Acta Neurol. Belg., 2009, **109**, 114-119

The Accuracy of Hospital Discharge Coding for Hemorrhagic Stroke

Matthew Anthony KIRKMAN, Wattana MAHATTANAKUL, Barbara A. GREGSON and A. David MENDELOW
Department of Neurosurgery, Regional Neurosciences Centre, Newcastle General Hospital, Newcastle upon Tyne, NE4 6BE, United Kingdom

Abstract

Background: Hospital discharge data is used in monitoring stroke epidemiology, and ensuring adequate resource allocation to treatment programs. Previous studies have reported variable accuracy levels for such data. We present the first study assessing the accuracy of International Classification of Diseases 10th Edition (ICD-10) discharge coding for hemorrhagic stroke in England.

Methods: We identified all patients with a primary diagnosis of intracerebral hemorrhage (ICH; ICD-10 code: I61.x) and subarachnoid hemorrhage (SAH; I60.x) admitted to the Newcastle upon Tyne Hospitals from 2002-2007. Positive predictive values (PPV) were calculated through validation with patient notes.

Results: Hospital discharge coding identified 978 ICH and 1169 SAH admissions over the six years. The number of diagnoses verified by patient notes was excellent for both ICH (n = 938) and SAH (n = 1123), with a PPV of 95.9% for ICH (95% confidence interval, CI = 94.5-97.0%) and 96.1% for SAH (95% CI = 94.8-97.0%). The coding errors observed were largely expected, with different types of stroke miscoded as ICH and SAH.

Conclusions: The accuracy of ICD-10 hospital discharge coding for hemorrhagic stroke was excellent. However, further research is needed to find ways to further improve its accuracy.

Key words: Epidemiology of stroke; intracerebral hemorrhage; subarachnoid hemorrhage; ICD-10; coding accuracy; administrative data.

Introduction

Stroke is the second leading cause of death worldwide and a major cause of long-term disability in survivors, with a marked socioeconomic and emotional impact on both patients and families (Feigin *et al.*, 2003). The two forms of haemorrhagic stroke - intracerebral haemorrhage (ICH) and subarachnoid haemorrhage (SAH) - are significant causes of morbidity and mortality; ICH affects 10-20 per 100,000 of the population annually and accounts for 10-20% of all stroke-related neurological deficits (Broderick *et al.*, 1999). Although SAH is less common than ICH, affecting around 10 per 100,000, it has high public health relevance since it often affects the young and middle aged (Feigin *et al.*, 2003). ICH and SAH are associated with significant one-month mortality (42% and 32%, respectively) (Feigin *et al.*, 2003). Over the past twenty years or so, the diagnostic accuracy of stroke and its subtypes has improved dramatically; before the advent of computed tomography (CT), all cerebrovascular diseases were grouped together as a single entity, affecting the quality of early research into risk factors and epidemiology. Monitoring trends in stroke rates is important in informing decision making about resource allocation to preventive and acute treatment programs, making accurate stroke epidemiology data paramount in importance.

Broadly speaking, the epidemiology of stroke can be measured through population-based stroke registers, or through hospital discharge data and mortality statistics. The use of hospital discharge data in quantifying trends in stroke has been criticized as inaccurate, with low sensitivity and specificity (Reker *et al.*, 2001; Tirschwell & Longstreth, 2002). Further, such data are unable to determine stroke severity; the most important short- and long-term prognostic indicator. Regardless, stroke coding has been reviewed previously and found to be useful for high-level comparisons, particularly when compared to other disease categories (Kokotailo & Hill, 2005).

The *International Classification of Diseases* (ICD) is the international standard for the classification of diagnoses, and its uses include: monitoring the provision of health care services nationally, research and the epidemiological monitoring of health trends and variations, local and national clinical audit and case-mix analysis, clinical governance and, in England, National Health Service (NHS) financial

planning and Payment by Results. ICD as a classification system is recognized as harboring ambiguities and inconsistencies (Surjan 1999), and numerous studies have reported inaccuracies in stroke coding using the ninth edition (ICD-9) (Hasan *et al.*, 1995; Truelsen *et al.*, 2001).

