Epilepsia, 51(1):62–69, 2010
doi: 10.1111/j.1528-1167.2009.02201.x

FULL-LENGTH ORIGINAL RESEARCH

How accurate is ICD coding for epilepsy?

*†Nathalie Jetté, *Aylin Y. Reid, †‡Hude Quan, *†Michael D. Hill, and *†Samuel Wiebe

*Division of Neurology, Department of Clinical Neurosciences, †Department of Community Health Sciences, and ‡Centre for Health and Policy Studies, University of Calgary, Alberta, Canada

Summary

Purpose: Assess the validity of ICD-9-CM and ICD-10 epilepsy coding from an emergency visit (ER) and a hospital discharge abstract database (DAD).
Methods: Two separate sources of patient records were reviewed and validated. (1) Charts of patients admitted to our seizure monitoring unit over 2 years (n = 127, ICD-10 coded records) were reviewed. Sensitivity (Sn), specificity (Sp), and positive and negative predictive values (PPV and NPV) were calculated. (2) Random sample of charts for patients seen in the ER or admitted to hospital under any services, and whose charts were coded with epilepsy or an epilepsy-like condition, were reviewed. Two time-periods were selected to allow validation of both ICD-9-CM (n = 486) and ICD-10 coded (n = 454) records. Only PPV and NPV were calculated for these records. All charts were reviewed by two physicians to confirm the presence/absence of epilepsy and compare to administrative coding.
Results: Sample 1: Sn, Sp, PPV, and NPV of ICD-10 epilepsy coding from the seizure monitoring unit (SMU) chart review were 99%, 70%, 85%, and 97% respectively. Sample 2: The PPV and NPV for ICD-9-CM coding from the ER database were, respectively, 99% and 97% and from the DAD were 98% and 99%. The PPV and NPV for ICD-10 coding from the ER database were, respectively, 100% and 90% and from the DAD were 98% and 99%. The epilepsy subtypes grand mal status and partial epilepsy with complex partial seizures both had PPVs >75% (ICD-9-CM and ICD-10 data).
Discussion: Administrative emergency and hospital discharge data have high epilepsy coding validity overall in our health region.
KEY WORDS: Surveillance, Administrative data, International classification of diseases, Validation.

Epilepsy is one of the most commonly reported neurologic conditions in primary health care after migraine, and accounts for 1% of the global burden of disease (Murray et al., 1994). In Canada, 16,000 people are diagnosed with epilepsy every year, and at any given time approximately 200,000 people have active epilepsy requiring medical attention (Tellez-Zenteno et al., 2004). It is expected that the prevalence of epilepsy will increase with aging populations. It is thus necessary to develop cost-effective and timely surveillance programs using newly validated tools to monitor and project future social and clinical demands, and management and outcomes of epilepsy.

Current and past surveillance programs for epilepsy have relied primarily on random telephone, door-to-door, or mailed population-based surveys (Wiebe et al., 1999; CDC, 2005). Prospective epilepsy surveillance programs are extremely rare, primarily because of the high cost of active surveillance (Olafsson et al., 2005). In recent years, administrative data have become a highly sought after source of data for passive disease surveillance, assessment of health resource utilization, and the evaluation of healthcare outcomes (Wennberg & Gittelsohn, 1973; May et al., 1991; Virnig & McBean, 2001; Kokotailo & Hill, 2005; Jetté et al., 2008). The benefits of administrative databases include their large population sizes, their cost-effectiveness, and often their rich content of historical population-based information (Deyo et al., 1994; Mitchell et al., 1994; Iezzoni, 1997). This is particularly relevant in the Canadian context, in which centralized care and single-payer systems have resulted in population-based administrative databases at regional, provincial, and national levels.

Currently and since 2002, inpatient facilities across Canada use a slight modification of the World Health

Accepted May 14, 2009; Early View publication July 20, 2009.
Address correspondence to Nathalie Jetté, Foothills Medical Centre, 1403-29th Street NW, Calgary, AB, Canada T2N 2T9. E-mail: njette@ucalgary.ca

Organization (WHO) International Classification of Diseases (ICD) system, ICD-10-CA, to code inpatient visits (CIHI 2008). There are no significant differences in epilepsy codes between the ICD-10 and ICD-10-CA systems at the third or fourth digit/character level; therefore, the term ICD-10 is used for the remainder of the article. ICD-9-CM and ICD-9 are still the primary systems used for coding mortality and morbidity in many countries including Pakistan, India, and others (WHO, 2008). Although ICD-10 is used for mortality coding in the United States, ICD-9-CM is still in use for morbidity coding in the United States (WHO, 2008).

