RESEARCH ARTICLE

Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review

Natalie McCormick[1,2], Vidula Bhole[2], Diane Lacaille[2,3,4], J. Antonio Avina-Zubieta[2,3,4]*

1 Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada,
2 Arthritis Research Canada, Richmond, British Columbia, Canada, 3 Division of Rheumatology,
Department of Medicine. University of British Columbia, Vancouver, British Columbia, Canada,
4 Cardiovascular Committee of the CANRAD Network, Richmond, British Columbia, Canada

* azubieta@arthritisresearch.ca

Abstract

Objective

To conduct a systematic review of studies reporting on the validity of International Classification of Diseases (ICD) codes for identifying stroke in administrative data.

Methods

MEDLINE and EMBASE were searched (inception to February 2015) for studies: (a) Using administrative data to identify stroke; or (b) Evaluating the validity of stroke codes in administrative data; and (c) Reporting validation statistics (sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), or Kappa scores) for stroke, or data sufficient for their calculation. Additional articles were located by hand search (up to February 2015) of original papers. Studies solely evaluating codes for transient ischaemic attack were excluded. Data were extracted by two independent reviewers; article quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies tool.

Results

Seventy-seven studies published from 1976–2015 were included. The sensitivity of ICD-9 430-438/ICD-10 I60-I69 for any cerebrovascular disease was $\geq$ 82% in most [$\geq$ 50%] studies, and specificity and NPV were both $\geq$ 95%. The PPV of these codes for any cerebrovascular disease was $\geq$ 81% in most studies, while the PPV specifically for acute stroke was $\leq$ 68%. In at least 50% of studies, PPVs were $\geq$ 93% for subarachnoid haemorrhage (ICD-9 430/ICD-10 I60), 89% for intracerebral haemorrhage (ICD-9 431/ICD-10 I61), and 82% for ischaemic stroke (ICD-9 434/ICD-10 I63 or ICD-9 434&436). For in-hospital deaths, sensitivity was 55%. For cerebrovascular disease or acute stroke as a cause-of-death on death certificates, sensitivity was $\leq$ 71% in most studies while PPV was $\geq$ 87%.

Conclusions

While most cases of prevalent cerebrovascular disease can be detected using 430-438/I60-I69 collectively, acute stroke must be defined using more specific codes. Most in-hospital deaths and death certificates with stroke as a cause-of-death correspond to true stroke deaths. Linking vital statistics and hospitalization data may improve the ascertainment of fatal stroke.

Introduction

Stroke imparts a substantial burden on patients, healthcare systems, and society, with stroke accounting for more than 6.6 million deaths in 2012 (11.9% of all deaths globally) [1]. Those who survive an acute stroke are often permanently disabled, with reduced work and social activities [2], and quality of life [3]. The economic consequences are also substantial; the annual costs of stroke were recently estimated at $33.6 billion in the United States [4] and £8.9 billion in the United Kingdom [5], with direct medical costs accounting for half of these expenditures. Although the incidence of stroke has been decreasing in high-income countries, this decrease is being offset by increasing rates in low- and middle-income countries [6], such that the world-wide burden of stroke is continuing to grow.

Administrative databases are increasingly being used for stroke research. These data sources, which link longitudinal health resource utilization data for hospitalizations, outpatient care, and, in some jurisdictions, dispensed medications, to individual-level demographic and vital statistics data, allow for more efficient analyses, and more generalizable findings. Unfortunately, as administrative databases are usually established for billing, and not research, purposes, the diagnoses contained within tend to be coded by non-medical staff and may not reflect the final diagnosis of the treating physician. But if these databases are to be used for stroke research, the diagnostic codes used to identify stroke must be valid. This means they must be able to distinguish those who have actually experienced a stroke (according to an accepted 'gold standard' reference diagnosis) from those who have not. These diagnostic codes must also allow researchers to distinguish the major subtypes of acute stroke, which differ from one another with regards to their incidence rates, risk factors, and outcomes. For example, haemorrhagic stroke occurs far less frequently than ischaemic stroke [4], but is associated with higher re-hospitalization rates [7,8] and earlier mortality [7,9–11], and greater short-term [8,10,12–21] and long-term [22] treatment costs.

While several validation studies of stroke codes have been conducted [23–26], these have varied widely with regards to their study populations, clinical and geographic settings, and the reference standards used. For example, while some assessed the validity of codes for just one subtype [23–25], others assessed broader groups of codes pertaining to cerebrovascular disease as a whole (including acute stroke, transient ischaemic attack, and stroke sequelae). To synthesize the current evidence, we, as part of a Canadian Rheumatology Network for establishing best practices in the use of administrative data for health research and surveillance (CANRAD) [27–31], conducted a systematic review of studies reporting on the validity of diagnostic codes for identifying cardiovascular diseases. Data from these studies were used to compare the validity of these codes, and evaluate whether administrative health data can accurately identify cardiovascular diseases for the purpose of capturing these events as covariates, outcomes, or complications in future research. We recently reported our findings on the validity of codes for

myocardial infarction [32] and heart failure [33]. In the current paper, we analyze studies reporting on the validity of stroke codes in administrative databases.

Methods

Literature Search

An experienced librarian (M-DW) undertook searches of the MEDLINE and EMBASE databases, from inception (1946 and 1974, respectively) for all available peer-reviewed literature. Two search strategies were used: (1) All studies where administrative data was used to identify cardiovascular diseases; (2) All studies reporting on the validity of administrative data for identifying cardiovascular diseases. Our MEDLINE and EMBASE search strategies are available as (S1, S2, S3, and S4 Texts). To identify additional studies, the authors hand-searched the reference lists of the key articles located. As well, the Cited-By tools in PubMed and Google Scholar were used to find relevant articles that had cited the articles located through the database search. The databases were originally searched from inception to November 2010, with the handsearch conducted up to February 2011. These searches were updated in February 2015.

Two reviewers independently screened the titles and abstracts of the located records for relevance to the study objectives. In the next step, full text publications were evaluated against the inclusion criteria. Any discrepancies were discussed until consensus was reached. When the conflict persisted a third reviewer (JAA-Z) was consulted. No protocol for this systematic review has been published, though more information is available in the following publication [27]. Our review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [34] statements, and our completed PRISMA checklist is provided as (S1 Checklist).

Inclusion Criteria

We considered full-length, English-language, peer-reviewed articles that used administrative data and either reported validation statistics for the International Classification of Diseases (ICD) codes of interest, or provided sufficient data for their calculation. We first included studies that evaluated at least one code pertaining to a subtype of acute stroke, being ICD-8/9 430 or ICD-10 I60 for subarachnoid haemorrhage (SAH), and ICD-8/9 431 or ICD-10 I61 for intracerebral haemorrhage (ICH). For ischaemic stroke, the main codes are ICD-8 433/434 and ICD-9 434 (occlusion of the cerebral arteries), and ICD-10 I63 (cerebral infarction).

Stroke is a heterogeneous disease that is not defined consistently by clinicians or researchers [35]. It has traditionally been distinguished from transient ischaemic attack (TIA) by way of duration (more or less than 24 hours) and the presence/absence of permanent brain infarction. Although advances in neuroimaging have resulted in many events that would previously have been labelled as TIA now being considered as minor strokes, this is an area of ongoing controversy [35]. As such, we took a conservative approach by not considering episodes of TIA as acute stroke, and so excluded studies that solely evaluated codes for TIA (ICD-9 435 or ICD-10 G45).

Although our focus was on the validity of codes for acute stroke-specifically (defined as SAH, ICH, or ischaemic stroke), we also included studies that evaluated a range of codes (ICD-8/9 430–438 or ICD-10 I60-I69) pertaining to a broader group of cerebrovascular diseases. Included in these ranges were the codes for acute stroke listed above, along with codes for acute but ill-defined stroke (ICD-9 436 and ICD-10 I64), other types of ill-defined stroke (ICD-9 437) and other cerebrovascular diseases (ICD-10 I67/68), other types of intracranial haemorrhage than ICH (ICD-9 432 and ICD-10 I62), TIA (ICD-9 435), and late effects of stroke or stroke sequelae (ICD-9 438 and ICD-10 I69). It was important to include these studies because,

while reviewing the literature, we observed that this broad range of codes for cerebrovascular disease is frequently used to identify cases of acute stroke.

Data Extraction

Two independent reviewers (NM and VB) examined the full text of each selected record and abstracted data using a standardized collection form (a copy is provided in **S5 Text**). Information was gathered on the study population, administrative data source, stroke codes and algorithm, validation process, and gold standard. Validation statistics comparing the codes to definite, probable, or possible cases of acute stroke, or the specific diagnoses of SAH, ICH, or ischaemic stroke in particular, were abstracted. These statistics included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and kappa. Wherever possible, we abstracted statistics for definite and probable cases of acute stroke. However, the number of categories available depended on the choice of gold standard. For example, under the World Health Organisation criteria, potential cases are categorized as either definite stroke or not stroke [36], while the National Survey of Stroke criteria [37] use the categories of definite, highly probable, and probable stroke. Statistics for each sex, for fatal versus non-fatal cases, and for each hospital discharge position (i.e. primary/principal and secondary diagnosis) were abstracted where reported. Data were independently abstracted by each reviewer who subsequently compared their forms to correct errors and resolve discrepancies, if any.

