrypted physician and patient identifiers, the date of service, the code for the service provided, and the patient's diagnosis. We used the service code to identify stroke-related diagnostic procedures.

Analysis

The audited data were linked with hospital administrative and physician billing data using unique encoded identifiers and analyzed at the Institute for Clinical Evaluative Sciences (ICES). In the analysis, we excluded patients younger than 18 years and older than 102 years.

We determined the reliability of hospital administrative data by calculating positive predictive value (PPV) for the reporting of ischemic stroke, ICH, and TIA. PPV is defined as the proportion of *Stroke Type X* recorded in administrative data that agrees with the stroke type reported in the audit. PPV is calculated for all hospitals, as well as for each of the four hospital types.

We calculated agreement between the audit and the hospital administrative data or physician billing data on the coding of risk factors (atrial fibrillation, diabetes, hyperlipidemia, hypertension) and stroke-related diagnostic procedures [CT of the brain, magnetic resonance imaging (MRI) of the brain, echocardiography, and carotid imaging (includes catheter angiography, carotid Doppler ultrasound, CT angiography and MR angiography of the carotid artery)]. Observed agreement is defined as concordance of the administrative data and the audit data on whether *Event X* (e.g., risk factor of diabetes, CT scan, etc.) was (or was not) coded in administrative and audit data. The denominator is the total number of audited records.

For the comparisons involving stroke-related diagnostic procedures, we calculated agreement under two scenarios. First, each hospital administrative data type (inpatient or ED) was individually compared with the audit. Second, we combined hospital data and physician billing data and compared with the audit (inpatient or ED + physician billing). To ensure, we captured preadmission diagnostic procedures received by inpatients, and we linked inpatient administrative data with the ED visit that precipitated their admission. Less than 2% of cases (1.8%; n = 170) could not be linked as the unique visit identifier in the inpatient database and did not match an ED record. Hospital administrative data were then linked to physician billing data using encrypted patient identifier and the hospital admission date (DAD) or registration date (NACRS) allowing for an absolute difference of 1 day between the hospital date and the service date entered in the physician claim.

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

Table 1. Characteristics of the OSA sample in the period between April 1, 2012 and March 31, 2013

Characteristic	Inpatient (n = 9,658)		ED (n = 4,850)	
	n	%	n	%
Sex				
Female	4,716	48.8	2,506	51.7
Male	4,942	51.2	2,344	48.3
Age group				
18–54 years	1,022	10.6	538	11.1
55–64 years	1,430	14.8	813	16.8
65–74 years	2,082	21.6	1,136	23.4
75–84 years	2,890	29.9	1,411	29.1
>85 years	2,234	23.1	952	19.6
Stroke subtype				
Ischemic	7,079	73.3	1,020	21.0
TIA	1,575	16.3	3,780	77.9
ICH	1,004	10.4	50	1.0
Diagnostic and therapeutic intervention				
CT scan of the brain	9,548	98.9	4,339	89.5
MRI scan of the brain	3,378	35.0	126	2.6
Carotid imaging	7,237	74.9	448	9.2
Echocardiogram	6,267	64.9	41	0.9
tPA	1,204	12.5	–	–
Hospital type				
RSC (n = 11)	3,853	39.9	1,560	32.2
DSC (n = 17)	3,022	31.3	1,389	28.6
Nondesignated hospital (n = 63)	2,317	24.0	1,656	34.1
Telestroke hospital[a] (n = 8)	466	4.8	245	5.1

– = Not applicable. [a] Hospital is also nondesignated.

The prevalence of stroke risk factors in the audited data and the inpatient administrative data was also determined. To evaluate differences in the prevalence of risk factor reporting in hospital administrative data compared to the audited data, we calculated the relative risk for binary matched pairs according to Agresti and Min [16]. Risk factors could be of any diagnosis type. Only risk factors reported in inpatient data were reviewed.

Kappa which adjusts for chance agreement was calculated for all comparisons. Kappa values of less than 0.2, 0.2–0.39, 0.4–0.59, 0.6–0.79, and 0.80–1.00 correspond to poor, fair, moderate, good, and very good agreement, respectively [17]. Where reported, 95% confidence intervals were calculated using the binomial approximation method.

