



Published in final edited form as:

Stroke. 2014 November ; 45(11): 3219–3225. doi:10.1161/STROKEAHA.114.006316.

Validity of Hospital Discharge Diagnosis Codes for Stroke: The Atherosclerosis Risk in Communities (ARIC) Study

Sydney A. Jones, BS¹, Rebecca F. Gottesman, MD, PhD², Eyal Shahar, MD³, Lisa Wruck, PhD¹, and Wayne D. Rosamond, PhD, MS¹

¹Gillings School of Global Public Health, University of North Carolina – Chapel Hill

²Johns Hopkins University School of Medicine

³University of Arizona College of Public Health

Abstract

Background and Purpose—Characterizing *International Classification of Disease* (ICD-9-CM) code validity is essential given widespread use of hospital discharge databases in research. Using the Atherosclerosis Risk in Communities (ARIC) Study we estimated the accuracy of ICD-9-CM stroke codes.

Methods—Hospitalizations with ICD-9-CM codes 430-438 or stroke keywords in the discharge summary were abstracted for ARIC cohort members (1987–2010). A computer algorithm and physician reviewer classified definite and probable ischemic stroke, intracerebral hemorrhage (ICH), and subarachnoid hemorrhage (SAH). Using ARIC classification as a gold standard, we calculated the positive predictive value (PPV) and sensitivity of ICD-9-CM codes grouped according to the American Heart Association/American Stroke Association (AHA/ASA) 2013 categories and an alternative code grouping for comparison.

Results—Thirty-three percent of 4,260 hospitalizations were validated as strokes (1,251 ischemic, 120 ICH, 46 SAH). The AHA/ASA code groups had PPV 76% and 68% sensitivity, compared to PPV 72% and 83% sensitivity for the alternative code groups.

The PPV of the AHA/ASA code group for ischemic stroke was slightly higher among African Americans, individuals <65 years, and at teaching hospitals. Sensitivity was higher among older individuals and increased over time. The PPV of the AHA/ASA code group for ICH was higher among African Americans, women, and younger individuals. PPV and sensitivity varied across study sites.

Conclusions—A new AHA/ASA discharge code grouping to identify stroke had similar PPV and lower sensitivity compared with an alternative code grouping. Accuracy varied by patient characteristics and study sites.

Address for Correspondence: Sydney A. Jones 137 East Franklin Street, Suite 306 Chapel Hill, NC 27514 Tel.: 512-762-8682 / Fax: 919-966-9800 SydneyJones@unc.edu.

Disclosures: none.

Keywords

ICD-9-CM; predictive value; sensitivity; administrative data; cerebrovascular disease

Introduction

Current data may be inadequate to monitor the national incidence of cerebrovascular disease,^{1,2} a leading cause of death and disability in the United States.³ One approach for national surveillance is capturing International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM) codes from hospital discharge and claims databases (“administrative data”).⁴⁻⁶ Characterizing the validity of ICD-9-CM codes is essential given their widespread use for surveillance, and for epidemiological and health services research.⁷⁻¹¹ Estimates of the accuracy of these codes can be used in sensitivity analyses to account for the misclassification of stroke events in administrative data.¹¹ Documenting coding accuracy over time is particularly important to understand the potential impact on temporal trends in stroke incidence estimated from administrative data.¹²

Estimates of the validity of ICD-9-CM codes for stroke vary depending on the codes investigated^{4,11} and by patient and hospital characteristics.¹³⁻¹⁷ In 2013, the American Heart Association/American Stroke Association (AHA/ASA) published an updated definition of stroke including ICD-9-CM codes grouped according to stroke subtypes (ischemic stroke, intracerebral hemorrhage [ICH], and subarachnoid hemorrhage [SAH]).¹⁸ The accuracy of these code groups for identifying stroke has not been reported. Also, the positive predictive value (PPV) of ICD-9-CM codes by patient sex and age has rarely been assessed,^{14,19} variation by race/ethnicity has not been explored, and most studies to date were conducted in a single geographic location.¹¹

Using the Atherosclerosis Risk in Community (ARIC) Study, we assessed the accuracy of hospital discharge ICD-9-CM coding of stroke. We estimated the PPV and sensitivity of the AHA/ASA code groups compared to previously validated alternative code groups^{11,20} overall and by stroke subtype (ischemic, ICH, and SAH). We characterized variation in ICD-9-CM code accuracy by patient sex, race, age, geographic location, hospital type (teaching vs. non-teaching), and ICD-9-CM code position (first vs. any position). We further investigated temporal trends in the accuracy of codes from 1991 to 2010.