Quality Scores

The design and methods employed in each study, including the rigour of the reference standard, and generalizability of the study population, could influence the resultant validity statistics. Hence, all studies were evaluated for quality, with the validation statistics stratified by level of study quality. An adaptation of the QUADAS tool (Quality Assessment of Diagnostic Accuracy Studies) [38] was used to evaluate study quality. Our group previously used the QUADAS in assessing the validity of codes for diabetes mellitus [30], myocardial infarction [32], heart failure [33] and osteoporosis and fractures [31].
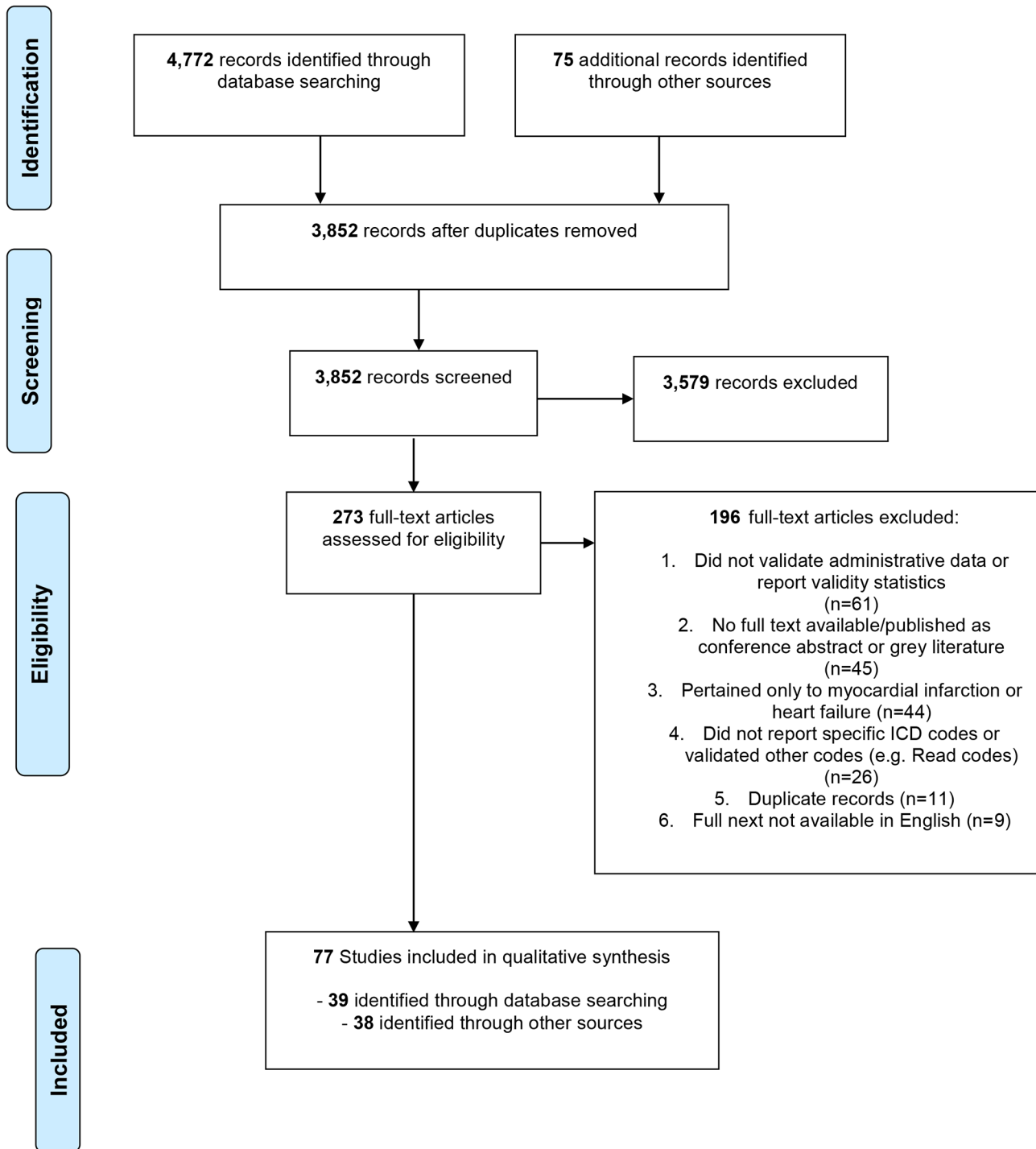
Statistical Analysis

All validation statistics were abstracted as reported. Where sufficient data were available we calculated 95% confidence intervals (95% CI) and validation statistics not directly reported in the original publication. Kappa values (a measure of agreement beyond that expected by chance) greater than 0.60 indicated substantial/perfect agreement, 0.21–0.60 were considered as fair/moderate agreement and those 0.20 or lower as light/poor agreement [39].

Results

Literature Search

We identified 1,587 citations through our original searches (inception to November 2010) of the MEDLINE and EMBASE databases, and an additional 2,160 citations in our updated searches of these databases (January 2010 to February 2015). All citations were screened for relevance to our study objectives, with 198 full-text articles assessed for eligibility (**Fig 1**), and 39 of these selected for inclusion. We also assessed 75 full-text articles for eligibility that were identified from hand searches, and selected 38 additional articles therein. Thus, a total of 273 articles were assessed for eligibility, from which 196 were excluded, mainly because they reported on the validity of other cardiovascular diseases (n = 44), or did not actually validate stroke diagnoses in administrative data (n = 61). Nine articles were excluded because they were

Identification

4,772 records identified through database searching

75 additional records identified through other sources

3,852 records after duplicates removed

Screening

3,852 records screened

3,579 records excluded

Eligibility

273 full-text articles assessed for eligibility

196 full-text articles excluded:

1. Did not validate administrative data or report validity statistics (n=61)
2. No full text available/published as conference abstract or grey literature (n=45)
3. Pertained only to myocardial infarction or heart failure (n=44)
4. Did not report specific ICD codes or validated other codes (e.g. Read codes) (n=26)
5. Duplicate records (n=11)
6. Full next not available in English (n=9)

Included

77 Studies included in qualitative synthesis

- **39** identified through database searching
- **38** identified through other sources

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097*

For more information, visit www.prisma-statement.org.

Fig 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-style Flowchart of Study Selection and Review.
ICD = International Classification of Diseases.

not published in English; their languages of publication were Danish, German, Italian, Japanese, Portuguese, Spanish (two articles), French, and Chinese. Ultimately 77 articles were included for the systematic review of acute stroke.

Study Characteristics

Of the 77 articles evaluating stroke diagnoses that were included in the final review, 31 (40%) were from the United States (USA), 26 (34%) were from Europe, 13 (17%) were from Canada, four were from Asia (5%), two (3%) were from Australia, and one (1%) was from Sri Lanka. Characteristics of these studies are presented in **Table 1**. Validation was the primary objective in all but ten [40–49] of these studies. Two articles [40,50] reported on the validity of stroke diagnoses exclusively in a paediatric population. Most studies evaluated the diagnostic codes in hospitalization databases though 16 studies [47,51–65] evaluated stroke as a cause-of-death on death certificates and one study [66] reported on outpatient data exclusively.

Gold standard

Chart reviews, sometimes in conjunction with unspecified diagnostic criteria, formed the basis of the gold standard in 35 studies, patient self-report was used in one [112], and national and regional stroke registries or clinical databases served as the gold standard in 12 [52,58–60,65,67,73,74,76,81,91,92]. One study [23] utilized two gold standards, with the reference diagnosis for some cases established upon prospective clinical examination by a neurologist, and for other cases, established after retrospective chart review by a different neurologist. The 28 remaining studies used a specific set of diagnostic criteria, most often the WHO criteria, to evaluate the stroke diagnosis.

Study quality was evaluated based on the QUADAS tool [38], with 54 of 77 studies (70%) categorized as high quality, and the remaining 23 studies as medium quality. A detailed breakdown of the quality assessment for each study is provided in **S1 Table**. Seven of the medium-quality studies [47,57,59,64,71,73,104] did not adequately describe the validation process or other key methodological aspects, while nine employed a selected study population [25,40,50,77,89,92,105,107,111] (e.g. atrial fibrillation cohort, kidney transplant recipients), and seven used a less-reliable gold standard [24,66,70,82,85,94,112], typically chart review by an individual other than a clinician or trained hospital coder.

Validity of Stroke Codes on Aggregate

The validation statistics reported by each of the included studies are provided in **S2** and **S3 Tables**. We located 36 papers examining the validity of the codes for cerebrovascular disease as an aggregate (ICD-9 430–438 or ICD-10 I60-I69); these codes were compared to diagnoses of any type of cerebrovascular disease (usually as a comorbidity) in 16 studies [56,61,63,66,71, 77,79,80,82,89,90,93,94,102,104,105], and to a diagnosis of acute stroke in particular in 21 [26,41,45,47,49,52,54,55,57,62,64,74,83,84,86,91,93,99,103,111,112]. The sensitivity of these codes for any type of cerebrovascular disease was ≥ 82% in seven of the 14 studies (range 32% to 100%). The PPV was ≥ 81% in seven of the 14 studies reporting this statistic (range 43% to 97%). Specificity, reported by ten studies [63,66,71,79,82,89,90,102,104,105], was ≥ 95% in nine of the ten (range 90% to 100%), while NPV was ≥ 95% in eight of the ten studies [63,71,79,82, 89,90,94,102,104,105] where this statistic was reported (range 84% to 100%). Kappa values, as reported by seven studies, ranged from 0.52 [82] to 0.76 [89] to 0.91 [79].

Eight of the 16 studies [56,61,63,79,80,90,93,102] were rated as high quality and the other eight [66,71,77,82,89,94,104,105] were rated as medium quality. There was little difference in the sensitivity values between the medium- and high-quality studies: sensitivity ranged from

Table 1. Characteristics of Included Studies.

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Aboa-Eboule [67], 2013	2004–2008	yes	France	903	residents of one community hospitalized for stroke at one teaching hospital	ICD-10 inpatient records	disease registry, using WHO criteria
Agrawal [40], 2009	1993–2003	no	USA (California)	1,307	children aged 0–19 years enrolled in the Kaiser Permanente Medical Care Program and participating in the Kaiser Paediatric Stroke Study	ICD-9 inpatient and outpatient records	CRMD
Appelros [65], 2011	1999–2000	yes	Sweden	377	residents of one community	ICD-10 inpatient and vital statistics records	disease registry, using WHO criteria
Arnason [68], 2006	1999–2000	yes	Canada (Ontario)	616	all patient discharged from one tertiary hospital with a bleeding-related or thromboembolic diagnosis	ICD-9 inpatient records	CRDC
Benesch [69], 1997	1992	yes	USA (Louisiana, Massachusetts, California, Iowa, Pennsylvania)	649	patients hospitalized at one of five academic medical centres and eligible for a telephone survey of persons at increased risk for major stroke	ICD-9 inpatient records	CRDC—WHO criteria
Birman-Deych [70], 2005	1998–1999	yes	USA (national)	23,657	Medicare beneficiaries (aged 20–105 years) on the National Registry of Atrial Fibrillation hospitalized for atrial fibrillation	ICD-9 inpatient records	chart review
Borzecki [66], 2004	1998–1999	yes	USA (national)	1,176	individuals regularly receiving care from one of 10 Veterans Affairs sites across the USA, random selection of 100 users from each site with hypertension and 20 without	ICD-9 outpatient records	chart review
Broderick [41], 1998	1993–1995	no	USA (Ohio, Kentucky)	733	all residents of one of five counties	ICD-9 inpatient and vital statistics records	CRDC–Rochester, Minnesota and National Institute of Neurological Disorders and Stroke
Brown [54], 2006	2000–2001	yes	USA (Texas)	186	participants aged 44 years and older enrolled in the Brain Attack Surveillance in Corpus Christi (BASIC) Project	ICD-10 vital statistics records	CRMD
Chen [71], 2009	2003	yes	Canada (Alberta)	4,008	general hospitalized population	ICD-10 inpatient records	chart review
Cheng [72], 2011	1999	yes	Taiwan	372	hospitalized patients aged 55 years and older	ICD-9 inpatient records	CRMD

(Continued)

Table 1. (Continued)

