Published in final edited form as:

Am J Cardiol. 2020 January 01; 125(1): 55–62. doi:10.1016/j.amjcard.2019.09.032.

Evaluation of Risk Prediction Models of Atrial Fibrillation (From the Multi-Ethnic Study of Atherosclerosis [MESA])

Joshua D. Bundy, PhD, MPH^{a,b}, Susan R. Heckbert, MD, PhD^c, Lin Y. Chen, MD, MS^d, Donald M. Lloyd-Jones, MD, ScM^b, Philip Greenland, MD^b

^aDepartment of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana

^bDepartment of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

^cDepartment of Epidemiology, University of Washington, Seattle, Washington

^dCardiovascular Division, Department of Medicine, University of Minnesota Medical School, Minneapolis, Minnesota

Abstract

Atrial fibrillation (AF) is prevalent and strongly associated with higher cardiovascular disease (CVD) risk. Machine learning is increasingly used to identify novel predictors of CVD risk, but prediction improvements beyond established risk scores are uncertain. We evaluated improvements in predicting 5-year AF risk when adding novel candidate variables identified by machine learning to the CHARGE-AF Enriched score, which includes age, race/ethnicity, height, weight, systolic and diastolic blood pressure, current smoking, use of antihypertensive medication, diabetes, and NT-proBNP. We included 3534 participants (mean age, 61.3 years; 52.0% female) with complete data from the prospective Multi-Ethnic Study of Atherosclerosis (MESA). Incident AF was defined based on study electrocardiograms and hospital discharge diagnosis ICD-9 codes, supplemented by Medicare claims. Prediction performance was evaluated using Cox regression and a parsimonious model was selected using LASSO. Within 5 years of baseline, 124 participants had incident AF. Compared with the CHARGE-AF Enriched model (c-statistic, 0.804), variables identified by machine learning, including biomarkers, cardiac magnetic resonance imaging variables, electrocardiogram variables, and subclinical CVD variables, did not significantly improve prediction. A 23-item score derived by machine learning achieved a c-statistic of 0.806, while a parsimonious model including the clinical risk factors age, weight, current smoking, NT-proBNP, coronary artery calcium score, and cardiac troponin-T achieved a c-

Corresponding author: Joshua D. Bundy, PhD, MPH, Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, Suite 2000, New Orleans, LA 70112, USA, Phone: (504) 988-3970; jrbundy1@tulane.edu.

Author Agreement/Declaration

We certify that this manuscript represents entirely original work that has not been presented or published before except in abstract form. All authors have seen and approved the final version of the manuscript being submitted.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

statistic of 0.802. This analysis confirms that the CHARGE-AF Enriched model and a parsimonious 6-item model performed similarly to a more extensive model derived by machine learning. In conclusion, these simple models remain the gold standard for risk prediction of AF, although addition of the coronary artery calcium score should be considered.

Keywords

cardiovascular disease; arrhythmias; risk factors; risk prediction

Atrial fibrillation (AF) is the most common clinically-significant cardiac arrhythmia and is strongly associated with higher risk of stroke and other cardiovascular disease (CVD) outcomes.^{1–5} Risk prediction scores for incident AF could be useful clinically to identify high-risk patients for additional surveillance or for participation in prevention trials.^{6–10} The validated Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) AF consortium model includes several important demographic and clinical variables.^{8,9,11,12} Machine learning techniques are becoming increasingly common and are proposed to improve risk prediction.¹³ In a machine learning study within the Multi-Ethnic Study of Atherosclerosis (MESA), researchers included >700 baseline variables in an attempt to improve risk prediction for several CVD outcomes, including AF. However, their potential prediction benefits beyond previous AF risk scores are unknown. The objectives of the current analysis are two-fold: to evaluate whether novel candidate variables improve prediction of 5-year risk of AF in MESA participants when added to the CHARGE-AF models and develop a parsimonious model in MESA.

