ARIC Manuscript Proposal # 1863

PC Reviewed: 11/8/11	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title:

Development and validation of prediction models for hemorrhagic and ischemic stroke in the Rotterdam Study, Cardiovascular Health Study, and Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters):

Prediction of stroke subtypes

2. Writing Group:
Writing group members:

Rotterdam Study investigators:
Bart S. Ferket
Bob J.H. van Kempen
Renske G. Wieberdink
Peter J. Koudstaal
Albert Hofman
Ewout W. Steyerberg
M.G. Myriam Hunink
M. Arfan Ikram

ARIC investigators:
Eyal Shahar
Rebecca Gottesman
Woody Chambless
Aaron R. Folsom

CHS investigators:
Jorge R. Kizer
Richard A Kronmal
Will Longstreth

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _BSF____ **[please confirm with your initials electronically or in writing]**

First author: Bart S. Ferket
Address:
Depts. of Epidemiology and Radiology
Room Ee-2130
Erasmus MC
P.O.Box 2040
3000 CA Rotterdam
The Netherlands

Phone: +31 10 7043481 Fax: +31 10 7044657
E-mail: b.ferket@erasmusmc.nl

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Aaron R. Folsom
Address: Division of Epidemiology & Community Health
1300 South Second Street, Suite 300
Minneapolis, MN 55454

Phone: 626-8862 Fax: 612-624-0315
E-mail: folso001@umn.edu

3. Timeline:

4 weeks data preparation, 4 weeks data analysis, 12 weeks writing

4. Rationale:

Multiple atherosclerotic risk factors that influence stroke risk are well established and can be used to estimate an individual's stroke incidence over a 5 to 10-year time period.[1-4] However, these risk predictions apply to ischemic stroke only or to a combined endpoint of ischemic and hemorrhagic stroke. In addition, these were all developed using standard Cox regression modeling. Standard survival analysis will generally overestimate the cumulative incidence because it fails to regard those who die of non-stroke causes as ineligible for development of stroke events. Methods to adjust for competing risks are now increasingly being used for cardiovascular risk prediction.[5-6]

Simultaneous prediction of stroke subtypes (e.g. any, hemorrhagic, ischemic stroke) may be valuable for several reasons. First, the predictors for the different stroke subtypes may vary or may have different weights.[7-9] Consequently, the likely effects of modifying predictors can vary for stroke subtypes. Second, for some preventive

interventions, a difference in efficacy has been demonstrated across stroke subtypes. For example, aspirin decreases the occurrence of ischemic stroke events, whereas it increases the risk of an intra-cerebral bleeding.[10] Therefore, decision-making for aspirin therapy can be improved by predicting these two stroke subtypes at the same time. Finally, a more refined communication of the risk of stroke to the individual and the public can be facilitated.

5. Main Hypothesis/Study Questions:

The aim of this project is to develop and validate separate prediction models for estimation of the 10-year cumulative incidences of hemorrhagic, ischemic and any stroke within different geographical regions. The research hypothesis that will be tested is the assumption that effects of predictors will differ between stroke types, making separate prediction models necessary. An additional hypothesis that will be explored is the assumption that predictor effects will not differ between the cohorts, and that differences in risk between cohorts can be explained by variation in baseline risk.

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).

The study design is a pooled prospective study. We will use data from three population-based cohort studies: the Rotterdam Study, representing a middle-aged and elderly European population, the Cardiovascular Health Study and the Atherosclerosis Risk in Communities Study, which will be combined and represent the middle-aged and elderly US population. Prediction models for 10-year stroke subtype risks will be derived from a merged dataset with stratified baseline risks for US and European subjects.

The subjects eligible for these analyses are study participants who did not have prior stroke, did not use anticoagulation, and did not have atrial fibrillation at baseline. The outcomes of interest are defined as: time to fatal/non-fatal stroke event with specification for subtype (intracerebral hemorrhage, ischemic stroke, unspecified stroke); and time to death by other causes than the event of interest. Subarachnoid hemorrhages were excluded as these are not atherosclerotic of origin. The following cardiovascular risk factors that are measured at the baseline visit will be considered as the potential predictors: age; sex; current smoking; systolic blood pressure; diastolic blood pressure; blood pressure lowering treatment; history of cardiovascular disease other than stroke; diabetes mellitus (defined as: fasting glucose $\geq$126 mg/dL (7 mmol/L) or non-fasting glucose $\geq$ 200 mg/dL (11.1 mmol/L) or self-reported use of diabetes medications); total cholesterol; high-density lipoprotein cholesterol; body mass index; waist-to-hip ratio; race (black/white); estimated glomerular filtration rate using serum creatinine.

