

ARIC Manuscript Proposal #2277

PC Reviewed: 12/10/13
SC Reviewed: _____

Status: A
Status: _____

Priority: 2
Priority: _____

1.a. Full Title: Burden of stroke in the US: Data from the National Inpatient Sample and the Atherosclerosis Risk in Communities (ARIC) study

b. Abbreviated Title (Length 26 characters): Burden of stroke in the US

2. Writing Group:

Silvia Koton
Andrea L.C. Schneider
Sunil K. Agarwal
Rebecca F. Gottesman
Sydney Jones
Jacqueline Wright
Eyal Shahar
Lisa Wruck
Joe Coresh
Wayne Rosamond
-Others welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. **S.K.**

First author: Silvia Koton, PhD
Address: 2024 E. Monument, Suite B-319
Baltimore, MD 21287
Phone: 443-287-1838
e-mail: skoton@jhsph.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located:

Name: **Josef Coresh, MD, PhD**
Address: 2024 E. Monument, Suite 2-630
Baltimore, MD 21287
Phone: 410-955-0495 Fax: 410-955-0476
E-mail: coresh@jhu.edu

3. Timeline: Starting immediately, manuscript to be completed within 3 months.

4. Rationale:

Stroke is a major cause of disability and death worldwide. In the US, stroke ranks fourth among all causes of death¹ and is a leading cause of serious physical and cognitive long-term disability². The estimated number of adult Americans who have had a stroke during 2007-2010 is 6.8 million. A 21.9% increase in the current 2.8% overall stroke prevalence is expected by 2030, adding 4 million people with stroke. According to the latest AHA report, almost 800,000 people have an incident or recurrent stroke each year, over 600,000 of them are first-ever events. Of all strokes, 87% are ischemic, 10% intracerebral hemorrhages, and 3% subarachnoid hemorrhages³. Stroke incidence differs by gender and ethnic group⁴⁻⁶. Studies on trends have generally shown a decline in stroke incidence in the last years, but this trend has been established only for the white population⁴⁻⁵.

The substantial health impact of stroke is reflected in the economic burden posed by the disease. In 2009, the mean cost for stroke care in the US was \$6018 per person; \$38.6 billion were invested to cover direct and indirect costs of care for persons with stroke. Moreover, the mean lifetime cost of ischemic stroke in the US is estimated at \$140 048³.

This significant health and economic burden emphasizes the importance of accurate appraisal of stroke morbidity and mortality. Hospital administrative data are available for large number of patients. Although originally designed to justify reimbursement, these data are increasingly being used for surveillance of stroke and for studies on stroke care quality and cost. In addition to the limitations of administrative data regarding the accuracy of reported codes⁷, using algorithms based on different combinations of diagnosis code positions influences the validity of reports on burden of stroke⁸.

Stroke is a heterogeneous disease that includes cerebral hemorrhages and several pathogenic subtypes of ischemic stroke⁹. Unlike other cardiovascular diseases, it is not consistently defined and reported in clinical practice or in studies on public health¹⁰. The use of the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes for brain ischemic events (43X.X1 for events with infarction and 43X.X0 for those without infarction) starting on October 1992 increased the proportion of patients correctly identified with acute ischemic stroke, but has not been successful for assignment of ischemic stroke subtypes¹¹.

As part of the efforts towards achieving a consistent categorization, an updated definition of stroke for the 21st century has been recently published by the Stroke Council of the American Heart Association/American Stroke Association (AHA/ASA). The new definition "incorporates clinical and tissue criteria and can be incorporated into practice, research, and assessments of the public health"¹⁰.

Data collected in the ARIC study are suitable for the validation of ICD administrative discharge diagnoses codes. A study based on the ARIC study cohort has shown higher PPV but lower sensitivity for the updated AHA/ASA stroke definition, compared with traditional codes (Jones et al, unpublished abstract). We aim to apply models which estimate the probability of acute stroke, (as well as the probability of particular type of stroke and subtype of ischemic stroke) for specific ICD codes in the primary and other positions to national data. Models using patient-level variables in the ARIC cohort will be examined and previous work (MP#2102, S Jones et al.) will be reviewed, as will the specific needs of an application to the National Inpatient Sample (NIS) database, which may require further modeling. Applying the probability of acute stroke to the NIS ICD codes will allow for national estimates of the frequency and rate of hospitalized stroke

(similar to work done in MP#1996, S. Agarwal et al.). We will also report patterns of care, length of stay (LoS), in-hospital procedures and costs related to LoS and procedures.

5. Main Hypothesis/Study Questions:

We hypothesize that the national rate of stroke admissions in the US is higher than estimated using discharge ICD codes for stroke reported in primary position. The secular trends similarly would be different than currently reported ones.

