

Accuracy of Administrative Data for the Coding of Acute Stroke and TIAs

Ruth Hall, Luke Mondor, Joan Porter, Jiming Fang, Moira K. Kapral

ABSTRACT: Objective: Administrative data validation is essential for identifying biases and misclassification in research. The objective of this study was to determine the accuracy of diagnostic codes for acute stroke and transient ischemic attack (TIA) using the Ontario Stroke Registry (OSR) as the reference standard. **Methods:** We identified stroke and TIA events in inpatient and emergency department (ED) administrative data from eight regional stroke centres in Ontario, Canada, from April of 2006 through March of 2008 using ICD-10-CA codes for subarachnoid haemorrhage (I60, excluding I60.8), intracerebral haemorrhage (I61), ischemic (H34.1 and I63, excluding I63.6), unable to determine stroke (I64), and TIA (H34.0 and G45, excluding G45.4). We linked administrative data to the Ontario Stroke Registry and calculated sensitivity and positive predictive value (PPV). **Results:** We identified 5,270 inpatient and 4,411 ED events from the administrative data. Inpatient administrative data had an overall sensitivity of 82.2% (95% confidence interval [CI_{95%}] = 81.0, 83.3) and a PPV of 68.8% (CI_{95%} = 67.5, 70.0) for the diagnosis of stroke, with notable differences observed by stroke type. Sensitivity for ischemic stroke increased from 66.5 to 79.6% with inclusion of I64. The sensitivity and PPV of ED administrative data for diagnosis of stroke were 56.8% (CI_{95%} = 54.8, 58.7) and 59.1% (CI_{95%} = 57.1, 61.1), respectively. For all stroke types, accuracy was greater in the inpatient data than in the ED data. **Conclusion:** The accuracy of stroke identification based on administrative data from stroke centres may be improved by including I64 in ischemic stroke type, and by considering only inpatient data.

RÉSUMÉ: Exactitude des données clinico-administratives dans l'encodage des accidents vasculaires cérébraux aigus et des ischémies cérébrales transitoires. Objectif: La validation des données clinico-administratives demeure essentielle si l'on veut déceler des biais et des erreurs de classification en matière de recherche. L'objectif de cette étude a été de déterminer l'exactitude des codes de diagnostic des accidents vasculaires cérébraux (AVC) aigus et des ischémies cérébrales transitoires (ICT) en utilisant le Registre de l'AVC de l'Ontario comme norme de référence. **Méthodes :** D'avril 2006 à mars 2008, nous avons répertorié des épisodes d'AVC et d'ICT à partir de données clinico-administratives obtenues auprès des centres régionaux ontariens de traitement des AVC, qu'elles concernent des patients hospitalisés ou des services d'urgence. Pour ce faire, nous avons utilisé les codes de la CIM-10-CA dans des cas d'hémorragie méningée (I60, en excluant I60.8), d'hémorragie intracérébrale (I61) et d'ICT (H34.1 et I63, en excluant I63.6). Lorsqu'incapables de déterminer s'il s'agissait d'un AVC, nous avons utilisé le code I64 alors que dans le cas d'une ICT, nous avons opté pour H34.0 et G45 en excluant G45.4. Nous avons ensuite associé ces données clinico-administratives au Registre de l'AVC de l'Ontario et calculé leur sensibilité et leur valeur prédictive positive (VPP). **Résultats :** À partir de ces données clinico-administratives, nous avons répertorié 5 270 patients hospitalisés et 4 411 épisodes survenus dans des services d'urgence. La sensibilité globale des données concernant les patients était de 82,2% (intervalle de confiance à 95% [IC95%] = 81,0 ; 83,3). La VPP de leurs données était de 68,8% (IC95% = 67,5 ; 70,0) en ce qui concerne le diagnostic d'un AVC, des différences manifestes étant observées selon les types d'AVC. En incluant I64, la sensibilité des données concernant les ICT est passée de 66,5 à 79,6%. Par ailleurs, la sensibilité et VPP des données clinico-administratives des services d'urgence dans des cas d'AVC étaient respectivement de 56,8% (IC95% = 54,8 ; 58,7) et 59,1% (IC95% = 57,1 ; 61,1). Pour tous les types d'AVC, les données fournies au sujet des patients hospitalisés se sont révélées davantage exactes que celles des services d'urgence. **Conclusions :** Sur la base des données clinico-administratives fournies par les centres régionaux ontariens de traitement des AVC, l'identification de ces derniers pourrait être améliorée en incluant le code I64 dans les types d'ICT et en ne considérant que les données des patients hospitalisés.

Keywords: Stroke, Diagnostic accuracy, Health administrative data, Health services research, Sensitivity, Predictive value

doi:10.1017/cjn.2016.278

Can J Neurol Sci. 2016; 43: 765-773

INTRODUCTION

Evaluation of the quality and outcomes of the care of patients with stroke or transient ischemic attack (TIA) typically relies on data from clinical stroke registries. However, there is increasing interest in the

use of administrative data to identify cohorts and provide follow-up information for epidemiological and comparative effectiveness studies.¹⁻⁹ As our health systems experience economic pressures and at the same time are expected to be accountable for the services provided, the need to rely on a comprehensive, cost-efficient and

From the Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada (RH, LM, JP, JF, MKK); Department of Medicine, University of Toronto, Toronto, Ontario, Canada (RH, MKK); Ontario Stroke Network, Toronto, Ontario, Canada (RH); Division of General Medicine and Toronto General Research Institute, University Health Network, Ontario, Canada (MKK).

RECEIVED JUNE 4, 2015. FINAL REVISIONS SUBMITTED JANUARY 21, 2016. DATE OF ACCEPTANCE FEBRUARY 10, 2016.

