Validity of International Classification of Disease Codes to Identify Ischemic Stroke and Intracranial Hemorrhage among Individuals with Associated Diagnosis of Atrial Fibrillation:

Thigpen et al: ICD-9 Stroke Case and Atrial Fibrillation Accuracy

Jonathan L. Thigpen, Pharm.D[1], **Chrisly Dillon, MD**[2], **Kristen B. Forster, MBA**[6], **Lori Henault, MPH**[3], **Emily K. Quinn, MA**[5], **Yorghos Tripodis, Ph.D**[4], **Peter B. Berger, MD**[6], **Elaine M. Hylek, MD**[3], and **Nita A. Limdi, PhD**[2]

[1]School of Pharmacy, Notre Dame of Maryland University, Baltimore, MD

[2]Department of Neurology, University of Alabama at Birmingham, Birmingham, AL

[3]Boston University Medical Center, Boston, MA

[4]Department of Biostatistics, Boston University, Boston, MA

[5]School of Public Health, Boston University, Boston, MA

[6]Geisinger Health System, Danville, PA

Abstract

Background—Due to its association with death and disability, stroke is a focus of outcomes in atrial fibrillation (AF) research. International Classification of Disease 9th (ICD-9) edition codes are commonly used to identify stroke in research, particularly in large administrative data. We sought to assess the validity of ICD-9 codes in stroke case ascertainment and for AF across three institutions.

Methods and Results—Participating centers included Boston Medical Center (safety net hospital), Geisinger Health System (rural Pennsylvania), and the University of Alabama (academic center in the southeastern "stroke belt"). ICD-9 codes for ischemic stroke (433-434, 436) and intracranial hemorrhage (ICH) (430-432) identified 1,812 stroke cases with an associated code for atrial fibrillation (427.31) from 2006-2010. Cases were vetted through chart review with final adjudication by a stroke neurologist.

Review deemed 94.2% of ICD-9 identified stroke cases valid with decreased accuracy for concurrent AF diagnosis (82.28%) and stroke attributable to AF (72.8%). Among events with "without infarction" modifiers, 7.2% were valid strokes. ICD-9 stroke code accuracy did not differ by stroke type or site. Stroke code 434 displayed higher accuracy than 433 (94.4% vs. 85.2%; p<0.01) and primary stroke codes were more accurate than non-primary codes (97.2% vs. 83.7%; p<0.0001).

Corresponding Author: Jonathan L Thigpen, Pharm.D, 4701 North Charles Street Baltimore, MD 21210, Phone: (410) 532-5064, Fax: (410) 532-5578, jthigpen@ndm.edu.

Conclusions—Using ICD-9 stroke and AF codes to identify patients with stroke plus AF resulted in inaccuracies. Given the expanded financial and policy implications of patient-oriented research, conclusions derived solely from administrative data without validation of outcome events should be interpreted with caution.

Keywords

ischemic stroke; intracranial hemorrhage; hemorrhagic stroke; atrial fibrillation; International Classification of Disease 9th Edition Codes

Stroke is a leading cause of death and long-term disability affecting approximately 795,000 people every year in the United States.[1] A major contributor to ischemic stroke (IS) risk, atrial fibrillation (AF), is associated with a five times higher risk. The attributable risk of IS among individuals with AF varies significantly with age. For those individuals age 50 to 59 years, the proportion of strokes attributable to AF is 1.5% while among those 80 years or older the attributable risk is 24%.[2, 3] In addition to increased risk of IS in AF, stroke in the context of AF is more debilitating and associated with a 30-day mortality of 24%.[4] Oral anticoagulants are highly efficacious in reducing the risk of IS in AF, however their effectiveness in clinical practice is challenged by hemorrhage, particularly intracranial hemorrhage (ICH), with its associated 46% fatality rate.[5]

As clinical and economic pressures mount, the expedient research strategy will be to rely on administrative data and forego primary data validation. This is potentially problematic because large administrative databases often capture outcomes using International Classification of Diseases 9th edition (ICD-9) codes. Therefore, in conducting stroke research, perhaps a more fundamental question involves the accuracy of ICD-9 codes to identify stroke events and associated AF. To date, there has been limited research on the accuracy of ICD-9 codes to identify IS and ICH in the setting of AF.[6, 7] Given the expanded implications of patient-oriented research on economic decisions, health policy, and performance measurement, we sought to define the validity of ICD-9 codes for stroke cases associated with AF across three health care systems.

