ARIC Publication Admin Use Only: ARIC Manuscript Proposal #4354 (revised)

ARIC Publication Admin Use Only: PC Reviewed: 12/12/23

1.a. Full Title: Atrial myopathy subtypes and risk of incident atrial fibrillation: the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters): atrial myopathy subtypes, and atrial fibrillation

2. Writing Group [please provide a middle name if available; EX: Adam Lee Williams]:

Writing group members: Ethan D. Moser, Ruoyu He, Jinwen Fu, Lap Sum Chan, Wendy Wang, Pratik Ramprasad, Daokun Sun, Nidhi Pai, Michael Zhang, Chunxiao Zhang, Scott D. Solomon, Amil M. Shah, Wei Pan, Lin Yee Chen, others welcome.

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

First author [please provide a middle name; EX: Adam Lee Williams]:

Ethan	Davis	Moser
First name	Middle name	Last name

Phone: 763-843-7048

E-mail: Moser250@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: 2231 6th Street SE, Room 4-165 CCRB, 2812 (Campus Delivery Code), Minneapolis, MN 55455

E-mail: chenx484@umn.edu

3. Timeline: Data analysis to begin immediately, and drafted manuscript to be completed 2024.

4. Rationale:

Atrial fibrillation (AF) is a serious public health problem because of its increasing prevalence in the aging population [1, 2] and its association with elevated risks of cardiovascular (CV) morbidity and mortality, especially stroke [3] [4]. **Atrial myopathy** (characterized by abnormalities in left atrial [**LA**] size and function) is a precursor of AF. Importantly, recent evidence indicates that atrial myopathy drives the risk of many sequelae of AF such as stroke, dementia, and heart failure [5-8]. Hence, atrial myopathy is a potential treatment target to prevent cardiovascular outcomes and dementia. A major obstacle to the discovery of prevention strategies is the heterogeneity of atrial myopathy. To facilitate the discovery of prevention strategies, it is crucial to resolve this heterogeneity and identify subtypes of atrial myopathy with differential risk.

Previous classification scheme (EHRAS classification) has been proposed to identify heterogeneity of atrial cardiomyopathies using histopathological characterization [9], but several weaknesses limit its application [10]. However, phenotype mapping, or phenomapping, may serve as a powerful tool in subtyping atrial myopathy. Phenomapping utilizes unsupervised machine learning to classify individuals into groups, or clusters, based on phenotypic data. This method has shown great clinical utility within cardiovascular disease [11-13]. Novel classifications of heart failure with preserved ejection fraction with distinct characteristics and differing risks of adverse outcomes were found using phenomapping [12, 13]. The ability of phenomapping to resolve heterogeneity and refine risk prediction has shown great promise.

We propose to use similar phenomapping methods to resolve the underlying heterogeneity of atrial myopathy. The ARIC study provides a unique opportunity to identify clusters, or subtypes, of atrial myopathy with the availability of dense phenotypic data. Identification of these clusters can be used to understand risk differences in cardiovascular outcomes such as AF.

5. Main Hypothesis/Study Questions:

- Clustering analysis will identify 2+ clusters of atrial myopathy subtypes.
- Atrial myopathy clusters will be differentially associated with risk of incident atrial fibrillation.

6. Design and analysis

a) study design

Prospective cohort study with visit 5 baseline until most recent follow-up.

b) inclusion/exclusion

Participants with atrial myopathy will be included in this study. Atrial myopathy will be sex specific defined as left atrial reservoir strain <28.6% for females and <26.4% for males based on previous ARIC study [14]. Participants with prevalent AF will be excluded from the analysis.

c) outcome and other variables of interest with specific reference to the time of their collection

Outcome: Incident AF (diagnosed from Zio®XT Patch and ARIC ascertainment: hospital discharge codes and linking with Medicare claims data) from visit 5 to most recent follow up.

An extensive list of variables will be used in initial clustering analysis. Variable selection within clustering analysis will reduce variables down to those that define clusters. Due to the large size, we will consider doing a separate clustering analysis including proteomics data with other clustering variables. All variables will be measured at visit 5.

Variables to be included in clustering analysis:

Demographic and clinical: Age, race, sex, height, weight, body mass index, systolic blood pressure (BP), diastolic BP, hypertension, type 2 diabetes, coronary heart disease, peripheral arterial disease, heart failure, chronic kidney disease (CKD).