Methods

The MESA is a prospective, population-based observational cohort study of 6814 men and women representing 4 racial/ethnic groups (Caucasian, African-American, Hispanic and Chinese-American), aged 45–84 years and free of clinical CVD at enrollment.¹⁵ As part of the baseline examination (2000–2002), study participants were recruited at 6 field centers in the US (Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles, CA; New York, NY; and St. Paul, MN). A total of 3534 participants with complete data on all candidate variables were included in the analysis. Institutional review boards of all field centers approved the study protocol, and all participants gave written informed consent.

Information on assessment of risk factors within MESA has been described previously,¹⁵ and a detailed description is provided in the Online Methods. We included biomarkers and measurements from questionnaires, demographics, anthropometry, medication use, blood biochemistry, magnetic resonance imaging (MRI) of the heart and aorta, coronary computed tomography, carotid ultrasound, and electrocardiography (ECG).

Several risk prediction models for AF are compared in Online Table 1.^{6–10} Candidate variables considered in this analysis included the individual components of the CHARGE-AF models:^{8,9} age, race/ethnicity, height, weight, systolic and diastolic blood pressure (BP), current smoking, use of antihypertensive medication, diabetes, and NT-proBNP. Because MESA enrolled participants without a history of CVD, history of heart failure and history of

myocardial infarction were not evaluated, consistent with previous validation of CHARGE-AF in MESA.¹⁰ We additionally considered variables identified as highly predictive of incident AF in a previous machine learning analysis in MESA, which used the random survival forests method:¹⁴ coronary artery calcium (CAC) score, ankle-brachial index, common carotid intima media thickness, internal carotid intima media thickness, serum creatinine, cardiac troponin-T, R amplitude in lead V₄, STJ amplitude in lead V₅, heart rate, estimate of overall heart rate variability (standard deviation of all normal-to-normal R-R intervals), QRS axis, end-systolic basal lateral wall thickness, and end-systolic midventricular anterior wall thickness. Variables identified as predictive in the machine learning analysis but missing at baseline among >30% of participants (left atrial ejection fraction, interleukin-2, and tumor necrosis factor- α) were not evaluated.

The primary end-point for the prediction models was incident AF. MESA participants or a proxy were contacted by telephone every 9 to 12 months to identify all new hospitalizations. Medical records were obtained, and trained staff abstracted discharge diagnostic and procedure codes from these hospitalizations. Incident AF was defined as presence of International Classification of Diseases, Ninth Revision (ICD-9) diagnosis codes for AF (427.31) or atrial flutter (427.32). We additionally included Medicare inpatient, outpatient, and carrier claims for participants enrolled in fee-for-service Medicare.¹⁶ Participants newly found to have AF by 12-lead ECG at the 2010–2012 study visit were classified as having AF as of the visit date. AF that occurred during a hospitalization with open cardiac surgery was not counted as an event.

We used Cox proportional hazards models to construct prediction models of 5-year risk of incident AF, including baseline traditional and non-traditional candidate variables. Time of follow-up was defined as time from the baseline exam to the first occurrence of incident AF, death, or last available contact through December 2014. Maximum follow-up time was censored at 5 years. In a secondary analysis, we evaluated the prediction performance of the models using all available follow-up time (mean 11.4 years; maximum 14.5 years).

The enriched CHARGE-AF model, which was previously validated in MESA,¹⁰ was selected as the base model to which novel candidate variables were added. We used MESA variable coefficients rather than CHARGE-AF coefficients to facilitate unbiased model comparisons within this cohort. Linearity of continuous variables was assessed visually using plots of Martingale residuals, and deviations were addressed with appropriate transformations (e.g. natural log). The LASSO (“least absolute shrinkage and selection operator”) method was used to identify a parsimonious set of predictors.¹⁷ Interactions of candidate variables with age, sex, and race/ethnicity were also considered in the LASSO analysis.