ICD-10-CA codes for risk factors, as well as CCI and physician billing codes used to identify diagnostic procedures are listed in the online supplementary appendix (for all online suppl. material, see www.karger.com/doi/10.1159/000449288). The analyses were performed using SAS, version 9.2 (SAS Institute, Cary, N.C., USA).

This study was approved by the Research Ethics Board of Sunnybrook Health Sciences Centre, Toronto, Ont., Canada.

Results

As shown in table 1, a total of 14,508 patient charts meeting the inclusion criteria of this study were audited of which two-thirds were inpatient cases (9,658, 66.6%). The majority of audited records in the inpatient setting were for ischemic stroke (73.3%), while in the ED,

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

Table 2. PPV and kappa agreement for stroke subtype, by all hospitals combined, hospital type, and service setting in the period between April 1, 2012 and March 31, 2013

| Stroke subtype	All hospitals (n = 99)		RSC (n = 11)		DSC (n = 17)		Nondesignated hospital (n = 63)		Telestroke hospital (n = 8)	
	PPV % (95% CI)	kappa (95% CI)	PPV % (95% CI)	kappa (95% CI)	PPV % (95% CI)	kappa (95% CI)	PPV % (95% CI)	kappa (95% CI)	PPV % (95% CI)	kappa (95% CI)
Inpatient										
Ischemic	97.3 (96.9–97.7)	0.88 (0.87–0.90)	98.2 (97.7–98.7)	0.91 (0.90–0.93)	97.3 (96.6–98.0)	0.89 (0.87–0.91)	96.8 (96.0–97.7)	0.86 (0.84–0.89)	93.0 (90.3–95.7)	0.75 (0.69–0.82)
TIA	89.5 (88.0–91.0)	0.88 (0.87–0.89)	89.5 (86.7–92.2)	0.89 (0.87–0.91)	91.0 (88.7–93.3)	0.90 (0.88–0.92)	88.5 (85.6–91.5)	0.87 (0.84–0.90)	83.7 (75.9–91.5)	0.77 (0.70–0.85)
ICH	91.9 (90.2–93.5)	0.93 (0.92–0.94)	93.8 (91.8–95.8)	0.95 (0.94–0.97)	91.7 (88.2–95.2)	0.92 (0.89–0.95)	88.6 (84.2–93.0)	0.90 (0.87–0.94)	81.1 (68.5–93.7)	0.82 (0.72–0.92)
ED										
Ischemic	78.8 (76.3–81.2)	0.75 (0.73–0.77)	79.2 (75.2–83.2)	0.76 (0.72–0.80)	79.2 (74.2–84.2)	0.74 (0.69–0.79)	76.5 (72.2–80.8)	0.76 (0.72–0.79)	92.3 (83.9–100)	0.70 (0.59–0.81)
TIA	95.3 (94.6–96.0)	0.77 (0.74–0.79)	95.4 (94.2–96.6)	0.78 (0.75–0.82)	94.9 (93.6–96.2)	0.75 (0.70–0.79)	96.3 (95.3–97.3)	0.77 (0.73–0.80)	90.6 (86.5–94.6)	0.73 (0.63–0.84)
ICH	86.3 (76.8–95.7)	0.87 (0.80–0.94)	81.0 (64.2–97.8)	0.81 (0.68–0.94)	100	0.96 (0.87–1.0)	85.7 (67.4–100)	0.89 (0.76–1.0)	a	a

a Suppressed due to small cell count.

EXTRA
Cerebrovascular Diseases

Cerebrovasc Dis Extra 2016;6:96–106

102

DOI: 10.1159/000449288 | © 2016 The Author(s). Published by S. Karger AG, Basel
www.karger.com/cee

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and
Their Care?

