International Journal of Epidemiology, 2019, 994–1003
doi: 10.1093/ije/dyz025
Advance Access Publication Date: 16 March 2019
Original article

Temporal trends in validated ischaemic stroke hospitalizations in the USA

Silvia Koton,[1,2]* Lisa Wruck,[3] Pedro Miguel Quibrera,[4] Rebecca F Gottesman,[2] Sunil K Agarwal,[2] Sydney A Jones,[5] Jacqueline D Wright,[6] Eyal Shahar,[7] Josef Coresh[2] and Wayne D Rosamond[5]

[1]Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, [2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, [3]Duke Clinical Research Institute, Duke University, Durham, NC, USA, [4]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [5]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [6]National Heart, Lung, and Blood Institute, Bethesda, MD, USA and [7]Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

*Corresponding author. The Stanley Steyer School of Health Professions, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: koton@tauex.tau.ac.il

Editorial decision 12 February 2019; Accepted 13 March 2019

Abstract

Background: Accurate assessment of the burden of stroke, a major cause of disability and death, is crucial. We aimed to estimate rates of validated ischaemic stroke hospitalizations in the USA during 1998–2011.

Methods: We used the Atherosclerosis Risk in Communities (ARIC) study cohort's adjudicated stroke data for participants aged ≥55 years, to construct validation models for each International Classification of Diseases (ICD)-code group and patient covariates. These models were applied to the Nationwide Inpatient Sample (NIS) data to estimate the probability of validated ischaemic stroke for each eligible hospitalization. Rates and trends in NIS using ICD codes vs estimates of validated ischaemic stroke were compared.

Results: After applying validation models, the estimated annual average rate of validated ischaemic stroke hospitalizations in the USA during 1998–2011 was 3.37 [95% confidence interval (CI): 3.31, 3.43) per 1000 person-years. Validated rates declined during 1998–2011 from 4.7/1000 to 2.9/1000; however, the decline was limited to 1998–2007, with no further decline subsequently through 2011. Validation models showed that the false-positive (∼23% of strokes) and false-negative rates of ICD-9-CM codes in primary position for ischaemic stroke approximately cancel. Therefore, estimates of ischaemic stroke hospitalizations did not substantially change after applying validation models.

Conclusions: Overall, ischaemic stroke hospitalization rates in the USA have declined during 1998–2007, but no further decline was observed from 2007 to 2011. Validated ischaemic stroke hospitalizations estimates were similar to published estimates of hospitalizations with ischaemic stroke ICD codes in primary position. Validation of national discharge data using prospective chart review data is important to estimate the accuracy of reported burden of stroke.

Key words: Atherosclerosis Risk in Communities, Nationwide Inpatient Sample, National Inpatient Sample, stroke, trends, epidemiology, cardiovascular

Key Messages

- Ischemic stroke hospitalization rates in the US have declined from 1998 to 2007 while no further decline was observed from 2007 to 2011.
- Following application of validation models, the number and rate of ischemic stroke hospitalizations for the US population 55 years and older did not noticeably differ from official estimates based ischemic stroke ICD codes in the primary position since false positive (~23% of strokes) and false negative rates cancel.
- Estimates of validated ischemic stroke hospitalizations in the US population 55 years and older were similar to published estimates based on reports with ischemic stroke ICD-codes in primary position. Nevertheless, validation of national discharge data using prospective chart review data is important to estimate the accuracy of reported burden of stroke.

Introduction

Stroke ranks fifth among causes of death in the USA,[1] is a leading cause of serious physical and cognitive long-term disability[2] and poses a substantial economic burden, costing the USA an estimated $40.1 billion each year in 2013 to 2014.[3] According to the American Heart Association report, each year about 795 000 people in the USA have a stroke, approximately 610 000 of which are first-ever events.[3] Stroke is a heterogeneous disease[4]; however, 87% of all stroke events are ischaemic strokes.[3] Studies on trends in stroke incidence have generally shown a decline in incidence in recent decades, but this trend has been established only among White[5,6] and older population groups.[7] Therefore, accurate appraisal of stroke morbidity and mortality is important.

Hospital administrative data are increasingly being used for stroke surveillance. Stroke hospitalization reports may include several International Classification of Disease, 9th Revision, Clinical-Modification (ICD-9-CM) codes for one single hospital discharge. In addition to the limitations of administrative data regarding the accuracy of reported codes,[8] the use of ICD codes in different positions (primary position only compared with any position of the relevant codes) influences the validity of reports on the burden of stroke.[9] Counting the number of hospitalizations with stroke ICD-9-CM codes in any position may overestimate the burden of stroke. Therefore, estimates of the burden of hospitalized stroke usually include only hospitalizations with an ICD-9-CM code for stroke in the primary position. Nevertheless, this strategy may underestimate the true number of stroke hospitalizations. Validation of each ICD-9-CM code for stroke in large administrative datasets is not possible. However, validation via chart review from large epidemiological studies can be used to develop models which relate demographic and clinical data to the probability of validated stroke by ICD-9-CM code. These models can be applied to administrative data to potentially improve estimates of stroke burden. The updated definition of stroke published in 2013 by the American Stroke Association[10] was previously validated using data in the Atherosclerosis Risk in Communities (ARIC) study.[11] In the present study, we used adjudicated data available in the ARIC study cohort[12] to build models for ischaemic stroke hospitalizations, and applied them to the large, longitudinal Nationwide Inpatient Sample (NIS) to study temporal trends in ischaemic stroke hospitalizations in the USA from 1998 to 2011.