Validation of ICD-coded data from hospital and emergency room (ER) discharge abstract databases is one of the first steps in developing a valid and accurate case definition for a condition of interest for future surveillance. Some fatal conditions, such as cancer, are coded more accurately than nonspecific conditions due largely to pathologic confirmation of the diagnosis (Whittle et al., 1991; Tennis et al., 1993). One group in the United States validated epilepsy ICD-9-CM codes from administrative data (Holden et al., 2005a, 2005b). This group used data from a U.S. Managed Care Organization. They developed a computer algorithm to identify epilepsy cases in this database. The best model correctly identified 90% of cases.

The objectives of the present study were: (1) to validate epilepsy ICD-9-CM and ICD-10 coding from an inpatient and an emergency discharge abstract database; and (2) to compare variations in coding validity between the two ICD systems and between the various hospital settings in a large Canadian health region.

Methods

Study sites

All adult and pediatric acute care sites in the Calgary Health Region in Alberta, Canada were included in the study. These sites serve a population of approximately 1.4 million people.

Administrative data sources

Three databases were used to identify our population: (1) a seizure monitoring unit (SMU) database; (2) the Ambulatory Care Classification System database (ACCS), and (3) the inpatient discharge abstract database (DAD). The SMU database captures all patients who were referred and admitted to the unit for diagnosis and treatment in the region. The ACCS database captures all ER and day procedure visits, and the DAD captures all hospitalizations. Because the DAD implies an inpatient stay with the associated neurologic workup and assessment, it has a greater depth of information compared to the ACCS database. Only ER visits were selected from the ACCS database. All three databases record demographic information (age, sex, and residence), diagnosis, procedure, and institution, as well as other information. These databases can be linked using the Alberta Health Care Insurance Plan Registry (AHCIP), which creates a unique personal health number for all subjects with healthcare coverage in the province of Alberta and captures more than 98% of the population. Coding of diagnoses in ACCS and DAD using ICD-9-CM (until 2002) and ICD-10 (since 2002) codes is routinely done by trained professional coders at each hospital.

Chart samples for validation

Ideally, a sample should be randomly or sequentially selected from all ACCS and DAD records, and then the presence or absence of epilepsy would be determined through a ''gold standard,'' which would usually be physician chart review for a coding validation study such as ours. However, because the prevalence of epilepsy in patients admitted to the ER or hospital is not high, it is not feasible to review a large volume of charts to set up a ''gold standard.'' Therefore, we used two different samples of patients in order to avoid reviewing a very large (>10,000) number of charts from the DAD and ACCS databases (Fig. 1).

Sample 1: Chart review of patients who were admitted to the seizure monitoring unit

All patients (n = 127) admitted to the SMU between 2005 and 2007 were identified, and their charts were reviewed by two physicians (NJ and AR) to confirm the presence or absence of epilepsy. Patients admitted to the SMU included patients with true epilepsy and those with nonepileptic conditions such as psychogenic seizures and syncope. Because all patients in the SMU underwent video-EEG (electroencephalography) monitoring, the certainty of their diagnosis is high. Chart review from the SMU admissions allowed us to calculate sensitivities (Sn), specificities (Sp), and negative and positive predictive values (NPVs and PPVs)—see statistical methods in subsequent text. These SMU hospitalizations are captured and coded in the DAD with ICD-10 codes.

Sample 2: Chart review of randomly selected patients from ACCS and DAD

For ICD-9-CM coded ACCS and DAD data, we selected records for patients of all ages for the 2000 fiscal year (April 1, 2000 through March 30, 2001) with the epilepsy ICD-9-CM code (345.x) in the primary diagnostic position as one stratum. Nonepilepsy records of diagnoses that may resemble epilepsy were also selected, including classical migraine (ICD-9-CM code 346.x), transient cerebral ischemia (TIA) (ICD-9-CM code 435.x), syncope (ICD-9-CM code 780.2), or convulsion (ICD-9-CM code 780.3, which in Canada is