We evaluated the performance of sequential models using measures of discrimination and calibration.¹⁸ Discrimination refers to the ability of a model to correctly identify those with and without the outcome and was assessed by estimating Harrell’s c-statistic¹⁹ and plotting survival receiver operating characteristic (ROC) curves.²⁰ Discrimination performance of sequential models was compared using the likelihood ratio test.²¹ Calibration refers to the agreement between observed outcomes and predictions provided by a given model and was

assessed by calculating the Greenwood-Nam-D'Agostino (GND) statistic.²² Additionally, calibration plots were created to visualize the observed vs. predicted risk across categories of predicted risk, defined according to recommended guidelines.²² We additionally conducted a post-hoc analysis to evaluate performance of the parsimonious model in an expanded analytic sample including all MESA participants with complete data for the candidate predictors selected in this model (age, weight, current smoking, NT-proBNP, CAC, and troponin-T; n=5502).

Results

Of the 6814 total participants in MESA, we excluded those with pre-baseline AF (n=66) and those missing data for the outcome (n=29) and candidate predictors (MRI variables, n=1958; serum biomarkers, n=819; ECG variables, n=238; subclinical CVD variables, n=119; traditional risk factors, n=51). Thus, a total of 3534 participants (mean age, 61.3 years; 52.0% female) with complete data were included in the analyses and were, on average, healthier compared with the full MESA cohort (Online Table 2). During an average 11.4-year follow-up, 436 participants had an AF event (incidence rate, 10.8 per 1000 person-years), 124 of which occurred during the first 5 years of follow-up (7.3 per 1000 person-years). Table 1 shows the prevalence and mean values of baseline characteristics by incident AF status within 5 years. Those who had an AF event were more likely to be older and taking antihypertensive medication. Additionally, those who had an AF event had, on average, higher systolic BP, CAC, cIMT, serum creatinine, NT-proBNP, troponin-T, and cardiac MRI wall thickness; and lower ankle-brachial index, STJ amplitude in lead V5, and QRS axis. Similar results were observed when comparing those who had an AF event over all follow-up with those who did not, although additional differences were noted, including for sex, race/ethnicity, and diabetes status (Online Table 3).

Table 2 provides discrimination and calibration statistics for several models. Compared with the CHARGE-AF Simple model, the CHARGE-AF Enriched model, which additionally included NT-proBNP, showed statistically-significant discrimination improvement and was well-calibrated. Further addition of biomarkers, MRI measurements, ECG measurements, and subclinical CVD variables did not significantly improve discrimination or calibration, nor did the addition of all potential predictors. A parsimonious model derived using LASSO included only age, weight, current smoking, NT-proBNP, CAC, and troponin-T (the “Novel MESA” score), and performed similarly to the model with all predictors. No interactions with age, sex, nor race/ethnicity were retained in the model. Compared with the CHARGE-AF Simple model, ROC curves for the CHARGE-AF Enriched, All Predictors, and Novel MESA models show improved discrimination, but largely overlapped (Figure 1).

Observed and predicted risks were close across the range of the scores for the CHARGE-AF Simple, CHARGE-AF Enriched, All Predictors, and Novel MESA models (Figure 2) and none showed evidence of poor calibration based on the GND statistics. However, the CHARGE-AF Simple model did not perform as well in intermediate predicted probability groups compared with other groups, resulting in poorer calibration overall compared with the other models. The CHARGE-AF Enriched model showed the best calibration (GND X² statistic, 2.4 [p=0.80]).

Table 3 provides estimated hazard ratios for the variables included in the CHARGE-AF Enriched, All Predictors, and Novel MESA models. Additionally, the baseline survival function and beta coefficients are provided that allow calculation of the predicted risk of AF within 5 years. Across all 3 models, baseline age, weight, current smoking, and NT-proBNP were significantly associated with AF risk. After employing LASSO, these variables were retained in the Novel MESA model along with CAC and troponin-T. While mid-ventricular anterior wall thickness was also a statistically-significant predictor in the All Predictors model, it was not retained in the Novel MESA model after LASSO selection.