Methods

Study Population

The ARIC Study design is well documented.²¹ Briefly, a population-based cohort of 15,792 individuals aged 45 to 64 years was recruited in 1987–1989 in four communities: Washington County, Maryland; suburbs of Minneapolis, Minnesota; Jackson, Mississippi; and Forsyth County, North Carolina. This analysis included eligible hospitalizations of ARIC cohort members occurring from enrollment through December 31, 2010, or date of last contact if deceased or lost to follow-up.

Identification of Stroke Events

Hospitalizations and deaths were ascertained via annual follow-up phone calls, study examinations, and surveillance of hospital discharges in ARIC communities.

Hospitalizations meeting one or more of the following criteria were eligible for medical record abstraction: 1) A discharge diagnosis ICD-9-CM code 430 through 438 (1987–1996) or 430 through 436 (since 1997); 2) One or more stroke-related keywords (see Supplemental Methods) in discharge summary; or 3) Diagnostic computed tomography (CT) or magnetic resonance imaging (MRI) scan with cerebrovascular findings or admission to the neurological intensive care unit. A trained nurse abstracted records for each eligible hospitalization, including up to 21 ICD-9-CM discharge codes (see Supplemental Methods).

A computer algorithm and physician reviewer independently classified each event according to criteria adapted from the National Survey of Stroke.²² A second physician reviewer adjudicated in cases where the computer and initial reviewer disagreed.

A definite or probable stroke was defined as a sudden and rapid onset of neurological symptoms lasting >24 hours or leading to death in the absence of evidence for a non-stroke cause (see Supplemental Material). Events that did not meet these criteria were classified as “possible stroke of undetermined type,” “out-of-hospital fatal stroke,” or “no stroke.” Definite and probable strokes were classified further as SAH, ICH, or ischemic stroke (including embolic and thrombotic brain infarction) (Supplemental Table I).²³

ICD-9-CM Code Groups

ICD-9-CM codes were grouped according to stroke subtype using two approaches. First, we used ICD-9-CM codes matched to stroke subtypes by the AHA/ASA in 2013.¹⁸ We excluded codes for spinal and retinal infarcts (336.1, 362.31, 362.32) because these events (N=3) were not validated in ARIC. Second, we used an alternative grouping of ICD-9-CM codes with high PPV and sensitivity in previous studies.^{5,11,20} The primary difference between the two code groupings was exclusion of ICD-9-CM codes 432 and 436 from the AHA/ASA code group.

Statistical Analysis

Using ARIC classification as a gold standard, we estimated the PPV and sensitivity of the AHA/ASA and alternative code groups. PPV was the proportion of validated strokes among all hospitalizations with a given ICD-9-CM code group. Sensitivity was the proportion of hospitalizations with a given code group among validated strokes. In total, 216 hospitalizations were identified for validation in ARIC based on keywords without any stroke-related ICD-9-CM codes. Of these, 18 (8%) were validated as strokes, and these were included in analyses as false negatives. Code 486 (pneumonia, organism unspecified) was the most common primary ICD-9-CM code among these 216 events.

We calculated the PPV and sensitivity of code groups for ischemic stroke and ICH stratified by patient and hospital characteristics. Stratified analyses were not conducted for SAH because of a small number of events. Patient characteristics were sex, age at event (<65 years, 65 years), race (black, white), and study center. Analyses stratified by race included

only hospitalizations of white participants and black participants from Forsyth Co. and Jackson (N=4,235). Hospitalization characteristics included ICD-9-CM code position (first, any position), incident vs. recurrent event, and teaching status. Among hospitalizations with symptoms present for 24 hours (N=3,250). Incident events were those with no history of stroke or transient ischemic attack (TIA) recorded in the medical record. Hospital teaching status was determined by the presence of full-time internal medicine residents for hospitalizations (N=3,936) that occurred at 31 hospitals located within ARIC Study communities.

To assess temporal trends, we calculated the PPV and sensitivity of the code groups from 1987–2010. Binomial regression was used to estimate the age-adjusted trend in the PPV of the AHA/ASA code group for ischemic stroke by sex and race from 1991–2010 because no ischemic strokes with AHA/ASA-identified ICD-9-CM codes occurred prior to 1991. Confidence intervals were calculated using the exact method. Analyses performed using SAS 9.3 (Cary, NC).