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Davenport [73], 1996	n/a	yes	Scotland	97,515	hospitalized patients at one university teaching hospital	ICD-9 inpatient records	disease registry: Lothian Stroke Register
de Faire [63], 1976	1961–73	yes	Sweden	1,156	10,000 pairs of twins enrolled in the Swedish Twin Registry and born during 1901–1925	ICD (1965 edition) vital statistics records	CRMD
Derby [42], 2000	1980–1991	no	USA (Rhode Island, Massachusetts)	3,811	residents of two communities aged 35–74 years	ICD-9 inpatient records	CRMD
Ellekjaer [74], 1999	1994–1996	yes	Norway	759	hospitalized patients aged 15 years and older	ICD-9 inpatient records	disease registry, using WHO criteria
Gaist [43], 2000	1977–1995	no	Denmark	191	patients hospitalized at one university hospital or two other hospitals within one county	ICD-8 and ICD-10 inpatient records	CRMD
Ghia [75], 2010	2003–2007	yes	Australia	570	hospitalized patients admitted through the emergency department and diagnosed upon admission with TIA	ICD-10 inpatient records	chart review
Goldstein [25], 1998	1995–1997	yes	USA (North Carolina)	175	hospitalized patients at one Veterans Affairs Medical Center	ICD-9 inpatient records	CRDC—TOAST criteria
Golomb [50], 2006	1999–2004	yes	USA (Indiana)	663	all inpatients and outpatients seen at one Children's Hospital	ICD-9 inpatient and outpatient records	CRMD
Haesebaert [76], 2013	2006–2007	yes	France	329	patients ≥ 18 years of age admitted to one of four university hospitals	ICD-10 inpatient records	disease registry: AVC69 cohort
Hasan [77], 1995	1993	yes	Wales	166	patients admitted to the Department of the Care of the Elderly at one of four hospitals within one health unit	ICD-9 inpatient records	CRMD
Heckbert [78], 2004	1994–2000	yes	USA (national)	34,016	women participating in the Women's Health Initiative clinical and observational studies	ICD-9 inpatient records	CRDC—Women's Health Initiative criteria
Henderson [79], 2006	1998–99, 2000–01	yes	Australia	14,635	all hospitalized patients (excluding same-day chemotherapy and dialysis)	ICD-10 inpatient records	chart review: charts were re-coded by professional coders
Hennessy [80], 2010	2002–2007	yes	Canada	1,292	patients hospitalized at one of four hospitals	ICD-10 inpatient records	chart review: charts were re-coded by nurses with coding experience

(Continued)

Table 1. (*Continued*)

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Holick [44], 2009	2003–2007	no	USA	132	new users of atomoxetine or stimulant ADHD medications, and general population controls, identified from a health insurance database for a study assessing the association between atomoxetine and stroke in adults	ICD-9 inpatient records	CRMD
Hsieh [81], 2013	2006–2008	yes	Taiwan	1,736	patients hospitalized at one tertiary referral centre	ICD-9 inpatient records	disease registry (Taiwan Stroke Registry) and CRMD
Humphries [82], 2000	1994–1995	yes	Canada (British Columbia)	817	patients hospitalized for percutaneous coronary intervention	ICD-9 inpatient records	chart review
Iso [57], 1990	1970 & 1980	yes	USA (Minnesota)	214	residents of the study area aged 30–74 years who died in hospital, identified as part of the Minnesota Heart Survey	ICD-8 and ICD-9 vital statistics records	CRDC–National Survey of Stroke
Ives [62], 1995	1989–1992	yes	USA (California, Maryland, North Carolina, Pennsylvania)	5,201	participants in the population-based Cardiovascular Health Study aged 65 years or older	ICD-9 inpatient and vital statistics records	CRMD
Johnsen [83], 2002	1993–1999	yes	Denmark	565	participants in a population-based cohort study on diet and cancer development aged 50–64 years at enrollment	ICD-10 inpatient records	CRDC—WHO criteria
Jones [84], 2014	1987–2010	yes	USA (Maryland, Minnesota, Mississippi, North Carolina)	4,260	members of the population-based Atherosclerosis Risk in Community (ARIC) Study cohort, aged 45–64 years at the time of study enrollment	ICD-9 inpatient records	CRDC–National Survey of Stroke, AHA/ASA
Kirkman [24], 2009	2002–2007	yes	United Kingdom	2,147	all hospitalized patients residing in the study area	ICD-10 inpatient records	chart review: mentioned in records
Klatsky [45], 2005	1978–1996	no	USA (California)	3,441	members of a prepaid healthcare program who supplied data on voluntary health examinations	ICD-9 inpatient records	CRMD
Kokotalio [85], 2005	2000–2003	yes	Canada (Alberta)	717	hospitalized patients at three centres	ICD-9 and ICD-10 inpatient records	chart review
Koster [60], 2013	2004	yes	Sweden	3,534	residents aged 20 years and older of two Swedish counties covered by the MONICA register	ICD-10 inpatient, outpatient, and vital statistics records	disease registry —MONICA

(*Continued*)

Table 1. (*Continued*)

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Krarup [86], 2007	1998–1999	yes	Denmark	236	enrollees in the population-based Copenhagen City Heart Study	ICD-10 inpatient records	CRDC—WHO criteria
Kumamaru [87], 2014	2003–2009	yes	USA (national)	15,089	participants in the REasons for Geographic And RacialDifferences in Stroke (REGARDS) study, aged ≥ 65 years with at least one month of Medicare eligibility	ICD-9 inpatient records	CRDC–WHO criteria
Lakshminarayan [46], 2009	1980, 1985, 1990, 1995, 2000	no	USA (Minnesota)	6,032	general population aged 30–74 years	ICD-9 inpatient records	CRDC–WHO and Minnesota Stroke Survey criteria, and neuroimaging
Lakshminarayan [88], 2014	1993–2007	yes	USA (national)	48,877	participants enrolled in the observational Women's Health Initiative studies aged 50–79 years at enrollment with Medicare fee-for-service coverage	ICD-9 inpatient records	CRDC–Women's Health Initiative criteria
Lambert [89], 2012	2002–2006	yes	Canada (Quebec)	1,982	patients hospitalized with MI as a principal diagnosis, or who underwent PCI or CABG, at one of 13 primary, secondary, or tertiary hospitals	ICD-9 inpatient records	chart review; mentioned in records
Lee [90], 2005	1997–1999	yes	Canada (Ontario)	1,592	hospitalized individuals <105 years of age coded with a primary/most responsible diagnosis of heart failure	ICD-9 inpatient records	CRDC—Charlson comorbidity index
Leibson [91], 1999	1970, 1980, 1984, 1989	yes	USA (Minnesota)	377	all hospitalized patients residing in the study area	ICD-A and ICD-9 inpatient hospitalizations	disease registry —Rochester Stroke Registry
Lentine [92], 2009	1991–2002	yes	USA (Missouri)	571	kidney transplant recipients aged ≥ 18 years who had Medicare as their primary insurer at transplant and the time of each clinical event	ICD-9 inpatient records	clinical database
Leone [93], 2004	1998	yes	Italy	1,126	hospitalized patients from the Neurology, Neurosurgery, General Medicine, Cardiac Surgery, and Intensive Care departments at one centre	ICD-9 inpatient records	CRDC—WHO criteria

(*Continued*)

Table 1. (*Continued*)

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Leppala [51], 1999	1985–1989, 1992	yes	Finland	593	male smokers enrolled in a population-based, randomized controlled trial of alpha-tocopherol and beta-carotene supplementation, aged 50–69 years at registration	ICD-8 and ICD-9 inpatient and vital statistics records	CRDC—National Survey of Stroke and MONICA criteria
Levy [94], 1999	1994	yes	Canada (Quebec)	224	individuals aged ≥ 65 years discharged alive with a primary diagnosis of myocardial infarction	ICD-9 inpatient records	chart review: mentioned in records
Lindblad [53], 1993	1977–1987	yes	Sweden	413	participants in a hypertension registry (hypertensive cases, normotensive participants, randomly-selected controls), recruited from a geographical half of one Swedish county aged 40–70 years at registration	ICD-9 inpatient and vital statistics records	CRMD
Liu [26], 1999	1990–1991	yes	Canada (Saskatchewan)	1,494	patients hospitalized at one of three tertiary-care, or three community, hospitals	ICD-9 inpatient records	CRDC—National Survey of Stroke criteria
Mayo [95], 1993	n/a	yes	Canada (Quebec)	96	patients hospitalized at five teaching, affiliated, and community hospitals	ICD-9 inpatient records	CRMD
Olson [96], 2014	2001–2009	yes	USA (Colorado)	4,689	enrollees in a managed care organization aged 18 years and older	ICD-9 inpatient and outpatient records	CRDC–Rochester, Minnesota Stroke Study criteria
Newton [97], 1999	1992–1995	yes	USA (Washington)	471	enrollees of a Health Maintenance Organization with diabetes aged 18 years and older	ICD-9 inpatient and outpatient records	chart review
Palmieri [47], 2007	1997–1999	no	Italy	8,000	general population aged 35–74 years residing in one of eight regions of Italy	ICD-9 inpatient and vital statistics records	CRDC—MONICA criteria
Phillips [55], 1993	1988–1989	yes	Canada	301	patients hospitalized at one teaching hospital	ICD-9 inpatient and vital statistics records	CRDC–WHO criteria
Piriyawat [98], 2002	2000	yes	USA (Texas)	815	as part of the Brain Attack Surveillance in Corpus Christi (BASIC) Project, patients ≥ 45 years admitted to one of six area hospitals	ICD-9 inpatient records	CRDC–MONICA criteria
Ramalle-Gomara [99], 2013	2009	yes	Spain	400	patients hospitalized at one of two public hospitals	ICD-9 inpatient records	CRMD

(*Continued*)

Table 1. (*Continued*)

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Rampatage [61], 2013	2006–2008	yes	Sri Lanka	648	deaths occurring at three large hospitals in/near the capital city	ICD-10 vital statistics records	CRMD
Rao [56], 2007	2002	yes	China	2,917	deaths occurring in health facilities located in one of six large cities	ICD-10 vital statistics records	CRMD
Reker [100], 2001	1998–1999	yes	USA (national)	671	patients hospitalized at 11 Veterans Affairs Medical Centres	ICD-9 inpatient records	CRDC—WHO criteria
Reggio [64], 1995	1985–1992	yes	Italy	193	deaths occurring among residents of one municipality	ICD-9 vital statistics records	CRDC–WHO criteria
Rinaldi [23], 2003	1999	yes	Italy	233	hospitalized patients at one centre	ICD-9 inpatient records	prospective clinical examination and retrospective CRDC—WHO criteria
Rosamond [49], 1999	1987–1995	no	USA (Maryland, Minnesota, Mississippi, North Carolina)	1,185	members of the population-based Atherosclerosis Risk in Community (ARIC) Study cohort, aged 45–64 years at the time of study enrollment	ICD-9 inpatient records	CRDC–National Survey of Stroke
Roumie [101], 2008	1999–2003	yes	USA (Tennessee)	231	Medicaid enrollees aged 50–84 years, identified as part of a larger retrospective cohort study on the relationship between NSAID use and stroke	ICD-9 inpatient records	CRMD
Shahar [48], 1995	1980,1985, 1990	no	USA (Minnesota)	2,939	general population aged 30–74 years	ICD-9 inpatient records	CRDC—WHO, Minnesota Stroke Survey
Singh [102], 2012	2007–2009	yes	USA (Minnesota)	240	retrospective cohorts of patients $\geq$ 18 years of age admitted to the intensive care unit	ICD-9 inpatient and outpatient records	CRMD
Sinha [103], 2008	1993–2003	yes	United Kingdom	250	residents in one community aged 40–79 years and enrolled in a population-based study of the determinants of chronic disease	ICD-10 inpatient records	CRDC—WHO criteria
So [104], 2006	2003	yes	Canada (Alberta)	193	patients hospitalized for myocardial infarction	ICD-9 and -10 inpatient records	chart review
Soo [105], 2014	2003	yes	Scotland	3,219	participants in a population-based cohort study of chronic kidney disease	ICD-10 inpatient records	CRMD
Spolaore [106], 2005	1999	yes	Italy	4,015	general hospitalized population	ICD-9 inpatient records	CRDC—MONICA criteria