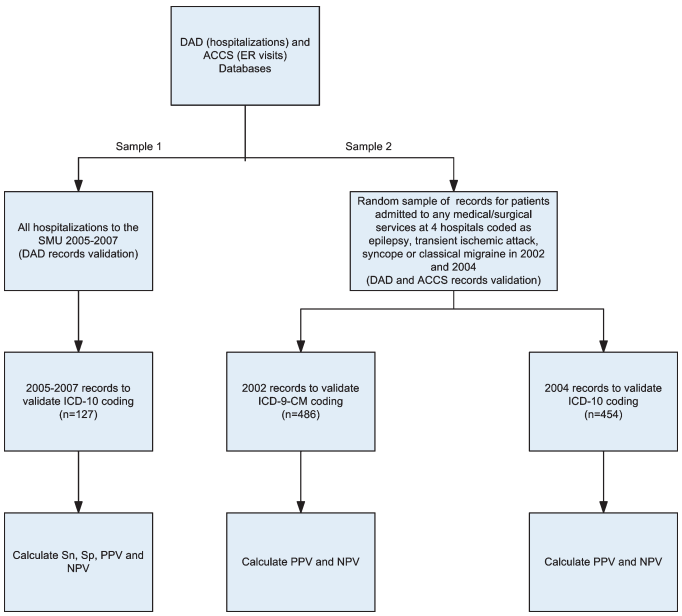

Figure 1.
Diagram outlining the two different sample sources of charts reviewed and respective statistical analyses done.
Epilepsia © ILAE

defined as general convulsions that are not associated with epilepsy). A random sample of records (approximately 10% based on a sample calculation) was selected from the epilepsy and nonepilepsy groups and each database, resulting in 486 records.

For ICD-10 coded ACCS and DAD data, we selected records for patients of all ages for the 2004 fiscal year (April 1, 2004 through March 30, 2005) with the epilepsy ICD-10 code (G40.x or G41.x) in the primary diagnostic position. Nonepilepsy records of conditions that may resemble epilepsy were also identified, including classical migraine (ICD-10 code G43.1), transient cerebral ischemia (TIA) (ICD-10 code G45.x), syncope (ICD-10 code R55), or convulsion (ICD-10 code R56.0 or R56.8, which are intended to be used for organic convulsions but not for epilepsy). Again, a sample was randomly selected from each group (epilepsy and conditions resembling epilepsy) in each database (ACCS and DAD) resulting in 454 records. These enriched populations from the ACCS and DAD databases, including patients with diagnoses other than epilepsy, were chosen specifically because these conditions can resemble seizures.

Charts from the randomly selected records of epilepsy and conditions resembling epilepsy (in total 940) from the ACCS and DAD, described previously, were identified using a combination of patient chart number and date of admission. An epileptologist (NJ) and a neurology resident (AR) reviewed each chart. Before formal chart review, the two reviewers (NJ and AR) independently assigned an ICD code to a sample of 20 charts. The discrepancies between the two reviewers were examined and resolved by discussion. This sample chart review was done to assess the appropriateness of the data extraction forms and to identify areas of possible disagreement among reviewers. Chart review of the randomly selected epilepsy and epilepsy-like conditions records from the ACCS and DAD allowed us to calculate NPV and PPV (see statistical methods in subsequent text).

The two reviewers (NJ and AR) then read all charts independently (selected from all databases), reviewing all notes from the physician history and physical examination, physician and nursing progress notes, paramedic notes, imaging reports (if available), electroencephalography (EEG) or SMU reports (if available), and any other pertinent investigations (e.g., electrocardiography). On the basis of this information, the two reviewers (NJ and AR) assigned the presence or absence of epilepsy, transient ischemic attack (TIA), migraine, syncope, convulsion to each chart (all databases) using ICD-9-CM or ICD-10 codes, depending on the year (ICD-9-CM for the year 2000, ICD-10 for year the 2004). The original ICD-9-CM or ICD-10 codes assigned by professional coders were unknown (blinded) to the reviewers. If the reviewers (NJ or AR) felt that the diagnosis was different from those listed above, then the chart was coded as ''other condition.'' Approval for this study was obtained from the local ethics board.

Statistical analysis

We described characteristics of study populations and then calculated Sn, Sp, PPV, and NPV and their 95% confidence intervals (95% CIs). Sn refers to the proportion of epilepsy cases in the SMU charts that have been recorded

as epilepsy cases in the DAD. Sp refers to the proportion of nonepilepsy cases identified from the SMU charts that were recorded in the DAD as nonepilepsy cases. The PPV refers to the proportion of epilepsy cases identified in administrative databases (ACCS and DAD) that were deemed ''true'' epilepsy cases on the basis of chart review data. The NPV refers to the proportion of nonepilepsy cases in administrative databases (ACCS and DAD) that were deemed ''true'' nonepilepsy cases based on chart review. The PPV was also calculated for the epilepsy subtype codes, referring to the proportion of each subtype of epilepsy (the fourth digit of the code) identified in the databases that were deemed ''true'' cases on the basis of chart review data.