Results

A total of 4,318 stroke-eligible hospitalizations among 2,533 persons were identified. Fifty-eight were excluded: 19 out-of-hospital fatal strokes, 24 possible strokes of undetermined type, 2 transfers from acute care facilities, and 13 hospitalizations for which ICD-9-CM codes were not available. Of 4,260 remaining hospitalizations (among 2,516 persons), 1,417 (33%) were classified as definite or probable strokes in ARIC. By subtype, there were 1,251 ischemic strokes, 120 ICH, and 46 SAH. The remaining 2,843 events included re-hospitalizations for prior stroke, TIAs, and borderline events that did not meet ARIC clinical criteria.

The PPV of individual ICD-9-CM codes 430-438 in any position ranged from 2% to 79% (Table 1). Among hospitalizations with a stroke-related ICD-9-CM code in the first position (N=2,521), the PPV increased by an average of 7% (Supplemental Table II).

AHA/ASA Code Group

Thirty percent (1,275 of 4,260) of eligible hospital discharges included ICD-9-CM codes in the AHA/ASA code group (Table 2). The AHA/ASA code group had a PPV of 76% (range 57–76% by stroke subtype) and 68% sensitivity (range 64–93% by stroke subtype).

Six percent (276 of the 4260) of hospitalizations occurred among individuals with no new symptoms at admission, but new symptoms during hospitalization. The PPV of the AHA/ASA code group to correctly identify these “in-hospital” events was similar to events presenting with symptoms at admission, 74% and 77% respectively, and the sensitivity was identical (64%).

Alternative Code Group

In total, 39% (1,649 of 4,260) of hospital discharges included ICD-9-CM codes in the alternative code group (Table 2). The PPV for the alternative code group was 72% (range 40–75% by stroke subtype) with 83% sensitivity (range 80–93% by stroke subtype). For

both code groups, PPV was highest for ischemic stroke and lowest for ICH, while sensitivity was highest for SAH and lowest for ischemic stroke.

Subgroup Analysis

Ischemic stroke—Across patient and hospital subgroups, the PPV of the AHA/ASA code group for ischemic stroke ranged from 68–85% with sensitivity between 24–93% (Table 3). PPV was higher for first compared to any position ICD-9-CM code, among African American and younger patients, and at teaching hospitals. Sensitivity was higher for ICD-9-CM codes in any compared to the first position, and among older patients (Table 3). Across the ARIC communities, PPV ranged from 68–80%, and sensitivity from 61–70%. Both sensitivity and PPV were similar for incident and recurrent strokes. Patterns in the PPV and sensitivity of the alternative ICD-9-CM code grouping were similar (Supplemental Table III).

Intracerebral Hemorrhage—The AHA/ASA code group for ICH had higher PPV for first compared to any position ICD-9-CM code, and among women, African American, and younger patients (Table 4). Across study sites, PPV ranged from 31–80% and sensitivity varied from 69–90%.

Temporal Trends

Sensitivity of the AHA/ASA and alternative code groups for ischemic stroke increased over time (Table 3, Supplemental Table III). There was no consistent temporal trend in the PPV for either the AHA/ASA code group (Figure 1) or the alternative code group (data not shown), including after adjustment for age. Adjusting for age did not significantly change the pattern of increasing sensitivity (data not shown).

Discussion

We expanded a preliminary validation of ICD-9-CM codes for stroke in the ARIC Study (1987–1995)²³ to include hospitalizations of cohort members through December 31, 2010 and considered two approaches to grouping ICD-9-CM codes. The new code group published by the AHA/ASA in 2013 had slightly higher PPV and lower sensitivity than a previously validated alternative code group. Both PPV and sensitivity varied by patient and hospital characteristics.

The National Center for Health Statistics relies on ICD-9-CM codes 430–438 to estimate the number of hospital discharges for stroke and stroke related mortality in the United States.²⁴ The PPV of these codes in our study was approximately 10–15% lower than previous estimates.^{11,25,26} The PPV of the AHA/ASA and alternative code groups also were lower than previously reported for similar code groups.^{11,17} For example, the alternative code group for ischemic stroke had PPV 90% and 86% sensitivity among hospitalized patients in Seattle, Washington (1990–1996),²⁰ compared to PPV 75% and 80% sensitivity in our study. We compare our findings to other studies with caution as differences in PPV may be due to diagnostic criteria, disease prevalence, or both. PPV estimates differed by 29% using World Health Organization compared to Minnesota Stroke Survey definitions of

stroke.¹² Moreover, previous studies sampled hospitalized patients^{20,27} or post-menopausal women enrolled in Medicare,¹⁷ which may have higher prevalence of stroke.

For both code groups, sensitivity was lowest for ischemic stroke. Of 1,251 definite and probable ischemic strokes validated in ARIC, 436 (35%) did not include any of the ICD-9-CM codes in the AHA/ASA code group. Among these 436 events, 46% had an ICD-9-CM code 436, 17% had code 434.9, and 15% had code 435. Thus, low sensitivity of the AHA/ASA code group for ischemic stroke was primarily due to exclusion of ICD-9-CM code 436.