Table 3. Prevalence of stroke risk factors in the hospital administrative data and the OSA, kappa agreement, and relative risk in the period between April 1, 2012 and March 31, 2013

Risk factor[a]	Hospital administrative data		Audit		Kappa (95% CI)	Relative risk (95% CI)
	n	%	n	%		
Atrial fibrillation	2,112	21.9	1,945	20.1	0.60 (0.58–0.62)	0.92 (0.89–0.95)*
Diabetes	2,843	29.4	2,730	28.3	0.86 (0.85–0.87)	0.96 (0.94–0.98)*
Hyperlipidemia	1,175	12.2	4,460	46.2	0.16 (0.14–0.17)	3.8 (3.6–4.0)*
Hypertension	5,757	59.6	7,025	72.7	0.41 (0.40–0.43)	1.22 (1.20–1.24)*

Total sample number = 9,658. [a] ICD-10-CA code with any diagnosis type. * p value <0.05.

Fig. 1. Observed agreement and kappa agreement between inpatient administrative data and the OSA on whether the diagnostic procedure was provided, and agreement between inpatient administrative data combined with physician billing data and the OSA, April 1, 2012 to March 31, 2013.
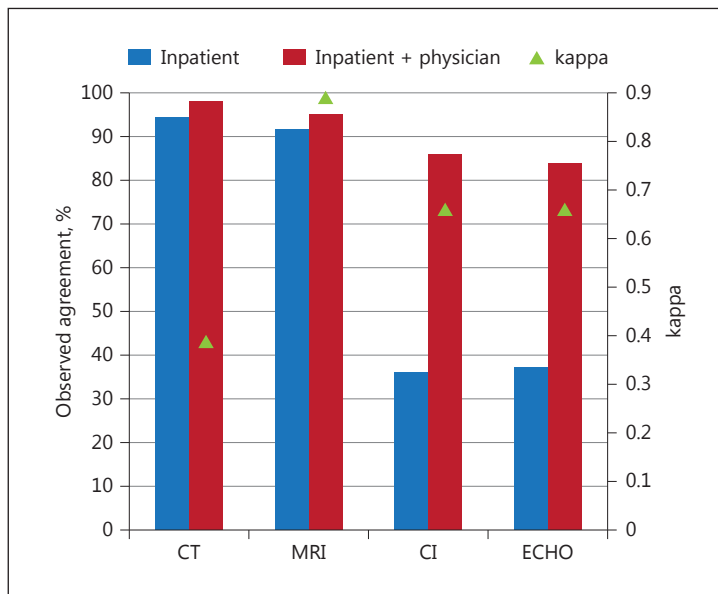
more than three-quarters of audited visits were for TIA (77.9%). According to the audit, for patients admitted to inpatient care, most received carotid imaging (74.9%) and echocardiography (64.9%), while 12.5% were administered tissue plasminogen activator (tPA).

The PPV for ischemic, ICH, or TIA in the inpatient setting among all hospital types ranged from 89.5% (TIA) to 97.3% (ischemic) (table 2). In the ED, PPV varied from 95.3% (TIA) to 78.8% (ischemic). When subset by hospital type, RSCs, DSCs, and nondesignated hospitals had higher PPVs for inpatient-coded ischemic stroke compared to Telestroke sites. In ED administrative data, Telestroke sites had a higher PPV for ischemic coded stroke than RSC and nondesignated hospitals.

The prevalence of stroke risk factors found in the hospital administrative data compared to the audited data is shown in table 3. Atrial fibrillation and diabetes prevalence in the administrative data was overreported compared to the audit, while the prevalence of hyperlipidemia and hypertension was underreported. The relative risk of the reporting of hyperlipidemia in the audit was approximately four times that found in administrative hospital data (RR = 3.8, 95% CI 3.6–4.0). The chance-corrected agreement between the audited and

E X T R A
Cerebrovascular Diseases

Cerebrovasc Dis Extra 2016;6:96–106

103

| DOI: 10.1159/000449288 | © 2016 The Author(s). Published by S. Karger AG, Basel www.karger.com/cee |

