

ARIC Manuscript Proposal #4384

PC Reviewed: 12/12/23
SC Reviewed: _____

Status: _____
Status: _____

Priority: 2
Priority: _____

1.a. Full Title: An Echocardiogram-Based Risk Score to Predict Atrial Fibrillation: The Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): Echo & AF

2. Writing Group:

Writing group members: Daokun Sun, Faye L. Norby, Riccardo M. Inciardi, Elsayed Z. Soliman, Alvaro Alonso, Scott D. Solomon, Amil M. Shah, Wei Pan, Lin Yee Chen, and others welcome

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

First author: Daokun Sun
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 South 2nd St, Suite 300
Minneapolis, MN 55455

Phone: (612) 625-4401
E-mail: sun00690@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: Lin Yee Chen
Address: Lillehei Heart Institute and Cardiovascular Division
University of Minnesota Medical School
420 Delaware St SE, MMC 508
Minneapolis, MN 55455

Phone: (612) 625-4401 Fax: (612) 626-4411
E-mail: chenx484@umn.edu

3. Timeline: Data analysis to begin immediately, anticipated draft completion Summer/Fall 2024.

4. Rationale:

Atrial fibrillation (AF), which affects 2.5 million people in the United States, is the most common type of cardiac arrhythmia and one of the major risk factors for stroke. Some people with AF may present with shortness of breath, heart palpitations, lightheadedness, and/or extreme fatigue. However, other people may have subclinical AF without overt symptoms. Subclinical AF is not uncommon. It is estimated that at least one-third of people with AF are asymptomatic.¹ In a previous study of 2,616 participants from ARIC who wore a leadless, ambulatory ECG monitor (Zio XT Patch) for up to 2 weeks, we found that the prevalence of subclinical AF was 2.5%.² Additionally, a meta-analysis of 54 studies showed that among patients with cardiac implantable electronic devices, subclinical AF was recorded in approximately 1/4 of patients.³ Overall, AF is associated with up to five fold greater risk of stroke.⁴ However, identification of people with subclinical/occult AF or who are at increased risk of AF is challenging.

Artificial intelligence (AI) has been shown as a promising tool for early identification of people at high risk of AF. By analyzing raw 12-lead electrocardiogram (ECG) sinus rhythm waveform data via AI techniques, previous studies reported relatively high accuracy of prediction of AF.^{1,5} The predicted risk of AF was significantly associated with structural and functional changes of the heart as indicated by measurements on two-dimensional transthoracic echocardiography,⁶ which confirms the importance of cardiac remodeling in the pathogenesis of AF.

Machine learning-enabled algorithms for diagnosis of subclinical AF and prediction of new-onset AF may also be developed by leveraging a priori measured indicators of cardiac structural and functional remodeling on echocardiography. Echocardiogram provides comprehensive assessment of the structural and functional features of cardiac chambers and may identify subclinical abnormalities related to AF. In this study, we propose to use supervised machine learning to derive an echocardiogram-based risk score to diagnose prevalent AF and predict new-onset AF using a priori measured echocardiographic parameters in the ARIC study. We will validate the new risk score in the Multi-Ethnic Study of Atherosclerosis (MESA).

5. Main Hypothesis/Study Questions:

Aim 1: Prevalent AF can be diagnosed using an echocardiogram-based risk score derived by supervised machine learning.

Aim 2: New-onset AF can be predicted using an echocardiogram-based risk score derived by supervised machine learning.

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 Design

Aim 1: Cross-sectional analysis at Visit 5.

Aim 2: Prospective cohort study using Visit 5 as the baseline and follow-up through 2019.

Inclusion/Exclusion

Aim 1: We will include participants who underwent echocardiography at visit 5. Participants who are neither African American nor white, and African Americans from the MN and MD centers will be excluded.

Aim 2: The ARIC visit 5 will be used as the baseline. We will include participants who underwent echocardiography at visit 5. Participants with prevalent AF by visit 5 will be excluded, as well as those who are neither African American nor white, and African Americans from the MN and MD centers.