Results

PPV and NPV for ICD-9-CM ACCS and DAD

A total of 486 charts were reviewed: 277 ER charts and 209 inpatient charts. The number of charts per hospital and subject demographics is presented in Table 1. The mean age was 41.9 (standard deviation, SD = 24.4) years, and 51% were female.

The PPV and NPV for records using ICD-9-CM codes are presented in Table 2. The PPV for epilepsy codes (code 345) in the ACCS was 98.9% (95% CI 94.2%–99.8%), in the DAD it was 97.7% (95% CI 92.1%–99.4%), and the overall PPV for both databases combined was 98.9% (95% CI 96.1%–99.7%). The NPV in the ACCS was 97.0% (95% CI 93.2%–98.7%), in the DAD it was 99.1% (95% CI 94.8%–99.8%), and the overall NPV for both databases combined was 97.4% (95% CI 94.8%–98.8%). There were no significant differences between sites (adult or children) or by visit type (ER or hospitalization) as shown in Table 2 under ICD-9-CM. When the convulsion code was added to the ICD-9-CM records (to determine if any of these nonepilepsy but organic convulsions were miscoded as epilepsy), the PPV dropped (Table 2 ICD-9-CM) to 84.0% (95% CI 78.4%–88.3%).

The epilepsy subtype (four-digit code) PPV was >75% for only grand mal status (83.9%, code 345.3) and partial epilepsy with impairment of consciousness (89.3%, code 345.4) (Table 3).

PPV and NPV for ICD-10 ACCS and DAD

A total of 454 charts were reviewed: 207 ER charts and 247 inpatient charts. The number of charts per hospital and demographic characteristics are presented in Table 1. The mean age was 43.3 (SD = 25.2) years, and 47% were females.

The PPV for epilepsy codes (G40-41) in the ACCS was 100% (95% CI 93.2%–100%), for the DAD it was 97.8% (95% CI 92.3%–99.4%), and the overall PPV for the ACCS and DAD combined was 98.6% (95% CI 95.1%–99.6%) (Table 2). The NPV for the ACCS was 89.5% (95% CI 83.1%–93.6%), for the DAD it was 98.5% (95% CI 94.7%–99.6%), and the overall NPV for the two databases combined was 94.0% (95% CI 90.5%–96.3%) (Table 2). There were no significant differences between hospital sites (adult or children) or by visit type (ER or hospitalization). When the convulsion code was added to the ICD-10 records (to determine if any of these nonepilepsy but organic convulsions were miscoded as epilepsy), the PPV dropped to 75.5% (95% CI 68.9%–81.1%) for the ACCS and DAD combined (Table 2).

The epilepsy subtype (four-digit code) PPV was >75% for the following only: localization-related symptomatic epilepsy and epileptic syndromes with complex partial seizures (78.3%, code G40.2), grand mal status epilepticus (100%, code G41.0), complex partial status epilepticus (83.3%, code G41.2), and febrile convulsions (90.9%, code R56.0) (Table 3).

Sn, Sp, PPV, and NPV for ICD-10–coded SMU hospitalization records

A total of 127 charts, comprising all inpatient charts for patients admitted to the SMU over a 2-year period (2005–2007) were reviewed. The mean age was 37.1 (SD = 11.9) years. Sn and Sp for the ICD-10 epilepsy codes (codes G40-41) were 98.8% (95% CI 93.3%–99.8%) and 69.6% (95% CI 55.2%–80.9%), respectively, whereas PPV and NPV were 84.9% (76.3–90.8) and 97.1% (85.1–99.5) respectively (Table 2).

Table 1. Number of charts reviewed by site and patient demographics (Total = 1,067)

	SMU database	ACCS and DAD databases (year 2000, ICD-9-CM)			ACCS and DAD databases (year 2004, ICD-10)		
	Inpatient only	ER	Inpatient	All	ER	Inpatient	All
Total charts	127	277	209	486	207	247	454
Adult hospitals	127	226	179	405	160	217	377
Children's hospital	0	51	30	81	47	30	77
Mean age (SD), years	37.13 (±11.9)	39.4 (±24.2)	45.3 (±24.4)	41.9 (±24.4)	33.8 (±22.4)	49.9 (±24.0)	43.3 (±25.3)

ACCS, acute care clinical classification system; DAD, discharge abstract database; ER, emergency room; ICD, International Classification of Diseases and Related Disorders; SD, standard deviation; SMU, seizure monitoring unit.

