HHS Public Access

Author manuscript

Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2024 February 16.

Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2023 December ; 32(12): 1439–1445. doi:10.1002/pds.5675.

Validation of ICD-10 hospital discharge diagnosis codes to identify incident and recurrent ischemic stroke from a US integrated healthcare system

Abraelle M Shirley[1], **Kerresa L Morrisette, MPH**[1], **Soon Kyu Choi, MPP, MSc**[1], **Kristi Reynolds, PhD**[1,2], **Hui Zhou, PhD**[1,2], **Mengnan M Zhou**[1], **Rong Wei, MS**[1], **Yiyi Zhang, PhD**[3], **Pamela Cheng, DO**[4], **Eric Wong, MD**[4], **Navdeep Sangha, MD**[4,5], **Jaejin An, PhD**[1,2]

[1]Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

[2]Department of Health Systems Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

[3]Department of Medicine, Columbia University, New York, NY, United States

[4]Department of Neurology, Los Angeles Medical Center, Southern California Permanente Medical Group, Los Angeles, CA, United States

[5]Department of Clinical Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

Abstract

Purpose: This study validated incident and recurrent ischemic stroke identified by International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10) hospital discharge diagnosis codes.

Methods: Using electronic health records (EHR) of adults (≥18 years) receiving care from Kaiser Permanente Southern California with ICD-10 hospital discharge diagnosis codes of ischemic stroke (I63.x, G46.3, G46.4) between October 2015 and September 2020, we identified 75 patients with both incident and recurrent stroke events (total 150 cases). Two neurologists independently evaluated validity of ICD-10 codes through chart reviews.

Results: The positive predictive value (PPV, 95% CI) for incident stroke was 93% (95% CI: 88%, 99%) and the PPV for recurrent stroke was 72% (95% CI: 62%, 82%). The PPV for recurrent stroke improved after applying a gap of 20 days (PPV of 75%; 95% CI: 63%, 87%) or

Correspondence: Jaejin An, PhD, Department of Research & Evaluation, Kaiser Permanente, 100 S. Los Robles Ave., Pasadena, CA 91101, jaejin.x.an@kp.org.

Ethics statement

The study was approved by the KPSC Institutional Review Board.

Conflict of interest

Dr. An reports contracts from AstraZeneca and Bayer AG outside the scope of this study.

Dr. Reynolds reports research funding received through her institution from Novartis and CSL Behring, LLC outside the scope of this study.

removing hospital admissions related to stroke-related procedures (PPV of 78%; 95% CI: 68%, 88%).

Conclusion: The ICD-10 hospital discharge diagnosis codes for ischemic stroke showed a high PPV for incident cases, while the PPV for recurrent cases were less optimal. Algorithms to improve the accuracy of ICD-10 codes for recurrent ischemic stroke may be necessary.

Keywords

recurrent stroke; ICD-10 diagnosis codes; positive predictive value

Introduction

Stroke is a leading cause of morbidity and mortality in the US and worldwide. [1] Though previous studies validated International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9) stroke codes,[2-4] few studies have assessed the validity of identifying ischemic stroke across International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10) codes.[5-9] Furthermore, to date, no studies have validated ICD-10 codes for recurrent ischemic stroke. This study validated ICD-10 hospital discharge diagnosis codes of incident and recurrent ischemic stroke.

Methods

Study population

The study population included adult members with an ischemic stroke diagnosis within the Kaiser Permanente Southern California (KPSC) integrated healthcare delivery system. KPSC provides services to over 4.8 million members at 233 medical offices throughout Southern California. Using electronic health records (EHR) data, we limited our sample size to members with ≥1 principal hospital discharge ICD-10 diagnosis codes of ischemic stroke, I63.x (cerebral infarction), G46.3 (brain stem stroke syndrome), or G46.4 (cerebellar stroke syndrome) from October 1, 2015 to September 30, 2020. We further limited the study population to members with ≥2 principal hospital discharge ICD-10 codes of ischemic stroke that were at least two days apart. We identified 1,347 patients with 1,422 hospital admissions for potential stroke events. Among the 1,347 patients, 75 patients had ≥2 hospital admissions for potential stroke events (all had I63.x codes). We conducted chart review of first (incident) and second (recurrent) potential stroke events of these 75 patients (total 150 cases).