When expanding AF risk prediction through all available follow-up (mean, 11.4 years; maximum, 14.5 years; 436 AF events), discrimination and calibration were reduced compared with 5-year risk prediction (Online Table 4). Similar to the 5-year risk analyses, most novel variables did not add significantly to prediction performance. However, the addition of subclinical CVD markers, including CAC, ankle-brachial index, common cIMT, and internal cIMT, significantly improved discrimination compared with CHARGE-AF Enriched. In particular, the addition of CAC alone achieved equivalent discrimination to the model including the other subclinical CVD markers (c-statistic, 0.784; 95% CI 0.765–0.803). Online Table 5 provides estimated hazard ratios and risk score calculation information for the variables included in the models for the expanded follow-up period.

We conducted a post-hoc analysis among the 5502 participants with complete data for the predictors included in the Novel MESA model. A total of 746 incident AF events were identified over a mean follow-up of 11.3 years, 224 of which occurred within the first 5 years (Online Table 6). Results were mostly similar compared with those in the derivation sample of 3534 participants, although the association of current smoking with AF risk was weaker.

Discussion

In this analysis of 3534 participants from a multi-ethnic cohort, we found that the addition of novel blood biomarkers, MRI measurements, ECG measurements, and subclinical CVD variables identified by machine learning to the base CHARGE-AF Enriched model did not significantly increase 5-year AF risk prediction ability. A parsimonious model including only age, weight, current smoking, NT-proBNP, CAC, and troponin-T performed as well as the 23-item risk score including all candidate predictor variables. These findings provide important guidance on the utility of novel measurements in risk prediction and implicate several important variables in predicting the risk of AF. This study also provides useful insights into the value of machine learning techniques for risk prediction in the clinical setting.

The CHARGE-AF consortium developed a simple score for predicting 5-year risk of AF by pooling individual-level data from 18556 participants from 3 US cohorts and validated the score in 7672 participants from 2 European cohorts.⁸ The simple model, which includes several traditional CVD risk factors, discriminated AF events well in the current analysis, but was not as well calibrated compared with the other models. A follow-up analysis in CHARGE-AF reflected the importance of including NT-proBNP in the model, forming the

CHARGE-AF Enriched model,⁹ which was previously validated in MESA.¹⁰ We corroborated evidence that the biomarker-enriched model performed better than the simple model in terms of both discrimination and calibration in the diverse MESA sample. Because the CHARGE-AF Enriched model was developed in a large, pooled cohort and has been validated in several other populations, it remains valuable in predicting the risk of AF in various settings, such as the clinic or for study recruitment. Additionally, it is an appropriate standard to which novel measurements can be compared for improvement in prediction ability, such as in the current analysis.

In 2017, an analysis in MESA used the random survival forests machine learning method to identify the 20 most predictive variables for several cardiovascular diseases, including AF.¹⁴ Ambale-Venkatesh et al. ranked the relative importance of >700 variables in predicting the risk of AF. Most of the top-20 variables identified are not included in risk prediction models of AF.^{6–10} Thus, our goal was to directly compare these findings to established risk scores to evaluate potential improvements in risk prediction. Ambale-Venkatesh et al. compared their results with traditional scores for the prediction of coronary heart disease and heart failure, and found performance improvements defined by higher c-statistics and lower Brier scores. However, their findings were not directly compared with previously-validated risk scores for AF. Our findings indicate that the novel variables identified by machine learning within the MESA cohort did not improve prediction performance beyond the CHARGE-AF Enriched model. Importantly, the machine learning methods successfully identified several key predictors that were retained in our novel LASSO-selected model, including NT-proBNP, CAC, and troponin-T. Future work should continue to evaluate new predictors as data become available.

Our findings offer several implications, both for the burgeoning field of machine learning within the context of clinical CVD research and for AF risk prediction. As noted by Ambale-Venkatesh et al, machine learning techniques may offer a robust variable selection method when facing a large number of potential predictors.¹⁴ However, it is important to evaluate findings from such data-driven techniques against previously-developed and validated risk scores already employed by clinicians, especially when potentially recommending measurements not routinely collected in a clinical setting. Findings from both our analyses complement previous work demonstrating that NT-proBNP is strongly and independently associated with higher risk of AF.^{23–26} Additionally, subclinical CVD markers, particularly CAC, afforded prediction performance improvements beyond CHARGE-AF Enriched when expanding the analysis to all available follow-up time. While CAC is primarily considered a subclinical marker of atherosclerosis,²⁷ it has been associated with AF in previous studies.^{28–30} It is possible that CAC may represent an accumulation of CVD risk factor burden or cardiac structural changes and vascular injury that may directly explain its role in higher risk of AF. In populations free of clinical CVD at baseline, such as MESA participants, CAC may be particularly valuable for AF risk prediction in lieu of history of myocardial infarction and history of heart failure, which are included in the CHARGE-AF scores.