Low sensitivity and PPV for identifying ICH were surprising. Low sensitivity was due in part to miscoding of 12% of ICH events as ischemic strokes (ICD-9-CM codes 433, 434, or 436), an error identified previously.²⁸ Hemorrhagic infarctions are classified as ischemic strokes in ARIC but may receive ICH-related ICD-9-CM codes, possibly contributing to lower PPV for ICH. Additionally, ARIC adjudicated events were assigned a single classification, such that events with combined pathology were assigned the subtype believed to be primary by the ARIC adjudicators. This may contribute to classification of events with hemorrhage-related ICD-9-CM codes as ischemic strokes in ARIC and vice versa.

In addition to the knowledge of coders and quality of the medical chart, the accuracy of ICD-9-CM coding depends on the specificity of coding criteria and regional or departmental variation in diagnosis and coding practices.^{10,16,19,29} Previous studies documented differences in coding accuracy by hospital department (emergency vs. neurology),^{10,15,16} and urbanicity.³⁰ The accuracy of ICD-9-CM code groups varied by hospital teaching status and ARIC community. Use of evidence-based diagnosis measures and updated diagnostic criteria may be more common at teaching hospitals and vary by region. The sensitivity of the alternative code group for ischemic stroke was more similar for teaching and non-teaching hospitals, likely due to inclusion of the non-specific, commonly used ICD-9-CM code 436.

The accuracy of code groups for ischemic stroke and ICH also varied by patient race, age, and over time. Higher PPV among blacks compared to whites was likely due to a higher prevalence of stroke among African Americans, but also may suggest differential usage of diagnostic tools. Among hospitalizations where a CT or MRI was performed, PPV of the AHA/ASA code group for ischemic stroke was more similar across races (data not shown).

Higher PPV among younger compared to older patients was unexpected given increasing stroke prevalence with age. Previous studies reported no difference in ICD coding accuracy by patient age,^{16,19,31} or higher PPV among older adults.^{14,15} One explanation for our findings may be a higher prevalence of comorbidities and recurrent symptoms from prior stroke among older adults, complicating diagnosis and increasing the misapplication of stroke-related ICD-9-CM codes

In contrast to PPV, the sensitivity of the AHA/ASA code group for ischemic stroke was higher among older patients. Disease severity was associated with higher sensitivity of ICD-9-CM coding among Medicare beneficiaries,²⁸ and it is possible that more severe strokes among patients over age 65 contributed to differences in sensitivity. Differences by

age were less pronounced using the alternative compared to AHA/ASA code group for ischemic stroke. ICD-9-CM code 436 increased the sensitivity of the alternative relative to AHA/ASA code group among all patients, but had a greater effect among younger patients. However, in the context of a prospective cohort study, temporal trends in coding accuracy complicate the interpretation of age-stratified estimates.

Few previous studies have investigated temporal changes in code accuracy. A systematic review identified no substantial variation in PPV or sensitivity comparing studies conducted before and after 2000;¹¹ and a Minnesota Stroke Survey study reported no consistent trend in PPV from 1980–2000.¹² Similar to our findings, the Rochester Stroke Registry documented no trend in PPV and increasing sensitivity of ICD-9-CM codes 430–438 from 1970 to 1989.¹⁰ Thus, interpretation of national trend data from administrative databases may need to consider the possibility of increasing sensitivity of discharge codes to correctly identify stroke events. However, given fluctuating PPV and lacking data on specificity, it is difficult to predict the effect of temporal increases in sensitivity.

Given evolving diagnostic procedures and definitions, temporal changes in ICD-9-CM coding accuracy are expected. For example, increased prevalence of imaging technology was expected to increase coding accuracy^{16,26} and was linked to decreased use of non-specific ICD-9-CM codes 436–437 in the Pawtucket Heart Health Program from 1980–1991.⁵ An important change between 1987–2010 was the 1992 addition of a fifth-digit clinical modification code indicative of cerebral infarction to ICD-9-CM codes 433 and 434.²⁰ We investigated the impact of this change by restricting the analysis to hospitalizations since 1992, which slightly increased sensitivity estimates for ischemic stroke using the AHA/ASA (70%) and alternative code groups (84%).

