ARIC Manuscript Proposal Form

ARIC Publication Admin Use Only

Publication Committee Review Date:
_12/12/23__
ARIC Manuscript Proposal Number: **#4377**

1.a. Full Title:Leveraging ECG Phenotypes to Improve Prediction of Dementia

b. Abbreviated Title (Length 26 characters): **ECG Prediction of Dementia**

2. Writing Group [please provide a middle initial if available; EX: Adam L Williams]:
Writing group members: Deling Chen, Yuchen Yao, Ethan D. Moser, Wendy Wang, Elsayed Soliman, Thomas Mosley, Wei Pan, and others welcome

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

First author [please provide a middle initial; EX: Adam L Williams]:

Deling		**Chen**
First name	Middle initial	Last name

Address: Division of Biostatistics and Health Data Science, University of Minnesota School of Public Health
2221 University Ave SE, Suite 200
Minneapolis, MN 55414]

Phone: (612)-626-2705
E-mail: chen8682@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 (The ARIC author should be involved enough in ARIC to be able to point the lead author to appropriate ancillary study PIs and to be able to search ARIC manuscript proposals if the lead author doesn't have the access needed to do such a search).

Name: **Wei Pan, PhD**
Address: Division of Biostatistics and Health Data Science, University of Minnesota School of Public Health
2221 University Ave SE, Suite 200

Minneapolis, MN 55414]

Phone: (612)-626-2705
E-mail: panxx014@umn.edu

3. Timeline: Statistical analysis to be completed in 3 months. Manuscript to be submitted in 6 months

4. Rationale: Alzheimer's Disease (AD) and AD-Related Dementias (ADRD) are anticipated to afflict 115 million people globally by 2050. Vascular contributions to cognitive impairment and dementia (VCID), a primary type of ADRD, is a major research focus for the NIH. There is a need to discover novel non-invasive, non-CNS organ-related (e.g., heart) markers to predict dementia. Established cardiovascular risk factors for incident dementia include hypertension, diabetes, physical inactivity, and obesity[1-4]. Cardiovascular disease such as coronary heart disease, atrial fibrillation, and heart failure are also known risk factors for dementia[5-7].

The 12-lead ECG is a ubiquitous, non-invasive, and inexpensive clinical tool. Alterations in the waves, durations, axes, and intervals on the ECG fundamentally reflect and aggregate the effects of years, if not decades, of exposure to cardiovascular risk factors and disease. Therefore, it is not surprising that many studies have shown ECG parameters to be linked to higher incidence of dementia.[8] One meta-analysis reported that ECG parameters such as abnormal P wave axis, prolonged P wave duration, and left ventricular hypertrophy are associated with greater cognitive decline.[8] Prior studies, however, have only considered ECG parameters individually, and no study has considered the predictive power of considering ECG parameters in combination.

Furthermore, in recent years, many studies have demonstrated the power of applying artificial intelligence (AI) and deep learning (DL) approaches to the 12-lead ECG to improve prediction of cardiovascular outcomes.[9,10] AI and DL methods have the power to discern and aggregate subtle variations within ECGs that are imperceptible to the naked eye that reflect underlying cardiac structure and function. As of now, no study has leveraged AI-ECG to improve prediction of dementia.

Hence, in this proposed study, we aim to use the 12-lead ECGs in the ARIC study to construct an ECG risk score and develop an AI-ECG model to enhance prediction of dementia.

5. Main Hypothesis/Study Aims:

1) Construct and validate an ECG risk score to predict dementia
2) Exploratory aim: Train and validate an AI-ECG model to predict dementia

6. Design and analysis - please address the following aspects:

a) inclusion/exclusion
b) study design
c) outcome and other variables of interest with specific reference to the time of their collection
d) summary of data analysis
e) Any anticipated methodologic limitations or challenges if present

Aim 1

Inclusion
We conduct this analysis using 2 different baselines: Visit 4 and Visit 5
All participants with interpretable ECGs at ARIC Visit 4
All participants with interpretable ECGs at ARIC Visit 5

Exclusion
Participants with prevalent dementia, missing covariates, those whose race is other than Black or White and non-White individuals in the Minneapolis and Washington County centers due to small numbers.

Study design
Prospective cohort with baseline at Visit 4 and follow up through 2019.
Prospective cohort with baseline at Visit 5 and follow up through 2019.

Exposures
Candidate ECG parameters are selected from Imahori *et al*.[8] and will be obtained from Visit 4 and Visit 5.