Case Adjudication

Two neurologists (PC, EW) independently reviewed EHR data associated with the 150 incident and recurrent ischemic stroke cases using the 2013 American Heart Association/ American Stroke Association[1] case definition of ischemic stroke. Ischemic stroke is defined as an episode of neurological dysfunction caused by focal cerebral, spinal, or retinal infarction based on 1) pathological, imaging, or other objective evidence of cerebral, spinal cord, or retinal focal ischemic injury in a defined vascular distribution or 2) clinical evidence of cerebral, spinal cord, or retinal focal ischemic injury based on symptoms persisting ≥24

hours or until death, and other etiologies excluded.[10] All associated admission records were examined: progress notes, discharge summary, magnetic resonance imaging, computerized tomography scan, medications, and laboratory values. A third neurologist (NS) reviewed discrepant adjudications to finalize the result. Adjudicators provided rationale for incorrect stroke diagnoses.

Statistical Analysis

We evaluated inter-rater reliability by calculating Gwet's agreement coefficient (AC1) by dividing the total number of cases by the number of cases in agreement between two reviewers.[11] Positive predictive value (PPV) for incident and recurrent ischemic stroke was calculated by dividing the number of confirmed cases by the total number of cases reviewed, respectively. PPVs were stratified by each ICD-10 code and by days between the first and second stroke events (20-, 30-, 45-, 60-, 90-, 180- or 365-day gap). Statistical significance of differences in PPVs was determined using Fisher's exact tests. As an exploratory analysis, we calculated PPVs after removing hospital admissions related to stroke-related procedures (CPT codes 35301 (thromboendarterectomy), 37215 and 37217 (transcatheter placement of intravascular stent).

Results

Inter-Rater Reliability

In the adjudication, there were 6 (8%) discrepant cases between reviewers for incident stroke (Appendix 1) and 17 (23%) discrepant cases for recurrent stroke (Appendix 2). Gwet's AC1 results showed comparison between the two reviewers on their determination of ischemic stroke: the AC1 for incident stroke was 0.89 (95% CI: 0.80, 0.97) and recurrent stroke was 0.72 (95% CI: 0.73, 0.90) (Appendix 3).

Case Adjudication

Through case adjudication of 75 incident stroke cases, 72 cases (96%) were confirmed as any stroke and 70 cases (93%) were confirmed as ischemic stroke (Appendix 1). Most (85%, 64 cases) met both objective and clinical ischemic stroke adjudication criteria. Through chart review of the 75 recurrent stroke cases, 55 (73%) were confirmed cases as ischemic stroke, and 54 (72%) were confirmed as recurrent ischemic stroke (Appendix 2a, 2b). The median number of days between the first and second stroke events was 48 (IQR:16, 249). Most recurrent stroke cases (70%, 52 cases) occurred after 21 days from the first event.

Adjudicators identified 2 (3%) incorrect incident and 21 (28%) incorrect recurrent ICD-10 stroke diagnosis cases. The most common reasons for incorrect recurrent stroke diagnoses were stroke related procedures (7 cases) including carotid endarterectomy (CEA), transcarotid artery revascularization (TCAR), carotid artery stenting (CAS), late effect of previous stroke (5 cases), and recrudescence of stroke (4 cases).

Positive Predictive Value (PPV)

Overall, the PPV for incident stroke was 93% (95% CI: 88%, 99%) and the PPV for recurrent stroke was 72% (95% CI: 62%, 82%) (Table 1). The PPV calculation per ICD-10

code showed that I63.9 (cerebral infarction, unspecified) had a PPV of 89% (95% CI: 80%, 99%) for incident stroke, and a PPV of 62% (95% CI: 44%, 80%) for recurrent stroke. When considering gaps between two stroke events, the PPV for gaps of under 20 days and over 20 days, were 65% (95% CI: 46%, 85%) and 75% (95% CI: 63%, 87%), respectively with an increase in PPV of 10% (95% CI: −13%, 33%). The PPV before and after excluding hospitalizations with stroke-related procedures from recurrent stroke cases were 72% (95% CI: 62%, 82%) and 78% (95% CI: 68%, 88%), respectively with an increase in PPV of 6% (95% CI: −8%, 20%).

Discussion

This study validated ICD-10 hospital discharge diagnosis codes to identify incident and recurrent ischemic stroke from a US integrated healthcare system. It benefitted from access to robust EHR data, and the expertise of trained neurologists performing double adjudication. The neurologist reviewers had high agreement on incident stroke cases identified via ICD-10 codes. These results suggest that ICD-10 hospital discharge principal diagnosis codes for incident ischemic stroke are reliable with a PPV of 93%, consistent with previous findings ranging from 70%-98.6% PPV. [5-9]