Strengths of this study include the application of a consistent validation methodology to over 4,000 hospitalizations across 24 years and four communities. We are the first to estimate the PPV and sensitivity of new ICD-9-CM code groups proposed by the AHA/ASA in 2013, and to compare these to previously validated code groups. We addressed limitations in the existing literature including using imaging data in validation; calculating PPV by stroke subtype; and, investigating variation by population subgroup for hemorrhagic events.¹¹ Unlike previous studies,^{5,13,20} we did not exclude recurrent strokes or multiple hospitalizations per individual. We report no difference in the PPV and sensitivity for incident vs. recurrent ischemic strokes, nor did eliminating multiple hospitalizations per individual substantially change our results (data not shown), in contrast to previous studies.^{11,17,32} These findings are important given that up to a third of in-hospital strokes may be recurrent.¹³

Limitations to this analysis include differences between the ARIC definition of stroke and the AHA/ASA 2013 definition of stroke.¹⁸ In particular, spinal and retinal infarctions were classified as ischemic strokes by the AHA/ASA but were not validated in the ARIC Study. However, ICD-9-CM codes for these events were rare in this dataset (N=3). Subgroup analyses for validity of ICD-9-CM coding for SAH were not conducted due to the small number of events, and no analyses were stratified by indicators of disease severity, comorbidity, or outcomes, which may impact code accuracy.^{6,20} Additional research is

needed to explore the validity of codes for strokes resulting from in-hospital procedures relative to strokes present at admission as code position alone may not be a sufficient proxy for this distinction.¹³

New groups of ICD-9-CM codes proposed by the AHA/ASA to identify stroke subtypes had similar PPV and lower sensitivity compared to previously validated ICD-9-CM code groups. Both sensitivity and PPV varied by patient characteristics including age, by geographic region, and over time. Given their affordability and ubiquity, administrative data are likely to remain an important source for surveillance and health services research.^{6,29} With the expansion of electronic health record systems, future studies should focus on the identification and collection of information in addition to ICD-9-CM codes that is required for accurate stroke surveillance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank the staff and participants of the ARIC study for their important contributions.

Funding Sources The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). SAJ was supported by NHLBI T32 training grant HL-007055 and the UNC Royster Society of Fellows.

References

1. Barrett-Connor, E.; Ayanian, JZ.; Brown, ER.; Coultas, DB.; Francis, CK.; Goldberg, RJ., et al. Institute of Medicine. The National Academies Press; Washington, DC: 2011. A nationwide framework for surveillance of cardiovascular and chronic lung diseases.
2. Sidney S, Rosamond WD, Howard VJ, Luepker RV, National Forum for Heart Disease and Stroke Prevention. The "heart disease and stroke statistics--2013 update" and the need for a national cardiovascular surveillance system. *Circulation*. 2013; 127:21–23. [PubMed: 23239838]
3. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, et al. Executive summary: Heart disease and stroke statistics--2014 update: A report from the American Heart Association. *Circulation*. 2014; 129:399–410. [PubMed: 24446411]
4. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011; 64:821–829. [PubMed: 21194889]
5. Derby CA. 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]
6. Reker DM, Rosen AK, Hoenig H, Berlowitz DR, Laughlin J, Anderson L, et al. The hazards of stroke case selection using administrative data. *Medical care*. 2002; 40:96–104. [PubMed: 11802082]
7. Allen NB, Holford TR, Bracken MB, Goldstein LB, Howard G, Wang Y, et al. Geographic variation in one-year recurrent ischemic stroke rates for elderly Medicare beneficiaries in the USA. *Neuroepidemiology*. 2010; 34:123–129. [PubMed: 20068358]
8. Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, Ali F, Sholley C, Worrall C, et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. *JAMA*. 2010; 304:411–418. [PubMed: 20584880]