Several potential limitations must be considered. First, a large amount of data was missing for many of the novel risk factor groups, particularly cardiac MRI measurements. Thus, our

complete case analysis may decrease generalizability, since inclusion in the analysis required participants to be able and willing to have biomarker, MRI, ECG, and computed tomography measurements. However, the sample included in this analysis still represents a population of clinical and public health relevance. Furthermore, evaluation of the parsimonious model in an expanded sample of participant revealed similar prediction performance. Second, ascertainment of AF relied on hospitalization and CMS claims data. Thus, undiagnosed AF or AF in those not seeking treatment is not identified, but these same limitations apply to all previous AF prediction studies. Additionally, a diagnosis of AF may arise because of various biologic mechanisms. Further research is warranted to investigate prediction of specific subtypes of AF, which was not possible in the current analysis. Third, relatively few AF events occurred within 5 years (N=124), which may limit statistical power to identify significant improvements in prediction performance. However, results and conclusions were similar when expanding the prediction horizon to all available follow-up (mean, 11.4 years; 436 AF events). Finally, we chose to focus our analysis on 5-year risk of AF, which is a short-term prediction. However, short-term risk prediction may be particularly attractive for screening and primary prevention purposes.

In conclusion, compared with the CHARGE-AF score, novel candidate variables identified by machine learning did not add significantly to AF risk prediction. We also found that a parsimonious score containing only age, weight, current smoking, NT-proBNP, CAC, and troponin-T performed as well as the 23-item score including risk factors identified by machine learning. These findings confirm the utility of existing risk scores for AF prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Sources of Funding:

This research was supported by grant R01HL127659 from the National Heart, Lung, and Blood Institute. The Multi-Ethnic Study of Atherosclerosis (MESA) was supported by contracts HHsN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants U11-TR-000040, UL1-TR-001079, and UL1-TR-001420 from NCATS. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at <http://www.mesa-nhlbi.org>. Dr. Bundy is supported by the National Heart, Lung, and Blood Institute Cardiovascular Epidemiology training grant T32HL069771. Dr. Chen is supported by R01HL126637 and R01HL141288 from the National Heart, Lung, and Blood Institute.

References

1. Wolf PA, Mitchell JB, Baker CS, Kannel WB, D'Agostino RB. Impact of Atrial Fibrillation on Mortality, Stroke, and Medical Costs. *Arch Intern Med* 1998;158:229–234. [PubMed: 9472202]
2. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby J V., Singer DE. Prevalence of Diagnosed Atrial Fibrillation in Adults. *JAMA* 2001;285:2370–2375. [PubMed: 11343485]
3. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Das SR, Delling FN, Djousse L, Elkind MS V, Ferguson JF, Fornage M, Jordan LC, Khan SS, Kissela BM, Knutson KL, Kwan TW, Lackland DT, Lewis TT, Lichtman JH, Longenecker CT, Loop MS, Lutsey PL, Martin SS, Matsushita K, Moran AE, Mussolino ME,