Identifying recurrent ischemic stroke via ICD-10 diagnosis codes seems more challenging, particularly when the recurrent stroke occurs immediately after the incident stroke. For recurrent ischemic stroke, compared to incident stroke, adjudicators encountered more false positives, with the PPV of 72%. Epidemiological literature suggests applying a gap of 21-28 days between the incident and recurrent stroke.[12,13] When applying a 20-day gap between two stroke events, the PPV increased from 65% for events occurring within 20 days to 75% for events occurring more than 20-days apart, although this increase did not reach statistical significance. When exploring alternative gaps between two stroke events and stroke-related procedures, increasing the gap between two events did not significantly improve the PPVs compared with not considering any gap. Additionally, when excluding hospitalizations with stroke-related procedures from recurrent stroke cases, the PPV for recurrent stroke increased to 78%; this increase was not statistically significant. While the PPV of 78% is a lower range of previously reported PPVs for incident stroke, it is higher than reported PPVs (12.9% to 72.5%) for transient ischemic attack.[14]

Though the PPV of ICD-10 codes for ischemic incident stroke has been examined, [6] to date, no studies have examined the PPV of ICD-10 codes for recurrent ischemic stroke. This study has some limitations. Since the results focus only on one healthcare system in a specific region, results may not be generalizable to other settings. We did not include claims data from external hospitals. While findings for incident ischemic stroke are reassuring, future studies investigating a larger array of healthcare networks would be useful in validating ICD-10 codes across settings. Additionally, we were not able to differentiate recrudescence (the return of ischemic stroke symptoms previously resolved) from recurrent stroke using structured data. Though we were able to exclude hospitalizations with stroke-related procedures using codes, recrudescence events were only identified through chart reviews and often incorrectly coded as a new stroke event. Finally, we limited our review to first and second stroke cases, though a few patients had three or more stroke events.

In conclusion, we found a high PPV for incident ischemic stroke using ICD-10 hospital discharge diagnosis codes, but a less optimal PPV for recurrent ischemic stroke. Algorithms to improve ICD-10 code accuracy, such as removing stroke-related procedures to identify recurrent ischemic stroke, may be necessary.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported by a research project grant from the National Heart, Lung and Blood Institute of the National Institutes of Health (R01HL155081).

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation. 2020;141:e139–e596. doi: 10.1161/cir.0000000000000757 [PubMed: 31992061]
2. Ammann EM, Leira EC, Winiecki SK, Nagaraja N, Dandapat S, Carnahan RM, Schweizer ML, Torner JC, Fuller CC, Leonard CE, et al. Chart validation of inpatient ICD-9-CM administrative diagnosis codes for ischemic stroke among IGIV users in the Sentinel Distributed Database. Medicine. 2017;96:e9440. doi: 10.1097/md.0000000000009440 [PubMed: 29384925]
3. 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. 2014;45:3219–3225. doi: 10.1161/strokeaha.114.006316 [PubMed: 25190443]
4. Olson KL, Wood MD, Delate T, Lash LJ, Rasmussen J, Denham AM, Merenich JA. Positive predictive values of ICD-9 codes to identify patients with stroke or TIA. The American journal of managed care. 2014;20:e27–34. [PubMed: 24738552]
5. Hsieh MT, Hsieh CY, Tsai TT, Wang YC, Sung SF. Performance of ICD-10-CM Diagnosis Codes for Identifying Acute Ischemic Stroke in a National Health Insurance Claims Database. Clinical epidemiology. 2020;12:1007–1013. doi: 10.2147/clep.S273853 [PubMed: 33061648]
6. Hsieh MT, Huang KC, Hsieh CY, Tsai TT, Chen LC, Sung SF. Validation of ICD-10-CM Diagnosis Codes for Identification of Patients with Acute Hemorrhagic Stroke in a National Health Insurance Claims Database. Clinical epidemiology. 2021;13:43–51. doi: 10.2147/clep.S288518 [PubMed: 33469381]
7. Sedova P, Brown RD Jr., Zvolsky M, Kadlecova P, Bryndziar T, Volny O, Weiss V, Bednarik J, Mikulik R. Validation of Stroke Diagnosis in the National Registry of Hospitalized Patients in the Czech Republic. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association. 2015;24:2032–2038. doi: 10.1016/j.jstrokecerebrovasdis.2015.04.019 [PubMed: 26139454]
8. 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. doi: 10.1161/01.STR.0000174293.17959.a1 [PubMed: 16020772]
9. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA. Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review. PloS one. 2015;10:e0135834. doi: 10.1371/journal.pone.0135834 [PubMed: 26292280]
10. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, Elkind MS, George MG, Hamdan AD, Higashida RT, 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 [PubMed: 23652265]