9. Mercaldi CJ, Ciarametaro M, Hahn B, Chalissery G, Reynolds MW, Sander SD, et al. Cost efficiency of anticoagulation with warfarin to prevent stroke in Medicare beneficiaries with nonvalvular atrial fibrillation. *Stroke*. 2011; 42:112–118. [PubMed: 21148442]
10. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. *Stroke*. 1994; 25:2348–2355. [PubMed: 7974572]
11. 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. *Pharmacoepidemiology and drug safety*. 2012; 21(Suppl 1):100–128. [PubMed: 22262598]
12. Lakshminarayan K, Anderson DC, Jacobs DR Jr, Barber CA, Luepker RV. Stroke rates: 1980–2000: The Minnesota Stroke Survey. *Am J Epidemiol*. 2009; 169:1070–1078. [PubMed: 19318614]
13. Rounie 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. *Pharmacoepidemiology and drug safety*. 2008; 17:20–26. [PubMed: 17979142]
14. Aboa-Eboule C, Mengue D, Benzenine E, Hommel M, Giroud M, Bejot 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 Neurology*. 2012; 260:605–613.
15. 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*. 2005; 36:1031–1034. [PubMed: 15790948]
16. Johnsen SP, Overvad K, Sorensen HT, Tjonneland A, Husted SE. Predictive value of stroke and transient ischemic attack discharge diagnoses in the Danish National Registry of patients. *J Clin Epidemiol*. 2002; 55:602–607. [PubMed: 12063102]
17. 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. [PubMed: 24525955]
18. 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. [PubMed: 23652265]
19. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta Neurologica Belgica*. 2009; 109:114–119. [PubMed: 19681442]
20. Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. *Stroke*. 2002; 33:2465–2470. [PubMed: 12364739]
21. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989; 129:687–702. [PubMed: 2646917]
22. The National Survey of Stroke. National Institute of Neurological and Communicative Disorders and Stroke. *Stroke*. 1981; 12:11–91. [PubMed: 7222163]
23. 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*. 1999; 30:736–743. [PubMed: 10187871]
24. Hall MJ, Levant S, DeFrances CJ. Hospitalization for stroke in U.S. Hospitals, 1989–2009. *NCHS data brief*. 2012:1–8.
25. Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and occurrence of total (first-ever and recurrent) stroke. *Stroke*. 1999; 30:2523–2528. [PubMed: 10582972]
26. 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]
27. 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]

28. 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]
29. De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: Report from an international consortium. *BMC health services research*. 2006; 6:77. [PubMed: 16776836]
30. Yiannakoulis N, Svenson LW, Hill MD, Schopflocher DP, Rowe BH, James RC, et al. Incident cerebrovascular disease in rural and urban Alberta. *Cerebrovascular diseases*. 2004; 17:72–78. [PubMed: 14534379]
31. Hsieh, CY.; Chen, CH.; Li, CY.; Lai, ML. [Accessed August 18, 2014] Validating the diagnosis of acute ischemic stroke in a national health insurance claims database. *J Formos Med Assoc*. 2013. published online ahead of print date October 13, 2013 <http://doi.org/10.1016/j.jfma.2013.09.009>
32. Ramalle-Gomara E. Validity of discharge diagnoses in the surveillance of stroke. *Neuroepidemiology*. 2013; 41:185–188. [PubMed: 24051447]

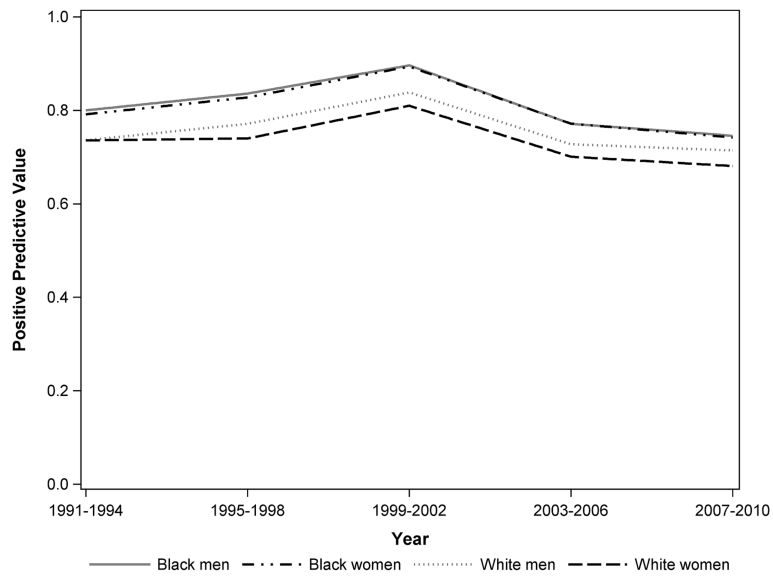


Figure 1. Trend in age-adjusted positive predictive value of the AHA/ASA ICD-9-CM code group for ischemic stroke by sex and race: the ARIC Cohort Study, 1991–2010. Abbreviations: AHA/ASA, American Heart Association/American Stroke Association; ARIC, Atherosclerosis Risk in Communities Study; ICD-9-CM, International Classification of Disease, 9th Revision, Clinical Modification.