Correspondence to: Ruth Hall, Institute for Clinical Evaluative Sciences, G1 06, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada. Email: Ruth.hall@ices.on.ca.

sustainable data source will become more important. In Canada, a country with universal publicly funded coverage of hospital-based services, administrative data offer an accessible and population-based source of information associated with each patient encounter. The validity of administrative data in identifying discrete health conditions, like stroke and TIA, is fundamental to the utility and ultimately the quality of the research based on these data.

Prior studies of the validity of coding for stroke and TIA in administrative data have been conducted in many jurisdictions, typically in the subgroup of patients admitted to hospital, with medical record review as the reference standard and in many cases in earlier eras when access to imaging and specialized stroke centres was limited.¹⁰ In a systematic review of methods for identifying stroke events using administrative and claims data, Andrade et al. (2012)¹⁰ compiled 26 articles that met their criteria for evaluation. The selected validation studies employed administrative data from as early as 1970 and up to 2006, though more than half of those included for review were based on data collected prior to 2000. Only one of the reviewed studies used ICD-10-coded data, 1 of the 26 studies based its validation on a stroke registry, and the remainder utilized review of the medical record. Validation of outpatient administrative data for stroke was evaluated in one study using a paediatric population. TIA has received less attention, and the Andrade review identified only seven studies specific to validation of TIAs. These TIA validation studies employed data from 1992 to 2006, while one study was based on the ICD-10 standard, and one reported outpatient results of the validation. We used the Ontario Stroke Registry (the OSR, formerly known as the Registry of the Canadian Stroke Network) as the reference standard for validation of administrative data for diagnoses of acute stroke and TIA, for identification of vascular risk factors, and for diagnostic and treatment interventions among inpatients with stroke or TIA.

METHODS

Data Sources

Registry

In Ontario, Canada, a province-wide system of stroke care management was launched in 2000 and fully implemented by 2006, the details of which are reported elsewhere.^{11,12} As part of the implementation of the stroke system, a registry was established, and it utilized an active method of identifying potentially eligible patients seen in the emergency department or admitted to any of the 11 regional stroke centres (with resources similar to American comprehensive stroke centres). In fiscal year 2007/2008 (April 1, 2007–March 31, 2008), 31% of all acute stroke and TIA events in Ontario were managed at these regional stroke centres.¹³ Specifically, onsite-trained neurology nurse research coordinators used a variety of recruitment strategies, including a review of lists of potential stroke patients generated by emergency and inpatient wards and medical records departments. The charts were then reviewed by research coordinators, eligibility assessed through review of ED and/or neurology consultation notes, as well as diagnostic imaging reports, and eligible patients were entered into the OSR.

Chart abstractors for the registry received intensive training by a research nurse and two physicians specializing in stroke. As part of their training, abstractors were required to abstract ten test charts of various levels of complexity. Interrater discrepancies

identified during the test chart abstraction were discussed and resolved. Once abstractors were in the field, interrater reliability was periodically assessed. In addition, once a month the research physicians teleconferenced with abstractors for the purpose of adjudicating clinical scenarios that had not been accounted for during training. In late 2006, abstractors from all 11 sites attended a one-day training workshop covering such topics as an overview of neuroimaging and review of the Canadian Neurological Score, the National Institutes of Health Stroke Scale and Oxfordshire Community Stroke Project scoring. In the 2007 reliability testing, excellent agreement (100%) was found for key variables, including age, sex, stroke type and use of thrombolysis. Cases were comprehensively documented using a combination of prospective data collection and chart review after the patient was discharged. Each stroke or TIA event represents one record in the registry and includes information about the patient's symptoms, stroke severity, medical history, diagnostic and treatment services provided, complications, and functional ability at discharge. Data collection for the OSR is done without patient consent since the OSR is housed at the Institute for Clinical Evaluative Studies (ICES), an organization designated as a prescribed entity under provincial privacy legislation.

Administrative Databases

Two administrative databases were utilized in this validation study: (1) the National Ambulatory Care Reporting System (NACRS) database, which includes information on all visits to hospital emergency departments (ED); and (2) the Discharge Abstract Database (DAD), a repository of all inpatient hospitalizations in Ontario. These databases are developed and maintained by the Canadian Institute for Health Information. For both administrative data sources, clinical and demographic information are abstracted from the hospital chart by trained health records technicians following a patient's discharge. Abstracted data elements include the main problem determined at the end of the ED visit (the diagnosis identified by the provider as being the most clinically significant reason for the visit), and, among admitted patients, the most responsible diagnosis (defined as the single diagnosis that contributes the most to the patient's length of stay or consumes the majority of resources during admission). Other recorded diagnoses are conditions that existed prior to or occurred during hospitalization and that affected the patient's treatment and management during hospitalization. The International Classification of Diseases, Tenth Revision, Canada (ICD-10-CA) and the Canadian Classification of Interventions (CCI) coding standards were employed to capture diagnoses and interventions, respectively. Up to ten diagnoses and ten interventions can be recorded in the ED database, and up to 25 diagnoses and 20 interventions can be recorded in the admitted patient database. In both the ED and inpatient databases, the procedure or intervention considered the most clinically significant is entered as the main intervention. For inpatient data, each diagnosis is assigned a type according to the temporal relationship it has with the admission date. Type 1 diagnoses are pre-admission comorbid conditions, while type 2 diagnoses represent conditions that develop during the admission. Age and sex variables were obtained from the Registered Persons Database, a file maintained by the provincial health authority and containing demographic information about all persons who have received a health card number.