(*Continued*)

Table 1. (*Continued*)

First Author, Year of Publication	Year(s) of Data Collection	Primary Validation Study?	Country	Records Evaluated (N)	Source Population	Type of Administrative Data	Gold Standard
Stegmayr [58], 1992	1985–1989	yes	Sweden	6,000	residents of the two provinces included in the Northern Sweden MONICA study aged 25–74 years	ICD-9 inpatient and vital statistics records	disease registry —MONICA
Szczesniewska [59], 1990	1984–1986	yes	Poland	213	residents of two city districts covered by the POL-MONICA Warsaw Project aged 25–64 years	ICD-9 vital statistics records	disease registry —MONICA
Thigpen [107], 2015	2006–2010	yes	USA (Alabama, Massachusetts, Pennsylvania)	1,812	patients with atrial fibrillation hospitalized at one of three medical centres	ICD-9 inpatient records	CRDC–WHO, AHA
Tirschwell [108], 2002	1990–1996	yes	USA (Washington)	206	general hospitalized population ≥ 20 years of age	ICD-9 inpatient records	CRMD
Tolonen [52], 2007	1993–1998	yes	Finland	3,633	general population aged 25 years and older	ICD-9 and -10 inpatient and vital statistics records	disease registry —FINMONICA/ FINSTROKE register
Tu [109], 2013	2011	yes	Canada	5,000	individuals aged ≥ 20 years seen by a family practice physician using the EMRALD EMR system	ICD-10 inpatient and outpatient records	CRMD
Wahl [110], 2010	2002–2004	yes	USA (national)	200	commercially-insured individuals in a large health claims database, identified as part of a larger retrospective observational cohort study on the risk of serious adverse events among users of selective coxibs and non-over-the-counter NSAIDs	ICD-9 inpatient records	CRMD
Wildenschild [111], 2014	2009–2010	yes	Denmark	228	Part 1: individuals ≥ 18 years admitted to hospital; Part 2: patients discharged from one of four neurologic wards	ICD-10 inpatient records	CRDC–WHO criteria
Wu [112], 2014	2004–2005	yes	Taiwan	15,574	individuals aged ≥ 12 years whose households were randomly selected for participation in the 2005 Taiwan National Health Interview Survey	ICD-9 inpatient and outpatient records	patient self-report

CRDC = Chart Review, Diagnostic Criteria–the charts of potential cases were reviewed, and a formal set of diagnostic criteria were applied when evaluating cases; CRMD = Chart Review, Medical Doctor–the charts of potential cases were reviewed by a physician, who evaluated cases using their clinical judgment or an otherwise unspecified set of criteria; AHA = American Heart Association; ASA = American Stroke Association; CABG = coronary artery bypass graft; EMRALD = Electronic Medical Record Administrative Data Linked Database; ICD = International Classification of Diseases; MONICA = MONItoring Trends and Determinants in CArdiovascular Disease; NSAID = non-steroid anti-inflammatory drug; PCI = percutaneous coronary intervention; TOAST = Trial of ORG 10172 in Acute Stroke Treatment; TIA = transient ischaemic attack; WHO = World Health Organization

doi:10.1371/journal.pone.0135834.t001

43% to 100% among the medium-quality studies, and from 32% to 96% among the high-quality studies. There was, however, more of a difference in the PPVs, which ranged from 43% to 83% among the medium-quality studies, and from 52% to 97% among the high-quality studies (with PPV ≥ 88% in five of the seven high-quality studies reporting on PPV).

The sensitivity of these codes for the narrower diagnosis of acute stroke (SAH, ICH, or ischaemic stroke), which was reported by ten studies [52,54,57,62,64,74,93,99,111,112], was ≥ 66% in seven of the ten (range 42% to 96%). The PPV for the reference diagnosis of acute stroke was generally lower than that for the broader category of any cerebrovascular disease, being ≤ 68% in 12 of the 21 studies reporting this statistic (range 28% to 98%). Indeed, in one study that evaluated the PPV against both acute stroke and 'any cerebrovascular disease', the PPV for 'any cerebrovascular disease' (93%) was higher than that for acute stroke (66%) [93].

Subarachnoid Haemorrhage

Twenty-seven papers reported on the validity of codes for SAH (ICD-9 430 or ICD-10 I60), and the PPV was ≥ 86% in 16 of the 26 studies where this was reported (**S2 Table**). Fifteen papers compared the SAH code to any type of acute stroke (meaning a case coded for SAH was considered a true-positive if the diagnosis assigned upon validation was SAH, ICH, or ischaemic stroke), and the PPV was ≥ 86% in eight of these studies (range 33% to 100%). Thirteen studies compared the SAH code to the diagnosis of SAH in particular, and the PPV was ≥ 93% in seven of the 13 (range 46% to 100%). The sensitivity of these codes for SAH, reported by only four studies [52,56,84,93], ranged from 35% [93] to 95% [52]. Kappa, reported by only one study [108], was 0.88.

Intracerebral Haemorrhage

Thirty-four studies evaluated the validity of codes for ICH (ICD-9 431/432 or ICD-10 I61/62) (**S2 Table**). Twenty-six evaluated the main ICH codes 431/I61, and the PPV was ≥ 87% in 16 of the 25 studies reporting on PPV. The PPV, when compared to any type of acute stroke, was ≥ 87% in ten of 15 studies (range 78% to 100%), and when compared to the diagnosis of ICH in particular, the PPV was ≥ 89% in six of the 12 studies (range 63% to 100%). The sensitivity of these codes for ICH, as reported by three studies, ranged from 57% [93] to 69% [56] to 95% [52]. The kappa value for ICD-9 431, as reported by one study [108], was 0.82, while that for ICD-9 430 and 431 combined, was 0.84 [88]. The lesser-used ICH codes (ICD-9 432 or ICD-10 I62) were evaluated in 15 studies [26,41,47,49,55,62,74,78,84,91,93,95,96,100,106], in which the PPV was ≤ 67% in all but two [62,95]. The PPV of ICD-9 431 and 432 combined, available from 13 studies [26,41,42,47,49,53,55,57,74,84,91,93,95], ranged from 33% to 99%. In sum, the PPV of the main codes for ICH (ICD-9 431/ICD-10 I61) was ≥ 87% in most studies.

Ischaemic Stroke

We located 39 papers that examined the validity of codes for ischaemic stroke (**S2 Table**). From these, 28 evaluated the main code for ischaemic stroke (ICD-9 434 or ICD-10 I63) and the PPV was ≥ 82% in 20 of the 27 studies reporting this statistics (range 62% to 100%). The PPV was ≥ 83% in 13 of the 19 papers [26,41,42,47,49,50,53,55,62,69,74,78,84,91,93,95,96,100,106] where the gold standard diagnosis was any type of acute stroke (SAH, ICH, or ischaemic stroke) (range 62% to 100%), and ≥ 82% in 10 of the 13 papers [23,25,49,50,52,53,76,83,84,86,93,96,107] where the gold standard diagnosis was iscahemic stroke in particular (range 52% to 100%). The sensitivity of these codes for ischaemic stroke, available from six papers [23,52,56,76,84,93], ranged from 2% [23] to

80% [52]. ICD-9 433 was evaluated in 19 papers [25,26,40,41,47,49,50,53,55,69,74,84,91,93, 95,96,100,106,107], with the PPV being ≤ 71% in 14 of the 19. Eight papers [25,49,50,53, 84,93,96,107] reported on the PPV of ICD-9 433 for ischaemic stroke in particular, and it was ≤ 79% in six of the eight. Just two studies reported on the sensitivity of ICD-9 433 for ischaemic stroke, which was 2% [93] in one and 9% [84] in the other.

The combination of ICD-9 433 and 434 (occlusion of the precerebral or cerebral arteries) was reported on by 20 studies [26,40,41,47,49,51–53,55,57,69,72,74,81,84,93,95,99,107,110], in which the PPV for any stroke ranged from 23% to 100%, and that for ischaemic stroke ranged from 40% to 100%. The sensitivity of these codes for ischaemic stroke, as reported by five studies [52,81,84,93,99], was ≥ 76% in four of these. The code for acute but ill-defined stroke (ICD-9 436 or ICD-10 I64) was evaluated in 23 studies [23,25,26,40,41,47,49,50,53,55,62,69,74,78,83, 84,91,93,95,96,100,106,107], with a PPV for any stroke of ≥ 75% in 12 of 19 studies (range 48% to 95%) and a PPV for ischaemic stroke of ≥ 75% in five of seven studies [23,25,49,50, 84,96,107] where this statistic was reported (range 50% to 87%).

Nineteen studies [25,26,40,41,47,49,52,55,69,74,84,85,87,88,93,95,101,107,108] employed a broader case definition (ICD-9 433/434/436 or ICD-10 I63/64); the PPV was ≥ 77% in 12 of the 19 (range 46% to 94%). Two studies reported on kappa, which, for ICD-9 433, 434, and 436 together, was 0.82 [108] in one study and 0.85 [88] in the other. Sixteen studies [23,26,40,41, 47,49,53,55,62,69,74,78,84,93,95,107] examined the validity of ICD-9 434 and 436 as a pair, and the PPV was ≥ 82% in ten of the 16 (range 66% to 94%). Of interest, Leone *et al* [93] compared the sensitivity and PPV of different codes for ischaemic stroke, and found that ICD-9 434 and 436 combined had a higher sensitivity (43% versus 35%) for ischaemic stroke, but similar PPV (87% versus 90%), than did ICD-9 434 alone. In sum, the PPV of the main codes for ischaemic stroke (ICD-9 434/ICD-10 I63) was ≥ 82% in most studies, and the PPV of codes for acute but ill-defined stroke (ICD-9 436 or ICD-10 I64) was ≥ 75%.