Table 1

Number of hospitalizations by ICD-9-CM discharge code and percent validated as definite or probable stroke: the ARIC Cohort Study, 1987–2010

ICD-9-CM		ARIC Classification				Total	PPV*	(95% CI)
Code and Classification		ISC	ICH	SAH	No Stroke			
430	Subarachnoid hemorrhage	1	3	40	12	56	79	(66, 88)
431	Intracerebral hemorrhage	11	95	3	50	159	69	(61, 76)
432	Other intracerebral hemorrhage	1	6	0	83	90	8	(3, 15)
433	Occlusion of precerebral arteries	116	1	0	1080	1197	10	(8, 12)
434	Occlusion of cerebral arteries	822	9	0	253	1084	77	(74, 79)
435	Transient cerebral ischemia	64	0	0	581	645	10	(8, 12)
436	Acute but ill-defined cerebrovascular disease	203	4	0	82	289	72	(66, 77)
437 [†]	Other ill-defined cerebrovascular disease	9	0		130	142	8	(4, 14)
438 [†]	Late effects of cerebrovascular disease	8	0	0	374	382	2	(1, 4)
Other [‡]	None of above	16	2	0	198	216	8	(5, 13)
Total		1251	120	46	2843	4260	33	(32, 35)

* Percent of hospitalizations classified as definite or probable stroke.

[†] Collected 1987–1996.

[‡] Events without ICD-9-CM codes 430–438, eligible for review by keywords in discharge abstract Abbreviations: ARIC, Atherosclerosis Risk in Communities Study; CI, confidence interval; ICD-9-CM, *International Classification of Disease, 9th Revision, Clinical Modification*; ICH, intracranial hemorrhage; ISC, ischemic stroke; PPV, positive predictive value; SAH, subarachnoid hemorrhage.

Table 2

Number of hospitalizations, positive predictive value, and sensitivity of two ICD-9-CM code groupings, overall and by stroke subtype: the ARIC Cohort Study, 1987–2010.

Stroke Subtype	ICD-9-CM Code (any position)	ARIC Classification					PPV* (95% CI)	Sensitivity† (95% CI)
		Total	ISC	ICH	SAH	No Stroke		
AHA/ASA Code Group‡		1275	815	107	46	307	76 (73, 78)	68 (66, 71)
ISC	433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.01, 434.11, 434.91	1048	801	8	0	239	76 (74, 79)	64 (61, 67)
ICH	431	168	13	96	3	56	57 (49, 65)	80 (72, 87)
SAH	430	59	1	3	43	12	73 (60, 84)	93 (82, 99)
Alternative Code Group		1649	1018	117	46	468	72 (69, 74)	83 (81, 85)
ISC	433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434, 434.01, 434.11, 434.91, 436	1335	1002	12	0	321	75 (73, 77)	80 (78, 82)
ICH	431, 432	255	15	102	3	135	40 (34, 46)	85 (77, 91)
SAH	430	59	1	3	43	12	73 (60, 84)	93 (82, 99)

Abbreviations: AHA/ASA, American Heart Association/American Stroke Association; CI, confidence interval; ICD-9-CM, *International Classification of Disease, 9th Revision, Clinical Modification*; ICH, intracranial hemorrhage; ISC, ischemic stroke; PPV, positive predictive value; SAH, subarachnoid hemorrhage.

* Percent of hospitalizations with group-specific ICD-9-CM codes classified in ARIC as definite or probable ISC, ICH, or SAH, respectively.

† Percent of definite or probable ISC (N=1251), ICH (N=120), or SAH (N=46) in ARIC with ICD-9-CM discharge codes in each code group.

‡ Sacco RL, et al. *Stroke*. 2013;44:2064–2089.¹⁸

Table 3

Positive predictive value and sensitivity of the AHA/ASA ICD-9-CM code group for ischemic stroke by patient and hospital characteristics: the ARIC Cohort Study, 1987–2010

	Ischemic Stroke* (N)	PPV [†]		Sensitivity [‡]	
		(%)	95% CI	(%)	95% CI
ICD-9-CM Code Position					
First	687	82	79, 84	55	52, 58
Any	801	76	74, 79	64	61, 67
By Patient & Hospital Characteristics (any position code)					
Sex					
Men	392	78	74, 81	63	59, 67
Women	409	75	71, 79	65	61, 69
Race [§]					
Black	319	80	75, 84	62	57, 66
White	477	74	71, 78	66	62, 69
Sex by race [§]					
Black men	126	81	74, 87	60	53, 66
Black women	193	79	73, 84	63	57, 68
White men	265	76	71, 81	65	60, 69
White women	212	72	67, 77	67	61, 72
Age (at event)					
<65 years	152	83	76, 88	42	37, 47
65 years	649	75	72, 78	73	70, 76
Community					
Forsyth Co, NC	180	78	72, 83	67	61, 73
Jackson, MS	282	80	75, 84	61	57, 66
Minneapolis, MN	163	78	72, 84	70	64, 76
Washington Co, MD	176	68	62, 74	60	54, 66
Hospital Type					
Teaching	274	81	76, 85	71	66, 75
Non-teaching	466	75	71, 78	60	56, 63
Incidence [#]					
Incident	548	80	76, 83	64	60, 67
Recurrent	253	79	74, 83	65	60, 70
Year of event					
1987–1990	0	NA		NA	
1991–1994	38	78	63, 88	24	17, 31
1995–1998	142	79	72, 85	61	55, 68
1999–2002	174	85	80, 90	61	55, 67
2003–2006	201	74	69, 79	83	77, 87

