1.a. Full Title: Predictive value of incorporating genetic variants into a risk score for atrial fibrillation: the CHARGE consortium

b. Abbreviated Title (Length 26 characters): Genetic score for AF

2. Writing Group: Steven A. Lubitz, Alvaro Alonso, Dan E. Arking, David Couper, Eric Boerwinkle, and investigators from other CHARGE cohorts (FHS, CHS, AGES, Rotterdam).

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

First author: Steven A. Lubitz, MD
Address: Massachusetts General Hospital
Cardiac Arrhythmia Service
GRB 109, 55 Fruit St., Boston, MA 02114
Phone: 617-701-3321 Fax: 617-726-5806
E-mail: slubitz@partners.org

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: Alvaro Alonso, MD, PhD
Address: University of Minnesota
1300 S 2nd St, Suite 300
Minneapolis, MN 55454
Phone: 612 626 8597 Fax: 612 624 0315
E-mail: alonso@umn.edu

3. Timeline:
Upon approval, we expect to have preliminary results in 2-3 months, and a first draft of the manuscript soon after.

4. Rationale:
Accurate identification of individuals at risk for atrial fibrillation (AF) may enhance efforts to prevent morbidity attributable to this arrhythmia. Prediction of AF is feasible though imperfect.\(^1\text{-}^3\) AF is heritable,\(^4\text{-}^6\) and several common genetic variants have been associated with the arrhythmia.\(^7\text{-}^9\) Furthermore, familial AF has been associated with AF
even after adjustment for known clinical risk factors for AF,\textsuperscript{10,11} suggesting that much of the heritability of AF remains unexplained. We therefore seek to determine whether a risk prediction score for AF comprised of genetic variables enhances the prediction of AF beyond that including clinical variables in individuals of European descent in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) AF consortium.\textsuperscript{12}

5. **Main Hypothesis/Study Questions:**
We hypothesize that incorporating genetic variants associated with AF into an established risk model will enhance the prediction of incident AF.

*\textit{Aim}:* Determine the incremental predictive contribution of genetic variants to CHARGE AF clinical and biomarker prediction model.

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).**

**Study participants**
Variables required for the proposed analysis include: age, sex, race/ethnicity, study center, body mass index, smoking, alcohol intake, height, systolic blood pressure, diastolic blood pressure, hypertension treatment, total cholesterol, HDL cholesterol, triglycerides, diabetes, ECG-based left ventricular hypertrophy, PR interval, heart rate, history of myocardial infarction, history of heart failure, C-reactive protein levels, natriuretic peptides levels, select genetic variant genotypes from GWAS of atrial fibrillation (CHARGE consortium), prevalence of atrial fibrillation at baseline, incidence of atrial fibrillation during follow-up.

All participants of European descent with available information in the predictive variables will be included in the analysis. Only those who did not provide consent to use of genetic information will be excluded from relevant analyses.

The proposed ancillary study will use available data from the five cohorts included in the CHARGE consortium. Details of the consortium and its cohorts have been published elsewhere.\textsuperscript{12}

**AF ascertainment**
All CHARGE cohorts have determined history of AF at baseline and identified incident AF cases during follow-up. AF ascertainment is different from cohort to cohort. In ARIC, prevalent AF was identified by 12-lead ECG at baseline. Incident AF was identified by the first occurrence of AF on study electrocardiograms, the first occurrence of a hospital discharge ICD-9 code for AF, or a death certificate listing AF as a cause of death. Cases of AF occurring in the same hospitalization as cardiac procedures, including bypass surgery, will not be considered as AF events.

**Clinical variables assessment**
The variables included in the CHARGE AF score are: age, body mass index, height, systolic blood pressure, diastolic blood pressure, hypertension treatment, smoking, diabetes, heart failure, and myocardial infarction.
At baseline, ARIC participants answered standard questionnaires assessing some of the previously mentioned risk factors, including smoking, alcohol intake, diabetes, and history of diverse forms of cardiovascular disease. A physical exam obtained information on height, weight, medication use, seated blood pressure measurement and 12-lead ECG. Blood tests included fasting glucose, cholesterol, HDL cholesterol, and serum creatinine.

Measurement of biomarkers
C-reactive protein (CRP) and NT-proBNP are biomarkers that, when added to clinical risk factors, enhance the risk prediction of AF. The present proposal will examine whether genetic variants further enhance risk prediction. Both CRP and NTproBNP have been measured in ARIC participants attending visit 4.