Validity of Sets of Stroke-Specific Codes

Thirty-six papers examined the validity of a set of stroke-specific codes for identifying any type of stroke (**S3 Table**) though in two [70,92], the reference diagnosis was stroke/TIA and not just acute stroke. There was some variability in the codes included in each set, but they generally excluded codes pertaining to occlusion of the precerebral arteries, intracranial haemorrhage other than ICH, late effects of stroke/stroke sequalae, and cerebrovascular disorders stemming from other conditions. The code for acute but ill-defined stroke was included in some of these sets. One study, by Reker *et al* [100], used both a high-sensitivity algorithm for detecting stroke (sensitivity = 91%, PPV = 52%) and high-specificity algorithm (sensitivity = 54%, PPV = 75%). Lakshminarayan *et al* [46] compared two diagnostic criteria, those from the WHO and Minnesota Stroke Survey (MSS), and found the PPV against the WHO criteria was 98%, while that against the stricter MSS criteria was 68% [46].

The sensitivity of these sets of codes for stroke was ≥ 82% in 13 of 22 studies [40,41,59,62, 65,67,70,73–75,78,87,88,92,93,97–101,109,111] where this was reported (range 34% to 97%), and the PPV was ≥ 86% in 18 of the 34 studies [40,41,44,46–48,51,53,58–60,62,65,67,68,70,73– 75,78,85,87,88,91,93,95,97–101,107,109,111] that reported this statistic (range 32% to 98%). Amongst these 34 studies, we observed a tendency towards lower PPV when codes for intracranial haemorrhage (ICD-9 432) or ill-defined stroke (ICD-9 437) were included in the search (**S3 Table**). Kappa ranged from 0.79 [108] to 0.81 [78] to 0.87 [88] in the three studies where this was reported.

Validity by Subgroups

The 77 studies included in this review were published over a 40-year period (1976–2015), though 81% of these (n = 62) were published from 1999-onwards. Few studies reported on any longitudinal changes in sensitivity, and few longitudinal trends in PPV were observed after the 77 studies were stratified by period of publication. For instance, amongst the 27 studies reporting on the PPV of ICD-9 434/ICD-10 I63, the PPV ranged from 64% to 100% in the eight-earliest studies (published from 1993 to 1998), from 72% to 100% in the ten middle studies (published from 1999 to 2004), and from 62% to 100% in the nine most-recent studies (published from 2005 to 2015). And among the 26 studies reporting on ICD-9 431/ICD-10 I61, the PPV ranged from 66% to 100% in the 12 earlier studies (published from 1993–2002) and from 65% to 100% in the 13 more-recent studies (published from 2004–2014). Still, several investigators collected data over ten or more years, and some improvements were observed in the PPV for stroke over time. In one study [91], the PPV of ICD 430–438 increased from 48% to 58% between 1970 and 1980, though by 1989 it had decreased slightly, to 54%. Moreover, Derby *et al* [42] found that the PPV of ICD-9 431, 432, 434, 435, 436, or 437 for stroke increased by 20% between 1980 and 1990, while Lakshminarayan *et al* [46] found that the PPV of ICD-9 431, 432, 434, 436, and 437 for stroke increased in this same period by 27% (from 55% in 1980 to 70% in 1990). These studies provide evidence for the accuracy of codes for acute stroke having improved over time.

The accuracy of fatal and non-fatal stroke diagnoses were compared in nine studies [47,51–53,58,60,62,65,91], and amongst these, the accuracy of fatal diagnoses was similar, and often slightly higher, than non-fatal diagnoses (**S4 Table**). Fifteen studies examined cerebrovascular disease or acute stroke as a cause-of-death on death certificates, and the PPV was ≥ 87% in ten of these (range 50% to 100%). However, the sensitivity of vital statistics data for deaths from stroke was ≤ 71% in six of the 10 studies reporting on this (range 32% to 96%), and the sensitivity of hospitalization data for fatal strokes—in the single study where this statistic was reported—was 55% (compared to 68% for non-fatal hospitalizations for stroke) [91].

Most studies examined codes from the ICD 8th and 9th revisions, though 22 studies examined codes from the 10th revision. Just one of these studies [54] was conducted in the United States. Separate validation statistics for ICD-9 and ICD-10 codes were provided by three articles [52,85,104], and, overall, there were few differences in the accuracy of codes from the two revisions. One study was conducted exclusively on males [51], and two were conducted exclusively on females [78,88], and their findings were consistent with those of most studies where both sexes were represented. Sex-stratified statistics were provided by six studies [47,52,84,87,93,105], from which only minor differences in the accuracy of codes for males and females were observed.

Discussion

In performing what is (to our knowledge) the broadest systematic review ever conducted on the validity of stroke diagnoses in administrative data, we observed high PPVs for codes pertaining to the different subtypes of acute stroke. The PPV of SAH codes for an SAH diagnosis was ≥ 93% in most studies, that for the main ischaemic stroke codes was ≥ 82%, and the PPV of the main ICH codes for an ICH diagnosis was ≥ 89%. For diagnoses of fatal stroke, the PPV was ≥ 87% in most studies. The validity of the group of ICD codes corresponding to cerebrovascular disease in general (ICD-9 430–438 and ICD-10 I60-69) was also generally good; sensitivity was ≥ 82% in half the studies where this was reported, specificity and NPV were ≥ 95%, and the PPV of these codes against the broader reference standard of 'any cerebrovascular disease' was ≥ 81% in most studies. However, the PPV was lower (≤ 68% in 12 of 21 studies)

when the reference standard was restricted to acute stroke (defined as SAH, ICH, or ischaemic stroke). Given these findings, we conclude that most diagnoses of fatal stroke in administrative data correspond to true stroke deaths, and that the presence of any code from 430–438 or I60-I69 can be used to rule-in the diagnosis of prevalent cerebrovascular disease. We also conclude that administrative data can be used to identify cases of acute stroke, as long as extraneous codes (i.e. ICD-9 432, 435, 437, and 438; ICD-10 I62, I67, I68, and I69) are excluded.

Only a few studies evaluated the sensitivity of individual codes for stroke but from these, it appears the sensitivity of the main ICD-9 code for ischaemic stroke (434) is suboptimal. However, findings from some studies included in this review suggest that adding ICD-9 code 433 (occlusion of precerebral arteries), and/or 436 (acute but ill-defined stroke) to the search algorithm can help capture more cases of ischaemic stroke at little cost to the PPV. Further support is provided by the fact that the PPV of ICD-9 436 for ischaemic stroke in most applicable studies was $\geq$ 75%. Cases of ischaemic stroke appear to be coded as "acute but ill-defined stroke" much more often than haemorrhagic strokes are. For example, of all strokes that were coded initially as ill-defined, Krarup et al [86] re-classified 57% of these as ischaemic, and just 6% as haemorrhagic. We believe this is because it is harder to make a conclusive diagnosis of ischaemic stroke: neuroimaging can identify bleeds and haemorrhagic lesions more easily than brain infarction, especially within the first twelve hours of onset [35]. Despite the WHO [36] and other criteria calling for cases that fulfill the clinical criteria of stroke, but whose CT is negative for recent brain lesions, to be classified as ischaemic stroke (ICD-9 433 or 434), conservative clinicians and coders may still be inclined to classify these as acute, but ill-defined stroke (ICD-9 436).

The findings of our review are consistent with those of a systematic review of algorithms for identifying acute stroke or TIA in administrative data that was published in 2012 [113]. That review, conducted as part of the US Food and Drug Administration's Mini-Sentinel Program, had a limited scope compared with ours, as it was restricted to evaluations of US and Canadian databases published from 1990-onwards. Still, consistent with our findings, that review found that when individual ICD-9 stroke codes were examined, the PPV's were highest amongst 430, 431, and 434, and much lower when any other code from 430–438 was included in the search algorithm. Further, the authors recommend that codes 433.x1, 434 (excluding 434.x0), and 436 be used when searching for cases of acute ischaemic stroke [113]. They also suggest that when using administrative data to evaluate drug safety, the outcome of interest should be a definite subtype of stroke rather than the broader endpoint of 'any cerebrovascular disease' as defined by ICD-9 430–438 [113].

Regional and Temporal Trends

The PPV of diagnostic codes for stroke appears to have improved over the decades, increasing by 20% in one study (where about 51% of stroke discharges identified in 1980, and about 61% of stroke discharges identified in 1990, were confirmed as stroke) [42], and by 27% in another [46] (from 55% in 1980 to 70% in 1990). Many papers included in this review [42,46,53,93,103] attribute these improvements to advances in neuroimaging technology, and increased use and availability of CT and MRI scanners in medical facilities. For example, the 20% increase in PPV reported by Derby et al [42] (in the northeastern US) was accompanied by a 120% increase in the proportion of potential cases who underwent CT or MRI, and a 60% increase in the proportion seen by a neurologist. CT rates have increased in other countries as well, in Finland from 18% in 1983 to 60% in 1989 [114], and in Sweden from 88% in 1995 to 98% in 2010 [115]. In fact, the introduction of CT scanning in the 1970's is thought to have improved not only the detection of acute strokes, but also the ability of clinicians to make a more precise

diagnosis [53]. For instance, when Lindblad *et al* [53] retrospectively reviewed the stroke codes assigned to cases over 1977–1987, the subtype was reclassified, often on the basis of CT findings, in 27% of the confirmed stroke cases [53]. More strokes were classified as thromboembolic, and fewer as ill-defined, over time in the Derby *et al* study [42], and similar trends have been reported in Denmark, where the incidence of unspecified stroke decreased from 1997 to 2009, while the incidence of ischaemic stroke increased [116].

Regional differences in access-to and use-of neuroimaging may explain other disparities that were observed in the validity of strokes diagnoses, even amongst studies conducted more recently (i.e. 1990 or later). Tolonen *et al* [52] observed that those aged 75 years and older were coded more often with non-specific stroke codes. They attributed this to older individuals being less likely to undergo CT and MRI, which could otherwise aide in making a more precise diagnosis. They also found that the sensitivity and PPV were lower at district hospitals than university hospitals [52], and attributed this to the limited neurological expertise and neuroimaging available at the district hospitals. Liu *et al* [26] also observed higher PPV's amongst the tertiary hospitals than community hospitals, and attributed this to the greater level of testing performed in the tertiary hospitals. As neuroimaging services continue to become more widespread, and more technologies like CT angiography and perfusion imaging [117] emerge, we expect the validity of stroke codes to increase alongside.

It is possible that changes in billing and reimbursement practices, including the introduction of Diagnosis-Related Groups (DRGs) in the US Medicare program in the 1980's, may also have contributed to the assignment of more precise stroke codes over time. To investigate this, Derby *et al* [118] compared how strokes were classified before and after DRGs were introduced, but their findings (increased use of ICD-9 434 and decreased use of ICD-9 436) did not directly correspond with any financial incentives. In addition, while the Medicare DRG system was implemented only in the US, this temporal trend was observed in other countries [115,116] as well.