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

Fig. 2. Observed agreement and kappa agreement between ED administrative data and the OSA on whether the diagnostic procedure was provided, and agreement between ED administrative data combined with physician billing data and the OSA, April 1, 2012 to March 31, 2013.
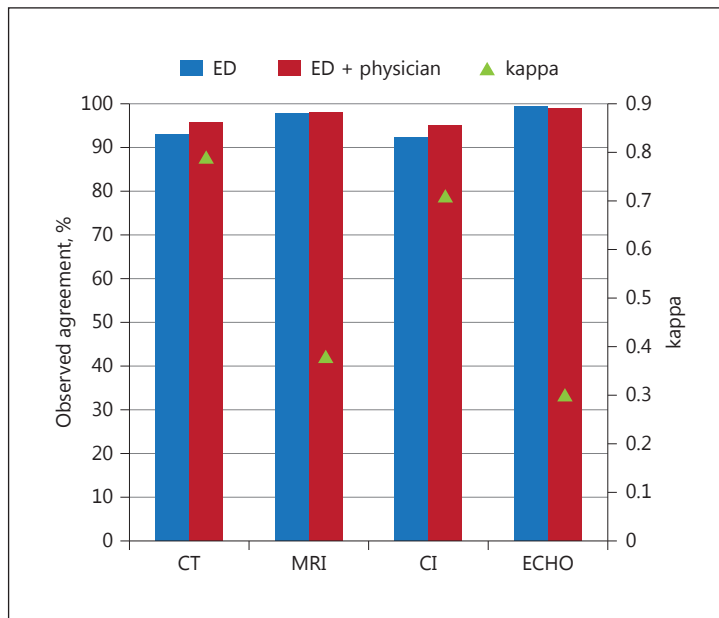
administrative data was good for atrial fibrillation (k = 0.60) and very good for diabetes (k = 0.86).

Agreement between the audit and hospital administrative data, with and without physician billing data, for select diagnostic procedures is shown in figures 1 and 2. Combining the inpatient administrative data with physician billing data (i.e., the procedure reported in either data) improved CT and MRI scan agreement rates minimally and more than doubled the agreement for CI (86.0) and echocardiography (84.0). Observed agreement on diagnostic procedures performed in the ED was over 90% for all procedures regardless of whether physician billing data were included while kappa agreement was good for CT (0.79) and CI (0.71) (fig. 2).

Discussion

In a population-based audit of acute stroke and TIA reported in administrative data, we found close alignment between the ICD-10-CA diagnosis reported in the administrative data and the type of stroke adjudicated in the audit. Compared to other jurisdictions that have assessed stroke and TIA using the ICD-10 classification standard, Johnsen et al. [18] reported lower PPV for ischemic stroke (87.6%), ICH (65.7%), and TIA (60.4%) in a Danish population than we found, while Kirkman et al. [19] using data from the UK reported a higher PPV for ICH (95.1%) compared to our study. One Canadian study reported similar PPVs to those in the current work for ischemic stroke (85%), ICH (98%), and TIA (97%) which may reflect the influence of national standards for health record coding that exist in Canada [20].

In our previous work involving validation of an earlier vintage of administrative data (2006–2008) with a stroke registry, we found lower accuracy in administrative data coding at the stroke type level [21]. Specificity of stroke type reporting in the administrative data has improved in the 4 years between the studies, perhaps due in part to the Canadian Stroke Strategy directives that circulated in 2010 [11].

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

When we examined the reliability of coding at the level of discrete hospital groups, nondesignated and Telestroke hospitals had lower PPVs for all stroke types compared to RSCs and DSCs. In the ED setting, no clear patterns emerged where a single hospital type was consistently more reliable at stroke type identification. The relatively high levels of reliability of stroke type coding regardless of the level of available stroke services is encouraging as previous research from another Canadian jurisdiction found that rural hospitals coded stroke with less specificity than urban-based hospitals [22].

The accuracy of risk factor reporting in administrative data with the exception of diabetes was poor and has shown very little change from our earlier work [21]. The underreporting of comorbidities in hospital administrative data is in keeping with findings from other studies that have examined this issue [23–25]. Lee et al. [23] found improvement in comorbidity identification by using a look-back window of 1-year to find comorbidities from all hospitalizations. Algorithms for identifying chronic conditions, for example diabetes and hypertension, which combined hospital administrative and physician billing data were found to have good predictive ability [26, 27]. In a study examining reliability of comorbid coding in hospital data in another Canadian province, the prevalence of hyperlipidemia and hypertension showed significant underreporting in the hospitalization database as was found in our study [28]. The authors attributed poor coding of chronic conditions such as hyperlipidemia to asymptomatic presentation. In an Australian study that audited hospital discharge records for 43 ICD-10 diagnoses of interest to the researchers, hypertension prevalence of 4.6% and kappa 0.67 were found substantially lower and higher, respectively, than what was found in our study [29]. In Ontario, health records technicians are required to code only those conditions that affect the admission being abstracted. While hyperlipidemia may be an important risk factor for stroke, it is unlikely to have a significant impact on the hospital stay. In our audit cohort, the average number of secondary diagnoses recorded on the inpatient abstract was 4.2 (median 3).