Gene variants
In ARIC, genotyping of the white participants has been done using a high-density Affymetrix 6.0 chip. A cohort-specific GWAS was performed in CHARGE using an additive genetic model adjusting for age, sex, and site. Cohort-specific results were meta-analyzed and will be submitted for publication shortly. For the present manuscript proposal we will create a genetic score based on the top uncorrelated SNPs identified in the CHARGE AF meta-analysis with p<1x10⁻³ and determine if they improve risk prediction in addition to standard clinical factors and biomarkers.

Samples:
- European ancestry
- Derivation (ARIC, CHS, FHS), Validation (Rotterdam Study (RS), AGES)
- Samples will be limited to those used in the derivation of a CHARGE clinical/biomarker AF risk prediction model for comparability

Outcome:
- 5-year risk of incident AF

Analysis:
For the three cohorts included in the derivation study, individual-level data will be pooled (part of ancillary study 2008.12 – Epidemiologic study of risk factors and biomarkers of atrial fibrillation), and analyses will be conducted at Boston University.

Risk score derivation
- **Weighted linear model comprised of top SNPs from AFGen**
  - The top 10 SNP genotypes tagging distinct AF susceptibility loci from the CHARGE AFGen meta-analysis (pending publication) will be selected
  - We will use proportional hazards regression to examine the associations between a linear combination of weighted genotype dosages and incident AF using a log-additive genetic model. Weighting will be based on the log relative risk for the respective SNP from the AFGen meta-analysis.
  - Models will be adjusted for covariates from the CHARGE AF clinical/biomarker risk score.

- **Liberal alpha model for genomewide profiling (preliminary plan)**

¹⁴⁻¹⁶
The top uncorrelated SNPs from the AFGen meta-analysis associated with AF at $p<1 \times 10^{-3}$

Uncorrelated SNPs will be selected using a clumping algorithm in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/clump.shtml)

Uncorrelated SNPs will be the most strongly associated SNPs at any given 250kb locus, that are not in linkage disequilibrium ($r^2<0.1$) with another more strongly associated SNP at the same 250kb locus. We have estimated that there are 1141 uncorrelated SNPs meeting these criteria in the AFGen analysis.

We will use proportional hazards regression to examine the associations between a linear combination of weighted genotype dosages and incident AF using a log-additive genetic model. Weighting will be based on the log relative risk for the respective SNP from the AFGen meta-analysis.

Models will be adjusted for covariates from the CHARGE AF clinical/biomarker risk score.

Models using only SNPs at various different significance thresholds (e.g., $p<1 \times 10^{-3}$, $<1 \times 10^{-4}$, $<1 \times 10^{-5}$,...) will be tested and compared using penalized model fit statistics such as Akaike’s Information Criterion, Bayes Information Criterion, etc.

Performance between the CHARGE AF clinical / biomarker model alone and with genetic variants will be assessed by comparing c-statistic$^{17}$, the net reclassification improvement index, and the integrated discrimination improvement index.$^{18,19}$

Risk score validation Both the top SNP approach and the best fitting model from the liberal alpha approach will be carried forward for validation.

Models will be validated in AGES and the Rotterdam Study separately using proportional hazards regression.

**Modeling considerations:**

- Proportional hazards regression
- Person-time begins at DNA collection or baseline enrollment (for ARIC it will start on visit 4 date, when biomarker data is available).
- Censoring at death or loss to follow-up
- Robust variance estimators will be used to account for potential relatedness of participants (in Framingham).
- Models will be stratified by cohort and site to account for potential population stratification and differing baseline hazards of AF
- Calibration will be assessed using the Hosmer-Lemeshow chi-squared test statistic

**Potential limitations:**

- European ancestry
- No family history for comparability / estimation of heritability explained
- No consideration of epistasis, gene environment interactions
- Important clinical confounders may be missing (e.g. valvular heart disease)
- Selected SNPs may serve as proxy for ‘causal’ SNPs
- Samples from CHS, FHS, and ARIC were used in the meta-analysis that established associations between included SNPs and AF, which will result in overoptimism of risk metrics
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?  _X_ Yes  
   ____ 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”?  _X_ 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.csc.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)?
   No manuscripts on genetic score and AF risk.

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
   _____ A. primarily the result of an ancillary study (list number* __________)
   _X_  B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 2008.12)

   *ancillary studies are listed by number at http://www.csc.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:

18. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to