Methods

Identification of Stroke Cases

Participating centers included Boston Medical Center, Geisinger Health System in Pennsylvania, and the University of Alabama at Birmingham. Boston Medical Center is a teaching hospital and considered the major safety net hospital for the city of Boston. Geisinger is a highly integrated health care system that serves a predominantly rural population. The University of Alabama is part of the southeastern "stroke belt" and cares for a diverse patient population. The study was approved by the Institutional Review Board at each respective site.

Stroke events were identified using discharge ICD-9 codes (Table 1) for IS ("with infarction" 433-434, 436) or ICH (430-432) from hospital admissions over a 5-year period (2006-2010). To ensure the most comprehensive search, we included ICD-9 codes identified in any position-primary (first position) or other (any position other than the first). Among

these ICD-9 identified stroke events, those cases associated with an AF ICD-9 code (427.31) were subject to in-depth medical record review. To validate the AF diagnosis, we required electrocardiogram (ECG) evidence of AF during the stroke admission or within 6 months of the stroke admission if the AF was not permanent. If neither of these criterion was fulfilled, we sought ECG evidence within 90 days of the stroke discharge. A valid IS was defined as a focal neurologic deficit of sudden onset that persisted for >24 hours, corresponded to a vascular territory, and was not explained by other etiologies.[8-10] ICH was defined by hemorrhage on imaging (CT or MRI) not associated with major trauma (ICD-9 codes 852.1, 852.3, 852.5, 853.1) or a surgical complication. Strokes associated with prosthetic heart valves were excluded. For patients experiencing multiple strokes within the 5-year study period, only the first valid stroke event was included. Stroke cases were adjudicated by site investigators (PB, EH, NL) based on review of the medical record. Questionable cases were resolved by consensus of the adjudication committee and a stroke neurologist. All validated IS and ICH cases were assigned a modified Rankin score by the site neurologist.[11] ICD-9 codes were used to initially identify stroke and AF. Covariates such as hypertension, diabetes, and heart failure were abstracted manually while most demographic covariates were automatically populated from electronic records. To further assess errors of ICD-9 coding, we also determined the accuracy of the "without infarction" modifier: 433.00, 433.10, 433.20, 433.30, 434.00, 434.10, and 434.90.

Statistical Analyses

Analyses were stratified according to event type, i.e., IS or ICH, and event position, primary or other. Validation rates were compared across the three clinical sites and across calendar years. For IS, we also assessed the accuracy of specific ICD-9 codes (433, 434, and 436). We subsequently assessed the validity of the AF diagnosis ICD-9 code (427.31) based on ECG. Final validation assessment excluded IS due to a different mechanism (i.e. strokes determined by medical record review to be attributable to vascular procedures, tumors, infection, or vasculitis) and ICH secondary to a surgical procedure or major trauma that had not been initially excluded due to lack of an accompanying ICD-9 trauma code. To further assess errors of ICD-9 coding, we also determined the accuracy of the "without infarction" modifier.

Data are displayed as counts and percentages. Bivariate analyses were performed and differences assessed with Chi-square tests for event type, event coding position, clinical site, and IS code. We calculated the positive predictive value (PPV) for each of these categories. All analyses were performed using SAS software (version 9.3; SAS Institute, Cary, NC).

Results

Our electronic search identified a total of 1,812 stroke events with AF diagnosis (1,358 IS and 454 ICH). Subsequent steps in the case-validation algorithm are displayed in the Figure. Manual review confirmed 1,706 (94.2%) events to be valid stroke events. After excluding patients without ECG-confirmed AF (n=217), 1,489 patients with manually confirmed stroke and AF remained (1,136 IS and 353 ICH). The PPV of ICD-9 stroke codes alone was 94.2%, and when combined with validation of ICD-9 AF codes, the total accuracy dropped

to 82.2%. Additional manual review excluded IS events due to a different mechanism (e.g., vascular procedure, endocarditis, tumor, sepsis, vasculitis) or ICH resulting from major head trauma with skull fracture. These additional trauma-related ICH cases were not initially filtered as they lacked the corresponding trauma codes. This final step yielded a total of 1,320 strokes (1,000 IS and 320 ICH) confirmed in the setting of AF (Table 2). The accuracy of ICD-9 stroke codes following these validation exercises fell to 72.8%. The final cohort included 33 events with no mutually exclusive coding position (i.e. had a combination of IS and ICH codes).