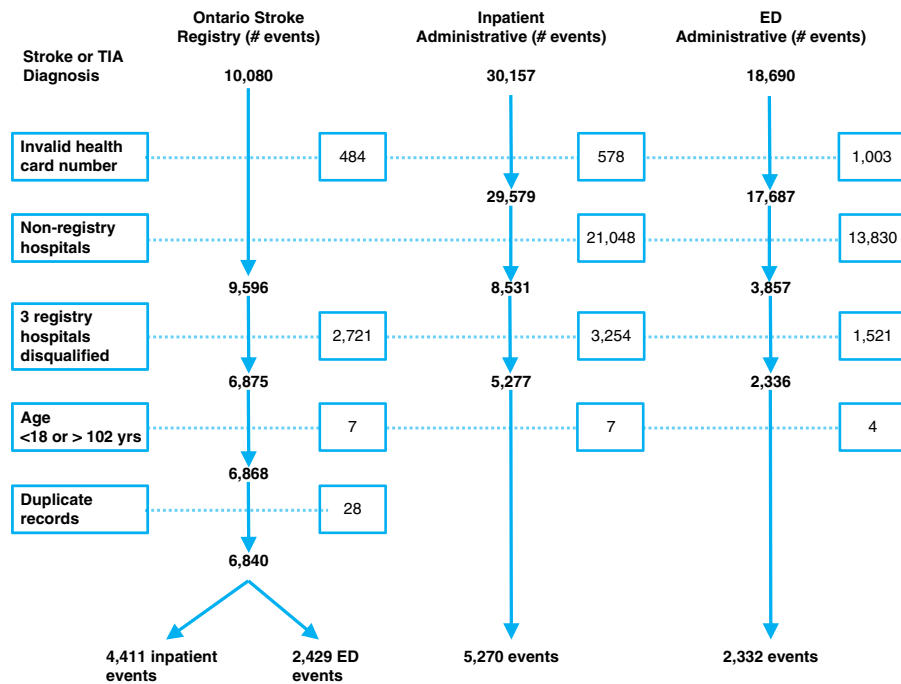


Figure 1: Flowchart of exclusions in the Ontario Stroke Registry and administrative data.

These datasets were linked using unique encoded identifiers and analyzed at the ICES. Our study was approved by the Sunnybrook Health Sciences Centre Research Ethics Board.

Study Population

We identified all patients recorded in the ED or inpatient administrative data with an ICD-10-CA diagnosis code of a subarachnoid haemorrhage (160.x, excluding 160.8), intracerebral haemorrhage (161.x), ischemic stroke (163.x and H34.1, excluding 163.6), a stroke not specified as haemorrhage or infarction (164.x) or TIA (H34.0 and G45.x, excluding G45.4), and with a service date (for ED visits) or discharge date (for inpatients) between April 1, 2006, and March 31, 2008 (see Figure 1). An individual may appear more than once in the study dataset if they experienced two or more strokes or TIAs over the observation period.

In both the registry and administrative data, we excluded events where the patient was younger than 18 or older than 102 years of age, as well as those where the health card number on the record was invalid, if the ED visit was a scheduled appointment, or if the stroke or TIA was a result of a post-admission complication. For the ED data, we excluded events that resulted in admission to an acute hospital, as these would already be captured in the inpatient group (see Figure 1). Three regional stroke centres were excluded from the analysis. Two centres were multi-site corporate entities but reported under a single hospital identifier in the administrative database. The third centre had incomplete registry data collection for a portion of the period under review, leaving registry data from eight regional stroke centres to compare with administrative data (see Figure 1). Stroke events recorded in the administrative data of these three hospitals were also excluded. Hospitals were grouped according to the peer groups defined by

the Ontario Joint Policy and Planning Committee.¹⁴ Teaching hospitals are acute hospitals with membership on the Council of Academic Hospitals of Ontario and that provide complex patient care, are affiliated with a medical or health sciences school, and have significant research activity and postgraduate training. Community hospitals are defined as large hospitals that do not meet the definition of a teaching hospital.¹⁵

Analysis

We linked events from the administrative databases to a registry record using encrypted patient identifier, institution, and date and time of registration in the ED, or, for those admitted, date of discharge. We allowed a 24-hour absolute difference in ED registration time and the registry, and a one-day absolute difference between discharge dates recorded in the inpatient administrative data and the registry.

We evaluated the validity of administrative data from eight regional stroke centres in identifying acute stroke and TIA events (excluding in-hospital strokes) in three ways. First, we compared events with an exact code match of acute stroke or TIA at the level of the main problem (ED) or most responsible diagnosis code (inpatient). Second, we created two stroke groups based on the reported main problem or most responsible diagnosis. The ischemic stroke group consisted of ischemic stroke (I63) and stroke not specified as haemorrhage or infarction (I64), while the haemorrhagic group was a combination of intracerebral haemorrhage and subarachnoid haemorrhage.¹⁶ Third, we compared events with stroke or TIA appearing in any diagnosis position and excluding those that occurred post-admission. We calculated sensitivity and positive predictive value (PPV), with sensitivity defined as the percentage of stroke and TIA events in the registry that linked to an administrative record, and PPV as the percentage

of acute stroke or TIA identified in administrative records that linked to an event in the registry. We also calculated agreement using Cohen's kappa methodology, which corrects for chance agreement. Kappa values <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

For the secondary objectives, we calculated the agreement between inpatient administrative data reporting of risk factors (hypertension, hyperlipidemia, diabetes, atrial fibrillation), stroke-related diagnostics (computed tomography [CT] of the brain, magnetic resonance imaging [MRI] of the brain, carotid imaging [includes catheter angiography, carotid Doppler ultrasound, CT angiography and MR angiography of the carotid artery], and echocardiography), and the use of tissue plasminogen activator (tPA) with what was documented in the registry. We excluded ED data from the risk factor analysis due to the minimal reporting of diagnoses beyond the main diagnosis

(median number of diagnoses = 0, mean = 0.4). The ICD-10-CA and CCI codes used in this analysis are included in Appendix 1.