The finding that inpatient hospital administrative data alone were unreliable for identifying carotid imaging and echocardiogram is not unprecedented. Quan et al. [30] looked at inpatient reporting of 29 common procedures among a random selection of general medicine and general surgery admissions. The authors concluded that procedures performed in operating rooms or invasively (e.g., lumbar puncture, dialysis) were more likely to be reported than less invasive ones and those occurring in radiology departments or on the ward. In our study, the combination of hospital administrative and physician billing data greatly improved the identification of procedures relevant to stroke care (fig. 1).

Evaluation of the completeness of stroke-related diagnostic procedure reporting in ED administrative data is a novel feature of our study. Although few procedures were performed (excluding CT scans) in those who were not admitted, the procedures we examined appear to be comprehensively reported in the administrative data and adding physician billing data did little to improve their identification. It should be noted that coding of diagnostic imaging in the ED administrative data is mandatory while optional in inpatient abstraction unless the procedure is the sole intervention performed in the presence of an anesthetist [A. Cote, pers. commun.].

Several limitations of this study are worth noting. First, the audit abstractors were not blinded to the discharge summary sheet in the hospital record which could result in greater agreement than what would have been observed had they been blinded to this information. Second, the definition of TIA used in this study includes criteria that effectively rule out a stroke but do not rule out other transient neurological conditions that mimic TIA, for example, migraine. Thus, some of the TIA patients included in this study may have been misclassified which could bias the reported results. Finally, agreement that includes physician billing data may be less than expected if shadow billing was not adhered to among physicians receiving

EXTRA
Cerebrovascular Diseases

Cerebrovasc Dis Extra 2016;6:96–106

105

DOI: 10.1159/000449288 | © 2016 The Author(s). Published by S. Karger AG, Basel
www.karger.com/cee

Porter et al.: How Reliable Are Administrative Data for Capturing Stroke Patients and Their Care?

compensation under alternative payment plans. In research that examined the effect of non-FFS payment on the detection of diabetes, 23.7% of patients with diabetes identified through prescription drug data remained undetected in physician service claims or hospital administrative data after up to 9 years of follow-up [31]. It is unknown whether diagnosis underreporting among non-FFS physicians correlates with procedure underreporting.

Conclusion

The ability of administrative data to accurately identify health conditions and their related diagnostic procedures is integral to health system evaluation. In the absence of a single comprehensive source of data that accounts for all relevant aspects of individuals' encounters with the health system, researchers may elect to use a combination of administrative data to obtain a more complete picture.

Acknowledgements

References

1 De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH, Johansen H, Lix LM, Luthi JC, Ma J, Romano PS, Roos L, Sundararajan V, Tu JV, Webster G, Ghali WA: Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. BMC Health Serv Res 2006;6:77.

2 Austin PC, Daly PA, Tu JV: A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario. Am Heart J 2002;144:290–296.

3 Roos LL, Gupta S, Soodeen RA, Jebamani L: Data quality in an information-rich environment: Canada as an example. Can J Aging 2005;24(suppl 1):153–170.

4 van Walraven C, Austin P: Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol 2012;65:126–131.

5 Andrade SE, Harrold LR, Tjia J, Cutrona SL, Saczynski JS, Dodd KS, Goldberg RJ, Gurwitz JH: A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. Pharmacoepidemiol Drug Saf 2012;21(suppl 1):100–128.

6 McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA: Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. PLoS One 2015;10:e0135834.