Table 2. Validity of ICD epilepsy coding in administrative databases

	Sample 1: Validity of records from SMU admissions (coded in DAD)				Sample 2: Validity of records from ER visits and hospitalizations to any medical/surgical services (coded in DAD and ACCS databases)							
	ICD-10				ICD-9-CM		ICD-10		ICD-9-CM		ICD-10	
					Positive predictive value % (95% CI)		Positive predictive value % (95% CI)					
	Sensitivity % (95% CI)	Specificity % (95% CI)	Positive predictive value % (95% CI)	Negative predictive value % (95% CI)	Epilepsy codes only (345)	Epilepsy plus convulsion code (345 + 780.3)	Epilepsy codes only (G40-41)	Epilepsy plus convulsion code (G40-41 + R56)	Negative predictive value % (95% CI)		Negative predictive value % (95% CI)	
All charts	98.8 (93.3–99.8)	69.6 (55.2–80.9)	84.9 (76.3–90.8)	97.1 (85.1–99.5)	98.9 (96.1–99.7)	84.0 (78.4–88.3)	98.6 (95.1–99.6)	75.5 (68.9–81.1)	97.4 (94.8–98.8)		94.0 (90.5–96.3)	
Inpatient visits	98.8 (93.3–99.8)	69.6 (55.2–80.9)	84.9 (76.3–90.8)	97.1 (85.1–99.5)	97.7 (92.1–99.4)	83.5 (75.1–89.4)	97.8 (92.3–99.4)	78.1 (69.6–84.7)	99.1 (94.8–99.8)		98.5 (94.7–99.6)	
Emergency visits	n/a	n/a	n/a	n/a	98.9 (94.2–99.8)	83.6 (75.6–89.4)	100 (93.2–100)	71.6 (60.5–80.6)	97.0 (93.2–98.7)		89.5 (83.1–93.6)	
Adult hospital visits	98.8 (93.3–99.8)	69.6 (55.2–80.9)	84.9 (76.3–90.8)	97.1 (85.1–99.5)	98.4 (94.4–99.6)	79.4 (72.3–85.0)	98.1 (93.3–100)	72.3 (64.4–79.1)	97.6 (94.9–98.9)		97.0 (94.0–98.6)	
Children's hospital visits	n/a	n/a	n/a	n/a	100 (93.5–100)	96.5 (88.1–99.0)	100 (91.2–100)	85.1 (72.3–92.6)	95.8 (79.8–99.3)		70.0 (52.1–83.3)	

Note: Nonepilepsy refers to conditions that may resemble seizures such as syncope, transient ischemic attack, classical migraine, psychogenic seizures. ICD-10 epilepsy codes refer to codes G40.XX-G41.XX, whereas ICD-9-CM codes refer to codes 345.XX.

Databases: ACCS, ambulatory care classification system database; DAD, discharge abstract database; CA, Canadian modification; CI, confidence interval; CM, clinical modification; ER, emergency room; ICD, International Classification of Diseases and Related Disorders; SMU, seizure monitoring unit.

Discussion

This study used chart data to assess the accuracy and validity of ICD-9-CM and ICD-10 administrative data for coding epilepsy. Our results show that inpatient and ER administrative databases have a very high validity at identifying a cohort of individuals with epilepsy. Performance between ICD-9-CM and ICD-10 coding was very similar. However, some epilepsy cases are likely to be missed due to miscoding as a convulsion (organic, but not epileptic).

Epilepsy coding is similar in the ICD-9-CM versus the ICD-10 coded databases

The ICD-9-CM and the ICD-10 ACCS and DAD combined data had similar PPVs (98.9% and 98.6%, respectively). The ICD-9-CM system was in use for a much longer period than the ICD-10 system at the time the charts were professionally coded, and the coders were more familiar with the ICD-9-CM system. However, the ICD-10 system had already been in use for 2 years by the time the charts in our study were coded by the professional coders. We would assume this 2-year period would be sufficient to ensure coder's familiarity with the ICD-10 epilepsy codes. It is very encouraging that the coding for ICD-9-CM and ICD-10 data was so accurate, as there are large databases nationally with patient information using both of these coding systems. Information from these databases is valuable for studying epilepsy, as patients captured in these databases will have long follow-up periods.