Where reported, 95% confidence intervals ($CI_{95\%}$) were calculated using the binomial approximation method. Data management and statistical analyses were performed using SAS software (v. 9.2, SAS Institute, Cary, NC).

RESULTS

The characteristics of patients with acute stroke or TIA in the administrative data and registry are shown in Table 1. Of the various stroke types, ischemic stroke represented the largest percentage of events in the inpatient setting (51.8% in administrative data and 68.9% in the registry), while TIA represented the largest percentage of events in the ED (65.8% in administrative data and 61.9% in the registry). Both inpatient and ED administrative data sources had higher percentages of stroke of undetermined type

Table 1: Characteristics of Stroke Events in the Inpatient and Emergency Department Administrative Database and Ontario Stroke Registry, April 1, 2006-March 31, 2008

	Inpatient		ED	
	Administrative	Registry	Administrative	Registry
	<i>n</i> = 5,270	<i>n</i> = 4,411	<i>n</i> = 2,332	<i>n</i> = 2,429
Characteristic	<i>N</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>N</i> (%)
Sex				
Female	2,596 (49.3)	2,145 (48.6)	1,193 (51.2)	1,206 (49.7)
Male	2,674 (50.7)	2,266 (51.4)	1,139 (48.8)	1,223 (50.3)
Age group				
18-54	811 (15.4)	543 (12.3)	269 (11.5)	346 (14.2)
55-64	809 (15.4)	603 (13.7)	381 (16.3)	417 (17.2)
65-74	1,085 (20.6)	942 (21.4)	519 (22.3)	534 (22.0)
75-84	1,649 (31.3)	1,472 (33.4)	804 (34.5)	781 (32.1)
85+	916 (17.4)	851 (19.3)	359 (15.4)	351 (14.5)
Year of event				
April 1, 2006-March 31, 2007	2,598 (49.3)	2,145 (48.6)	1,165 (50.0)	1,151 (47.4)
April 1, 2007-March 31, 2008	2,672 (50.7)	2,266 (51.4)	1,167 (50.0)	1,278 (52.6)
Stroke type				
Ischemic (H34.1, I63.x, excluding I63.6)	2,727 (51.8)	3,041 (68.9)	93 (4.0)	563 (23.2)
Intracerebral haemorrhage (I61.x)	680 (12.9)	542 (12.3)	78 (3.3)	56 (2.3)
Subarachnoid haemorrhage (I60.x, excluding I60.8)	414 (7.9)	117 (2.7)	67 (2.9)	29 (1.2)
Transient ischemic attack (H34.0, G45.x, excluding G45.4)	773 (14.7)	633 (14.4)	1,535 (65.8)	1,505 (61.9)
Not specified as haemorrhage or ischemic (I64.x)	676 (12.8)	78 (1.8)	559 (24.0)	277 (11.4)
Stroke group				
Ischemic/not specified (H34.1, I64.x, I63.x excluding I63.6)	3,403 (64.6)	3,119 (70.7)	652 (28.0)	840 (34.6)
Hemorrhagic (I61.x, I60.x excluding I60.8)	1,094 (20.8)	659 (14.9)	145 (6.2)	85 (3.5)
Hospital type				
Community	1,693 (32.1)	1,492 (33.8)	582 (25.0)	643 (26.5)
Teaching	3,577 (67.9)	2,919 (66.2)	1,750 (75.0)	1,786 (73.5)
Number of unique patients	5,058	4,257	2,195	2,292
Mean number of stroke/TIA events (range)	1.04 (1-4)	1.07 (1-6)	1.06 (1-4)	1.1 (1-5)

Table 2: Diagnostic Accuracy of Stroke and Transient Ischemic Attack (TIA) Coded in Administrative Data Compared to the Ontario Stroke Registry, by Service Setting, Stroke ICD-10-CA Code and Stroke Group, April 1, 2006-March 31, 2008