The validation rates of ICD-9 stroke codes were similar across event type, IS (94.0%) and ICH (94.7%) (p=0.55; Table 3). Validation rates were also similar across the three clinical sites: Boston Medical Center (94.9%), Geisinger Health System (93.3%), and University of Alabama (94.3%) (p = 0.55), and year of stroke admission. Approximately 15% of events designated as 433 were not valid IS. Compared to stroke code 434, this difference was statistically significant (85.2% vs. 94.4%; p=0.003). Accuracy of stroke case ascertainment was influenced by the ICD-9 coding position (primary vs. other). Events possessing the stroke ICD-9 code in the primary position (n=1368) were more likely to be valid strokes than events recorded in a non-primary (or other) position (n=399) (97.2% vs. 83.7%; p<0.0001). Overall 20% of valid acute strokes were coded in the non-primary position.

In addition to the 1,812 stroke events identified in our initial electronic search, as a separate exercise, we identified an additional 458 potential events with the "without infarction" ICD-9 codes (433.00, 433.10, 433.20, 433.30, 434.00, 434.10, and 434.90). Of these, 33 (7.2%) were deemed to be valid strokes in the setting of AF after review of primary data. The position of the ICD-9 code did not significantly influence the accuracy of these codes (p=0.42).

Discussion

In this study, we found less than optimal accuracy of ICD-9 coded stroke events when compared to the gold standard of medical record primary data review. For both IS "with infarction" and ICH, the proportion of events meeting the respective definitions was 94.2%. This was remarkably consistent across three sites. The accuracy of case identification was higher (97.2%) for events with the stroke related ICD-9 code in the primary position compared to those that were identified with codes in the non-primary position (83.7%). However, reliance on a search strategy that identifies cases based solely on ICD-9 codes in the primary position could miss a substantial proportion (20%) of valid strokes.

When assessing the accuracy of ICD-9 AF diagnosis, we determined an additional 12% of stroke events to have no ECG-confirmed AF at the time of stroke admission, during the hospitalization, within the 6-month period prior to the stroke or 90-day period post discharge. This lack of ECG-confirmed AF further lowered the total PPV of the ICD-9 identified population to 82.2%. We further reviewed stroke cases so as to only include IS cases attributable to AF (e.g., excluding strokes that occurred within 24 hours of a vascular procedure) and ICH cases that were not secondary to major head trauma. The resulting low ICD-9 code predictive value of 72.8% warrants caution for researchers seeking to solely use

ICD-9 screening as case-validation in this specific population. Furthermore, among the 458 individuals with the ICD-9 modifier "without infarction," we adjudicated 33 (7.2%) of these as valid stroke events based on medical record review. The composite of false positives and false negatives found in this study should raise some question over the validity of conclusions from research based exclusively on administrative data. ICD-9 codes can be a viable screening tool if used correctly but should not be used as the sole method of case-validation. The ramifications of these observations are significant. Stroke research identifies the gaps in care that drive health policy and resource allocation. Given the emphasis on performance measures and public reporting, the financial implications are tangible, particularly for those institutions without sufficient margin to invest in the data systems and personnel training that are requisite to improve accuracy.

For ICD-9 stroke codes, a PPV ≥85% has been suggested to be adequately accurate for research purposes.[12] Assuming nondifferential misclassification, similarly high PPVs (i.e. >90%) for outcome capture will pose little bias on the relative estimates[13] and may be considered a target threshold for use of administrative data for comparative effectiveness research for stroke prevention in AF. Bias in the risk estimate is also linked to a test's specificity more so than sensitivity with decreased specificity resulting in increased estimate bias.[13-15] Most bias is observed when specificity is <85%.[14] Furthermore, decreasing the PPV will reduce specificity thereby increase bias in the estimate.[13] We can also assume that even a minimal amount of misclassification leads to bias in utilization studies and outcome incidence estimates.