Epilepsy coding from the ACCS and DAD is improved when codes for convulsions (excludes epileptic convulsions) are excluded

Epilepsy ICD-9-CM and ICD-10 coding by professional coders for ER and inpatient visits is good when looking at other mimics of epilepsy or seizures, such as migraine, TIA, syncope, or convulsions. However, agreements are much better when codes for convulsions are excluded. The ''convulsions'' code is supposed to be used only in situations where a seizure has not occurred as a result of epilepsy. Therefore, epileptic convulsions should not be coded with the general ''convulsions'' code. In our study, many cases were assigned an epilepsy code by the epileptologist and neurology resident, but were coded as ''convulsions'' by the professional coder. Therefore, one would not be able to capture all patients with epilepsy from an inpatient or ER administrative database using only the ''true'' epilepsy codes, as many would be missed due to miscoding as general ''convulsions'' (which are organic but not deemed to be associated with epilepsy). However, the epilepsy cases that would be identified by the databases would indeed be capturing epilepsy patients, due to the high specificities, even with convulsions codes included. We validated the regionally used ACCS

Table 3. Positive predictive value (PPV) for ICD-9-CM and ICD-10 epilepsy and convulsion subcodes

Condition	Code	Definition	Positive predictive value % (95% CI)
Epilepsy ICD-9-CM	345.0	Generalized nonconvulsive epilepsy	n/a
	345.1	Generalized convulsive epilepsy	17.1 (8.1–32.7)
	345.2	Petit mal status	n/a
	345.3	Grand mal status	83.9 (67.4–92.9)
	345.4	Partial epilepsy, with impairment of consciousness	89.3 (72.8–96.3)
	345.5	Partial epilepsy, without mention of impairment of consciousness	31.3 (14.2–55.6)
	345.6	Infantile spasms	n/a
	345.7	Epilepsia partialis continua	n/a
	345.8	Other forms of epilepsy	n/a
	345.9	Epilepsy, unspecified	51.7 (39.3–63.8)
Convulsions ICD-9-CM	780.3	Convulsions, excluding epileptic convulsions and convulsions of newborn	45.2 (33.4–57.5)
Epilepsy ICD-10	G40.0	Localization-related idiopathic epilepsy and epileptic syndromes	n/a
	G40.1	Localization-related symptomatic epilepsy and epileptic syndromes with simple partial seizures	18.8 (6.6–43.0)
	G40.2	Localization-related symptomatic epilepsy and epileptic syndromes with complex partial seizures	77.6 (64.1–87.0)
	G40.3	Generalized idiopathic epilepsy	33.3 (9.7–70.0)
	G40.4	Other generalized epilepsy	n/a
	G40.5	Special epileptic syndromes	11.1 (2.0–43.5)
	G40.6	Grand mal seizures, unspecified	37.5 (21.2–57.3)
	G40.7	Petit mal, unspecified	n/a
	G40.8	Other epilepsy	n/a
	G40.9	Epilepsy unspecific	23.1 (8.2–50.3)
	G41.0	Grand mal status epilepticus	100 (43.8–100)
	G41.1	Petit mal status epilepticus	n/a
	G41.2	Complex partial status epilepticus	83.3 (43.6–97.0)
	G41.8	Other status epilepticus	20.0 (3.6–62.4)
	G41.9	Status epilepticus, unspecified	6.3 (1.1–28.3)
Convulsions ICD-10	R56.0	Febrile convulsions	90.9 (62.3–98.4)
	R56.8	Other and unspecified convulsions	32.9 (23.0–44.5)

CI, confidence interval; CM clinical modifications; ICD, International Classification of Diseases and Related Disorders; n/a, not available.

(emergency) database as well as the hospital discharge abstract database (inpatient), which is used in all Canadian hospitals. Our findings support the notion that population-based hospital discharge administrative data can be used to accurately identify an epilepsy cohort, but in order to capture all of those with epilepsy, the general convulsion code would need to be included in the final case definition, with appropriate adjustment for the prevalence of epilepsy cases coded with this code.

Accuracy of epilepsy coded in administrative data is high across all hospital settings

We showed that accuracy of epilepsy coding by professional coders is high, whether the encounter (patient visit) was an ER visit or a hospitalization, and whether it was a pediatric or an adult site. This may be related to the fact that coders in these settings were trained and supervised by one manager and that they were required to follow the same coding guidelines throughout the health region. The management style might contribute to the high accuracy across various settings. Although the validated databases were used for patients in only one health region, all professional coders in Canada undergo standardized training. We would expect similar findings across the country. However, although the coders have the same coding training they may be presented with different standards of charting from the physicians in their hospital, which may or may not enable them to come to the appropriate code at the same rate in all parts of the country. One surprising finding was the lower Sp (with false positive rate of 30%) of the SMU codes compared to the non-SMU records reviewed. This is very likely due to the fact that patients without epilepsy in the SMU have conditions that mimic epilepsy extremely closely, or have both epilepsy and nonepilepsy events, making it difficult for coders to be accurate. The SMU is a highly select and diagnostically difficult population.