	TP (n)	FP (n)	FN (n)	Sensitivity	PPV	Kappa*
Emergency department administrative data				(CI _{95%})	(CI _{95%})	(CI _{95%})
Stroke/TIA coded in any diagnosis position	1,399	1,122	1,030	57.6 (55.6, 59.6)	55.5 (53.6, 57.4)	†
Main problem, stroke/ TIA	1,379	953	1,050	56.8 (54.8, 58.7)	59.1 (57.1, 61.1)	†
Main problem, stroke group						
Ischemic/not specified (H34.1, I64.x, I63.x, excluding I63.6)	287	365	553	34.2 (31.0, 37.4)	44.0 (40.2, 47.8)	0.62 (0.58, 0.67)
Hemorrhagic (I61.x, I60.x, excluding I60.8)	31	114	54	36.5 (26.2, 46.7)	21.4 (14.7, 28.1)	0.86 (0.77, 0.94)
Main problem diagnosis code						
Ischemic (H34.1, I63.x, excluding I63.6)	39	54	524	6.9 (4.8, 9.0)	41.9 (31.9, 52.0)	0.15 (0.10, 0.19)
TIA (H34.0, G45.x, excluding G45.4)	844	691	660	56.1 (53.6, 58.6)	54.9 (52.5, 57.5)	0.64 (0.60, 0.69)
Intracerebral haemorrhage (I61.x)	23	55	33	41.1 (28.2, 54.0)	29.5 (19.4, 39.6)	0.80 (0.69, 0.92)
Subarachnoid haemorrhage (I60.x, excluding I60.8)	7	60	22	24.1 (8.6, 39.7)	10.4 (3.1, 17.8)	0.93 (0.80, 1.0)
Inpatient administrative data						
Stroke/TIA coded in any diagnosis position	3,740	1,995	671	84.8 (83.7, 85.8)	65.2 (63.9, 66.4)	†
Most responsible diagnosis stroke/TIA	3,624	1,646	787	82.2 (81.0, 83.3)	68.8 (67.5, 70.0)	†
Most responsible diagnosis stroke group						
Ischemic/not specified (H34.1, I64.x, I63.x, excluding I63.6)	2,483	920	636	79.6 (78.2, 81.0)	72.9 (71.5, 74.5)	0.80 (0.78, 0.82)
Haemorrhagic (I61.x, I60.x excluding I60.8)	465	629	194	70.6 (67.1, 74.0)	42.5 (39.6, 45.4)	0.93 (0.91, 0.94)
Most responsible diagnosis code						
Ischemic (H34.1, I63.x, excluding I63.6)	2,021	706	1,020	66.5 (64.8, 68.1)	74.1 (72.5, 75.8)	0.59 (0.57, 0.62)
TIA (H34.0, G45.x, excluding G45.4)	386	387	247	60.9 (57.2, 64.8)	49.9 (46.4, 53.5)	0.74 (0.70, 0.77)
Intracerebral haemorrhage (I61.x)	365	315	177	67.3 (63.4, 71.3)	53.7 (49.9, 57.4)	0.89 (0.87, 0.92)
Subarachnoid haemorrhage (I60.x, excluding I60.8)	83	331	34	70.9 (62.7, 79.2)	20.0 (16.2, 23.9)	0.88 (0.84, 0.93)

*Based on linked records: $n = 3,624$ (inpatient) and $n = 1,379$ (ED).

†Value of κ cannot be calculated, as true negatives are not known.

TP = true positive; FP = false positive; FN = false negative.

compared to the registry (12.8 vs. 1.8% of inpatient events and 24.0 vs. 11.4% of ED events).

As shown in Table 2, when stroke or TIA (ignoring stroke type) is in the main diagnosis position, the sensitivity of the inpatient administrative data reached 82.2%, with a PPV of 68.8%. When all diagnosis positions were considered, sensitivity increased to 84.8% but PPV decreased to 65.2%. Events coded with ischemic stroke as the most responsible reason for hospitalization had poor sensitivity (66.5%), though when combined with UTD stroke (I64) sensitivity improved (79.6%), with only a small reduction in PPV. Subarachnoid haemorrhagic (SAH) stroke demonstrated the highest sensitivity (70.9%) among the various stroke types, and the lowest PPV (20.0%). For stroke or TIA events assessed in the ED and discharged to the community, the sensitivity and PPV for all stroke types were low, ranging from a sensitivity of 6.9% (ischemic) to 56.1% (TIA) and a PPV of 10.4% (SAH) to 54.9% (TIA). Although not shown, we investigated the sensitivity and PPV of stroke type stratified by service setting and teaching and community hospital status and found similar results for both institution types for stroke and TIA collectively, as well as for ischemic stroke type combined with unspecified stroke type.

We also reviewed the distributions of false positive strokes and TIA and found that ischemic stroke was frequently coded

as stroke-not specified and TIA as ischemic, and in the case of haemorrhagic strokes, subarachnoid was substituted for intracerebral (results not shown). Similar patterns are reported in other studies.^{18,19}

Agreement between the administrative data and registry on documentation of risk factors, diagnostic procedures and treatment interventions is shown in Table 3. Among the risk factors examined, agreement was very good for diabetes ($\kappa = 0.83$), good for atrial fibrillation ($\kappa = 0.60$), fair for hypertension ($\kappa = 0.32$) and poor for hyperlipidemia ($\kappa = 0.13$). For diagnostic and therapeutic interventions provided to inpatients, agreement was good for both CT ($\kappa = 0.64$) and MRI ($\kappa = 0.77$) but poor for carotid imaging ($\kappa = 0.03$) and echocardiography ($\kappa = 0.02$). Agreement for thrombolysis administration was moderate ($\kappa = 0.47$). In the ED setting, CT scan ($\kappa = 0.77$) and MRI scan ($\kappa = 0.66$) had good agreement while carotid imaging had poor agreement ($\kappa = 0.15$).

DISCUSSION

We found inpatient administrative data from regional stroke centres to be a valid data source for identifying stroke or TIA as well as for identifying the combined group of ischemic stroke and

Table 3: Prevalence and Agreement of Diagnostic and Therapeutic Interventions in the Administrative Data Record as Compared to the Ontario Stroke Registry Among Records that Linked (April 1, 2006-March 31, 2008)