The first phase of validation in our study, ICD-9 stroke code accuracy, (94.2%) is similar to that previously reported.[6, 7, 12, 16-34] This is not necessarily surprising given that our study had characteristics associated with high ICD-9 accuracy including using hospital discharge codes, having a majority of events coded in the primary position, having an overwhelming majority of IS (93%) coded with the more accurate 434 stroke code as opposed to the 433 code, and excluding "without infarct" modifier codes.[35] In addition, given that the three participating centers in this study are all recognized stroke centers, it is reasonable to surmise that the accuracy of coding found outside centers of expertise may be considerably less. Previous research using similar methodology has led to highly variable IS event capture accuracy, varying from 62%[29] to 85%.[12] Additionally, informed interpretation of previous IS research is challenged by inclusion of transient ischemic attack or the nonspecific ICD-9 code 438 "late effects of cerebrovascular disease."

The second phase of our validation sought confirmation of the AF diagnosis. Despite AF being listed among the associated diagnoses, medical record documentation of AF or ECG-confirmation of AF was lacking in 12% of stroke cases, indicating that AF ICD-9 codes had a relatively high PPV of 88% in patients with confirmed stroke. Possible explanations for this finding include error in ECG interpretation, propagation of the AF code following a remote transient episode, or physician use of this code as a "rule out AF" to justify an ECG in the setting of an irregular pulse. A recent systematic review described the PPV of AF ICD-9 codes that ranged from 70 to 96%.[36] Accuracies of AF validation methods differed according to factors such as the number of documented AF codes required, the period of time searched for AF codes, and whether ECGs were used to confirm AF. In general, more

strict validation methods resulted in lower PPVs of AF ICD-9 codes. Although not validated in the literature, a validation algorithm incorporating inpatient and outpatient AF diagnosis codes with ECG confirmation may be best to confirm incident or prevalent AF.[36] Strategies based on multiple diagnoses or ECGs potentially could miss the heightened risk of new onset AF.

Accuracy of both stroke and AF codes is critically important for performance measurement and medication post-marketing surveillance, e.g., recently introduced novel oral anticoagulants (apixaban, rivaroxaban, and dabigatran).[37-39] Rigorously conducted comparative effectiveness studies of these agents in clinical practice will be contingent upon validation of AF, IS, and ICH. In our population, the 72.8% PPV for ICD-9 codes describing stroke associated with AF indicates modest usefulness of ICD-9 codes alone; however, a higher PPV (and increased usefulness) may be expected when assessing these codes in a population entirely comprised of AF patients on oral anticoagulants with prior instances of ICD-9 confirmed AF. Interpretation of results across different patient populations will require adjustment for key covariates which raises additional challenges. A previous study reports additional stroke covariates such as hypertension (85%), diabetes (97%), and coronary artery disease (88%) have high predictive values, while more difficult to obtain information such as history of cerebrovascular accident (59%) or tobacco use (58%) have lower accuracy.[12] Another study reported similarly high accuracies of coronary artery disease (96%), diabetes (98%), and hypertension (97%), but a low accuracy of deep vein thrombosis (72%).[6] Similar to stroke, the accuracy of ICD-9 codes for stroke covariates seems to be highly variable, which confirms our belief that as of currently, a more conservative approach using manual review in addition to ICD-9 codes in retrospective research is necessary to ensure more accurate stroke covariate ascertainment. Furthermore, manual verification is especially prudent when identifying fluctuating covariates like blood pressure or renal function and key confounding factors that are not reliably captured in electronic health records like nonprescription aspirin use, medication interruption for procedures, or medication adherence.

Newer ICD-10 codes are not currently used in hospital discharges in the United States. These codes are considered more specific and provide a more intuitive coding method compared to currently used ICD-9 codes. Regarding stroke, ICD-10 codes specify the hemorrhage location and source in ICH, differentiate between thrombotic and embolic IS, and include codes for intraoperative and post-procedural strokes. Data comparing the accuracy of ICD-9 and ICD-10 stroke codes is limited and research indicates the same accuracy for IS ICD-9 (85%) and ICD-10 (85%) codes and similar accuracies of ICH ICD-9 (97%) and ICD-10 (98%) and subarachnoid hemorrhage (SAH) ICD-9 (98%) and ICD-10 (91%) codes.[12] Another study shows ICD-10 codes for ICH and SAH having positive predictive values of approximately 96%.[40] Until ICD-10 stroke codes accuracies are further determined and compared to ICD-9 stroke codes, manual review of all or a subset of events appears to be warranted based on the results of our study.