The 10th revision (ICD-10) was introduced into the World Health Organization Member States from 1994 (World Health Organization 1992), which offers a more comprehensive classification system for diagnostic coding compared to its predecessor. The accuracy of stroke coding in England using ICD-10 has not been previously assessed. In this study we aimed to assess the accuracy of hospital discharge coding for hemorrhagic stroke, intracerebral hemorrhage (ICH) and subarachnoid hemorrhage (SAH), using six years of admissions data to the Newcastle upon Tyne Hospitals NHS Foundation Trust.

Materials and methods

Four hospitals make up the Newcastle upon Tyne Hospitals NHS Foundation Trust: the Freeman Hospital (FH), Newcastle General Hospital (NGH), Royal Victoria Infirmary (RVI), and Walkergate Hospital (WH). The Regional Neurosciences Centre, based at the NGH, offers specialist care for a large catchment area of the North of England, serving approximately three million people. Electronic data of all patient admissions, categorized by ICD-10 diagnoses, are maintained by Information Services at the RVI within the Patient Administration System (PAS).

Using data recorded within PAS, we identified all patients coded as being admitted with a primary diagnosis of ICH (ICD-10 code: I61.x) or SAH (I60.x) between 1st January and 31st December for the years 2002-2007. Utilising data from only the primary diagnosis has been shown to result in high specificity and positive predictive value (PPV) (Tirschwell & Longstreth, 2002). Our data included patients admitted to hospital wards as an inpatient and those presenting to Accident and Emergency (A&E) who either died or were discharged without admission. It excludes those attending outpatient clinics and those not presenting to medical attention. No specific diagnostic criteria for ICH and SAH were required for this study, since we relied solely on diagnoses already established by clinicians on the patient discharge summaries.

All of the patient notes for individuals identified by the ICD-10 codes I60 and I61 were retrieved, and the patient discharge summaries within the notes were verified for accuracy with the assigned ICD-10 discharge code. This allowed us to calculate the positive predictive value (PPV) of the ICD-10 PAS data in predicting the same diagnosis in a given patient's discharge summary. For consistency, only the final diagnosis within the discharge summary was looked at.

Results

For the six years of our study, the total number of ICH admissions to the Newcastle upon Tyne Hospitals identified under ICD-10 code I61.x was 978, and that of SAH (ICD-10 code I60.x) was 1169. The number of patients with an ICD code of I61 (ICH) or I60 (SAH) validated as having this same primary diagnosis in the patient notes was 938 and 1123, respectively (Tables II & III). The overall accuracy (PPV) of ICD-10 coding for ICH over the four years of the study was therefore excellent, at 95.9% (95% confidence interval, CI = 94.5-97.0%). The PPV for SAH coding was also excellent, at 96.1% (95% CI = 94.8-97.0). Neither age nor gender had any influence on the accuracy of the ICD-10 coded data. Table I shows the basic demographics of this group of patients.

Coding errors observed were largely expected, with different types of stroke miscoded in ICD-10 data as ICH or SAH. Most (n = 23; 58%) of the erroneous ICH diagnoses identified were actually subarachnoid hemorrhage (SAH), the next majority being ischemic stroke (n = 16; 40%). One (3%) error was due to miscoding of a transient ischemic attack (TIA) (Table II). The majority of miscoded SAH diagnoses (n = 40; 87%) were ICH; the others were ischemic stroke (n = 4; 9%) and TIA (n = 2; 4%) (Table III).

Discussion

The quality of administrative data in the English NHS has long been criticized; in 1982, Körner noted that 'the inaccuracy, lack of timeliness and certain inherent defects form the core of criticism leveled against NHS information' (Körner 1987). Problems with the validity of such data have even been acknowledged by the country's Department of Health (Department of Health 2000), and a British study of the accuracy of the previous version of the International Classification of Diseases (ICD-9) found diagnostic coding to be incorrect in 26% of stroke cases (Hasan *et al.*, 1995). The international picture is not dissimilar, with several numerous studies reporting variable levels of accuracy of ICD-9 and ICD-10 stroke discharge coding, occasionally with significant differences in accuracy between the different subtypes of stroke (Table IV) (Stegmayr & Asplund, 1992; Leibson *et al.*, 1994; Leppala *et al.*,

Table I

Hemorrhagic Stroke Admissions to Newcastle: Basic Demographics

	2002	2003	2004	2005	2006	2007	Total
Total ICH admissions	194	174	162	149	143	116	938
Male to female (M:F) ratio	109:85	92:82	84:78	80:69	74:69	58:58	497:441
Mean age (range)	65.4 (14-92)	68.1 (12-93)	66.7 (13-94)	64.8 (15-91)	67.4 (11-96)	65.9 (15-94)	66.4 (11-96)
Total SAH admissions	192	188	183	176	189	195	1,123
M:F	66:126	69:118	64:119	70:106	76:113	64:131	409:713
Mean age (range)	56.3 (2-90)	54.9 (8-91)	55.5 (12-90)	56.0 (13-89)	55.4 (12-85)	54.3 (16-88)	55.4 (2-91)

ICH = intracerebral hemorrhage; SAH = subarachnoid hemorrhage.