	Administrative	Registry	
Risk factor,[†] inpatient	n = 5,270	n = 4,411	Kappa* (CI_{95%})
	n (%)	n (%)	
Hypertension	2,469 (46.9)	2,994 (67.9)	0.32 (0.30, 0.35)
Hyperlipidemia	375 (7.1)	1,539 (34.9)	0.13 (0.11, 0.16)
Diabetes	1,234 (23.4)	1,095 (24.8)	0.83 (0.81, 0.85)
Atrial fibrillation	827 (15.7)	796 (18.1)	0.60 (0.56, 0.63)
Intervention, inpatient	n = 5,270	n = 4,411	Kappa* (CI_{95%})
	n (%)	n (%)	
CT scan	2,590 (49.2)	2,092 (47.4)	0.64 (0.61, 0.66)
MRI scan	1,364 (25.9)	1,187 (26.9)	0.77 (0.74, 0.79)
CT or MRI scan	3,434 (65.2)	2,948 (66.8)	0.61 (0.58, 0.64)
Carotid imaging	212 (4.0)	2,976 (67.5)	0.03 (0.03, 0.04)
Echocardiogram	71 (1.4)	2,567 (58.2)	0.02 (0.01, 0.02)
Thrombolysis	202 (3.8)	515 (11.7)	0.47 (0.42, 0.52)
Intervention, ED	n = 2,195	n = 2,292	Kappa* (CI_{95%})
	n (%)	n (%)	
CT scan	1,822 (78.1)	2,001 (82.4)	0.77 (0.73-0.82)
MRI scan	23 (1.0)	21 (0.9)	0.66 (0.41-0.92)
Carotid imaging	51 (2.2)	347 (14.3)	0.15 (0.09-0.21)
Echocardiogram	–	–	–

*Based on linked records.

Inpatient *n* = 3,624; ED *n* = 1,379.

[†]ICD-10-CA code of any diagnosis type.

CT = computed tomography scan, brain; MRI = magnetic resonance imaging scan, brain; – = suppressed due to small cell count.

Carotid imaging includes carotid catheter angiography, carotid Doppler ultrasound, CT angiography or MR angiography of the carotid artery.

stroke—not specified. In contrast, ED administrative data had a low predictive value for identifying stroke or TIA.

The sensitivity and PPV of the inpatient administrative data were maximized when all stroke types were combined with TIA and appeared in the most responsible diagnosis position (sensitivity = 82.2%, PPV = 68.8%). These findings are consistent with previous studies suggesting that inpatient administrative data can be used to identify patients with stroke.¹⁹⁻²¹ When expanded to include all diagnosis positions, sensitivity for overall stroke and TIA increased to 84.8%, but at the expense of PPV (65.2%), that is, the number of false positive stroke/TIA events increased. Other studies have found that, while PPV was lower when all diagnosis positions were utilized to identify stroke, 20% of valid cases would be missed by focusing on the main diagnosis exclusively.²² Tirschwell et al.²³ found higher sensitivity and PPV when all diagnosis positions were included rather than using the most responsible diagnosis alone; however, their analysis was based on a 1% sample of eligible cases from acute hospitals in Seattle, Washington.

We found that the validity of administrative data for identifying TIA was poor, with PPVs of 49.9% in inpatient and 54.9% in ED administrative data. This is consistent with previous studies¹⁰ that reported PPVs ranging from 28 to 97%. The limited and variable information about TIA validity suggests that caution

is needed when using ICD codes to create a TIA cohort and that one should consider including an active approach for TIA case identification.²¹

Our finding of poor validity of stroke coding in ED administrative data is consistent with the work of Johnsen et al.,²⁴ who found a PPV of 46.7% for TIA and even lower percentages for ischemic stroke, as well as for subarachnoid and intracerebral haemorrhage. This may be related to incomplete clinical investigations and/or documentation in the ED, as well as the challenges involved in selection of the main problem for the ED visit by the health records technician.

We found that the reporting of stroke risk factors in inpatient administrative data was limited, where diabetes was found to be very good ($\kappa=0.83$) and atrial fibrillation good ($\kappa=0.60$). Other important stroke risk factors, such as hypertension and hyperlipidemia, and a key intervention, thrombolysis, were poorly reported. This is in contrast to another Canadian study,¹⁰ where these same risk factors had better kappa agreement than what was found in our study. This discrepancy may be attributable to the specialty training received by the health records technician at the largest of the three participating hospitals, including access to a stroke team for advice in resolving coding issues during the administrative database abstraction process.

There was good agreement between administrative and registry data for identification of brain imaging. However, there was only moderate agreement for the reporting of thrombolysis and poor agreement for the use of carotid imaging and echocardiography. The moderate agreement for thrombolysis is not unexpected, given that specific intervention codes for tPA administered for stroke did not exist during the study period (a dedicated CCI code for tPA was introduced on April 1, 2010). The poor agreement for carotid imaging and echocardiography is likely attributable to the fact that, when these diagnostics are performed on inpatients, the associated costs are absorbed by hospital global budgets and are not captured in the discharge abstract. Although we did not evaluate this in our project, use of other linked administrative data—such as physician billing data—may allow for better identification of inpatient diagnostic procedures.

The validity of administrative data depends in part on the quality of the initial clinical documentation in the medical chart, the training of health records technicians to locate and interpret information, the diagnostic and clinical expertise available, and hospital-specific coding practices. In 2010, directives from the Canadian Stroke Strategy specifically addressed the overuse of code I64.x—“stroke not specified as haemorrhage or infarction.”¹⁶ The directive advised health records technicians to reduce the use of this code since most stroke patients seen in the ED receive brain imaging, allowing strokes to be categorized as ischemic or haemorrhagic. A recent evaluation of all acute hospitals in Ontario found that the prevalence of stroke—not specified among inpatient stroke and TIA patients has almost halved from 16.9% in 2010/2011 to 8.0% in 2012/2013, with a corresponding increase in the reported prevalence of ischemic stroke from 50.7% in 2010/2011 to 59.0% in 2012/2013.²⁵ Other efforts to improve the coding of administrative data include mandated collection of the date and time tPA is administered, an initiative introduced as of fiscal year 2012/2013. As part of the introduction of these new data elements, education workshops for health records technicians are provided with a focus on locating and interpreting chart information.