In our systematic review of the validity of myocardial infarction (MI) diagnoses in administrative data [32], we observed that the accuracy of MI as a cause-of-death on death certificates was generally lower than hospital discharge diagnoses. In contrast, findings from this review suggest that diagnoses of fatal stroke are as accurate, if not more accurate, than diagnoses of non-fatal stroke. One article included in this review, by Tolonen *et al* [52], suggests this is because autopsies can provide additional information that improves the accuracy of the diagnosis. In that study, the sensitivity and PPV for SAH and ICH were markedly high (sensitivity 100% and 90%, and PPV 100% and 100%, for SAH and ICH, respectively) amongst cases diagnosed from outpatient clinics, and all of these were fatal cases that were autopsied [52]. With these findings, researchers should feel confident that most diagnoses of fatal stroke in administrative data correspond to true stroke deaths. However, our findings do suggest that, individually, these databases have suboptimal sensitivity for detecting fatal stroke. In one paper, 79% of deaths that were listed in the mortality register as being from unspecified cerebrovascular disease were re-classified by the investigating physician as deaths from acute stroke [53]. In another paper, only 21% of confirmed stroke deaths in the vital statistics database appeared in the hospitalization database [51]. Thus, when investigating fatal stroke, researchers could improve the ascertainment of cases by linking vital statistics data with hospitalization data, and attributing to stroke any deaths occurring within a certain period (i.e. 28 days, as used in the WHO MONICA Project [36]) of a hospitalization coded for stroke. Sensitivity analyses investigating the impact of longer periods of time (i.e. 90 days) on case ascertainment should be conducted alongside.

Consequences of Using Broad Sets of Cerebrovascular Disease Codes

Many of the studies that reported on the validity of ICD-9 codes 430–438, or ICD-10 codes I60-69, as a group were examining how well these codes performed for detecting preexisting cerebrovascular disease as a comorbidity. Searching for any one of the codes in this group is appropriate when seeking to identify and adjust for preexisting cerebrovascular disease in the analyses of other clinical conditions. The high sensitivity we observed for these codes, and high PPV they had when compared to the reference diagnosis of 'any cerebrovascular disease', provides additional support for this use. However, when studying acute stroke as a primary outcome, this broad group of codes, which includes the codes for non-acute and ill-defined cerebrovascular disease, and the late effects of stroke, should not be used. It is far more difficult to identify risk factors for stroke from this mixture of recent-onset and prevalent cerebrovascular disease because it is unclear which characteristics may have increased the risk of *developing* stroke, versus *surviving* the stroke. Instead, in pharmacoepidemiologic studies and other analyses of risk factors where acute stroke is the primary outcome of interest, and diagnostic specificity is of upmost importance, only stroke-specific codes should be used.

Limitations

We acknowledge some limitations to our systematic review. There is the potential for a language bias as we could not consider articles whose full-text was not available in English. We were also conservative in our definition of acute stroke, and excluded studies that only reported on the validity of diagnostic codes for TIA. Another potential limitation stems from the fact that, even though our database searches were conducted by an experienced librarian, administrative databases are not well catalogued in MEDLINE and EMBASE (e.g. no MeSH term pertaining to "administrative database"). Although the majority of the included studies were located through database searches, our subsequent hand search turned up other relevant articles that had not been indexed under terms relating to Administrative Data or Validation. As a result, despite our extensive hand search, we may have missed some relevant articles if they were not indexed in MEDLINE or EMBASE under a term relating to administrative data or validation. Our findings are also subject to publication bias, wherein reports of stroke codes having poor validity may have been differentially withheld from publication. We feel this is unlikely, however, given that we did locate reports of case definitions (i.e. ICD-9 432 or 433 individually) whose sensitivity and PPV for acute stroke were suboptimal.

Conclusions and Recommendations

Following our analysis of the evidence, we conclude that the diagnostic codes for acute stroke in administrative databases are valid. In fact, advances in neuroimaging and the increased availability of CT scanners may have helped improve diagnosis and coding of acute stroke subtypes over time. However, it is apparent that researchers have been using a variety of codes to identify acute stroke, some of which have suboptimal validity. Based on current evidence, we provide researchers with several recommendations for the use of diagnostic codes to capture cases of stroke in administrative data. We believe the findings of our review will help guide researchers in their efforts to better understand and decrease the burden of stroke.

1. As a group, the range of codes for cerebrovascular disease (ICD-9 430-438/ICD-10 I60-I69) has good sensitivity ($\geq$ 82%), specificity ($\geq$ 95%), and PPV ($\geq$ 81%) for identifying the aggregate diagnosis of acute or preexisting cerebrovascular disease.

2. Codes that pertain to the diagnosis of acute stroke (ICD-9 430/ICD-10 I60, ICD-9 431/ICD-10 I61, ICD-9 434/I63, and ICD-9 436/ICD-10 I64) are highly predictive of true cases of acute stroke of any type, and of the particular subtype.

    ○ These are the codes that should be used when identifying acute stroke as an outcome, especially in pharmacoepidemiologic and other analyses of risk factors where diagnostic specificity is essential.

3. When searching for cases of ischaemic stroke, including both the code for ischaemic stroke (ICD-9 434/ICD-10 I63) and the code for acute-but-ill-defined stroke (ICD-9 463/ICD-10 I64) in the case definition should help capture more cases of ischaemic stroke with little impact on the PPV.

4. Whether identified from hospitalization or vital statistics data, diagnoses of fatal stroke generally correspond to true deaths from stroke.

5. Hospitalization and vital statistics databases should be linked and searched together in order to maximize the capture of stroke deaths.

Supporting Information

S1 Checklist. PRISMA Checklist.
(DOC)

S1 Text. MEDLINE search strategy (inception to November 2010).
(DOCX)

S2 Text. EMBASE search strategy (inception to November 2010).
(DOCX)

S3 Text. MEDLINE search strategy (January 2010 to February 2015).
(DOCX)

S4 Text. EMBASE search strategy (January 2010 to February 2015).
(DOCX)

S5 Text. Data Collection Form.
(DOC)

S1 Table. Item-by-Item QUADAS Breakdown for Each Study.
(DOCX)

S2 Table. Results of Studies Validating Diagnoses of Stroke in Administrative Data.
(DOC)

S3 Table. Results of Studies Validating Sets of Diagnostic Codes for Stroke in Administrative Data.
(DOC)

S4 Table. Results of Studies Validating Diagnoses of Fatal Stroke in Administrative Data.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: DL JAAZ. Performed the experiments: NM VB DL JAAZ. Analyzed the data: NM VB DL JAAZ. Wrote the paper: NM VB DL JAAZ. Final approval of the version of the manuscript to be published: NM VB DL JAAZ.

References

1. WHO | Global Health Estimates [Internet]. [cited 8 May 2014]. Available: http://www.who.int/healthinfo/global_burden_disease/en/

2. Skolarus LE, Burke JF, Brown DL, Freedman VA. Understanding Stroke Survivorship: Expanding the Concept of Poststroke Disability. Stroke J Cereb Circ. 2013; doi: 10.1161/STROKEAHA.113.002874

3. Sprigg N, Selby J, Fox L, Berge E, Whynes D, Bath PMW, et al. Very low quality of life after acute stroke: data from the efficacy of nitric oxide in stroke trial. Stroke J Cereb Circ. 2013; 44: 3458–3462. doi: 10.1161/STROKEAHA.113.002201

4. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart Disease and Stroke Statistics-2015 Update: A Report From the American Heart Association. Circulation. 2014; doi: 10.1161/CIR.0000000000000152

5. Saka O, McGuire A, Wolfe C. Cost of stroke in the United Kingdom. Age Ageing. 2009; 38: 27–32. doi: 10.1093/ageing/afn281 PMID: 19141506

6. Feigin VL, Lawes CMM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurol. 2009; 8: 355–369. doi: 10.1016/S1474-4422(09)70025-0 PMID: 19233729

7. Sun Y, Lee SH, Heng BH, Chin VS. 5-year survival and rehospitalization due to stroke recurrence among patients with hemorrhagic or ischemic strokes in Singapore. BMC Neurol. 2013; 13: 133. doi: 10.1186/1471-2377-13-133 PMID: 24088308

8. Christensen MC, Munro V. Ischemic stroke and intracerebral hemorrhage: the latest evidence on mortality, readmissions and hospital costs from Scotland. Neuroepidemiology. 2008; 30: 239–246. doi: 10.1159/000128323 PMID: 18446043

9. Goulart AC, Bensenor IM, Fernandes TG, Alencar AP, Fedeli LM, Lotufo PA. Early and one-year stroke case fatality in Sao Paulo, Brazil: applying the World Health Organization's stroke STEPS. J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc. 2012; 21: 832–838. doi: 10.1016/j.jstrokecerebrovasdis.2011.04.017

10. Yoneda Y, Okuda S, Hamada R, Toyota A, Gotoh J, Watanabe M, et al. Hospital cost of ischemic stroke and intracerebral hemorrhage in Japanese stroke centers. Health Policy Amst Neth. 2005; 73: 202–211. doi: 10.1016/j.healthpol.2004.11.016

11. Lee WC, Joshi AV, Wang Q, Pashos CL, Christensen MC. Morbidity and mortality among elderly Americans with different stroke subtypes. Adv Ther. 2007; 24: 258–268. PMID: 17565915

12. Gioldasis G, Talelli P, Chroni E, Daouli J, Papapetropoulos T, Ellul J. In-hospital direct cost of acute ischemic and hemorrhagic stroke in Greece. Acta Neurol Scand. 2008; 118: 268–274. doi: 10.1111/j.1600-0404.2008.01014.x PMID: 18384454

13. Christensen MC, Previgliano I, Capparelli FJ, Lerman D, Lee WC, Wainsztein NA. Acute treatment costs of intracerebral hemorrhage and ischemic stroke in Argentina. Acta Neurol Scand. 2009; 119: 246–253. doi: 10.1111/j.1600-0404.2008.01094.x PMID: 18771525

14. Christensen MC, Valiente R, Sampaio Silva G, Lee WC, Dutcher S, Guimarães Rocha MS, et al. Acute treatment costs of stroke in Brazil. Neuroepidemiology. 2009; 32: 142–149. doi: 10.1159/000184747 PMID: 19088487

15. Dodel RC, Haacke C, Zamzow K, Paweilik S, Spottke A, Rethfeldt M, et al. Resource utilization and costs of stroke unit care in Germany. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2004; 7: 144–152.