Variables

Predictors: Derived echocardiographic features at Visit 5, including but not limited to -

- Interventricular septum thickness (cm)
- Posterior wall thickness (cm)
- End-diastolic volume (ml)
- End-systolic volume
- Ejection fraction (%)
- Left ventricular (LV) mass index (g per m²)
- Maximal left atrial anterior-posterior diameter (cm)
- Left atrial (LA) volume index (ml per m²)
- Mitral regurgitation jet area (cm²)
- Peak E wave velocity (cm per sec)
- Peak A wave velocity (cm per sec)
- Lateral early diastolic myocardial velocity (cm per sec)
- Lateral late diastolic myocardial velocity (cm per sec)
- Septal early diastolic myocardial velocity (cm per sec)
- Septal late diastolic myocardial velocity (cm per sec)
- E/e' lateral ratio (cm per sec)
- Left ventricular outflow tract (LVOT) peak velocity (cm per sec)
- LVOT velocity time integral (VTI) (cm)
- LA reservoir strain, %
- LA conduit strain, %
- LA contraction strain, %

Outcome:

Aim 1: Prevalent AF

Aim 2: New onset AF

Both the prevalent and new-onset AF will be ascertained from electrocardiograms (ECGs) conducted during study visits, ICD codes from hospitalization records, or an ambulatory ECG monitoring device during follow-up.

Data analysis

Baseline characteristics of participants will be described using means and proportions.

Primary analysis for aim 1 will use logistic regression to develop a prediction model for prevalent AF.

Primary analysis for aim 2 will use the lasso method for variable selection in Cox proportional hazards model to develop a prediction model for new-onset AF among people without known history of AF.⁷

The following variables available at visit 5 will be considered as predictors in the model -
(1) Baseline demographics, including age, sex, and race
(2) Derived echocardiographic features as listed above.

Other machine learning techniques, such as random forest models, will be conducted as secondary analysis. Model performance of the algorithms derived through the primary analysis and sensitivity analysis will be evaluated using AUC (for aim 1) and Harrell's C-index (also called the concordance index) (for aim 2).⁸

We will compare the performance (C-statistic and calibration) of our risk score with other risk scores (CHARGE-AF score, C2HEST score).

Validation analysis (MESA) - The ARIC model will be applied to the MESA cohort. Calibration and discrimination of the model will be assessed.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ___ Yes ___x___ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ___ Yes ___ No

(The file ICTDER 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 current derived consent file ICTDER05 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.csc.unc.edu/aricproposals/dtSearch.html>

___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)?

#3907: Riccardo M. Inciardi ... Lin Yee Chen. Cardiac Structure and Function and the Risk of Incident Atrial Fibrillation in a Community Cohort of Elderly

#3750: Wendy Wang ... Lin Yee Chen. Association of Left Atrial Function with Neurocognitive Outcomes in the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS)

#3236: Faye L. Norby ... Aaron R. Folsom. Prediction of Atrial Fibrillation in an Elderly Cohort: The Atherosclerosis Risk in Communities (ARIC) study

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

11.b. If yes, is the proposal

A. primarily the result of an ancillary study (list number* 2008.06 (NCS), 2015.29)

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 <https://sites.csc.unc.edu/aric/approved-ancillary-studies>

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.csc.unc.edu/aric/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. Yuan N, Duffy G, Dhruva SS, et al. Deep Learning of Electrocardiograms in Sinus Rhythm From US Veterans to Predict Atrial Fibrillation. *JAMA Cardiol*. Published online October 18, 2023. doi:10.1001/jamacardio.2023.3701
2. Rooney MR, Soliman EZ, Lutsey PL, et al. Prevalence and Characteristics of Subclinical Atrial Fibrillation in a Community-Dwelling Elderly Population: The ARIC Study. *Circ Arrhythm Electrophysiol*. 2019;12(10):e007390.

3. Proietti M, Romiti GF, Vitolo M, et al. Epidemiology of subclinical atrial fibrillation in patients with cardiac implantable electronic devices: A systematic review and meta-regression. *Eur J Intern Med.* 2022;103:84-94.
4. McGrath ER, Kapral MK, Fang J, et al. Association of atrial fibrillation with mortality and disability after ischemic stroke. *Neurology.* 2013;81(9):825-832.
5. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet.* 2019;394(10201):861-867.
6. Verbrugge FH, Reddy YNV, Attia ZI, et al. Detection of Left Atrial Myopathy Using Artificial Intelligence-Enabled Electrocardiography. *Circ Heart Fail.* 2022;15(1):e008176.
7. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997;16(4):385-395.
8. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* 1982;247(18):2543-2546.