Our study has several strengths including its multi-site nature, diverse patient population, and the comprehensiveness of our manual medical record reviews. Stroke events were adjudicated by the site principle investigator with modified Rankin score assignment by the

site neurologist. Questionable stroke cases were vetted through our adjudication committee with ultimate assignment by a stroke neurologist. We conducted comprehensive searches of outpatient, inpatient, and transferring hospital ECGs to validate the AF code. Additional strengths include the comprehensive assessment of commonly used stroke codes and coding position, and chart reviews of the "without infarction" group. Furthermore, we reviewed and assessed ICD-9 accuracies for a large number of ICH (n=454) cases which significantly contributes to the current limited literature.

As mentioned previously, although we included three different and complementary sites, our findings may not reflect the coding patterns or coding accuracy of other settings. Although we searched all medical records within 6 months of stroke admission and 90 days after stroke discharge for ECG evidence of AF, subclinical AF cannot be ruled out as the underlying mechanism. Our patient cohort includes patients with concurrent stroke and AF ICD-9 codes and therefore, we cannot calculate PPV for stroke and AF codes independent from each other. Additionally, we sought to include the first valid stroke for patients with multiple events. Although not likely impactful, this may overestimate our PPV. We cannot calculate the sensitivity of ICD-9 stroke codes because we did not review patients with diagnosis that may contribute additional cases of false negatives (i.e. TIA). Similarly, we are unable to report meaningful specificities of stroke codes because our estimates of non-valid stroke cases were restricted to patients assigned with stroke related codes (430-432, 433-434, and 436). Lastly, we only assess PPV and therefore we cannot calculate the prevalence of stroke in our cohort. This limits our ability to infer how changes in disease prevalence may alter our PPV. In general, as disease prevalence increases, a test's resulting PPV increases.

In conclusion, screening ICD-9 codes are commonly used in stroke-related research to identify outcome events. We identified ICD-9 coded IS and ICH events across three different stroke centers, and then implemented a thorough manual chart review to deem if the stroke events were valid. The accuracy of combined ICD-9 codes for stroke and AF to identify valid stroke accompanied by AF was 72.8%. Furthermore, review of "without infarction" ICD-9 coded events deemed 7.2% to be valid acute stroke. These findings highlight the need for more rigorous methods to validate outcome events identified by ICD-9 codes in large administrative databases.

Acknowledgments

References

1. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Magid D, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER, Moy CS, Mussolino ME, Nichol G, Paynter NP, Schreiner PJ, Sorlie PD, Stein J, Turan TN, Virani SS, Wong ND, Woo D, Turner MB, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2013 update: a report from the American Heart Association. Circulation. 2013; 127:e6–e245. [PubMed: 23239837]

2. Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, Larson MG, Kannel WB, Benjamin EJ. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. JAMA. 2003; 290:1049–1056. [PubMed: 12941677]

3. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke. 1991; 22:983–988. [PubMed: 1866765]

4. Lin HJ, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, D'Agostino RB. Stroke severity in atrial fibrillation. The Framingham Study. Stroke. 1996; 27:1760–1764. [PubMed: 8841325]

5. Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med. 1994; 120:897–902. [PubMed: 8172435]

6. 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. [PubMed: 15838413]

7. Brophy MT, Snyder KE, Gaehde S, Ives C, Gagnon D, Fiore LD. Anticoagulant use for atrial fibrillation in the elderly. J Am Geriatr Soc. 2004; 52:1151–1156. [PubMed: 15209654]

8. Adams HP Jr, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks EF, American Heart Association/American Stroke Association Stroke Council, American Heart Association/American Stroke Association Clinical Cardiology Council, American Heart Association/American Stroke Association Cardiovascular Radiology and Intervention Council, Atherosclerotic Peripheral Vascular Disease Working Group, Quality of Care Outcomes in Research Interdisciplinary Working Group. Guidelines for the early management of adults with ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups: The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Circulation. 2007; 115:e478–534. [PubMed: 17515473]