Overall epilepsy coding had a high validity, but validity of epilepsy subtypes was not optimal

There was a high level of diagnostic accuracy using the first three-digit epilepsy codes. Overall, agreements between the professional coder and the epileptologist were very good, and often excellent. However, this will only

identify a patient as having a diagnosis of epilepsy. There was lower agreement for the four-digit codes, which specify epilepsy subgroups such as partial versus generalized epilepsy. This was likely due to many factors, such as lack of medical knowledge about seizure types by the professional coders, as well as incomplete information in the patient charts. The epileptologist was often able to provide a more specific epilepsy subgroup diagnosis compared to an ER physician or a general neurologist. Overall, the four-digit unadjusted agreement between the epileptologist and professional coder was better for the ICD-9-CM system, where there are only 10 epilepsy codes, than for the ICD-10 system, where there are 16 epilepsy codes. However, the PPV for epilepsy subtype was >75% for only two ICD-9-CM codes, and for four ICD-10 codes. Both systems were very good at identifying grand mal status and complex partial seizures, whereas the ICD-10 system was also very good at identifying complex partial status and febrile seizures. These databases can now be used to accurately identify patients with a general diagnosis of epilepsy, or one of grand mal status or complex partial seizures, but would not be able to accurately identify all patients with other subtypes of epilepsy.

Limitations to our work

First, our largest sample (data from the DAD and ACCS databases) was not randomly selected from the general population. Accordingly, our stratified method generated a higher prevalence of epilepsy than known population prevalence rates in Canada (Tellez-Zenteno et al., 2004, 2005). However, our findings of high levels of Sn and Sp in the entire SMU population lend support to the validity of the PPV and NPV derived from the stratified sample, even after accounting for the prevalence overestimation. Secondly, we assessed only ER and inpatient databases. A majority of epilepsy cases are treated in outpatient settings and would only be captured in the physician billing claims database. Generalizing our findings to outpatient databases (such as physician billings database) should not be done. This is predominantly because those individuals coding records from the Canadian inpatient databases (DAD and ACCS) are professionally trained, and it is thus presumed that coding of inpatient records will be of high quality nationally. However, there is a need to validate databases that capture outpatient visits such as the physician claims database, which are not coded by professional coders to ensure that data quality is also adequate. Third, our ''gold standard'' relied solely on chart documentation. However, when admitted to the ER department or in hospital, patients tend to have more extensive work-ups. Therefore, we believe that the charts most likely reflected the ''true diagnosis.'' Fourth, our study was conducted in one health region. Generalizability of the findings to the remaining Canadian geographic areas or to other countries is uncertain. Although the validated databases were used only for patients in the health region, all professional coders in Canada undergo standardized training. However, although the coders have the same coding training, they may be presented with different standards of charting from the physicians in their hospital, which may or may not enable them to come to the appropriate code at the same rate in all parts of the country. As has been documented in stroke, settings such as urban versus rural may have more variability in charting and coding practices (Yiannakoulias et al., 2003).

In conclusion, ICD-9-CM or ICD-10 coding for the identification of those with epilepsy who visit the emergency room or who are admitted in hospital in our health region is highly accurate. Coding validity improves greatly when convulsion codes (excluding epilepsy related convulsions) are excluded. Therefore, if one wanted to identify a cohort of individuals with epilepsy, the ''true'' epilepsy codes could be used with very high specificity. However, if one wanted to develop a surveillance program, the more general convulsion code would have to be included, and adjustment would be required for the small sample of subjects coded with this convulsion code who do not necessarily have epilepsy. Now that these databases have been validated they can be used to answer many questions about health care in hospitalized (inpatient or ER) individuals with epilepsy. The development of passive surveillance programs for epilepsy may be possible in the future but will require validation of the physician billing claims database to ensure accuracy of coding of outpatient visits and certainty that most of those with epilepsy are captured in a population-based fashion.