Table II

The Accuracy (Positive Predictive Value) and Errors Identified in ICD-10 Coding of Intracerebral Hemorrhage

Year	Number identified by ICD-10 coding	Number validated by patient discharge summary	Percentage correct – PPV (%)	95% Confidence Interval (%)	Errors identified	
					Number of patients coded erroneously	Actual diagnosis from patient discharge summary
2002	204	194	95.1	91.2-97.3	10	6 – AIS, 4 – SAH
2003	179	174	97.2	93.6-98.8	5	2 – AIS, 3 – SAH
2004	170	162	95.3	91.0-97.6	8	3 – AIS, 4 – SAH, 1 – TIA
2005	152	149	98.0	94.4-99.3	3	1 – AIS, 2 – SAH
2006	151	143	94.7	90.0-97.3	8	3 – AIS, 5 – SAH
2007	122	116	95.1	90.0-97.7	6	1 – AIS, 5 – SAH
Total	978	938	95.9	94.5-97.0	40	16 – AIS, 23 – SAH, 1 – TIA

AIS = acute ischemic stroke; SAH = subarachnoid hemorrhage; TIA = transient ischemic attack.

Table III

The Accuracy (Positive Predictive Value) and Errors Identified in ICD-10 Coding of Subarachnoid Hemorrhage

Year	Number identified by ICD-10 coding	Number validated by patient discharge summary	Percentage correct – PPV (%)	95% Confidence Interval (%)	Errors identified	
					Number of patients coded erroneously	Actual diagnosis from patient discharge summary
2002	201	192	95.5	91.7-97.6	9	7 – ICH, 1 – AIS, 1 – TIA
2003	198	188	95.0	91.0-97.2	10	9 – ICH, 1 – TIA
2004	191	183	95.8	92.0-97.9	8	8 – ICH
2005	180	176	97.8	94.4-99.1	4	4 – ICH
2006	198	189	95.5	91.6-97.6	9	7 – ICH, 2 – AIS
2007	201	195	97.0	93.6-98.6	6	5 – ICH, 1 – AIS
Total	1,169	1,123	96.1	94.8-97.0	46	40 – ICH, 4 – AIS, 2 – TIA

ICH = intracerebral hemorrhage; AIS = acute ischemic stroke; TIA = transient ischemic attack.

Table IV

Summary of Previous Studies into Accuracy of Discharge Coding for Stroke

Reference	Country	ICD version	Stroke type	Accuracy (PPV)
4	USA	ICD-9	AIS ICH SAH	88% 89% 94%
5	Canada	ICD-9 ICD-10	AIS ICH SAH TIA AIS ICH SAH TIA	85% 97% 98% 70% 85% 98% 91% 97%
8	UK	ICD-9	All	74%
16	Canada	ICD-9	All	78%* 90%**
18	Denmark	ICD-10	SAH ICH AIS TIA	48.3% 65.7% 87.6% 60.4%
19	Italy	ICD-9	SAH ICH	76% 78%

ICD = International Classification of Diseases (ICD-9 = ninth edition, ICD-10 = tenth edition); PPV = positive predictive value; AIS = acute ischemic stroke; ICH = intracerebral hemorrhage; SAH = subarachnoid hemorrhage; TIA = transient ischemic attack; * = community-based hospitals; ** = tertiary-care hospitals.

1999; Liu *et al.*, 1999; Mahonen *et al.*, 2000; Johnsen *et al.*, 2002; Spolaore *et al.*, 2005).