2D-echocardiographic: LA volume index, LA function indices (reservoir strain, conduit strain, and contraction strain), left ventricular (LV) structure (mass, wall thickness, relative wall thickness, systolic and diastolic diameters and volumes), LV ejection fraction, LV diastolic function (E wave, A wave, E/e'), LV deformation (average longitudinal strain, average circumferential strain), pulmonary artery systolic pressure, pulmonary vascular resistance, arterial elastance (EA), LV end-systolic elastance (EES), EA/EES ratio.

Pulse wave velocity (PWV): carotid-femoral PWV and femoral-ankle PWV
ECG: P-wave duration, P-wave axis, advanced inter-atrial block, and P-wave terminal force in lead V1 [15]

Laboratory measures: (a) Plasma: high sensitivity troponin T, NT-proBNP, galectin-3, high sensitivity CRP, albumin, creatinine, uric acid, hemoglobin, WBC, platelet, glucose, insulin, hemoglobin A1C, TSH, total cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein; (b) Estimated glomerular filtration rate based on 2021 creatinine-cystatin C formula [16], albumin-to-creatinine ratio.

Proteomics: SOMAscan assay data from visit 5. Will use 44 proteins significantly associated with LA reservoir strain within ARIC that were previously found by Michael Zhang.

d) summary of data analysis

Cluster Analysis

Sample for clustering analysis will be divided into discovery and validation set to assess performance of clustering. The ARIC study will serve as discovery cohort and the Multi-Ethnic Study of Atherosclerosis will serve as validation cohort. Variables with high proportion of missing will be removed from analysis (>15% missing). Missing values for other variables will be imputed using weighted K-nearest neighbor imputation [17]. Random forest prediction will be considered for nonparametric imputation using the R package missForest [18]. Next, dimension reduction will be conducted with iCluster+ [19] latent variable analysis to find low dimension latent variables that represent original higher dimension data. Variable selection will be done using a penalized log-likelihood function to improve clustering performance and interpretability.

After obtaining latent variables model-based clustering will be conducted to obtain final clusters. Number of clusters will be determined by BIC with cross validation being explored [20]. Clustered data will be plotted to visualize underlying structure. Final clusters of atrial myopathy will be used as primary exposure for incident atrial fibrillation analysis.

Incident AF Analysis

Participant baseline characteristics will be reported as mean (standard deviation) or proportions stratified by exposures and cohort (e.g., discovery and validation). For the discovery cohort, we will use multivariable adjusted Cox proportional hazards model to calculate hazard ratios (HR) with 95% confidence intervals and conduct the logrank test of clusters (derived from cluster analysis) for AF. Clustering group with the lowest comorbidities will be selected as reference group for model. We will fit models adjusting for CHARGE-AF [21] and relevant medications.

For sensitivity analysis, we will adjust for medications listed above. We will test the proportional hazards assumption by conducting goodness-of-fit tests, calculating Schoenfeld residuals, comparing log(-log) survival curves, and testing time*cluster interactions. Additionally, we will evaluate the impact of non-CV death as a potential competing risk using subdistribution hazard models.

To determine improvement in predictive ability and risk classification for AF, we will calculate C-statistic, net reclassification improvement, and integrated discrimination improvement with CHARGE-AF as the benchmark for comparison.

Participants in the validation cohort will be assigned to clusters with model-based clustering as derived in discovery cohort cluster analysis by using the predict function within *mclust* in R[22, 23]. We will evaluate the association of clusters in the validation cohort with AF using the same approach as the discovery cohort.

All analyses will have race and sex interactions tested with sex- and race- stratified analyses conducted when appropriate.

e) any anticipated methodologic limitations or challenges if present

f) Will the author need Limited data to complete the proposed manuscript? ☒ Yes, Limited data is needed. ☐ No, De-identified data will be sufficient.

*Please note, Limited dataset access is strict and rarely provided. Limited data includes identifiable information such as dates (birthdays, visit dates, etc.). CMS, Genomic, Geocoded, and Proteomics/Somalogic data all fall under the limited data category. De-identified data does not include dates. All dates are date adjusted to "Days since Visit 1".

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ☐ Yes ☒ 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 is distributed to ARIC PIs annually, 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 ☐ 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: https://aric.cscc.unc.edu/aric9/proposalsearch [ARIC Website→Publications→Proposal Search]

☒ 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 Proposal #3956 – Charles B. Eaton, HF phenomapping
- MS Proposal #3123 – Scott D. Solomon, Phenotyping in heart failure

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* __)**

☒ **B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* R01HL126637 data used for monitor-detected atrial fibrillation as an outcome. K24HL155813: Clustering analysis. LA function data: R01HL141288)**