Acknowledgments

This study was fully supported by grants from the MSI Foundation and Alberta Heritage Foundation for Medical Research to Dr Samuel Wiebe. Dr Michael Hill is funded by the Alberta Heritage Foundation for Medical Research and the Heart & Stroke Foundation of Alberta/NWT/NU. Dr Nathalie Jetté is supported by a Population Health Investigator Award from the Alberta Heritage Foundation for Medical Research and by a New Investigator Award from the Canadian Institutes of Health Research Canada. Dr Hude Quan is supported by a Population Health Investigator Award from the Alberta Heritage Foundation for Medical Research and by a New Investigator Award from the Canadian Institutes of Health Research Canada.

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Disclosure: None of the authors has any conflict of interest to disclose.

References

CDC. (2005) Prevalence of epilepsy and health-related quality of life and disability among adults with epilepsy–South Carolina, 2003 and 2004. *MMWR Morb Mortal Wkly Rep* 54:1080–1082.

CIHI (2008) *ICD-10-CA*. Canadian Institute for Health Information, Ottawa.

Deyo RA, Taylor VM, Diehr P, Conrad D, Cherkin DC, Ciol M, Kreuter W. (1994) Analysis of automated administrative and survey databases to study patterns and outcomes of care. *Spine* 19:2083S–2091S.

Holden EW, Grossman E, Nguyen HT, Gunter MJ, Grebosky B, Von Worley A, Nelson L, Robinson S, Thurman DJ. (2005a) Developing a computer algorithm to identify epilepsy cases in managed care organizations. *Dis Manag* 8:1–14.

Holden EW, Thanh Nguyen H, Grossman E, Robinson S, Nelson LS, Gunter MJ, Von Worley A, Thurman DJ. (2005b) Estimating prevalence, incidence, and disease-related mortality for patients with epilepsy in managed care organizations. *Epilepsia* 46:311–319.

Iezzoni LI. (1997) Assessing quality using administrative data. *Ann Intern Med* 127:666–674.

Jetté N, Quan H, Faris P, Dean S, Li B, Fong A, Wiebe S. (2008) Health resource use in epilepsy: significant disparities by age, gender, and aboriginal status. *Epilepsia* 49:586–593.

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

May DS, Kelly JJ, Mendlein JM, Garbe PL. (1991) Surveillance of major causes of hospitalization among the elderly, 1988. *MMWR CDC Surveill Summ* 40:7–21.

Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Muhlbaier LH, Wiesman JM, Young WW, Epstein RS, Javitt JC. (1994) Using Medicare claims for outcomes research. *Med Care* 32:JS38–JS51.

Murray CJ, Lopez AD, Jamison DT. (1994) The global burden of disease in 1990: summary results, sensitivity analysis and future directions. *Bull World Health Organ* 72:495–509.

Olafsson E, Ludvigsson P, Gudmundsson G, Hesdorffer D, Kjartansson O, Hauser WA. (2005) Incidence of unprovoked seizures and epilepsy in Iceland and assessment of the epilepsy syndrome classification: a prospective study. *Lancet Neurol* 4:627–634.

Tellez-Zenteno JF, Pondal-Sordo M, Matijevic S, Wiebe S. (2004) National and regional prevalence of self-reported epilepsy in Canada. *Epilepsia* 45:1623–1629.

Tellez-Zenteno JF, Matijevic S, Wiebe S. (2005) Somatic comorbidity of epilepsy in the general population in Canada. *Epilepsia* 46:1955–1962.

Tennis P, Bombardier C, Malcolm E, Downey W. (1993) Validity of rheumatoid arthritis diagnoses listed in the Saskatchewan Hospital Separations Database. *J Clin Epidemiol* 46:675–683.

Virnig BA, McBean M. (2001) Administrative data for public health surveillance and planning. *Annu Rev Public Health* 22:213–230.

Wennberg J, Gittelsohn A. (1973) Small area variations in health care delivery. *Science* 182:1102–1108.

Whittle J, Steinberg EP, Anderson GF, Herbert R. (1991) Accuracy of Medicare claims data for estimation of cancer incidence and resection rates among elderly Americans. *Med Care* 29:1226–1236.

WHO (2008) *ICD Implementation by Countries World Health Organization*.

Wiebe S, Bellhouse DR, Fallahay C, Eliasziw M. (1999) Burden of epilepsy: the Ontario Health Survey. *Can J Neurol Sci* 26:263–270.

Yiannakoulias N, Svenson LW, Hill MD, Schopflocher DP, James RC, Wielgosz AT, Noseworthy TW. (2003) Regional comparisons of inpatient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. *Chronic Dis Can* 24:9–16.