11. Gwet K. Inter-Rater Reliability: Dependency on Trait Prevalence and Marginal Homogeneity. Statistical Methods For Inter-Rater Reliability Assessment Series. 2002;2:1–9.
12. Coull AJ, Rothwell PM. Underestimation of the early risk of recurrent stroke: evidence of the need for a standard definition. Stroke. 2004;35:1925–1929. doi: 10.1161/01.Str.0000133129.58126.67 [PubMed: 15192248]
13. Sarti C, Stegmayr B, Tolonen H, Mähönen M, Tuomilehto J, Asplund K. Are changes in mortality from stroke caused by changes in stroke event rates or case fatality? Results from the WHO MONICA Project. Stroke. 2003;34:1833–1840. doi: 10.1161/01.Str.0000081224.15480.52 [PubMed: 12855832]
14. Kandiyali R, Lasserson DS, Whiting P, Richards A, Mant J. Predictive values of referrals for transient ischaemic attack from first-contact health care: a systematic review. The British journal of general practice : the journal of the Royal College of General Practitioners. 2017;67:e871–e880. doi: 10.3399/bjgp17X693677 [PubMed: 29158247]

Key Points

- Little is known of the validity of ICD-10 codes for recurrent ischemic stroke. Using ICD-10 codes, we identified incident and recurrent ischemic stroke cases of 75 patients through electronic health records and neurologists adjudicated these stroke cases through chart reviews.
- Our results show that ICD-10 diagnosis codes were correctly identified for most incident ischemic stroke cases but did not perform as well for recurrent ischemic stroke cases. Removing stroke-related procedures improved the validity of ICD-10 codes for recurrent ischemic stroke.
- Algorithms to further improve the accuracy of ICD-10 codes for recurrent ischemic stroke may be necessary.

Table 1.

Positive Predictive Values (PPVs) for Incident and Recurrent Stroke

	Confirmed Cases (Yes/No)	PPV (95% CI)
Incident Stroke		
All codes	Yes=70, No=5	0.93 (0.88, 0.99)
I63.0x	Yes=1, No=0	1.0 (1.0, 1.0)
I63.1x	Yes=3, No=1	0.75 (0.33, 1.0)
I63.2x	Yes=4, No=0	1.0 (1.0, 1.0)
I63.3x	Yes=2, No=0	1.0 (1.0, 1.0)
I63.4x	Yes=12, No=0	1.0 (1.0, 1.0)
I63.5x	Yes=12, No=0	1.0 (1.0, 1.0)
I63.8x	Yes=2, No=0	1.0 (1.0, 1.0)
I63.9	Yes=34, No=4	0.89 (0.80, 0.99)
Recurrent Stroke		
All codes	Yes=54, No=21	0.72 (0.62, 0.82)
I63.0x	Yes=3, No=1	0.75 (0.33, 1.0)
I63.1x	Yes=2, No=0	1.0 (1.0, 1.0)
I63.2x	Yes=4, No=8	0.33 (0.07, 0.60)
I63.3x	Yes=0, No=0	N/A
I63.4x	Yes=17, No=0	1.0 (1.0, 1.0)
I63.5x	Yes=9, No=1	0.90 (0.71, 1.0)
I63.8x	Yes=1, No=0	1.0 (1.0, 1.0)
I63.9	Yes=18, No=11	0.62 (0.44, 0.80)
Recurrent Stroke Stratified by Gaps between the First and Second Stroke		
≤20 Days	Yes=15, No=8	0.65 (0.46, 0.85)
>20 Days	Yes=39, No=13	0.75 (0.63, 0.87)
≤30 Days	Yes=21, No=9	0.70 (0.54, 0.86)
>30 Days	Yes=33, No=12	0.73 (0.60, 0.86)
≤45 Days	Yes=26, No=11	0.70 (0.56, 0.85)
>45 Days	Yes=28, No=10	0.74 (0.60, 0.88)
≤60 Days	Yes=30, No=13	0.70 (0.56, 0.83)
>60 Days	Yes=24, No=8	0.75 (0.60, 0.90)
≤90 Days	Yes=36, No=13	0.73 (0.61, 0.86)
>90 Days	Yes=18, No=8	0.69 (0.51, 0.87)
≤180 Days	Yes=40, No=14	0.74 (0.62, 0.86)
>180 Days	Yes=14, No=7	0.67 (0.47, 0.87)
≤365 Days	Yes=44, No=17	0.72 (0.61, 0.83)
>365 Days	Yes=10, No=4	0.71 (0.48, 0.95)
Recurrent Stroke after Removing Hospital Admissions related to Stroke-Related Procedures *	Yes=50, No=14	0.78 (0.68, 0.88)

All reviewed cases included ICD codes I63.x. There was no recurrent stroke case identified by G46.3 or G46.42 codes. *Carotid endarterectomy, transcarotid artery revascularization, or carotid artery stenting procedures (CPT codes: 35301, 37215, 37217)