ARIC Manuscript Proposal # 1578

PC Reviewed: 11/10/09	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Prediction of atrial fibrillation in the community: the CHARGE consortium

b. Abbreviated Title (Length 26 characters): Prediction of atrial fibrillation

2. Writing Group:
Alvaro Alonso, David Couper, other ARIC investigators welcome.
CHS: Susan Heckbert, Richard Kronmal; Framingham: Emelia Benjamin, Martin Larson, Michael Pencina; Rotterdam: Bruno Stricker, Charlotte van Noord; AGES: Thor Aspelund, Vilmundur Gudnason.

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

First author: **Alvaro Alonso**
Address: Division of Epidemiology & Community Health
Univ. of Minnesota
1200 S 2nd St, Suite 300
Minneapolis, MN 55454
Phone: 612 626 8597 Fax: 612 624 0315
E-mail: alonso@umn.edu

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: **David Couper**
Address:

Phone: Fax:
E-mail:

3. Timeline:
Data request / DDA preparation: 1 month.
Analysis: 4-6 months, a first draft is expected in approximately 4 months
Submission for publication: 6-8 months after DDA approval.

4. Rationale:

Atrial fibrillation (AF) is an important public health problem. The prevalence of AF doubles for each advancing decade of life, affecting more than 10% of individuals over the age of 80 years.[1] The lifetime risk of AF is about 25%.[2] In addition, with the aging of the population, and increased survival with cardiovascular disease (CVD), the prevalence of AF is increasing over time.[3] Furthermore, AF is a major source of CVD morbidity and mortality. Risk factors for AF are multi-factorial and include CVD and its risk factors.[1, 4] However, the ability to accurately predict risk of AF in the individual has been limited.

The Framingham Heart Study (FHS) has recently published the first instrument for the prediction of incident AF.[5] However, the generalizability of the AF model to other populations remains unknown. The CHARGE cohorts (AGES, ARIC, Cardiovascular Health Study (CHS), FHS and Rotterdam study(RS)) have joined together for the present manuscript with the goals of conducting pooled analyses to markedly improve the ability to predict the onset of AF and advance statistical approaches to risk prediction.

We seek to improve the FHS risk prediction algorithm using a larger, more diverse population. In future proposals, we will additionally determine the added value of biomarkers (NT-proBNP, C-reactive protein) and genetic variants in the prediction of AF.

5. Main Hypothesis/Study Questions:

The main objective of this manuscript proposal is to validate, recalibrate and potentially modify the FHS AF clinical risk prediction model in diverse communities using widely available clinical factors.

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).

We aim to conduct the proposed research in the context of the CHARGE consortium. The CHARGE consortium started in 2008 to facilitate meta-analysis of GWAS and replication opportunities between a group of well-phenotyped cohort studies.[6] For the last year and a half the investigators have harmonized phenotypes (AF) and analytical approaches to conduct meta-analyses of GWAS data imputed to over 2.5 million single nucleotide polymorphisms (SNPs) from the HapMap.

<u>Ascertainment of atrial fibrillation</u>

The main outcome of interest in this proposal will be the incidence of AF, defined as the occurrence of paroxysmal, persistent or permanent AF, or atrial flutter, in individuals without AF at baseline. AF ascertainment has been done already in all

the participating cohorts. Methodology for the case ascertainment is similar across cohorts.

Specifically in ARIC, we determined AF from three sources: electrocardiograms at study visits, hospital discharge records and death certificates (latter two reviewed by trained abstractor, ICD-9 code 427.31 or 427.3; ICD-10 I48). Electrocardiogram-diagnosed AF in ARIC was confirmed by a cardiologist.[7] Incidence of AF was identified through 2005 as the first occurrence of AF by any of the sources in individuals who did not present AF in ECGs performed at baseline. A pilot study to determine the validity of hospital discharge codes as a method of AF ascertainment confirmed 111 (89%) diagnosis of a sample of 125 AF cases after review of available medical records corresponding to that hospitalization.[8]

Clinical variables

The five CHARGE cohorts included detailed assessments at baseline, gathering information on sociodemographic variables and lifestyle factors (smoking, alcohol intake), previous history of cardiovascular disease, and use of medication. Additionally, detailed physical exams collected anthropometric measures (height, weight) and seated blood pressure. Blood samples were obtained and analyzed to determine glucose, cholesterol (and its fractions), and triglycerides levels. In ARIC, CHS, FHS and RS, participants were examined in several occasions during the follow-up period.