Resting heart rate (continuous)
Abnormal P-wave terminal force in lead V1 (aPTFV1) (binary)
Prolonged P wave duration (PPWD) (binary)
Abnormal P wave axis (aPWA) (binary)
Advanced interatrial block (aIAB) (binary)
Left ventricular hypertrophy (LVH): sex specific Cornell voltage criteria (SV3 + RaVL > 28mm for men and 22mm for women) (binary)
Corrected QT interval (QTc) (continuous)[8]

Outcome
Dementia will be ascertained in 3 ways: (1) Adjudicated dementia cases identified from in-person evaluations at ARIC Neurocognitive Study (ARIC-NCS), (2) Among participants who were alive, but did not attend ARIC-NCS visits, the modified Telephone Interview of Cognitive Status (TICSm) or 6-item screener (SIS) were used to determine cognitive status or an informant interview was conducted; and (3) *International Classification of Diseases* (*ICD*) hospitalization discharge codes or death certificate codes were used to identify additional dementia cases

Other variables (to be used for other dementia risk scores)
UKDRS variables: age (years), sex (male/female), BMI (kg/m^2), area deprivation index (quintiles), smoker status (ever, current, never), drinker status (grams per week), current antidepressant medication use, aspirin use, history of stroke, TIA, atrial fibrillation, or diabetes.[11]

CAIDE variables: age (lowest tertile: 0 points, middle tertile: 3 points and highest tertile: 4 points), sex (men: 1 point), educational years (highest tertile: 0 points, middle tertile: 2 points, lowest tertile: 3 points), hypertension (SBP> 140 mmHg: 2 points), body mass index (> 30

kg/m^2: 2 points), cholesterol (> 6.5 mmol/L: 2 points), and physical activity (inactivity: 1 point [active: engaged in physical activity at least twice a week, lasting at least 20–30 min each time]).[12] For Visit 4 baseline, we will obtain physical activity from Visit 3 (physical activity data are not available at Visit 4).

Statistical analysis

At each baseline, we will split the sample into a derivation or training sample (80%) and validation sample (20%). The first stage variable selection will use only the derivation sample. We will construct the most parsimonious multivariable Cox regression model for prediction of dementia using a backward selection method (keeping variables with $p<0.10$ in the model). We will also consider a least absolute shrinkage and selection operator (LASSO) Cox regression to identify a parsimonious model with dementia as the outcome.[11] As an alternative, we will also conduct the analysis using random survival forest model (RSF model). The RSF model, an ensemble tree-based method for the analysis of right-censored data, will be used to predict and efficiently select variables.[13,14] We will include the top ranked variables in the prediction model. A locally weighted scatter smoothing curve and bar plot in a nonparametric regression will be used to assess nonlinear associations with survival probability.[15]

Duration of follow-up will be calculated as time between baseline and either date of dementia, death, or censoring date. We will construct models for different prediction time horizons (e.g., 15 years, 20 years for Visit 4 baseline; 1 year, 3 years, 5 years for Visit 5 baseline).

The performance of the ECG risk score will be compared with the UK dementia risk score (UKDRS) and Cardiovascular Risk Factors, Ageing, and Dementia score (CAIDE). We will also compare our model with a model consisting of chronological age only, to examine the added predictive value of the ECG risk score. The model discrimination will be evaluated using the area under the curve (AUC). DeLong's test will be used for comparison of the AUCs of the ECG risk score with the other risk models. We will use risk calibration to assess the agreement between the observed proportion of dementia cases and predicted probabilities of developing dementia as calculated from the risk score.

Aim 2 (Exploratory Aim)

Inclusion

All participants with interpretable ECGs at ARIC Visit 4
All participants with interpretable ECGs at ARIC Visit 5

Exclusion

Participants with prevalent dementia, missing covariates, those whose race is other than Black or White and non-White individuals in the Minneapolis and Washington County centers due to small numbers.

Study design

Prospective cohort with baseline at Visit 4 and follow up through 2019.
Prospective cohort with baseline at Visit 5 and follow up through 2019.

Exposures
12-lead ECG digital images

Outcome

Dementia will be ascertained in 3 ways: (1) Adjudicated dementia cases identified from in-person evaluations at ARIC Neurocognitive Study (ARIC-NCS), (2) Among participants who were alive, but did not attend ARIC-NCS visits, the modified Telephone Interview of Cognitive Status (TICSm) or 6-item screener (SIS) were used to determine cognitive status or an informant interview was conducted; and (3) *International Classification of Diseases* (*ICD*) hospitalization discharge codes or death certificate codes were used to identify additional dementia cases. Dementia outcome will be treated as a binary variable rather than a survival outcome.

Overview of the AI Model
At each baseline, we will split the sample into a training (80%) and internal validation sample (20%). We will implement a convolutional neural network (CNN) using Python. The 12-lead ECG is recorded using eight physical leads and four augmented leads (i.e., 12 leads by 10-second duration sampled at 250 Hz for Visit 4 ECGs and 500 Hz for Visit 5 ECGs). The long axis represents the temporal axis and most of the convolutions will be used on it to allow the model to extract morphological and temporal features, while the short axis represents the lead or spatial axis and will be only used on layer to fuse the data from all the leads. We will implement a residual neural network using Keras and Python. The network will consist of 4 residual blocks, following the architecture in Ribeiro *et al*.[16]

The performance of the ECG risk score will be compared with the UK dementia risk score (UKDRS) and Cardiovascular Risk Factors, Ageing, and Dementia score (CAIDE). Measures of diagnostic performance will include the ROC AUC, accuracy (i.e., a weighted average of sensitivity and specificity indicating the percentage of patients whose labels were predicted correctly), sensitivity, specificity, and the F1 score (i.e., the harmonic mean of the sensitivity and positive predictive value).