Some study limitations merit comment. We were unable to calculate specificity or negative predictive value because of the manner in which events were identified in the registry. Only those events presenting at a centre’s ED and suggestive of stroke or TIA were adjudicated, and, as a result, true negatives are not known. Some patients with true positive TIA or mild stroke may also have been missed. Benchimol et al.²⁶ found in their review of administrative data validation studies that the reference standard cohort in many studies did not include patients without disease, precluding the calculation of specificity. In addition, research nurses abstracting for the registry had the option of continuing to complete the chart as new information about the patient became available. Thus, the research nurse may have waited for a diagnostic report that was unavailable at the time of discharge before finalizing the stroke diagnosis in the registry, an option not available to the health records technician abstracting the administrative record. Using an active approach to identify admitted stroke or TIA patients, Piriawat et al.²¹ found that the majority (over 75%) of cases missed were due to admission terms not suggestive of stroke or TIA.

Additionally, our results were based on 2007 and 2008 data and may not reflect contemporary coding practices, diagnostic resources and clinical documentation. Furthermore, the hospitals

participating in the registry are regional referring centres where there are stroke expertise and diagnostic resources, which may limit the generalizability of our findings to other hospital types. To this point, a study²⁷ using primary care electronic medical records as the reference standard to assess the validity of physician claims and hospitalization data to identify prevalent stroke and TIA found that 45% of false positive cases associated with the best algorithm for capturing prevalent stroke/TIA were due to administrative data miscoding. Specifically, patients were coded as having a stroke before the investigation was complete and, when completed, were found not to have suffered a stroke.

Despite these limitations, our study contributes to the growing body of research on the validity of ICD-10-CA-coded stroke and TIA in administrative data and the importance of reporting observational research consistently and transparently to allow for interprovincial/territorial and international comparisons.^{18,19,21,24-28}

CONCLUSION

Routinely collected administrative inpatient data at regional stroke centres in Ontario, Canada, are accurate for identifying inpatients with stroke and TIA combined, and ischemic stroke when combined with stroke of undetermined type. Administrative emergency department data have lower accuracy for identification of stroke and TIA. As advances are made in stroke management and treatment, combined with health record technological improvements and the fact that facility use of administrative databases expands beyond resource utilization to system performance and capacity planning, evaluation of the validity of administrative data for identifying stroke and TIA will need to continue.

ACKNOWLEDGMENTS AND FUNDING

This study was supported by the Ontario Stroke Network (OSN) and the Institute for Clinical Evaluative Sciences (ICES), which are funded by a grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results and conclusions reported herein are those of the authors and are independent of the funding sources. No endorsement by the OSN, the ICES or the MOHLTC is intended or should be inferred. Parts of this work are based on data and information compiled and provided by Canadian Institute for Health Information (CIHI). However, the analyses, conclusions, opinions and statements expressed herein are those of the authors, and not necessarily those of CIHI.

Moira Kapral is supported by a Career Investigator Award from the Heart and Stroke Foundation (Ontario Provincial Office).

DISCLOSURES

Ruth Hall, Luke Mondor, Joan Porter, Jiming Fang and Moira Kapral hereby state that they have nothing to disclose.

REFERENCES

1. Appelros P, Jonsson F, Åsberg S, Asplund K, Glader EL, Åsberg KH, et al. Trends in stroke treatment and outcome between 1995 and 2010: observations from Riks-Stroke, the Swedish stroke register. *Cerebrovasc Dis.* 2014;37(1):22-9. Epub ahead of print Dec 17, 2013.
2. Cadilhac DA, Lannin NA, Anderson CS, Levi CR, Faux S, Price C, et al. Protocol and pilot data for establishing the Australian Stroke Clinical Registry. *Int J Stroke.* 2010;5(3):217-26.