16. Wei JW, Heeley EL, Jan S, Huang Y, Huang Q, Wang J-G, et al. Variations and determinants of hospital costs for acute stroke in China. PloS One. 2010;5: doi: 10.1371/journal.pone.0013041

17. Asil T, Celik Y, Sut N, Celik AD, Balci K, Yilmaz A, et al. Cost of acute ischemic and hemorrhagic stroke in Turkey. Clin Neurol Neurosurg. 2011; 113: 111–114. doi: 10.1016/j.clineuro.2010.09.014 PMID: 21036465

18. Rossnagel K, Nolte CH, Muller-Nordhorn J, Jungehulsing GJ, Selim D, Bruggenjurgen B, et al. Medical resource use and costs of health care after acute stroke in Germany. Eur J Neurol. 2005; 12: 862–868. doi: 10.1111/j.1468-1331.2005.01091.x PMID: 16241975

19. Wang G, Zhang Z, Ayala C, Dunet DO, Fang J, George MG. Costs of Hospitalization for Stroke Patients Aged 18–64 Years in the United States. J Stroke Cerebrovasc Dis. 2013; doi: 10.1016/j.jstrokecerebrovasdis.2013.07.017

20. Reed SD, Blough DK, Meyer K, Jarvik JG. Inpatient costs, length of stay, and mortality for cerebrovascular events in community hospitals. Neurology. 2001; 57: 305–314. PMID: 11468317

21. Holloway RG, Witter DM Jr, Lawton KB, Lipscomb J, Samsa G. Inpatient costs of specific cerebrovascular events at five academic medical centers. Neurology. 1996; 46: 854–860. PMID: 8618712

22. Lee WC, Christensen MC, Joshi AV, Pashos CL. Long-term cost of stroke subtypes among Medicare beneficiaries. Cerebrovasc Dis Basel Switz. 2007; 23: 57–65. doi: 10.1159/000096542

23. Rinaldi R, Vignatelli L, Galeotti M, Azzimondi G, de Carolis P. Accuracy of ICD-9 codes in identifying ischemic stroke in the General Hospital of Lugo di Romagna (Italy). Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. 2003; 24: 65–69. doi: 10.1007/s100720300074

24. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The accuracy of hospital discharge coding for hemorrhagic stroke. Acta Neurol Belg. 2009; 109: 114–119. PMID: 19681442

25. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke: effect of modifier codes. Stroke J Cereb Circ. 1998; 29: 1602–1604.

26. 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 Basel Switz. 1999; 9: 224–230. 15960

27. Bernatsky S, Lix L, O'Donnell S, Lacaille D, CANRAD Network. Consensus statements for the use of administrative health data in rheumatic disease research and surveillance. J Rheumatol. 2013; 40: 66–73. doi: 10.3899/jrheum.120835 PMID: 23118109

28. Widdifield J, Labrecque J, Lix L, Paterson JM, Bernatsky S, Tu K, et al. Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. Arthritis Care Res. 2013; 65: 1490–1503. doi: 10.1002/acr.21993

29. Barber C, Lacaille D, Fortin PR. Systematic review of validation studies of the use of administrative data to identify serious infections. Arthritis Care Res. 2013; 65: 1343–1357. doi: 10.1002/acr.21959

30. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. PloS One. 2013; 8: e75256. doi: 10.1371/journal.pone.0075256 PMID: 24130696

31. Hudson M, Avina-Zubieta A, Lacaille D, Bernatsky S, Lix L, Jean S. The validity of administrative data to identify hip fractures is high—a systematic review. J Clin Epidemiol. 2013; 66: 278–285. doi: 10.1016/j.jclinepi.2012.10.004 PMID: 23347851

32. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. PloS One. 2014; 9: e92286. doi: 10.1371/journal.pone.0092286 PMID: 24682186

33. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. PloS One. 2014; 9: e104519. doi: 10.1371/journal.pone.0104519 PMID: 25126761

34. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6: e1000097. doi: 10.1371/journal.pmed.1000097 PMID: 19621072

35. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An Updated Definition of Stroke for the 21st Century: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2013; 44: 2064–2089. doi: 10.1161/STR.0b013e318296aeca PMID: 23652265

36. MONICA Manual—Stroke event registration data component [Internet]. [cited 4 Jan 2014]. Available: http://www.thl.fi/publications/monica/manual/part4/iv-2.htm#s2-2

37. Walker AE, Robins M, Weinfeld FD. The National Survey of Stroke. Clinical findings. Stroke J Cereb Circ. 1981; 12: I13–44.

38. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003; 3: 25. PMID: 14606960

39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159–174. PMID: 843571

40. Agrawal N, Johnston SC, Wu YW, Sidney S, Fullerton HJ. Imaging data reveal a higher pediatric stroke incidence than prior US estimates. Stroke J Cereb Circ. 2009; 40: 3415–3421. doi: 10.1161/STROKEAHA.109.564633

41. Broderick J, Brott T, Kothari R, Miller R, Khoury J, Pancioli A, et al. The Greater Cincinnati/Northern Kentucky Stroke Study: preliminary first-ever and total incidence rates of stroke among blacks. Stroke J Cereb Circ. 1998; 29: 415–421.

42. Derby CA, Lapane KL, Feldman HA, Carleton RA. Trends in Validated Cases of Fatal and Nonfatal Stroke, Stroke Classification, and Risk Factors in Southeastern New England, 1980 to 1991 : Data From the Pawtucket Heart Health Program. Stroke. 2000; 31: 875–881. doi: 10.1161/01.STR.31.4.875 PMID: 10753991

43. Gaist D, Vaeth M, Tsiropoulos I, Christensen K, Corder E, Olsen J, et al. Risk of subarachnoid haemorrhage in first degree relatives of patients with subarachnoid haemorrhage: follow up study based on national registries in Denmark. BMJ. 2000; 320: 141–145. PMID: 10634731

44. Holick CN, Turnbull BR, Jones ME, Chaudhry S, Bangs ME, Seeger JD. Atomoxetine and cerebrovascular outcomes in adults. J Clin Psychopharmacol. 2009; 29: 453–460. doi: 10.1097/JCP.0b013e3181b2b828 PMID: 19745645

45. Klatsky AL, Friedman GD, Sidney S, Kipp H, Kubo A, Armstrong MA. Risk of hemorrhagic stroke in Asian American ethnic groups. Neuroepidemiology. 2005; 25: 26–31. doi: 10.1159/000085310 PMID: 15855802

46. Lakshminarayan K, Anderson DC, Jacobs DR, Barber CA, Luepker RV. Stroke rates: 1980–2000: the Minnesota Stroke Survey. Am J Epidemiol. 2009; 169: 1070–1078. doi: 10.1093/aje/kwp029 PMID: 19318614

47. Palmieri L, Barchielli A, Cesana G, de Campora E, Goldoni CA, Spolaore P, et al. The Italian register of cardiovascular diseases: attack rates and case fatality for cerebrovascular events. Cerebrovasc Dis Basel Switz. 2007; 24: 530–539. doi: 10.1159/000110423

48. Shahar E, McGovern PG, Sprafka JM, Pankow JS, Doliszny KM, Luepker RV, et al. Improved survival of stroke patients during the 1980s. The Minnesota Stroke Survey. Stroke J Cereb Circ. 1995; 26: 1–6.

49. Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. Stroke J Cereb Circ. 1999; 30: 736–743.

50. Golomb MR, Garg BP, Saha C, Williams LS. Accuracy and yield of ICD-9 codes for identifying children with ischemic stroke. Neurology. 2006; 67: 2053–2055. doi: 10.1212/01.wnl.0000247281.98094.e2 PMID: 17159120

51. Leppälä 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. PMID: 10204645

52. Tolonen H, Salomaa V, Torppa J, Sivenius J, Immonen-Räihä P, Lehtonen A, et al. The validation of the Finnish Hospital Discharge Register and Causes of Death Register data on stroke diagnoses. Eur J Cardiovasc Prev Rehabil Off J Eur Soc Cardiol Work Groups Epidemiol Prev Card Rehabil Exerc Physiol. 2007; 14: 380–385. doi: 10.1097/01.hjr.0000239466.26132.f2

53. Lindblad U, Råstam L, Ranstam J, Peterson M. Validity of register data on acute myocardial infarction and acute stroke: the Skaraborg Hypertension Project. Scand J Soc Med. 1993; 21: 3–9. PMID: 8469941

54. Brown DL, Senani F Al-, Lisabeth LD, Farnie MA, Colletti LA, Langa KM, et al. Defining cause of death in stroke patients: The Brain Attack Surveillance in Corpus Christi Project. Am J Epidemiol. 2007; 165: 591–596. doi: 10.1093/aje/kwk042 PMID: 17158473

55. Phillips S, Cameron K, Chung C. Stroke surveillance revisited. Can J Cardiol. 1993; 9: 124D.

56. Rao C, Yang G, Hu J, Ma J, Xia W, Lopez AD. Validation of cause-of-death statistics in urban China. Int J Epidemiol. 2007; 36: 642–651. doi: 10.1093/ije/dym003 PMID: 17329316

57. Iso H, Jacobs DR Jr, Goldman L. Accuracy of death certificate diagnosis of intracranial hemorrhage and nonhemorrhagic stroke. The Minnesota Heart Survey. Am J Epidemiol. 1990; 132: 993–998. PMID: 2239914

58. 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. PMID: 1291884

59. Szczesniewska D, Kurjata P, Broda G, Polakowska M, Kupsc W. Comparison of official mortality statistics with data obtained from myocardial infarction and stroke registers. Rev Dépidémiologie Santé Publique. 1990; 38: 435–439.

60. Köster M, Asplund K, Johansson Å, Stegmayr B. Refinement of Swedish administrative registers to monitor stroke events on the national level. Neuroepidemiology. 2013; 40: 240–246. doi: 10.1159/000345953 PMID: 23364278

61. Rampatige R, Gamage S, Peiris S, Lopez AD. Assessing the reliability of causes of death reported by the Vital Registration System in Sri Lanka: medical records review in Colombo. HIM J. 2013; 42: 20–28. PMID: 24067238

62. Ives DG, Fitzpatrick AL, Bild DE, Psaty BM, Kuller LH, Crowley PM, et al. Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study. Ann Epidemiol. 1995; 5: 278–285. PMID: 8520709

63. de Faire U, Friberg L, Lorich U, Lundman T. A validation of cause-of-death certification in 1,156 deaths. Acta Med Scand. 1976; 200: 223–228. PMID: 970229

64. Reggio A, Failla G, Patti F. Reliability of death certificates in the study of stroke mortality. A retrospective study in a Sicilian municipality. Ital J Neurol Sci. 1995; 16: 567–570. PMID: 8613419

65. Appelros P, Terént A. Validation of the Swedish inpatient and cause-of-death registers in the context of stroke. Acta Neurol Scand. 2011; 123: 289–293. doi: 10.1111/j.1600-0404.2010.01402.x PMID: 21361878

66. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? Am J Med Qual Off J Am Coll Med Qual. 2004; 19: 201–206.

67. Aboa-Eboulé C, Mengue D, Benzenine E, Hommel M, Giroud M, Béjot Y, et al. 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: 605–613. doi: 10.1007/s00415-012-6686-0 PMID: 23076827

68. Arnason T, Wells PS, van Walraven C, Forster AJ. Accuracy of coding for possible warfarin complications in hospital discharge abstracts. Thromb Res. 2006; 118: 253–262. doi: 10.1016/j.thromres.2005.06.015 PMID: 16081144

69. Benesch C, Witter DM Jr, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology. 1997; 49: 660–664. PMID: 9305319

70. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care. 2005; 43: 480–485. PMID: 15838413

71. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. BMC Med Res Methodol. 2009; 9: 5–2288–9–5.

72. Cheng C-L, Kao Y-HY, Lin S-J, Lee C-H, Lai ML. Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. Pharmacoepidemiol Drug Saf. 2011; 20: 236–242. doi: 10.1002/pds.2087 PMID: 21351304

73. Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. Health Bull (Edinb). 1996; 54: 402–405.