7   Koton S, Schneider AL, Rosamond WD, Shahar E, Sang Y, Gottesman RF, Coresh J: Stroke incidence and mortality trends in US communities, 1987 to 2011. JAMA 2014;312:259–268.

8   Public Health Agency of Canada: Tracking heart disease and stroke in Canada – stroke highlights. Ottawa, Public Health Agency of Canada, 2011.

9   Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa, Broderick JP, Kleindorfer DO: Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology 2012;79:1781–1787.

10 Smajlovic D: Strokes in young adults: epidemiology and prevention. Vasc Health Risk Manag 2015;11:157–164.

11   CSS Information and Evaluation Working Group: Canadian stroke strategy core performance indicator update. 2010. www.strokebestpractices.ca/wp-content/uploads/2012/07/Stroke_Core_ENG.pdf.

12 Hall R, Porter J: Report to the Ministry of Health and Long-Term Care on the quality of reporting of CIHI 340 stroke data elements by Ontario acute hospitals. 2015. http://ontariostrokenetwork.ca/blog/report-to-the-ministry-of-health-and-long-term-care-on-the-quality-of-reporting-of-cihi-340-stroke-data-elements-by-ontario-acute-hospitals/.

13 du Plessis V, Beshiri R, Bollman RD, Clemenson HA: Definitions of 'rural'. Agriculture and Rural Working Paper Series, Catalogue no. 21-601-MIE. Ottawa, Statistics Canada, 2002.

14 Hall R, Khan F, O'Callaghan C, Kapral MK, Levi J, Cullen A, Wu J, Fang J, Bayley M: Evaluation report 2014: on target for stroke prevention and care. Toronto, Institute for Clinical Evaluative Sciences, 2014.

15 Canadian Institute for Health Information: Physician services benefit rates report, Canada, 2012–2013. Ottawa, Canadian Institute for Health Information, 2014.

16 Agresti A, Min Y: Effects and non-effects of paired identical observations in comparing proportions with binary matched-pairs data. Stat Med 2004;23:65–75.

17 Altman DG: Practical statistics for medical research. London, Chapman and Hall, 1991.

18 Johnsen SP, Overvad K, Sorensen HT, Tjonneland A, Husted SE: Predictive value of stroke and transient ischemic attack discharge diagnoses in The Danish National Registry of Patients. J Clin Epidemiol 2002;55:602–607.

19 Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD: The accuracy of hospital discharge coding for hemorrhagic stroke. Acta Neurol Belg 2009;109:114–119.

20 Kokotailo RA, Hill MD: Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. Stroke 2005;36:1776–1781.

21 Hall R, Mondor L, Porter J, Fang J, Kapral MK: Accuracy of administrative versus registry data for the coding of acute stroke and transient ischemic attack. Can J Neurol Sci 2016, Epub ahead of print.

22 Yiannakoulias N, Svenson LW, Hill MD, Schopflocher DP, Rowe BH, James RC, Wielgosz AT, Noseworthy TW: Incident cerebrovascular disease in rural and urban Alberta. Cerebrovasc Dis 2004;17:72–78.

23 Lee DS, Donovan L, Austin PC, Gong Y, Liu PP, Rouleau JL, Tu JV: Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. Med Care 2005;43:182–188.

24 Quan H, Parsons GA, Ghali WA: Validity of information on comorbidity derived from ICD-9-CCM administrative data. Med Care 2002;40:675–685.

25 Romano PS, Mark DH: Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32:81–90.

26 Hux JE, Ivis F, Flintoft V, Bica A: Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care 2002;25:512–516.

27 Tu K, Mitiku T, Lee DS, Guo H, Tu JV: Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD). Can J Cardiol 2010;26:e225–e228.

28 Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM, Spinelli JJ: Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? J Clin Epidemiol 2000;53:343–349.

29 Henderson T, Shepheard J, Sundararajan V: Quality of diagnosis and procedure coding in ICD-10 administrative data. Med Care 2006;44:1011–1019.

30 Quan H, Parsons GA, Ghali WA: Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. Med Care 2004;42:801–809.

31 Alshammari AM, Hux JE: The impact of non-fee-for-service reimbursement on chronic disease surveillance using administrative data. Can J Public Health 2009;100:472–474.