9. Hatano S. Experience from a multicentre stroke register: a preliminary report. Bull World Health Organ. 1976; 54:541–553. [PubMed: 1088404]

10. Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T, American Heart Association/American Stroke Association Council on Stroke, Council on Cardiovascular Radiology and Intervention, American Academy of Neurology. Guidelines for prevention of stroke in patients with ischemic stroke or transient ischemic attack: a statement for healthcare professionals from the American Heart Association/American Stroke Association Council on Stroke: co-sponsored by the Council on Cardiovascular Radiology and Intervention: the American Academy of Neurology affirms the value of this guideline. Circulation. 2006; 113:e409–49. [PubMed: 16534023]

11. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project--1981-86. 2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral and subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry. 1990; 53:16–22. [PubMed: 2303826]

12. 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. [PubMed: 16020772]

13. White E. The effect of misclassification of disease status in follow-up studies: implications for selecting disease classification criteria. Am J Epidemiol. 1986; 124:816–825. [PubMed: 3532767]

14. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol. 1977; 105:488–495. [PubMed: 871121]

15. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. Int J Epidemiol. 2002; 31:1246–1252. [PubMed: 12540730]

16. 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. [PubMed: 16081144]

17. 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. [PubMed: 9305319]

18. 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. [PubMed: 10753991]

19. Ellekjaer H, Holmen J, Kruger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. Stroke. 1999; 30:56–60. [PubMed: 9880388]

20. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke: effect of modifier codes. Stroke. 1998; 29:1602–1604. [PubMed: 9707200]

21. 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. [PubMed: 17159120]

22. Heckbert SR, Kooperberg C, Safford MM, Psaty BM, Hsia J, McTiernan A, Gaziano JM, Frishman WH, Curb JD. 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. [PubMed: 15583367]

23. 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. [PubMed: 19745645]

24. 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. [PubMed: 10785564]

25. 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. [PubMed: 15855802]

26. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke. 1994; 25:2348–2355. [PubMed: 7974572]

27. 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. 2004; 25:281–288. [PubMed: 15624086]

28. Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. Cerebrovasc Dis. 1999; 9:224–230. [PubMed: 10393410]

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

30. Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, Copper LS, Shahar E. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. Stroke. 1999; 30:736–743. [PubMed: 10187871]

31. 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. [PubMed: 17979142]

32. Spolaore P, Brocco S, Fedeli U, Visentin C, Schievano E, Avossa F, Milan G, Toso V, Vanuzzo D, Pilotto L, Pessina AC, Bonita R. Measuring accuracy of discharge diagnoses for a region-wide surveillance of hospitalized strokes. Stroke. 2005; 36:1031–1034. [PubMed: 15790948]

33. Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. Stroke. 2002; 33:2465–2470. [PubMed: 12364739]

34. Williams LS, Rotich J, Qi R, Fineberg N, Espay A, Bruno A, Fineberg SE, Tierney WR. Effects of admission hyperglycemia on mortality and costs in acute ischemic stroke. Neurology. 2002; 59:67–71. [PubMed: 12105309]

35. Andrade SE, Harrold LR, Tjia J, Cutrona SL, Saczynski JS, Dodd KS, Goldberg RJ, Gurwitz JH. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. Pharmacoepidemiol Drug Saf. 2012; 21(Suppl 1):100–128. [PubMed: 22262598]

36. Jensen PN, Johnson K, Floyd J, Heckbert SR, Carnahan R, Dublin S. A systematic review of validated methods for identifying atrial fibrillation using administrative data. Pharmacoepidemiol Drug Saf. 2012; 21(Suppl 1):141–147. [PubMed: 22262600]

37. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu J, Diaz R, Lewis BS, Darius H, Diener HC, Joyner CD, Wallentin L, RE-LY Steering Committee and Investigators. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med. 2009; 361:1139–1151. [PubMed: 19717844]

38. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, Becker RC, Nessel CC, Paolini JF, Berkowitz SD, Fox KA, Califf RM, ROCKET AF Investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med. 2011; (365):883–891. [PubMed: 21830957]

39. Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, Bahit MC, Diaz R, Easton JD, Ezekowitz JA, Flaker G, Garcia D, Geraldes M, Gersh BJ, Golitsyn S, Goto S, Hermosillo AG, Hohnloser SH, Horowitz J, Mohan P, Jansky P, Lewis BS, Lopez-Sendon JL, Pais P, Parkhomenko A, Verheugt FW, Zhu J, Wallentin L, ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med. 2011; 365:981–992. [PubMed: 21870978]

40. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The accuracy of hospital discharge coding for hemorrhagic stroke. Acta Neurol Belg. 2009; 109:114–119. [PubMed: 19681442]
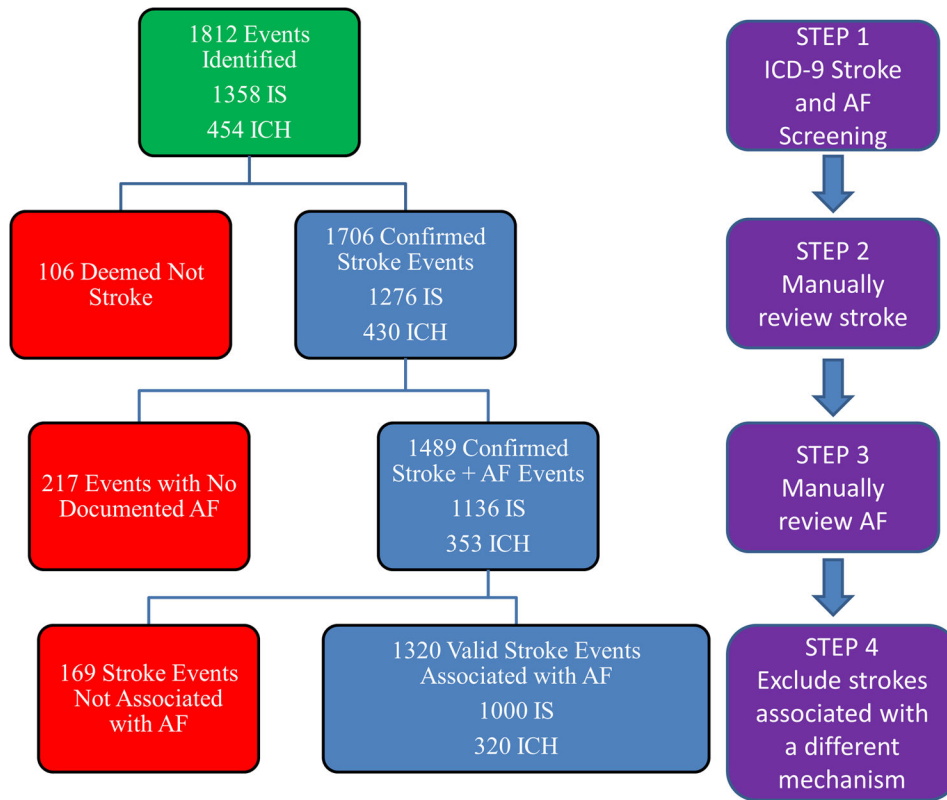
Figure. Flowchart Describing Manual Validation Algorithm of Hospital ICD-9 Stroke and Atrial Fibrillation Codes

Step 1: Initial electronic screening using ICD-9 codes for eligible stroke events in the setting of atrial fibrillation.

Step 2: Excluding events in which manual review confirmed no stroke occurred. Remaining events are valid stroke with ICD-9 documented atrial fibrillation.

Step 3: Excluding events in which manual review confirmed no active and/or recent atrial fibrillation. Remaining events are valid stroke in the setting of confirmed atrial fibrillation.

Step 4: Excluding ischemic stroke events in which manual review determined stroke cases were not due to atrial fibrillation (i.e. sepsis, tumor emboli, post-procedure). Excluding non-spontaneous intracranial hemorrhage events and those due to hospital procedures and/or complications. Remaining events are valid ischemic strokes due to atrial fibrillation or valid intracranial hemorrhage associated with atrial fibrillation.