According to our results, ICD-10 discharge coding of hemorrhagic stroke in England offers an excellent and consistently high level of accuracy in terms of PPV. This is corroborated by a Canadian study which found ICD-10 discharge coding of ICH and SAH to have a PPV of 98% and 91% across three hospital sites, respectively (Kokotailo & Hill, 2005); that study, like ours, also found the commonest coding errors of ICH to be SAH and ischemic stroke, and coding errors of SAH to be ICH.

A prospective study looking at new admissions would provide a slightly more accurate picture of the validity and reliability of ICD-10 data in PAS; this clearly requires further study. Limiting our sampling frame to only those with ICH or SAH as the primary diagnosis improved the specificity but limited the sensitivity of our results; we may have missed ICH or SAH classified under different ICD-10 diagnoses and those admitted with comorbidities resulting in ICH or SAH being a secondary diagnosis. Importantly, the high level of accuracy identified in our study may not be reflected in discharge coding for other conditions, or for hemorrhagic stroke in other hospitals.

From an epidemiological point of view, using datasets based on ICD-10 diagnostic coding can reduce the problems of bias such as non-response and recall associated with primary research, but can introduce other biases, namely detection and misclassification bias. The use of ICD-10 hospital discharge coding in assessing trends in stroke is cautioned for two reasons. First, since data in an administrative system only includes patients admitted to hospital or seen within A&E, and excludes those not seeking medical attention, it is inherently inaccurate; there may be many patients with less severe symptoms not presenting to hospital and hence excluded from such data (Truelsen *et al.*, 2001), and many who do not survive long enough to make it to hospital.

Second, the quality of coded data is highly dependent on the quality of the original patient notes and the experience and expertise of the coder. This probably accounts for the wide variations in reported accuracy of hospital discharge coding. Hospitals have varying arrangements as to who exactly codes ICD-10 data; in English hospitals it is non-clinical coding staff. One would expect staff specializing in coding diagnoses from a specific department to code more accurately than those dealing with data from several departments, and a Canadian study found

staff in rural hospitals to code stroke using more general codes, with those in urban hospitals coding more specifically (Yiannakoulias *et al.*, 2004). In our dataset, all data were coded in the same department of the same hospital so, whilst there would be individual differences between coding staff, they would presumably have been trained to a similar standard.

Since classification systems such as ICD-10 group similar diagnoses together for the purpose of comparison, clinical detail is lost in the coding process. Traditionally, medical professionals have had little to do with ICD-10 data, which is seen as a largely administrative dataset. Since much of its content originates from patient notes, quality of the original data is important. There have been several national reports in England that have highlighted a lack of clinical interest from health professionals as a key reason for data errors in ICD-10 (Williams & Mann, 2002); if the validity of this data is not verified, then improvements cannot be made. Whilst it would seem logical that coding by clinicians would improve accuracy rates, such a resource-intensive activity may be difficult to implement.

Further research is needed into ways of improving coding accuracy. The quality of coded data, and clinical confidence in it, can be improved by closer working relationships between coding staff and clinicians. With more accurate coding, ICD-10 will become a very powerful epidemiological tool, saving time and resources. Accuracy of the data could also be improved through a standardized in-depth training program for all coding staff, and regular auditing of coding for quality control purposes. In the future, the development of more sophisticated health records underpinned by clinical terminologies (such as SNOMED CT; the Systemized Nomenclature of Medicine - Clinical Terms) will enable more detailed clinical information to be collected and retrieved, improving the flow of information between clinician and administration. Further, this should lead to more accurate epidemiological monitoring of stroke, more representative resource allocation for preventive and acute treatment programs and, ultimately, improved health outcomes for patients.

Acknowledgements and funding

None.

Conflicts of interest

ADM is a director of the Newcastle Neurosurgery Foundation and has received honoraria for serving on the advisory committees of Codman and NovoNordisk. The other authors report no conflicts of interest.

Ethical approval

This was a purely observational retrospective study and hence no ethical approval was required.

REFERENCES

Broderick JP, Adams HP Jr, Barsan W, Feinberg W, Feldmann E, Grotta J, Kase C, Krieger D, Mayberg M, Tilley B, Zabramski JM, Zuccarello M. Guidelines for the management of spontaneous intracerebral hemorrhage: a statement for healthcare professionals from a special writing group of the stroke council, American heart association. Stroke. 1999;30:905-915.

Department of Health. HES - the book. London: Department of Health; 2000.

Feigin VL, Lawes CM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. Lancet Neurol. 2003;2:43-53.