*ancillary studies are listed by number
https://aric.cscc.unc.edu/aric9/researchers/ancillary_studies/approved_ancillary_studies [ARIC Website→Ancillary Studies→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 https://aric.cscc.unc.edu/aric9/publications/policies_forms_and_guidelines [ARIC WebsiteàPublicationsàPublication Policies, Forms, and Guidelines].
http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References: _____ **)**

***To view publications materials, click "Log in" at the top right of the page. Click "Forgot Password" if you are experiencing issues with logging in.**

1. Go, A.S., et al., *Prevalence of Diagnosed Atrial Fibrillation in AdultsNational Implications for Rhythm Management and Stroke Prevention: the AnTicoagulation and Risk Factors In Atrial Fibrillation (ATRIA) Study*. JAMA, 2001. **285**(18): p. 2370-2375.
2. Miyasaka, Y., 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**(2): p. 119-25.

3. Odutayo, A., et al., *Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis.* BMJ, 2016. **354**: p. i4482.
4. Philip, A.W., et al., *Epidemiologic assessment of chronic atrial fibrillation and risk of stroke.* Neurology, 1978. **28**(10): p. 973.
5. Kamel, H., et al., *Atrial cardiopathy: a broadened concept of left atrial thromboembolism beyond atrial fibrillation.* Future Cardiol, 2015. **11**(3): p. 323-31.
6. Reyes, J.L., et al., *Association of abnormal p-wave parameters with brain MRI morphology: The atherosclerosis risk in communities neurocognitive study (ARIC-NCS)*.* Pacing and Clinical Electrophysiology, 2023. **46**(8): p. 951-959.
7. Wang, W., et al., *Association of Echocardiographic Measures of Left Atrial Function and Size With Incident Dementia.* Jama, 2022. **327**(12): p. 1138-1148.
8. Inciardi, R.M., et al., *Association of Left Atrial Structure and Function With Heart Failure in Older Adults.* J Am Coll Cardiol, 2022. **79**(16): p. 1549-1561.
9. Goette, A., et al., *EHRA/HRS/APHRS/SOLAECE expert consensus on Atrial cardiomyopathies: Definition, characterisation, and clinical implication.* J Arrhythm, 2016. **32**(4): p. 247-78.
10. Guichard, J.-B. and S. Nattel, *Atrial Cardiomyopathy: A Useful Notion in Cardiac Disease Management or a Passing Fad?* Journal of the American College of Cardiology, 2017. **70**(6): p. 756-765.
11. Krittanawong, C., et al., *Machine learning prediction in cardiovascular diseases: a meta-analysis.* Scientific Reports, 2020. **10**(1): p. 16057.
12. Shah, S.J., et al., *Phenomapping for novel classification of heart failure with preserved ejection fraction.* Circulation, 2015. **131**(3): p. 269-79.
13. Segar, M.W., et al., *Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis.* Eur J Heart Fail, 2020. **22**(1): p. 148-158.
14. Inciardi, R.M., et al., *Association of Left Atrial Structure and Function With Heart Failure in Older Adults.* Journal of the American College of Cardiology, 2022. **79**(16): p. 1549-1561.
15. Chen, L.Y. and E.Z. Soliman, *P Wave Indices-Advancing Our Understanding of Atrial Fibrillation-Related Cardiovascular Outcomes.* Front Cardiovasc Med, 2019. **6**: p. 53.
16. Inker, L.A., et al., *New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race.* New England Journal of Medicine, 2021. **385**(19): p. 1737-1749.
17. Liao, S.G., et al., *Missing value imputation in high-dimensional phenomic data: imputable or not, and how?* BMC Bioinformatics, 2014. **15**(1): p. 346.
18. Stekhoven, D.J. and P. Bühlmann, *MissForest--non-parametric missing value imputation for mixed-type data.* Bioinformatics, 2012. **28**(1): p. 112-8.
19. Mo, Q., et al., *Pattern discovery and cancer gene identification in integrated cancer genomic data.* Proc Natl Acad Sci U S A, 2013. **110**(11): p. 4245-50.
20. Schwarz, G., *Estimating the Dimension of a Model.* The Annals of Statistics, 1978. **6**(2): p. 461-464, 4.
21. Alonso, A., et al., *Simple Risk Model Predicts Incidence of Atrial Fibrillation in a Racially and Geographically Diverse Population: the CHARGE‐AF Consortium.* Journal of the American Heart Association, 2013. **2**(2): p. e000102.
22. Scrucca, L., et al., *mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.* R j, 2016. **8**(1): p. 289-317.
23. Fraley, C. and A.E. Raftery, *Model-Based Clustering, Discriminant Analysis, and Density Estimation.* Journal of the American Statistical Association, 2002. **97**(458): p. 611-631.