3. Cloud G, Hoffman A, Rudd A, Intercollegiate Stroke Working Party. National sentinel stroke audit 1998–2011. *Clin Med (Lond)*. 2013;13(5):444-8.
4. Fonarow GC, Reeves MJ, Smith EE, Saver JL, Zhao X, Olson DW, et al. Characteristics, performance measures, and in-hospital outcomes of the first one million stroke and transient ischemic attack admissions in get with the guidelines-stroke. *Circ Cardiovasc Qual Outcomes*. 2010;3(3):291-302. Epub ahead of print Feb 22.
5. Heuschmann PU, Biegler MK, Busse O, Elsner S, Grau A, Hasenbein U, et al. Development and implementation of evidence-based indicators for measuring quality of acute stroke care: the Quality Indicator Board of the German Stroke Registers Study Group (ADSR). *Stroke*. 2006;37(10):2573-8.
6. Iguchi Y, Kimura K, Sone K, Miura H, Endo H, Yamagata S, et al. Stroke incidence and usage rate of thrombolysis in a Japanese urban city: the Kurashiki stroke registry. *J Stroke Cerebrovasc Dis*. 2013;22(4):349-57. Epub ahead of print Nov 2.
7. Meretoja A, Roine RO, Kaste M, Linna M, Juntunen M, Erilä T, et al. Stroke monitoring on a national level: PERFECT Stroke, a comprehensive, registry-linkage stroke database in Finland. *Stroke*. 2010;41(10):2239-46.
8. Tu JV, Nardi L, Fang J, Liu J, Khalid L, Johansen H, et al. National trends in rates of death and hospital admissions related to acute myocardial infarction, heart failure and stroke, 1994-2004. *CMAJ*. 2009;180(13):E118-25.
9. Yip PK, Jeng JS, Lee TK, Chang YC, Huang ZS, Ng SK, et al. Subtypes of ischemic stroke: a hospital-based stroke registry in Taiwan (SCAN-IV). *Stroke*. 1997;28(12):2507-12.
10. Andrade SE, Harrold LR, Tjia J, Cutrona SL, Dodd KS, Goldberg RJ, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21(Suppl 1):129-40.
11. Fang J, Kapral MK, Richards J, Robertson A, Stampelcoski M, Silver FL. The Registry of Canadian Stroke Network: an evolving methodology. *Acta Neurol Taiwan*. 2011;20(2):77-84.
12. Kapral MK, Fang J, Silver FL, Hall R, Stampelcoski M, O'Callaghan C, et al. Effect of a provincial system of stroke care delivery on stroke care and outcomes. *CMAJ*. 2013;185(10):E483-91. Epub ahead of print May 27.
13. Hall R, Khan F, O'Callaghan C, Meyer S, Fang J, Hodwitz K. Ontario Stroke Evaluation Report 2011: Improving System Efficiency by Implementing Stroke Best Practices. Toronto: Institute for Clinical Evaluative Sciences; 2011. Available at: <http://www.ices.on.ca/Publications/Atlases-and-Reports/2011/Ontario-Stroke-Evaluation-Report-2011>.
14. Tu JV, Donovan L, Lee DS, Austin P, Wang J, Newman A. Quality of Cardiac Care in Ontario. Toronto: Institute for Clinical Evaluative Sciences; 2004. Available at: <http://www.ices.on.ca/Publications/Atlases-and-Reports/2004/Quality-of-cardiac-care-in-Ontario>.
15. Canadian Institute for Health Information. Hospital Report 2007: Acute Care; 2007. Available at: <https://secure.cihi.ca/estore/productFamily.htm?pf=PFC861&lang=en&media=0>.
16. CSS Information and Evaluation Working Group. Canadian Stroke Strategy Core Performance Indicator, Update 2010; 2010. Available at: http://www.strokebestpractices.ca/wp-content/uploads/2012/07/Stroke_Core_ENG.pdf.
17. Altman DG. *Practical Statistics for Medical Research*. London: Chapman & Hall; 1991.
18. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta Neurol Belg*. 2009;109(2):114-9.
19. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke*. 2005;36(8):1776-81. Epub ahead of print Jul 14.
20. Aboa-Eboule C, Mengue D, Benzenine E, Hommel M, Giroud M, Béjot Y, Quantin C. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. *J Neurol*. 2013;260(2):605-13. Epub ahead of print Oct 18, 2012.
21. Piriyawat P, Smajsova M, Smith MA, Pallegar S, Al-Wabil A, Garcia NM, et al. Comparison of active and passive surveillance for cerebrovascular disease: the Brain Attack Surveillance in Corpus Christi (BASIC) Project. *Am J Epidemiol*. 2002;156(11):1062-9.
22. Thigpen JL, Dillon C, Forster KB, Henault L, Quinn EK, Tripodis Y, et al. Validity of international classification of disease codes to identify ischemic stroke and intracranial hemorrhage among individuals with associated diagnosis of atrial fibrillation. *Circ Cardiovasc Qual Outcomes*. 2015;8(1):8-14. Epub ahead of print Jan 13.
23. Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. *Stroke*. 2002;33(10):2465-70.
24. Johnsen SP, Overvad K, Sørensen HT, Tjønneland 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(6):602-7.
25. Hall R, Khan F, O'Callaghan C, Kapral MK, Cullen A, Levi J, et al. Evaluation Report 2014: On Target for Stroke Prevention and Care. Toronto: Institute for Clinical Evaluative Sciences; 2014. Available at: <http://www.ices.on.ca/Publications/Atlases-and-Reports/2014/Ontario-Stroke-Evaluation-Report-2014>.
26. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011;64(8):821-9. Epub ahead of print Dec 30.
27. Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, et al. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMERALD as a reference standard. *Can J Cardiol*. 2013;29(11):1388-94. Epub ahead of print Sep 26.
28. Bennett DA, Brayne C, Feigin VL, Barker-Collo S, Brainin M, Davis D, et al. Explanation and elaboration of the Standards of Reporting of Neurological Disorders Checklist: a guideline for the reporting of incidence and prevalence studies in neuroepidemiology. *Neuroepidemiology*. 2015;45(2):113-37. Epub ahead of print Sep 22.

APPENDIX 1

Table A1: ICD–10–CA Diagnostic Codes and CCI Intervention Codes Associated with the Assessment and Treatment of Stroke and TIAs

	Code
Stroke type (ICD–10–CA)	
Subarachnoid haemorrhage	I60.x excluding 160.8
Intracerebral haemorrhage	I61.x
Ischemic	H34.1, I63.x excluding 163.6
Transient ischemic attack	G45.x, H34.0 excluding G45.4
Stroke–not specified as haemorrhage or ischemic	164.x
Risk factor[†] (ICD–10–CA)	
Hypertension	I10.x–I13.x, I15.x
Hyperlipidemia	E78.x
Diabetes	E10.x–E14.x
Atrial fibrillation	I48.x
Diagnostic and therapeutic intervention (CCI)	
CT	3AN20, 3ER20
MRI	3AN40, 3ER40
Carotid imaging (catheter angiography, US, CTA, MRA)	3JE10, 3JE20, 3JE30, 3JE40
Echocardiogram	3IP30
Thrombolysis	1xx35HxC1

ICD–10–CA code of any diagnosis type.

CT = computed tomography scan, brain; MRI = magnetic resonance imaging scan, brain; US = carotid Doppler ultrasound; CTA = computed tomography angiography of carotid artery; MRA = magnetic resonance angiography of carotid artery.