Statistical analysis

Our main goal is to develop a new risk score for the prediction of AF pooling data from three cohorts (ARIC, CHS and FHS). This analysis will be done at Boston University under the supervision of Dr. Michael Pencina. We understand the need to prepare and get approved a DDA before any analysis can be done. AGES and RS will serve as an external validation data set (data transfer precluded by consent issues).

We will follow the methodology previously used in the development of risk scores in FHS. First, variables considered as potential predictors of AF will be those studied in the original FHS AF risk score and CHS AF risk factors publications:[4, 5] age, sex, race (in ARIC and CHS), smoking, alcohol consumption, body mass index, waist circumference, height, blood pressure indicators (systolic and diastolic blood pressure, and treatment for hypertension), cholesterol concentration, use of lipid lowering medication, prevalence of diabetes mellitus, ECG features (left ventricular hypertrophy, PR interval, and heart rate), and indicators of heart disease (history of heart failure, myocardial infarction, valve surgery, or CABG). Unadjusted and age-, sex-, and cohort-adjusted hazard ratios will be estimated for each potential risk factor. Risk factors that reach statistical significance and are considered clinically relevant will be considered candidates for a multivariable model. With large numbers of person-examinations available for analysis, statistical significance alone is not sufficient to identify important risk

factors. A multivariable Cox proportional hazards regression model will be estimated including risk factors identified as important in the first step. Calibration and discrimination will be evaluated for the multivariable model. Risk factors that do not improve calibration or discrimination may be removed from the model. Meaningful statistical interactions will be explored and judged important based on improvements in discrimination and calibration. Once a final model is developed, its performance in important groups (e.g. within cohorts, in men and women, in persons <65 years versus ≥65 years of age, etc) will be examined informally. Once a final multivariable model is developed, a points scoring system will be developed and disseminated. The points system is developed as follows:

(1) The Cox proportional hazard including all predictors of interest, censoring individuals at 10 years of follow-up is finalized.
(2) Organize the risk factors into categories (for continuous variables) and determine reference values for each risk factor.
(3) Determine the referent risk factor profile. This referent category for each risk factor will be assigned 0 points in the scoring system.
(4) Determine how far each category is from the base category in regression units.
(5) Set the fixed multiplier or constant B, that is, the number of regression units that reflect 1 point in the final points system. For example, FHS usually sets up the constant to be equivalent to the increase in risk associated with a 5-year increase in age. We will tentatively use the same approach.
(6) Determine the number of points for each of the categories of each risk factor.
(7) Determine risks associated with point totals.

Replication

We will examine the risk prediction equation derived from the pooled derivation cohorts in AGES and RS using methods described above.[9]

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)?

MS #1453 (Chamberlain) Development of a risk score for predicting atrial fibrillation in a bi-racial cohort: the Atherosclerosis Risk in Communities Study. This manuscript focuses on developing a risk score using data from ARIC, while the current proposal is more inclusive, with many more cases of atrial fibrillation. Alvaro Alonso, first author in the present manuscript proposal, is senior author in MS #1453.

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

11.b. If yes, is the proposal

_X_ A. primarily the result of an ancillary study (list number* 2008.12)

___ 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. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study. *JAMA.* 1994;271:840-844.
2. Lloyd-Jones DM, Wang TJ, Leip EP, et al. Lifetime risk for development of atrial fibrillation. The Framingham Heart Study. *Circulation.* 2004;110:1042-1046.
3. Miyasaka Y, Barnes ME, Gersh BJ, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation.* 2006;114:119-125.

4. Psaty BM, Manolio TA, Kuller LH, et al. Incidence of and risk factors for atrial fibrillation in older adults. *Circulation.* 1997;96:2455-2461.

5. Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet.* 2009;373:739-745.

6. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation Cardiovascular Genetics.* 2009;2:73-80.

7. Soliman EZ, Prineas RJ, Case D, Zhang Z-M, Goff DC, Jr. Ethnic distribution of electrocardiographic predictors of atrial fibrillation and its impact on understanding the ethnic distribution of ischemic stroke in the Atherosclerosis Risk in Communities Study (ARIC). *Stroke.* 2009;40:1204-1211.

8. Alonso A, Agarwal SK, Soliman EZ, et al. Incidence of atrial fibrillation in whites and African-Americans: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J.* 2009;158:111-117.

9. D'Agostino RB, Sr., Grundy S, Sullivan LM, Wilson P, for the CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA.* 2001;286:180-187.