	Ischemic Stroke [*]	PPV [†]		Sensitivity [‡]	
		(N)	(%)	95% CI	(%)
2007–2010	246	72	66, 76	93	89, 96

Abbreviations: AHA/ASA, American Heart Association/American Stroke Association; CI, confidence interval; ICD-9-CM, *International Classification of Disease, 9th Revision, Clinical Modification*; PPV, positive predictive value.

* Definite or probable ischemic stroke in ARIC with AHA/ASA ischemic stroke code group ICD-9-CM hospital discharge code.

† Percent of hospitalizations with AHA/ASA-specified ischemic stroke ICD-9-CM codes classified in ARIC as definite/probable ischemic stroke.

‡ Percent of definite or probable ischemic stroke in ARIC (N=1251) with AHA/ASA-specified ischemic stroke ICD-9-CM discharge codes.

§ Hospitalizations among white individuals or black individuals from Forsyth or Jackson (N=4, 235).

// Hospitalizations occurring within ARIC Study catchment areas (N=3, 915)

Hospitalizations where symptoms lasted ≥ 24 hours (N=3, 250).

Table 4

Positive predictive value and sensitivity of the AHA/ASA ICD-9-CM code group for ICH by patient and hospital characteristics: the ARIC Cohort Study, 1987–2010

	ICH* (N)	PPV [†]		Sensitivity [‡]	
		(%)	95% CI	(%)	95% CI
ICD-9-CM Code Position					
First	92	71	62, 78	78	68, 84
Any	96	57	49, 65	80	72, 87
By Patient & Hospital Characteristics (any position code)					
Sex					
Men	41	49	38, 60	79	65, 89
Women	55	65	54, 76	81	70, 89
Race[§]					
Black	49	75	63, 85	84	73, 93
White	45	45	35, 55	75	62, 85
Sex by race[§]					
Black men	19	66	46, 82	76	55, 91
Black women	30	83	67, 94	91	76, 98
White men	21	39	26, 53	81	61, 93
White women	24	51	36, 66	71	53, 85
Age (at event)					
<65 years	35	74	60, 86	85	71, 94
65 years	61	50	41, 60	77	66, 86
Community					
Forsyth Co, NC	19	45	30, 61	90	70, 99
Jackson, MS	45	80	68, 90	83	71, 92
Minneapolis, MN	9	31	15, 51	69	39, 91
Washington Co, MD	23	56	40, 72	72	53, 86
Hospital Type					
Teaching	25	56	40, 70	83	65, 94
Non-teaching	62	59	49, 69	78	68, 87
Incidence[#]					
Incident	67	62	52, 71	77	67, 85
Recurrent	29	53	39, 66	88	72, 97
Year of event					
1987–1990	9	90	56, 100	75	43, 95
1991–1994	12	57	34, 78	80	52, 96
1995–1998	19	63	44, 80	100	82, 100
1999–2002	19	59	41, 76	86	65, 97
2003–2006	18	47	31, 64	72	51, 88

	ICH [*]	PPV [†]		Sensitivity [‡]	
	(N)	(%)	95% CI	(%)	95% CI
2007–2010	19	51	34, 68	70	50, 86

Abbreviations: AHA/ASA, American Heart Association/American Stroke Association; CI, confidence interval; ICD-9-CM, *International Classification of Disease, 9th Revision, Clinical Modification*; ICH, intracerebral hemorrhage; PPV, positive predictive value.

* Definite or probable ICH in ARIC with ICD-9-CM hospital discharge code 431.

† Percent of hospitalizations with ICD-9-CM discharge code 431 classified in ARIC as definite/probable ICH.

‡ Percent of definite or probable ICH in ARIC (N=120) with ICD-9-CM code 431.

§ Hospitalizations among white individuals or black individuals from Forsyth or Jackson (N=4, 235).

// Hospitalizations occurring within ARIC Study catchment areas (N=3, 915).

Hospitalizations where symptoms lasted ≥ 24 hours (N=3, 250).