Materials and Methods

Study setting and design

ARIC is a population-based study of 15 792 individuals aged 45–64 years at baseline (1987–89), recruited from Forsyth County, North Carolina; Jackson, Mississippi;

suburbs of Minneapolis, Minnesota; and Washington County, Maryland; four geographical locations with urban, suburban and rural settings in the US. The ARIC cohort study is considered a reliable source of information on stroke incidence, recurrence and mortality by professional organizations and health policy makers in the USA. Participants were prospectively followed through examinations, annual phone interviews, active surveillance of discharges from local hospitals, and linkage with the National Death Index.[13] The institutional review board at each study centre approved the methods, and all participants provided informed consent.

Data on stroke are collected in the ARIC study cohort, all stroke events are adjudicated by physician reviewers and classified by stroke type.[7,12] For the present analysis we included definite/probable strokes. A definite or probable stroke was defined as a sudden and rapid onset of neurological symptoms lasting >24 h or leading to death, in the absence of evidence for a non-stroke cause. The sample used for the derivation of models for validated ischaemic stroke hospitalization included 3344 stroke hospitalizations reported in ARIC in 1998–2011.

The NIS is the largest all-payer inpatient care database that is publicly available in the US. For the years before 2012, the NIS contained all discharge data from more than 1000 hospitals each year, a 20% stratified sample of US community hospitals. Criteria for the stratified sampling of hospitals included geographical region, location (urban or rural), teaching status, ownership and bed-size category. For the present study, national estimates for analysis of trends were created using NIS Trend Weights Files, available yearly.[14] Sampling methods were consistent between 1998 and 2011,[15,16] and therefore data allow for study of trends in hospitalizations over that period. The sample to which our models were applied consisted of 3 070 691 eligible hospitalizations (weighted number of hospitalizations: 15 085 315) reported in NIS in 1998–2011 among participants aged ≥55 years at discharge, corresponding to the age of ARIC participants. Stroke hospitalizations with ICD-9-CM codes 432, 433, 433.X, 433.X1, 434, 434.X, 434.X1, 435, 436 in any position, or codes 430, 431 in primary position, were considered eligible. We acknowledge that ICD-10 is now standard practice. However, historical datasets are important to understand long-term trends, and the validation of ICD-9 is still relevant to trend analysis. Future similar work should address the impact of ICD-10 on true burden in the population.

ICD-9-CM code-based definitions of stroke and model covariates

ICD-9-CM codes 433, 433.X, 433.X1, 434, 434.X, 434.X1 and 436 were defined as ischaemic stroke codes, and two mutually exclusive group definitions were created based on the code's position and applied consistently across the models' derivation (ARIC) and application (NIS) samples: group 1, ischaemic stroke code in primary position; and group 2, ischaemic stroke code in non-primary position. The first among 25 available positions was used to categorize the code position, with the first position used to denote the primary code. NIS data on covariates were extracted using ICD-9 codes. Only covariates available both in ARIC and NIS were considered. These included race, sex, age at discharge, teaching hospital status and selected additional variables defined by the presence of the following codes in any position: hypertension (ICD-9-CM 401–405), hyperlipidaemia (ICD-9-CM 272.0–272.2, 272.4), diabetes mellitus (ICD-9-CM 250), atrial fibrillation (ICD-9-CM 427.31, 427.32), coronary artery disease/ischaemic heart disease (ICD-9-CM 410–414) and heart failure (ICD-9-CM 428).

Statistical methods

For the derivation of validation models, first we estimated the unadjusted positive predictive value (PPV) of each discharge group. PPV was defined as the proportion of events categorized as stroke hospitalizations that were true (i.e. definite/probable) ischaemic stroke hospitalizations according to the ARIC adjudication criteria (gold standard). We then constructed corresponding validation models for each discharge code group/outcome combination, in which we used logistic regression to model the predictive value of each code group, including covariates for the gold standard stroke diagnosis. Similar methods were previously used to estimate rates and trends of acute decompensated heart failure hospitalization in the USA.[17] Race, sex and age at discharge were forced into each model. Other covariates were selected for inclusion in the model using step-wise backwards elimination, with variables retained if the likelihood ratio *P*-value was less than 0.20. We considered three parameterizations of age (linear, quadratic and categorical) as well as calendar time period as an alternative to age, and selected final models using the quadratic age parameterization based on area under the receiver operating characteristic curve (AUC) and Akaike information criterion (AIC). Hosmer-Lemeshow statistics were also calculated for each model. Temporal and geographical trends in PPV were tested by fitting final models with interactions by time period and ARIC study centre, respectively.