Hasan M, Meara RJ, Bhowmick BK. The Quality of Diagnostic Coding in Cerebrovascular Disease. Int J Qual Health Care. 1995;7:407-410.

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.

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.

Körner E. First report to the Secretary of State of the steering group on health services information. London: HMSO; 1987.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-174.

Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke. 1994;25:2348-2355.

Leppala JM, Virtamo J, Heinonen OP. Validation of stroke diagnosis in the National Hospital Discharge Register and the Register of Causes of Death in Finland. Eur J Epidemiol. 1999;15:155-160.

Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. Cerebrovasc Dis. 1999;9:224-230.

Mahonen M, Salomaa V, Keskimaki I, Moltchanov V, Torppa J, Molarius A, Tuomilehto J, Sarti C. The feasibility of combining data from routine Hospital Discharge and Causes-of-Death Registers for epidemiological studies on stroke. Eur J Epidemiol. 2000;16:815-817.

Reker DM, Hamilton BB, Duncan PW, Yeh SC, Rosen A. Stroke: who's counting what? J Rehabil Res Dev. 2001;38:281-289.

Spolaore P, Brocco S, Fedeli U, Visentin C, Schievano E, Avossa F, Milan G, Toso V, Vanuzzo D, Pilotto L, Pessina AC, Bonita R. Measuring accuracy of discharge diagnoses for a region-wide surveillance of hospitalized strokes. Stroke. 2005;36:1031-1034.

Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. Neuroepidemiology. 1992;11:204-213.

Surjan G. Questions on validity of international classification of diseases-coded diagnoses. Int J Med Inform. 1999;54:77-95.

Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. Stroke. 2002;33:2465-2470.

Truelsen T, Bonita R, Jamrozik K. Surveillance of stroke: a global perspective. Int J Epidemiol. 2001;30:511-516.

Williams JG, Mann RY. Hospital episode statistics: time for clinicians to get involved? Clin Med. 2002;2:34-37.

World Health Organization. ICD-10 – International Statistical Classification of Diseases and Related Health Problems. Geneva: World Health Organization; 1992.

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.

Matthew A. Kirkman,
Department of Neurosurgery,
Regional Neurosciences Centre,
Newcastle General Hospital,
Newcastle upon Tyne,
NE4 6BE (United Kingdom).
E-mail: matthew.kirkman@gmail.com

Appendix 1: ICD-10 codes relevant to our study

I60 Subarachnoid haemorrhage

Includes: ruptured cerebral aneurysm
Excludes: sequelae of subarachnoid haemorrhage (I69.0)

I60.0 Subarachnoid haemorrhage from carotid siphon and bifurcation
I60.1 Subarachnoid haemorrhage from middle cerebral artery
I60.2 Subarachnoid haemorrhage from anterior communicating artery
I60.3 Subarachnoid haemorrhage from posterior communicating artery
I60.4 Subarachnoid haemorrhage from basilar artery
I60.5 Subarachnoid haemorrhage from vertebral artery
I60.6 Subarachnoid haemorrhage from other intracranial arteries
Multiple involvement of intracranial arteries
I60.7 Subarachnoid haemorrhage from intracranial artery, unspecified
Ruptured (congenital) berry aneurysm NOS
Subarachnoid haemorrhage from:
- cerebral
- communicating

} artery NOS

I60.8 Other subarachnoid haemorrhage
Meningeal haemorrhage
Rupture of cerebral arteriovenous malformation
I60.9 Subarachnoid haemorrhage, unspecified
Ruptured (congenital) cerebral aneurysm NOS

I61 Intracerebral haemorrhage

Excludes: sequelae of intracerebral haemorrhage (I69.1)

I61.0 Intracerebral haemorrhage in hemisphere, subcortical
Deep intracerebral haemorrhage
I61.1 Intracerebral haemorrhage in hemisphere, cortical
Cerebral lobe haemorrhage
Superficial intracerebral haemorrhage
I61.2 Intracerebral haemorrhage in hemisphere, unspecified
I61.3 Intracerebral haemorrhage in brain stem
I61.4 Intracerebral haemorrhage in cerebellum
I61.5 Intracerebral haemorrhage, intraventricular
I61.6 Intracerebral haemorrhage, multiple localized
I61.8 Other intracerebral haemorrhage
I61.9 Intracerebral haemorrhage, unspecified