Secondary analyses will examine LoS and related costs for patients with different probabilities of stroke (high, medium, low) based on discharge diagnosis codes. We hypothesize that patients with a lower probability of stroke will have lower LoS and less reported in-hospital procedures than patients with a high probability of stroke (e.g. those with primary ICD codes), but much higher utilization than patients with codes for TIAs. We chose TIAs as the comparison group based on NIS data according to which 70% of all patients with TIA are admitted to hospitals in the US, providing a sample large enough for comparisons.¹²

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

ARIC cohort stroke hospitalizations related ICD-9-CM codes such as 433.X1, 434.X1, 436 for ischemic stroke, 430 for subarachnoid hemorrhage, and 431 for intracerebral hemorrhage, and others listed on the Appendix, Table1, in any position will be used. Codes for TIA and spinal and retinal infarction are not validated in ARIC, therefore will be excluded.

In total, 1,418 events have been validated as definite or probable stroke in the ARIC cohort, of them 1,252 ischemic strokes and 166 hemorrhagic strokes. Number of stroke hospitalizations reported in the NIS dataset for the same cohort is likely higher. Using the ARIC stroke classification as gold-standard, probability of stroke for each stroke type and ischemic stroke subtype code will be estimated. Sensitivity, specificity, PPV and NPV will be computed first for stroke codes in primary position and then using other positions. We will perform these analyses using both ICD-9-CM codes grouped according to the AHA/ASA updated stroke definition and the specific codes listed on Table 1.

Predictive model development: Predictive models for validated stroke in ARIC will be derived using variables available in **both** NIS and the ARIC cohort. This work will be done in collaboration with previous stroke modeling in ARIC (MP#2102), but tailored to the limitation of needing to apply the results to the NIS. Presence of a stroke ICD code and position of reported code will be the main model variables. Additional covariates are calendar time-period, teaching status of hospital, age group, gender, hypertension (ICD-9-CM code 401-405), diabetes (code 250) and coronary artery disease/ischemic heart disease (code 410-414). Other risk factors listed on Table 2 may also be considered. Interactions of predictors with time and geographic location will be explored since these may influence the extent to which ARIC stroke prediction models will generalize well. If meaningful interactions are found, different models will be studied in sensitivity analyses in deriving national estimates. Models' performance will be evaluated using calibration and discrimination, with more emphasis on calibration. Since models are expected to be generalizable, small models successfully performing will be preferred to avoid overfitting.

Application of acute stroke probability to NIS data: Best fitting models for all strokes, as well as for each stroke type and ischemic stroke subtype will be applied to the NIS 2011 dataset that includes information on 20% of hospitalizations from 1,051 hospitals in 45 states for 1998-2011. New national estimates of the number of stroke hospitalizations will be computed.

Our goal is not to derive individual level prediction, but estimates of the average probability (PPV) of stroke which will allow estimates of the number of acute strokes (total and by stroke type) among a group of people with a set of covariates (including ICD codes). Following the work in the acute decompensated heart failure (ADHF) modeling to estimate national burden, it is likely that using stratified models within broad ICD codes will be useful. We plan to group strokes as:

1. Primary position codes 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436, n (in ARIC)=882 (most commonly reported codes for ischemic stroke)
2. Primary position codes 430, 431, n=74 (hemorrhagic stroke- SAH and ICH codes)
3. Non-primary position codes 437.X, 438.X, 438.XX, n=774 (other and ill-defined cerebrovascular disease, and late effects of cerebrovascular disease)
4. Non-primary position codes 433, 433.X, 433.X1, 434, 434.X, 434.X1, 436, n=315 (ischemic stroke)
5. Non-primary position codes 430, 431, n=30 (hemorrhagic stroke- SAH and ICH codes)

We will use US census data as denominator to calculate rates. Models will be applied across years thus reporting recent estimates and secular trends from 1998 through 2011.

In-hospital mortality and health care utilization (LoS, procedures, and their related costs) will be evaluated based on the new estimates of number of hospitalizations for stroke. We will use summation of probabilities of stroke (phats) to get the health care utilization characteristics when a variable is categorical. For continuous variables we will multiply phat with the variable, for instance LoS, with phat to get fraction of LoS attributed to stroke and sum them up. We will discuss other possible methods to estimate health care utilization metrics.

Anticipated challenges/ limitations:

1. Limitations in the generalizability of ARIC models will be assessed and noted in the paper. To note, the lower the geographic and time variation in acute stroke probability conditional on the predictors, the more confidence we will have in the resulting national estimates.
2. There might be limited data for the study of the impact of health care system characteristics on estimates.
3. Data on severity of stroke, known to influence stroke outcome, length of stay and costs, are not available in the NIS dataset.
4. We expect the analysis by stroke type and ischemic stroke subtype to be particularly challenging.
5. Stroke cases in the outpatient settings or discharged from the ER or at home resulting in death before hospital admission will be missed.