Unadjusted PPV and best-fitting validation models for gold standard diagnosis of ischaemic stroke were applied to the NIS dataset to estimate the probability of ischaemic stroke for each hospitalization in the application sample. Stroke prevalence rates in ARIC and the US population for the same age groups are similar.[18] This similarity enhances the validity of applying PPVs from ARIC for validated stroke to the NIS eligible stroke codes for the relevant age groups. The sum of these probabilities yielded new national estimates of the number of validated ischaemic stroke hospitalizations. The variance in these estimates was obtained via the Delta Method. As recommended by NIS, all analyses were weighted to represent all US community hospitals, and variance estimates accounted for the stratified sampling design. Race, an important covariate in the validation models, was missing at a high rate in the NIS data (weighted percentage: 21.6%). Examination of the data, correspondence with Healthcare Cost and Utilization Project (HCUP) staff and review of NIS documentation made clear that the missingness mechanism occurs at the state level. We used multiple imputation methods to impute race, based on two different strategies[19]: (i) imputation was based on state-specific racial distributions as defined by the United States Census data; and (ii) imputation was based on state-specific models including age, gender, median household income for ZIP code of patient residence, expected primary payer, and rural/urban-teaching/urban-non-teaching hospital. Imputation models included covariates in the main model as well as NIS sample design variables. For states with nearly complete missingness of race, models were run within regions as defined by NIS. Imputations were run five times and yielded stroke hospitalization counts. We report mean counts over the imputations with standard errors that reflect the imputation procedure. Results were consistent between the two strategies, so only results of the first strategy were reported. US intercensal and postcensal population estimates were used as denominators to calculate annual rates of hospitalizations. These population estimates were treated as fixed in standard error calculations. Models were applied across years, yielding annual estimates and secular trends from 1998 through 2011. Average annual percentage change in number or rate of hospitalizations was estimated on the basis of regression models of annual log counts or rates, with inverse variance weighting. Temporal trends in the validated rates of ischaemic stroke hospitalizations were compared with officially reported rates in the NIS. All analyses were conducted using SAS 9.4 (SAS Institute Inc., Cary, NC).

Results

Eligible hospitalizations

During 1998–2011, the number of eligible admissions was 3344 in the ARIC study cohort and over 3 million (nationally weighted to over 15 million hospitalizations) in the NIS sample ≥55 year-old (the application sample). Characteristics of eligible hospitalizations in both samples are shown in Table 1. Patients in the ARIC study cohort were on average ∼2.5 years younger than in the NIS. Differences in risk factors between the ARIC and NIS varied by stroke type and by position of stroke code (Table 1). Among patients with an ischaemic stroke code in non-primary position, the most common primary position ICD-9-CM codes were for heart failure and coronary heart disease, both in ARIC participants (20.4%) and in the NIS (21.1%).

ARIC ischaemic stroke hospitalization validation

PPVs for validated ischaemic stroke hospitalizations are shown in Table 2. The PPV was higher for ischaemic stroke code in primary position (PPV: 79.2%), than for ischaemic stroke code in non-primary position (PPV: 50.9%). As expected, hospitalizations with a code corresponding to haemorrhagic stroke or other eligible cerebrovascular event had very low PPVs for validated ischaemic stroke (6.4% and 5.4%, respectively).

The validation models for each ICD code group for ischaemic stroke included race, sex and age at discharge. Additional covariates were entered as described in the Methods section. The final models are presented in Supplementary Table 1, available as Supplementary data at *IJE* online. The probability of validated ischaemic stroke based on the prediction models (50th; 2.5th, 97.5th percentile showing the predictive range) was 0.80; 0.69, 0.88 for ischaemic stroke codes in primary position and 0.47; 0.33, 0.80 for ischaemic stroke codes in non-primary position (Table 2).

Validated ischaemic stroke hospitalizations in the USA

Validated numbers and rates of stroke hospitalizations were calculated by applying the validation models for each ICD code group (Supplementary Table 1, available as Supplementary data at *IJE* online, with average PPVs shown in Table 2) to eligible hospitalizations (Table 3, top section). Among the 359 000 hospitalizations with ischaemic stroke in the primary position, 275 000 (77%) were estimated to be validated ischaemic stroke hospitalizations (Table 3, bottom section). However, an additional 85 000 of the hospitalizations without an ischaemic stroke code in primary position were also estimated to be ischaemic stroke hospitalizations (55 000 with ischaemic stroke in the non-primary position, 4000 with haemorrhagic stroke code in the primary position and 26 000 with other stroke eligible ICD codes). Therefore, the total estimated number of validated ischaemic stroke hospitalizations (360 000 hospitalizations) is almost identical to the annual average

Table 1. Characteristics of stroke eligible[a] hospitalizations (%) for the US population aged 55 years and older in the Atherosclerosis Risks in Communities (ARIC) study cohort (1998–2011) and the Nationwide Inpatient Sample (NIS), by code group (1998–2011)