f) Will the author need Limited data to complete the proposed manuscript? ☐ Yes, Limited data is needed (Provide a brief (2-3 sentences) justification for requesting PHI data) . ☒ 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, Proteomics/Somalogic, and other -omic 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? (Non-ARIC analysis means that the authors are not regarded as ARIC

investigators and the "ARIC author" is essentially just a facilitator rather than an integral part of the writing group.) ☐ **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 or the "sponsoring" ARIC 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 website 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)?
#2545, Gutierrez
#3678, Soliman EZ

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or does it use current [or ongoing] ancillary study data (this includes ACHIEVE)? ☐ **Yes** ☒ **No → Skip to question 12**

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

11.c. If yes to 11.a., list number* __

*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 Website Publications Publication Policies, Forms, and Guidelines]. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References: 1. Sierra C. Hypertension and the Risk of Dementia. *Front Cardiovasc Med*. 2020;7:5. doi: 10.3389/fcvm.2020.00005

2. Cholerton B, Baker LD, Montine TJ, Craft S. Type 2 Diabetes, Cognition, and Dementia in Older Adults: Toward a Precision Health Approach. *Diabetes Spectr*. 2016;29:210-219. doi: 10.2337/ds16-0041

3. Huang SY, Li YZ, Zhang YR, Huang YY, Wu BS, Zhang W, Deng YT, Chen SD, He XY, Chen SF, et al. Sleep, physical activity, sedentary behavior, and risk of incident dementia: a prospective cohort study of 431,924 UK Biobank participants. *Mol Psychiatry*. 2022;27:4343-4354. doi: 10.1038/s41380-022-01655-y

4. Ma Y, Ajnakina O, Steptoe A, Cadar D. Higher risk of dementia in English older individuals who are overweight or obese. *Int J Epidemiol*. 2020;49:1353-1365. doi: 10.1093/ije/dyaa099

5. Deckers K, Schievink SHJ, Rodriquez MMF, van Oostenbrugge RJ, van Boxtel MPJ, Verhey FRJ, Köhler S. Coronary heart disease and risk for cognitive impairment or dementia: Systematic review and meta-analysis. *PLoS One*. 2017;12:e0184244. doi: 10.1371/journal.pone.0184244

6. Bunch TJ. Atrial Fibrillation and Dementia. *Circulation*. 2020;142:618-620. doi: 10.1161/circulationaha.120.045866

7. Wolters FJ, Segufa RA, Darweesh SKL, Bos D, Ikram MA, Sabayan B, Hofman A, Sedaghat S. Coronary heart disease, heart failure, and the risk of dementia: A systematic review and meta-analysis. *Alzheimers Dement*. 2018;14:1493-1504. doi: 10.1016/j.jalz.2018.01.007

8. Imahori Y, Vetrano DL, Ljungman P, Qiu C. Electrocardiographic Predictors of Cognitive Decline and Dementia: A Systematic Review. *J Alzheimers Dis*. 2021;84:1303-1322. doi: 10.3233/jad-210606

9. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, Harrington LX, Wang X, Al-Alusi MA, Sarma G, et al. ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. *Circulation*. 2022;145:122-133. doi: 10.1161/circulationaha.121.057480

10. Yuan N, Duffy G, Dhruva SS, Oesterle A, Pellegrini CN, Theurer J, Vali M, Heidenreich PA, Keyhani S, Ouyang D. Deep Learning of Electrocardiograms in Sinus Rhythm From US Veterans to Predict Atrial Fibrillation. *JAMA Cardiol*. 2023. doi: 10.1001/jamacardio.2023.3701

11. Anatürk M, Patel R, Ebmeier KP, Georgiopoulos G, Newby D, Topiwala A, de Lange AG, Cole JH, Jansen MG, Singh-Manoux A, et al. Development and validation of a dementia risk score in the UK Biobank and Whitehall II cohorts. *BMJ Ment Health*. 2023;26. doi: 10.1136/bmjment-2023-300719

12. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol*. 2006;5:735-741. doi: 10.1016/s1474-4422(06)70537-3

13. Wright M, Dankowski T, Ziegler A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Statistics in Medicine*. 2017;36. doi: 10.1002/sim.7212

14. Nasejje JB, Mwambi H, Dheda K, Lesosky M. A comparison of the conditional inference survival forest model to random survival forests based on a simulation study as well as on two applications with time-to-event data. *BMC Medical Research Methodology*. 2017;17:115. doi: 10.1186/s12874-017-0383-8

15. Zhuang XD, Tian T, Liao LZ, Dong YH, Zhou HJ, Zhang SZ, Chen WY, Du ZM, Wang XQ, Liao XX. Deep Phenotyping and Prediction of Long-term Cardiovascular Disease: Optimized by Machine Learning. *Can J Cardiol*. 2022;38:774-782. doi: 10.1016/j.cjca.2022.02.008

16. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira Jr W, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*. 2020;11:1760. doi: 10.1038/s41467-020-15432-4

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