IS = ischemic stroke; ICH = intracranial hemorrhage; AF = atrial fibrillation

Table 1

List of ICD-9 Codes Used to Identify Stroke Events

Condition	ICD-9 Code	ICD-9 Code Description
	433	occlusion and stenosis of precerebral arteries
Ischemic stroke	434	occlusion of cerebral arteries
	436	acute but ill-defined cerebrovascular disease
	430	subarachnoid hemorrhage
Intracranial hemorrhage	431	intracerebral hemorrhage
	432	other and unspecified intracranial hemorrhage

Table 2

Clinical and Demographic Characteristics of Patients with Valid Stroke in the Setting of Atrial Fibrillation

Characteristic	Ischemic Stroke (n=1000)	Intracranial Hemorrhage (n=320)
Age (Mean±SD)	77.1 (11.1)	75.0 (11.4)
Gender (n, %)		
Male	439 (43.9%)	162 (50.6%)
Female	561 (56.1%)	158 (49.4%)
Race (n, %)		
White	742 (74.2%)	270 (84.4%)
Black	216 (21.6%)	35 (10.9%)
Hispanic	23 (2.3%)	7 (2.2%)
Other/Unknown	19 (1.9%)	8 (2.5%)
Atrial Fibrillation Type (n, %)		
New Onset	248 (24.8%)	69 (21.6%)
Paroxysmal	244 (24.4%)	75 (23.4%)
Persistent/Permanent	508 (50.8%)	176 (55.0%)
$CHADS_2$ (prior to stroke) (Mean±SD)	2.9 (1.4)	2.6 (1.4)
0	25 (2.5%)	10 (3.1%)
1	123 (12.3%)	62 (19.4%)
2	274 (27.4%)	105 (32.8%)
3	251 (25.1%)	71 (22.2%)
4	179 (17.9%)	38 (11.9%)
5	109 (10.9%)	26 (8.1%)
6	39 (3.9%)	8 (2.5%)
Event Coding Position (n, %) [†]		
Primary	869 (86.9%)	281 (87.8%)
Other	109 (10.9%)	28 (8.8%)
Study Site (n, %)		
Boston Medical Center	230 (23.0%)	70 (21.9%)
Geisinger Health System	370 (37.0%)	105 (32.8%)
University of Alabama	400 (40.0%)	145 (45.3%)

[†] There were 33 events with combinations of stroke codes recorded so there was no mutually exclusive stroke coding position.

Table 3

Percentage of Valid Stroke Events Using ICD-9 Coding Alone by Event Type, Coding Position, Clinical Site, and Ischemic Stroke Code

	Valid (%; CI) n=1706 (94.2%)	Not Valid (%; CI) n=106 (5.8%)	p-value
Event Type			0.55
Ischemic stroke			
n=1358 (74.9%)	1276 (94.0%; 92.6-95.1)	82 (6.0%; 4.9-7.4)	
ICH			
n=454 (25.1%)	430 (94.7%; 92.2-96.4)	24 (5.3%; 3.6-7.7)	
Event Coding Position			<0.0001
Primary			
n=1368 (77.4%)	1329 (97.2%; 96.1-97.9)	39 (2.8%; 2.1-3.9)	
Other			
n=399 (22.6%)	334 (83.7%; 79.8-87.0)	65 (16.3%; 13.0-20.2)	
Site			0.55
Boston Medical Center			
n=433 (23.9%)	411 (94.9%; 92.4-96.6)	22 (5.1%; 3.4-7.6)	
Geisinger Health System			
n=538 (29.7%)	502 (93.3%; 90.9-95.1)	36 (6.7%; 4.9-9.1)	
University of Alabama			
n=841 (46.4%)	793 (94.3%; 92.5-95.7)	48 (5.7%; 4.3-7.5)	
Ischemic Stroke Code[†]			0.0034
433			
n=61 (4.6%)	52 (85.2%; 74.3-92.0)	9 (14.8%; 8.0-25.7)	
434			
n=1265 (95.4%)	1194 (94.4%; 93.0-95.5)	71 (5.6%; 4.5-7.0)	

CI = confidence interval

[†]There were two events coded with ICD-9 ischemic stroke code 436 (50% valid). There were 30 valid ischemic stroke events coded with combinations of IS codes or ICH and IS codes. Comparing codes 433, 434, and 436 resulted in a Fisher's Exact Test p-value of 0.0103. We present the p-value for code 433 versus code 434