	ARIC stroke cohort				Nationwide Inpatient Sample[b]			
	Ischaemic stroke code[a] primary position $n = 1067$	Ischaemic stroke code[a] non-primary position $n = 289$	Haemorrhagic stroke codes[a] primary position $n = 157$	Other stroke eligible code[a] $n = 1831$	Ischaemic stroke code[a] primary position $n = 5.02$M	Ischaemic stroke code[a] non-primary $n = 1.38$M	Haemorrhagic stroke code[a] primary position $n = 0.91$M	Other stroke eligible code[a] $n = 7.78$M
Age, mean, years	72.2	72.5	71.4	72.7	74.6	74.9	73.1	74.2
Age group, years								
55–64	16.4	13.2	20.4	13.8	18.6	16.6	24.0	17.0
65–74	48.6	49.1	43.3	50.4	27.1	27.9	27.3	30.9
≥75	35.1	37.7	36.3	35.8	54.3	55.5	48.7	52.1
Teaching hospital	29.2	24.9	28.7	23.9	39.4	42.4	56.4	40.9
Male	46.3	41.2	47.8	52.1	45.4	45.2	45.2	48.6
African American[c]	39.3	30.5	36.3	22.5	14.1	13.4	12.1	7.9
Hypertension	67.1	50.5	64.3	60.7	64.4	42.5	64.4	60.9
Current smoking	16.1	14.2	11.5	18.6	13.9	8.1	11.7	17.9
Diabetes	40.2	43.6	20.4	36.9	32.3	28.7	22.0	29.4
Dyslipidaemia	41.3	34.6	19.8	45.7	31.6	17.9	19.2	37.4
Ischaemic heart disease[d]	28.4	46.4	17.8	47.4	27.5	38.2	20.0	42.3
Heart failure	10.9	23.5	7.0	17.5	13.6	27.2	9.7	14.1
Atrial fibrillation/Atrial flutter	15.7	27.0	8.3	18.1	22.0	26.3	18.0	15.8
CABG/PTCA	14.0	14.5	9.6	21.9	9.7	6.9	7.4	16.5
History of stroke/TIA	11.2	9.7	12.12	15.0	7.1	3.9	6.4	10.5
Chronic kidney disease	5.9	11.8	3.8	10.0	5.7	9.1	4.6	6.8

Numbers shown are percentages unless otherwise noted.

M, million; CABG/PTCA, coronary artery bypass graft surgery and percutaneous transluminal coronary angioplasty; TIA, transient ischaemic attack.

[a]Eligible sample of hospitalizations among the US population aged 55 years and older are defined by hospital ICD-9-CM codes 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436 for ischaemic stroke, ICD-9-CM codes 430 and 431 for haemorrhagic stroke (430 for subarachnoid haemorrhage and 431 for intracerebral haemorrhage), and ICD-9-CM codes 432.X and 435.X for other stroke codes.

[b]NIS data were weighted so that results are representative of the US population.

[c]Estimated among hospitalizations with non-missing race in NIS (the weighted percent of NIS eligible hospitalizations with race missing was 21.6%).

[d]Ischaemic heart disease includes acute myocardial infarction, other acute and subacute forms of ischaemic heart disease, old myocardial infarction, angina pectoris and other forms of chronic ischaemic heart disease.

Table 2. Validation of stroke hospitalizations in the Atherosclerosis Risk in Communities (ARIC) study, 1998–2011

	Ischaemic stroke ICD code, primary position	Ischaemic stroke ICD code, non-primary position	Haemorrhagic stroke ICD code, primary position	Other stroke eligible ICD code, any position
Total number of hospitalizations[a]	1067	289	157	1831
Validated ischaemic stroke PPV (95% CI)	0.792 (0.768, 0.816)	0.509 (0.451, 0.566)	0.064 (0.025, 0.102)	0.054 (0.044, 0.064)
Probability of validated ischaemic stroke based on prediction models, 50th (2.5th–97.5th) percentiles	0.80 (0.69, 0.88)	0.47 (0.33, 0.80)	0.03 (0.01, 0.40)	0.04 (0.02, 0.16)
Ischaemic stroke validation model AUC	0.578	0.671	0.773	0.680

[a]Hospitalizations with the following ICD-9-CM discharge code group definitions: ischaemic stroke code (ICD-9-CM 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436) in primary position, ischaemic stroke code (ICD-9-CM 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436) in non-primary position, haemorrhagic stroke code (ICD-9-CM 430, 431) in primary position and other stroke code (ICD-9-CM 432, 435) in any position. The first position of a code among 25 available code positions was used to categorize its position, with the first position used to denote the primary code.

PPV = positive predictive value, AUC = area under the receiver operating characteristics curve.