7.a. Will the data be used for non-CVD analysis in this manuscript? Yes No

8. a. Will the DNA data be used in this manuscript? Yes No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: <http://www.cscce.unc.edu/ARIC/search.php> Yes No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

- 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
- Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2004;160:259-269
- Agarwal SK, Coresh J, Chang PP, Fox E, Wright J, Wruck L, Rosamond W. Burden of heart failure hospitalization in the US based on the National Inpatient Sample and the Atherosclerosis Risk in Communities surveillance program (manuscript in preparation; MP#1996)
- Jones S, Gottesman RF, Shahar E, Wruck L, Rosamond W. Validity of hospital discharge diagnosis codes for stroke: Evaluation of new American Heart Association/American Stroke Association definitions (unpublished abstract; MP#2102)

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? Yes No

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from <http://publicaccess.nih.gov/> are posted in <http://www.cscce.unc.edu/atic/index.php>, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.

References

1. Centers for Disease Control and Prevention, National Center for Health Statistics. Compressed mortality file 1999-2009. CDC WONDER online database, compiled for compressed mortality file 1999–2009 series 20, no. 20, 2012. Underlying cause-of-death 1999–2009. [Http://wonder.Cdc.Gov/mortsql.Html](http://wonder.Cdc.Gov/mortsql.Html). Retrieved november 15, 2013.
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;58:421–426.
3. 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. Heart disease and stroke statistics—2013 update: A report from the American Heart Association. *Circulation.* 2013;127:e6-e245
4. Carandang R, Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Kannel WB, Wolf PA. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA.* 2006;296:2939-2946
5. Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. 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;41:1326-1331
6. Reeves MJ, Bushnell CD, Howard G, Gargano JW, Duncan PW, Lynch G, Khatiwoda A, Lisabeth L. Sex differences in stroke: Epidemiology, clinical presentation, medical care, and outcomes. *Lancet Neurol.* 2008;7:915-926
7. Sacco S, Pistoia F, Carolei A. Stroke tracked by administrative coding data: Is it fair? *Stroke.* 2013;44:1766-1768
8. Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. *Stroke.* 2002;33:2465-2470
9. 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
10. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, Elkind MS, George MG, Hamdan AD, Higashida RT, Hoh BL, Janis LS, Kase CS, Kleindorfer DO, Lee JM, Moseley ME, Peterson ED, Turan TN, Valderrama AL, Vinters HV. 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
11. 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
12. Qureshi AI, Adil MM, Zacharatos H, Suri MF. Factors associated with length of hospitalization in patients admitted with transient ischemic attack in United States. *Stroke.* 2013;44:1601-1605

APPENDIX

Table 1: Hospital discharge ICD-9-CM stroke codes used for identification of ARIC stroke events.

ICD-9-CM code	ICD-9-CM classification
430	Subarachnoid hemorrhage
431	Intracerebral hemorrhage
433.01	Occlusion and stenosis of basilar artery with cerebral infarction
433.11	Occlusion and stenosis of carotid artery with cerebral infarction
433.21	Occlusion and stenosis of vertebral artery with cerebral infarction
433.31	Occlusion and stenosis of multiple and bilateral precerebral arteries with cerebral infarction
433.81	Occlusion and stenosis of other specified precerebral artery with cerebral infarction
433.91	Occlusion and stenosis of unspecified precerebral artery with cerebral infarction
434.01	Occlusion of cerebral arteries- Cerebral thrombosis with cerebral infarction
434.11	Occlusion of cerebral arteries- Cerebral embolism with cerebral infarction
434.91	Occlusion of cerebral arteries- Cerebral artery occlusion unspecified with cerebral infarction
436	Acute but ill-defined cerebrovascular disease

ICD-9-CM codes 437.X (Other and ill-defined cerebrovascular disease) and 438.XX (Late effects of cerebrovascular disease) were dropped from the ARIC stroke eligible target codes starting on 1997. Experience validating stroke events during 1987-1996 showed <2% of validated cases from this code groups.

Table 2: Risks factors for stroke and their corresponding ICD-9-CM.

Risk factor	ICD-9-CM code
<i>A. Based on previous reports¹</i>	
Hypertension	401.0, 401.1, 401.9, 405.01, 405.09, 405.11, 405.19, 405.91, 405.99
Diabetes mellitus	250.XX
Coronary artery disease/Ischemic heart disease	410-414
History of TIA/stroke	V12.54
AFib	427.31, 427.32
Hyperlipidemia	272.0-272.2, 272.4
Renal failure	586
Tobacco use	305.1
<i>B. Others suggested</i>	
Dementia	290.0, 290.1X-290.4X, 290.8, 290.9
Cancer	140-209
Migraine	346
Coagulation disorders	286
Artificial heart valve	V43.3
Obesity	278.00, 278.01
CABG/PTCA	V45.81, V45.82

Reference:

1. 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