We will provide descriptive statistics on participants in the analysis stratified by sex. We will estimate predictor effects of cardiovascular risk factors on time to 1) intracerebral hemorrhage, 2) ischemic stroke, and 3) any stroke by using Cox proportional hazard models entering all predictors in the models. We assumed that the majority of unspecified stroke events would be ischemic of nature. Therefore we used a combined endpoint of classified ischemic and unspecified stroke events as a proxy for the true ischemic stroke hazard in order to avoid underestimation. Decisions on exclusion of predictors will be made upon the Akaike Information Criterion. Models will be developed in all subjects (cohorts merged) assuming that the baseline risk of the stroke types may differ between cohorts, but that predictor effects are similar. Interactions of included predictors with cohorts will be therefore tested. Effect modification will be examined by sex, and an interaction term for blood pressure lowering treatment with systolic blood pressure and diastolic blood pressure will be included if applicable. Subsequently, predictor effects will be compared amongst the 3 prediction models.

The standard Cox regression model, will generally overestimate the cumulative incidence because it fails to regard those who die of non-stroke causes as ineligible for development of stroke events. Because for clinical decision-making and risk communication, estimation of the cumulative incidence function (CIF) is required,[11] predictions will be adjusted for competing risk by non-stroke death. This will require estimation of effects of the predictors included in the 3 stroke models on 1) non-hemorrhagic stroke mortality, 2) non-ischemic stroke mortality and 3) non-stroke mortality by Cox regression modeling. The CIF can be estimated by adjusting the hazard of the event of interest for survival of all competing events. Prediction models estimating CIFs for hemorrhagic, ischemic and any stroke will be developed for a 10-year time horizon. Validation of the 3 prediction models will be performed in the different cohorts (Rotterdam Study[12] and ARIC / CHS combined) with respect to discrimination (assessed by the concordance index adjusted for competing risks), calibration (assessed by calibration plots, also adjusted for competing risks) and reclassification (assessed by the continuous net reclassification improvement, in the absence of established risk thresholds). In addition, models will be derived in one cohort (ARIC / CHS combined) and validated within the other cohort (Rotterdam Study) and vice versa, which is also known as cross-validation performance. We will also explore whether the predictions by the developed models are sensitive to race (white vs. black) and to age (e.g. < 65 years vs. ≥ 65 years).

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

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No

(This file ICTDER03 has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

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

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"?
____ 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.cscc.unc.edu/ARIC/search.php

__x____ 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)?

Potential overlap exists with:
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;160:259-69.

This study also uses established risk factors and a 10-year time horizon for prediction of stroke. It however only considers ischemic stroke and does not take into account the competing risk of death by other causes than ischemic stroke.

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

11.b. If yes, is the proposal
___ A. primarily the result of an ancillary study (list number* _____)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12. 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.

References

1. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. Stroke 1991;22:312-8.
2. Lumley T, Kronmal RA, Cushman M, Manolio TA, Goldstein S. A stroke prediction score in the elderly: validation and Web-based application. J Clin Epidemiol 2002;55:129-36.
3. 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;160:259-69.
4. Assmann G, Schulte H, Cullen P, Seedorf U. Assessing risk of myocardial infarction and stroke: new data from the Prospective Cardiovascular Munster (PROCAM) study. Eur J Clin Invest 2007;37:925-32.
5. Pencina MJ, D'Agostino RB, Sr., Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. Circulation 2009;119:3078-84.
6. Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. Bmj 2010;341:c6624.
7. Leppala JM, Virtamo J, Fogelholm R, Albanes D, Heinonen OP. Different risk factors for different stroke subtypes: association of blood pressure, cholesterol, and antioxidants. Stroke 1999;30:2535-40.
8. Hart CL, Hole DJ, Smith GD. The relation between cholesterol and haemorrhagic or ischaemic stroke in the Renfrew/Paisley study. J Epidemiol Community Health 2000;54:874-5.
9. Bots ML, Elwood PC, Nikitin Y, et al. Total and HDL cholesterol and risk of stroke. EUROSTROKE: a collaborative study among research centres in Europe. J Epidemiol Community Health 2002;56 Suppl 1:i19-24.
10. Antithrombotic Trialists C, Baigent C, Blackwell L, et al. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. Lancet 2009;373:1849-60.
11. Wolbers M, Koller MT, Witteman JC, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009;20:555-61.
12. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol 2009;24:553-72.