Table 3. Stroke hospitalizations in the US population aged 55 years and older, the Nationwide Inpatient Sample (NIS) data 1998–2011

	Ischaemic stroke ICD code, primary position	Ischaemic stroke ICD code, non-primary position	Hemorrhagic stroke ICD code, primary position	Other stroke eligible ICD code, any position	Total
Eligible hospitalizations[a]					
Total number (weighted), thousands	5024	1379	907	7776	15086
Annual average number (95% CI), thousands	359 (348, 369)	98 (95, 102)	65 (62, 68)	555 (537, 574)	1078 (1044, 1111)
Annual average rate[b] (95% CI), per 1000 person-years	3.31 (3.21, 3.41)	0.91 (0.88, 0.94)	0.60 (0.57, 0.62)	5.12 (4.95, 5.30)	9.94 (9.63, 10.25)
Validated ischaemic stroke hospitalizations					
Annual average number[c] (95% CI), thousands	275 (269, 282)	55 (53, 56)	4 (3, 4)	26 (24, 27)	360 (353, 366)
Annual average rate[b] (95% CI), per 1 000 person-years	2.57 (2.52, 2.63)	0.51 (0.50, 0.53)	0.03 (0.03, 0.04)	0.25 (0.23, 0.26)	3.37 (3.31, 3.43)

[a]Hospitalizations with the following ICD-9-CM discharge code group definitions: ischaemic stroke code (ICD-9-CM 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436) in primary position, ischaemic stroke code (ICD-9-CM 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436) in non-primary position, hemorrhagic stroke code (ICD-9-CM 430 and 431) in primary position and other stroke code (ICD-9-CM 432, 435) in any position. The first position of a code among 25 available code positions was used to categorize its position, with the first position used to denote the primary code.

[b]US inter-censal and post-censal population estimates were used as denominators to calculate rates of stroke hospitalizations

[c]Stroke hospitalizations in NIS were estimated after applying the relevant ARIC validation model to each stroke code group. Validation models (Supplementary Table 1, available as Supplementary data at *IJE* online) included age, age^2, race, sex for all models and the following additional variables by model: Ischaemic stroke code in primary position—atrial fibrillation; Ischaemic stroke code in non-primary position—HF or CHD code in primary position, teaching status of hospital; Haemorrhagic stroke code in primary position—atrial fibrillation; Other stroke code—hyperlipidaemia

number of eligible hospitalizations with ischaemic stroke code in the primary position (359 000, Table 3). Likewise, the annual average rate of validated hospitalizations per 1000 person-years was 3.37, similar to the rate of ischaemic stroke hospitalizations with ischaemic stroke ICD codes in the primary position (3.31 per 1000 person-years).

Temporal trends in eligible hospitalizations and in validated ischaemic stroke hospitalizations

Trends in stroke hospitalizations in the US ≥55-year-old population in 1998–2011 are shown in Figures 1 and 2 and Supplementary Table 2, available as Supplementary data at *IJE* online. Numbers of eligible hospitalizations in the USA decreased from 1 150 713 in 1998 to 1 010 111 in 2005, and then increased to 1 092 589 in 2011, whereas annual average numbers (and rates) of validated ischaemic stroke hospitalizations decreased from 431 998 (4.7/1000) in 1998 to 320 437 (2.8/1000) in 2007, increasing afterwards to 356 438 in 2011 (2.9/1000). The decrease in annual rates represents a relative percentage change of −38.3% (95% CI: −33.1%, −43.2%) from 1998 to 2011. The total number and pattern by year of the validated ischaemic stroke hospitalizations are very similar to those