74. Ellekjaer H, Holmen J, Krüger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. Stroke J Cereb Circ. 1999; 30: 56–60.

75. Ghia D, Thomas PR, Cordato DJ, Worthington JM, Cappelen-Smith C, Griffith N, et al. Validation of emergency and final diagnosis coding in transient ischemic attack: South Western Sydney transient ischemic attack study. Neuroepidemiology. 2010; 35: 53–58. doi: 10.1159/000310338 PMID: 20431303

76. Haesebaert J, Termoz A, Polazzi S, Mouchoux C, Mechtouff L, Derex L, et al. Can hospital discharge databases be used to follow ischemic stroke incidence? Stroke J Cereb Circ. 2013; 44: 1770–1774. doi: 10.1161/STROKEAHA.113.001300

77. Hasan M, Meara RJ, Bhowmick BK. The quality of diagnostic coding in cerebrovascular disease. Int J Qual Health Care J Int Soc Qual Health Care ISQua. 1995; 7: 407–410.

78. Heckbert SR, Kooperberg C, Safford MM, Psaty BM, Hsia J, McTiernan A, et al. Comparison of self-report, hospital discharge codes, and adjudication of cardiovascular events in the Women's Health Initiative. Am J Epidemiol. 2004; 160: 1152–1158. doi: 10.1093/aje/ PMID: 15583367

79. Henderson T, Shepheard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. Med Care. 2006; 44: 1011–1019. PMID: 17063133

80. Hennessy DA, Quan H, Faris PD, Beck CA. Do coder characteristics influence validity of ICD-10 hospital discharge data? BMC Health Serv Res. 2010; 10: 99. doi: 10.1186/1472-6963-10-99 PMID: 20409320

81. Hsieh C-Y, Chen C-H, Li C-Y, Lai M-L. Validating the diagnosis of acute ischemic stroke in a National Health Insurance claims database. J Formos Med Assoc Taiwan Yi Zhi. 2015; 114: 254–259. doi: 10.1016/j.jfma.2013.09.009 PMID: 24140108

82. Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM, Spinelli JJ. Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? J Clin Epidemiol. 2000; 53: 343–349. PMID: 10785564

83. 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: 602–607. PMID: 12063102

84. Jones SA, Gottesman RF, Shahar E, Wruck L, Rosamond WD. Validity of hospital discharge diagnosis codes for stroke: the Atherosclerosis Risk in Communities Study. Stroke J Cereb Circ. 2014; 45: 3219–3225. doi: 10.1161/STROKEAHA.114.006316

85. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. Stroke J Cereb Circ. 2005; 36: 1776–1781. doi: 10.1161/01.STR.0000174293.17959.a1

86. Krarup L-H, Boysen G, Janjua H, Prescott E, Truelsen T. Validity of stroke diagnoses in a National Register of Patients. Neuroepidemiology. 2007; 28: 150–154. doi: 10.1159/000102143 PMID: 17478969

87. Kumamaru H, Judd SE, Curtis JR, Ramachandran R, Hardy NC, Rhodes JD, et al. Validity of claims-based stroke algorithms in contemporary Medicare data: reasons for geographic and racial differences in stroke (REGARDS) study linked with medicare claims. Circ Cardiovasc Qual Outcomes. 2014; 7: 611–619. doi: 10.1161/CIRCOUTCOMES.113.000743 PMID: 24963021

88. Lakshminarayan K, Larson JC, Virnig B, Fuller C, Allen NB, Limacher M, et al. Comparison of Medicare Claims Versus Physician Adjudication for Identifying Stroke Outcomes in the Women's Health Initiative. Stroke. 2014; 45: 815–821. doi: 10.1161/STROKEAHA.113.003408 PMID: 24525955

89. Lambert L, Blais C, Hamel D, Brown K, Rinfret S, Cartier R, et al. Evaluation of care and surveillance of cardiovascular disease: can we trust medico-administrative hospital data? Can J Cardiol. 2012; 28: 162–168. doi: 10.1016/j.cjca.2011.10.005 PMID: 22230034

90. Lee DS, Donovan L, Austin PC, Gong Y, Liu PP, Rouleau JL, et al. Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research. Med Care. 2005; 43: 182–188. PMID: 15655432

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

92. Lentine KL, Schnitzler MA, Abbott KC, Bramesfeld K, Buchanan PM, Brennan DC. Sensitivity of billing claims for cardiovascular disease events among kidney transplant recipients. Clin J Am Soc Nephrol CJASN. 2009; 4: 1213–1221. doi: 10.2215/CJN.00670109 PMID: 19541817

93. Leone MA, Capponi A, Varrasi C, Tarletti R, Monaco F. Accuracy of the ICD-9 codes for identifying TIA and stroke in an Italian automated database. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. 2004; 25: 281–288. doi: 10.1007/s10072-004-0355-8

94. Levy AR, Tamblyn RM, Fitchett D, McLeod PJ, Hanley JA. Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction. Can J Cardiol. 1999; 15: 1277–1282. PMID: 10579743

95. Mayo N, Danys I, Carlton J, Scott S. Accuracy of hospital discharge coding for stroke. Can J Cardiol. 1993; 9: 121D.

96. Olson KL, Wood MD, Delate T, Lash LJ, Rasmussen J, Denham AM, et al. Positive predictive values of ICD-9 codes to identify patients with stroke or TIA. Am J Manag Care. 2014; 20: e27–34. PMID: 24738552

97. Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol. 1999; 52: 199–207. PMID: 10210237

98. Piriyawat P, Smajsová M, Smith MA, Pallegar S, Wabil A Al-, 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: 1062–1069. PMID: 12446264

99. Ramalle-Gomara E, Ruiz E, Serrano M, Bartulos M, Gonzalez M-A, Matute B. Validity of discharge diagnoses in the surveillance of stroke. Neuroepidemiology. 2013; 41: 185–188. doi: 10.1159/000354626 PMID: 24051447

100.   Reker DM, Hamilton BB, Duncan PW, Yeh SC, Rosen A. Stroke: who's counting what? J Rehabil Res Dev. 2001; 38: 281–289. PMID: 11392661

101.   Roumie CL, Mitchel E, Gideon PS, Varas-Lorenzo C, Castellsague J, Griffin MR. Validation of ICD-9 codes with a high positive predictive value for incident strokes resulting in hospitalization using Medicaid health data. Pharmacoepidemiol Drug Saf. 2008; 17: 20–26. doi: 10.1002/pds.1518 PMID: 17979142

102.   Singh B, Singh A, Ahmed A, Wilson GA, Pickering BW, Herasevich V, et al. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. Mayo Clin Proc. 2012; 87: 817–824. doi: 10.1016/j.mayocp.2012.04.015 PMID: 22958988

103.   Sinha S, Myint PK, Luben RN, Khaw K-T. Accuracy of death certification and hospital record linkage for identification of incident stroke. BMC Med Res Methodol. 2008; 8: 74. doi: 10.1186/1471-2288-8-74 PMID: 19000303

104.   So L, Evans D, Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. BMC Health Serv Res. 2006; 6: 161. PMID: 17173686

105.   Soo M, Robertson LM, Ali T, Clark LE, Fluck N, Johnston M, et al. Approaches to ascertaining comorbidity information: validation of routine hospital episode data with clinician-based case note review. BMC Res Notes. 2014; 7: 253. doi: 10.1186/1756-0500-7-253 PMID: 24751124

106.   Spolaore P, Brocco S, Fedeli U, Visentin C, Schievano E, Avossa F, et al. Measuring accuracy of discharge diagnoses for a region-wide surveillance of hospitalized strokes. Stroke J Cereb Circ. 2005; 36: 1031–1034. doi: 10.1161/01.STR.0000160755.94884.4a

107.   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: 8–14. doi: 10.1161/CIRCOUTCOMES.113.000371 PMID: 25587093

108.   Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. Stroke J Cereb Circ. 2002; 33: 2465–2470.

109.   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 EMRALD as a reference standard. Can J Cardiol. 2013; 29: 1388–1394. doi: 10.1016/j.cjca.2013.07.676 PMID: 24075778

110.   Wahl PM, Rodgers K, Schneeweiss S, Gage BF, Butler J, Wilmer C, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. Pharmacoepidemiol Drug Saf. 2010; 19: 596–603. doi: 10.1002/pds.1924 PMID: 20140892

111.   Wildenschild C, Mehnert, Frank W. Thomsen R, Iversen H, Vestergaard K, Ingeman, Annette A, et al. Registration of acute stroke: validity in the Danish Stroke Registry and the Danish National Registry of Patients. Clin Epidemiol. 2013; 27. doi: 10.2147/CLEP.S50449

112.   Wu C-S, Lai M-S, Gau SS-F, Wang S-C, Tsai H-J. Concordance between patient self-reports and claims data on clinical diagnoses, medication use, and health system utilization in Taiwan. PloS One. 2014; 9: e112257. doi: 10.1371/journal.pone.0112257 PMID: 25464005

113.   Andrade SE, Harrold LR, Tjia J, Cutrona SL, Saczynski JS, Dodd KS, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data: DETECTION OF CEREBROVASCULAR ACCIDENT AND TRANSIENT ISCHEMIC ATTACK IN CLAIMS. Pharmacoepidemiol Drug Saf. 2012; 21: 100–128. doi: 10.1002/pds.2312 PMID: 22262598

114.   Sarti C, Tuomilehto J, Sivenius J, Kaarsalo E, Narva EV, Salmi K, et al. Declining trends in incidence, case-fatality and mortality of stroke in three geographic areas of Finland during 1983–1989. Results from the FINMONICA stroke register. J Clin Epidemiol. 1994; 47: 1259–1269. PMID: 7722562

115.   Appelros P, Jonsson F, Åsberg S, Asplund K, Glader E-L, Åsberg KH, et al. Trends in stroke treatment and outcome between 1995 and 2010: observations from Riks-Stroke, the Swedish stroke register. Cerebrovasc Dis Basel Switz. 2014; 37: 22–29. doi: 10.1159/000356346

116.   Demant MN, Andersson C, Ahlehoff O, Charlot M, Olesen JB, Gjesing A, et al. Temporal trends in stroke admissions in Denmark 1997–2009. BMC Neurol. 2013; 13: 156. doi: 10.1186/1471-2377-13-156 PMID: 24171730

117.   Vagal A, Meganathan K, Kleindorfer DO, Adeoye O, Hornung R, Khatri P. Increasing Use of Computed Tomographic Perfusion and Computed Tomographic Angiograms in Acute Ischemic Stroke From 2006 to 2010. Stroke. 2014; 45: 1029–1034. doi: 10.1161/STROKEAHA.113.004332 PMID: 24569815

118.   Derby CA, Lapane KL, Feldman HA, Carleton RA. Possible Effect of DRGs on the Classification of Stroke : Implications for Epidemiological Surveillance. Stroke. 2001; 32: 1487–1491. doi: 10.1161/01.STR.32.7.1487 PMID: 11441190