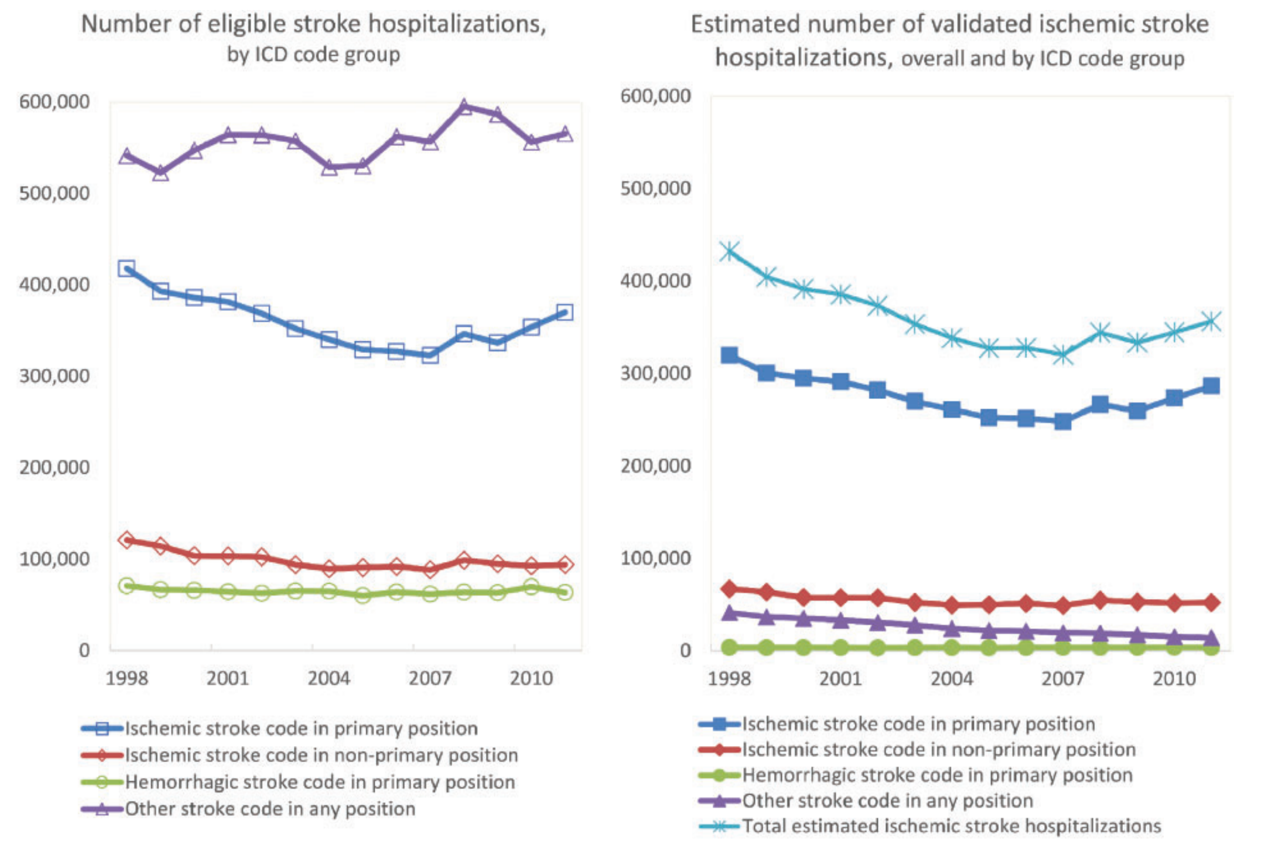

Figure 1. Temporal trends in number of eligible stroke hospitalizations and estimated number of validated ischaemic stroke hospitalizations in the US population aged 55 years and older, 1998–2011. Eligible hospitalizations were those identified in NIS with ICD-9-CM codes for stroke subtypes. The validated numbers of stroke hospitalizations were calculated by applying the validation models for each ICD code group to eligible hospitalizations.

for eligible hospitalizations with ischaemic stroke codes in primary position (Figures 1 and 2).

Discussion

Application of PPVs from validation models derived using the prospective ARIC data for estimating the probability of ischaemic stroke for each hospitalization in the NIS dataset yielded new national estimates of the number of ischaemic stroke hospitalizations. Application of the ARIC-based validation models for stroke hospitalization resulted in little net change in estimates of the burden of ischaemic stroke hospitalizations, compared with the simpler approach of counting strokes coded in primary position only. However, only 77% of the NIS hospitalizations with an ischaemic stroke code in primary position were validated ischaemic stroke hospitalizations, whereas the remaining 23% were in fact wrongly categorized as ischaemic stroke (false-positive rate of 23%). On the other hand, a similar number of validated ischaemic stroke hospitalizations (false-negatives) were missed if one assumed hospitalizations without a stroke code in the primary position were not strokes. The number of these missed strokes can be identified by applying validation models to hospitalizations with ischaemic stroke ICD codes in non-primary position (average PPV 50.9%), haemorrhagic stroke codes in primary position (PPV 6.4%) and other eligible stroke codes in any position (PPV 5.4%), which would represent the false-negative ischaemic stroke hospitalizations. The similar number of false-positive and false-negative strokes based on ICD codes in the primary position for ischaemic stroke is fortuitous, and stands in marked contrast to the situation in heart failure where the number of false-negatives is far larger than false-positives.[17]

As for temporal trends based on the new estimates, we observed decreasing number and rate of stroke hospitalizations for ischaemic stroke during 1998–2007, whereas no further decline was observed afterwards. We have previously reported a substantial decrease in total and ischaemic stroke incidence rates in 1987–2011 in ARIC, based on the study of trends in adjudicated stroke events.[7] Our present findings on stroke hospitalizations are consistent with previous reports on stroke hospitalizations overall,[20] as well as ischaemic stroke hospitalizations.[21,22] A small number of studies assessed trends in stroke hospitalizations based on hospital-based diagnoses codes in primary or secondary

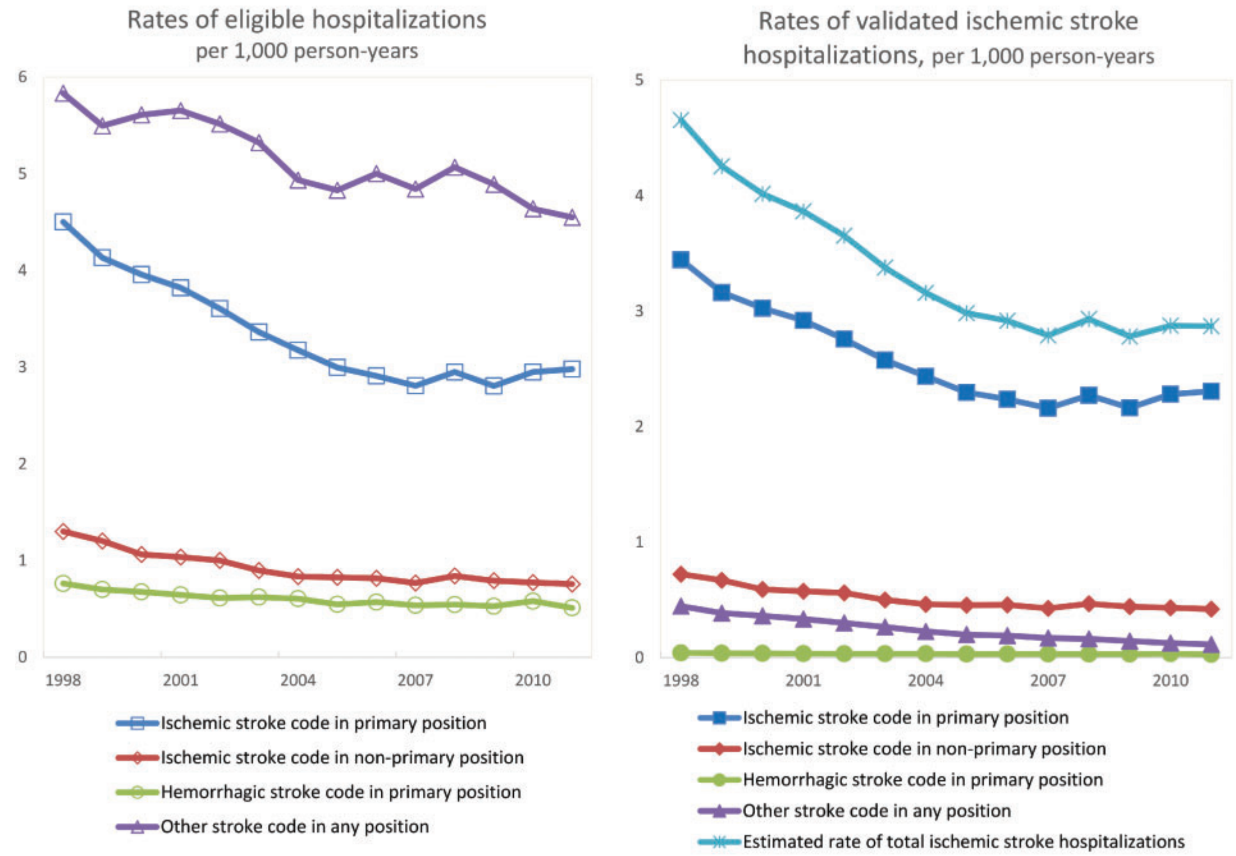

Figure 2. Temporal trends in rates of eligible hospitalizations and rates of validated ischaemic stroke hospitalizations (per 1000 person-years) in the US population aged 55 years and older, 1998–2011. US inter-censal and post-censal population estimates were used as denominators to calculate rates of stroke hospitalizations. Eligible hospitalizations were those identified in NIS with ICD-9-CM codes for stroke subtypes. The validated numbers of stroke hospitalizations were calculated by applying the validation models for each ICD code group to eligible hospitalizations.

position in the NIS, showing decreases in rates from 1997 to 2006.[23,24] However, unlike our present analysis, both studies included ICD-9-CM codes 437 (Other and ill-defined cerebrovascular disease) and 438 (Late effects of cerebrovascular disease). These codes were excluded from the ARIC stroke eligible target codes starting in 1997, since experience validating stroke events during 1987–96 showed that <2% of validated cases came from these code groups.[12] Consequently, our analysis did not include ICD-9-CM codes 437 and 438.

Using primary and all secondary ICD-9-CM stroke codes allows for identification of virtually all hospitalized strokes and provides a complete ascertainment of stroke hospitalizations.[25] However, it substantially overestimates the number of strokes.[23,25] Multiplying the number of discharges with each ICD-9-CM code by its estimated probability of stroke, derived from the ARIC dataset, corrected for potential over-ascertainment of stroke hospitalizations. We found that the probability of validated ischaemic stroke based on the prediction models was highest for hospitalizations with an ischaemic stroke code in primary position and lowest for those with other stroke ICD-9-CM codes in any position.

Our study allows for more valid estimates of the burden of ischaemic stroke hospitalizations in the USA, compared with estimates based on ICD codes only. The use of nationwide data, inclusion of both primary and secondary diagnosis codes, and validation of reported hospitalizations using probabilities derived in a population-based prospective cohort study, are important strengths of the study. However, there are some important considerations when interpreting the results. First, our study was restricted to hospital discharge data and did not include stroke events occurring outside the hospitals which, according to a previous data from the Northern Manhattan Stroke Study, account for about 5% of strokes.[26] Second, for 1998–2011, the NIS has lower percentage coverage for some regions, including the South, and therefore was less nationally representative than in recent years.[16] Third, as we aimed to assess the total burden of ischaemic stroke hospitalizations, our unit of analysis was hospitalization event. We included recurrent admissions and therefore our findings included all validated ischaemic strokes rather than incidence of new hospitalized ischaemic stroke. Also, our validation was restricted to ischaemic stroke hospitalizations among ≥55-year-old individuals. Validation of hospitalizations for other stroke types, as well as in patients <55 years old, are interesting topics for future studies. In addition, the study of hospitalizations with ischaemic strokes coded in secondary position, which comprise approximately 23% of the validated strokes, is warranted.

In conclusion, rates of validated ischaemic stroke hospitalization in the USA have declined from 1998 to 2007, but no further decline was observed from 2007 to 2011. Validation models using prospective chart review data can be applied to national hospital discharge data to more accurately estimate the burden of validated ischaemic stroke hospitalization in the USA. Fortuitously, for the US population aged 55 years and older, false-positive and false-negative ischaemic stroke numbers are similar, leading published estimates of ischaemic stroke in the primary position to be similar to the newly derived estimates of validated ischaemic stroke hospitalizations.

Supplementary Data

Supplementary data are available at *IJE* online.

ARIC study Funding

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. [HHSN268201700001I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I, HHSN2682017000021]. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Acknowledgements

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

Conflict of interest: None declared.

References

1. National Center for Health Statistics. *Health US, 2015: With Special Feature on Racial and Ethnic Health Disparities*. 2016. (28 February 2017, date last accessed).
2. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults: United States, 2005. *MMWR Morb Mortal Wkly Rep* 2009;**58**:421–26.
3. Benjamin EJ, Virani SS, Callaway CW *et al*. Heart disease and stroke statistics – 2018 update: a report from the American Heart Association. *Circulation* 2018;**137**:e493.
4. Sacco RL, Kargman DE, Gu Q, Zamanillo MC. Race-ethnicity and determinants of intracranial atherosclerotic cerebral infarction. The Northern Manhattan Stroke Study. *Stroke* 1995;**26**: 14–20.
5. Carandang R, Seshadri S, Beiser A *et al*. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA* 2006;**296**:2939–46.
6. Kleindorfer DO, Khoury J, Moomaw CJ *et al*. Stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. *Stroke* 2010;**41**:1326–31.
7. Koton S, Schneider AL, Rosamond WD *et al*. Stroke incidence and mortality trends in US communities, 1987 to 2011. *JAMA* 2014;**312**:259–68.
8. Sacco S, Pistoia F, Carolei A. Stroke tracked by administrative coding data: is it fair? *Stroke* 2013;**44**:1766–68.
9. Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. *Stroke* 2002;**33**:2465–70.
10. Sacco RL, Kasner SE, Broderick JP *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**:2064–89.
11. 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**: 3219–25.
12. Rosamond WD, Folsom AR, Chambless LE *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**:736–43.
13. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am J Epidemiol* 1989;**129**:687–702.
14. HCUP NIS Database Documentation. *Healthcare Cost and Utilization Project (HCUP)*, 2018. Rockville, MD: Agency for Healthcare Research and Quality. www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp (21 October 2018, date last accessed).
15. Agency for Healthcare Research and Quality. *2011 Introduction to the NIS. Healthcare Cost and Utilization Project (HCUP)*. 2016. (16 August 2016, date last accessed).
16. Houchens RL, Ross D, Elixhauser A, *Using the HCUP National Inpatient Sample to Estimate Trends*. 2015. HCUP Methods Series Report # 2006–05 ONLINE. 2016. (17 August 2017, date last accessed).
17. Agarwal SK, Wruck L, Quibrera M *et al*. Temporal trends in hospitalization for acute decompensated heart failure in the United States, 1998–2011. *Am J Epidemiol* 2016;**183**: 462–70.
18. Centers for Disease Control and Prevention (CDC). Prevalence of stroke – United States, 2006–2010. *Morb Mortal Wkly Rep* 2012;**61**:379–82.

19. Little RJA, Rubin DB. *Statistical Analysis With Missing Data,* 2nd edn. Hoboken, NJ: Wiley, 2002.
20. Fang J, Alderman MH, Keenan NL, Croft JB. Declining US stroke hospitalization since 1997: National hospital discharge survey, 1988–2004. *Neuroepidemiology* 2007;**29**:243–49.
21. Lee LK, Bateman BT, Wang S, Schumacher HC, Pile-Spellman J, Saposnik G. Trends in the hospitalization of ischaemic stroke in the United States, 1998–2007. *Int J Stroke* 2012;**7**:195–201.
22. Ramirez L, Kim-Tenser MA, Sanossian N *et al*. Trends in acute ischaemic stroke hospitalizations in the United States. *J Am Heart Assoc* 2016;**5**:e003233.
23. Ovbiagele B. National sex-specific trends in hospital-based stroke rates. *J Stroke Cerebrovas Dis* 2011;**20**:537–40.
24. Towfighi A, Markovic D, Ovbiagele B. Recent patterns of sex-specific midlife stroke hospitalization rates in the United States. *Stroke* 2011;**42**:3029–33.
25. Broderick J, Brott T, Kothari R *et al*. The Greater Cincinnati/Northern Kentucky Stroke Study: preliminary first-ever and total incidence rates of stroke among blacks. *Stroke* 1998;**29**:415–21.
26. Sacco RL, Boden-Albala B, Gan R *et al*. Stroke incidence among white, black, and Hispanic residents of an urban community: the Northern Manhattan Stroke Study. *Am J Epidemiol* 1998;**147**:259–68.