- O'Flaherty M, Pandey A, Perak AM, Rosamond WD, Roth GA, Sampson UKA, Satou GM, Schroeder EB, Shah SH, Spartano NL, Stokes A, Tirschwell DL, Tsao CW, Turakhia MP, VanWagner LB, Wilkins JT, Wong SS, Virani SS. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation* 2019;139:e56–e528. [PubMed: 30700139]
4. Magnani JW, Rienstra M, Lin H, Sinner MF, Lubitz SA, McManus DD, Dupuis J, Ellinor PT, Benjamin EJ. Atrial fibrillation: Current knowledge and future directions in epidemiology and genomics. *Circulation* 2011;124:1982–1993. [PubMed: 22042927]
 5. Soliman EZ, Lopez F, O'Neal WT, Chen LY, Bengtson L, Zhang ZM, Loehr L, Cushman M, Alonso A. Atrial fibrillation and risk of ST-segment-elevation versus non-ST-segment-elevation myocardial infarction: The atherosclerosis risk in communities (ARIC) study. *Circulation* 2015;131:1843–1850. [PubMed: 25918127]
 6. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, Newton-Cheh C, Yamamoto JF, Magnani JW, Tadros TM, Kannel WB, Wang TJ, Ellinor PT, Wolf PA, Vasan RS, Benjamin EJ. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet* 2009;373:739–745. [PubMed: 19249635]
 7. Chamberlain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R, Ambrose M, Alonso A. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). *Am J Cardiol* 2011;107:85–91. [PubMed: 21146692]
 8. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, Sinner MF, Sotoodehnia N, Fontes JD, Janssens ACJW, Kronmal RA, Magnani JW, Witteman JC, Chamberlain AM, Lubitz SA, Schnabel RB, Agarwal SK, McManus DD, Ellinor PT, Larson MG, Burke GL, Launer LJ, Hofman A, Levy D, Gottdiener JS, Kääb S, Couper D, Harris TB, Soliman EZ, Stricker BHC, Gudnason V, Heckbert SR, Benjamin EJ. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc* 2013;2:e000102.
 9. Sinner MF, Stepas KA, Moser CB, Krijthe BP, Aspelund T, Sotoodehnia N, Fontes JD, Janssens ACJW, Kronmal RA, Magnani JW, Witteman JC, Chamberlain AM, Lubitz SA, Schnabel RB, Vasan RS, Wang TJ, Agarwal SK, McManus DD, Franco OH, Yin X, Larson MG, Burke GL, Launer LJ, Hofman A, Levy D, Gottdiener JS, Kääb S, Couper D, Harris TB, Astor BC, Ballantyne CM, Hoogeveen RC, Arai AE, Soliman EZ, Ellinor PT, Stricker BHC, Gudnason V, Heckbert SR, Pencina MJ, Benjamin EJ, Alonso A. B-type natriuretic peptide and C-reactive protein in the prediction of atrial fibrillation risk: The CHARGE-AF Consortium of community-based cohort studies. *Europace* 2014;16:1426–1433. [PubMed: 25037055]
 10. Alonso A, Roetker NS, Soliman EZ, Chen LY, Greenland P, Heckbert SR. Prediction of atrial fibrillation in a racially diverse cohort: The multi-ethnic study of atherosclerosis (MESA). *J Am Heart Assoc* 2016;5:e003077.
 11. Pfister R, Brägelmann J, Michels G, Wareham NJ, Luben R, Khaw K-T. Performance of the CHARGE-AF risk model for incident atrial fibrillation in the EPIC Norfolk cohort. *Eur J Prev Cardiol* 2015;22:932–939. [PubMed: 25059930]
 12. Shulman E, Kargoli F, Agaard P, Hoch E, Biase L Di, Fisher J, Gross J, Kim S, Krumerman A, Ferrick KJ. Validation of the Framingham Heart Study and CHARGE-AF Risk Scores for Atrial Fibrillation in Hispanics, African-Americans, and Non-Hispanic Whites. *Am J Cardiol* 2016;117:76–83. [PubMed: 26589820]
 13. Deo RC. Machine learning in medicine. *Circulation* 2015 132:1920–1930. [PubMed: 26572668]
 14. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, Gomes AS, Folsom AR, Shea S, Guallar E, Bluemke DA, Lima JAC. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res* 2017;121:1092–1101. [PubMed: 28794054]
 15. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacobs DR, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *Am J Epidemiol* 2002;156:871–881. [PubMed: 12397006]

16. Heckbert SR, Wiggins KL, Blackshear C, Yang Y, Ding J, Liu J, McKnight B, Alonso A, Austin TR, Benjamin EJ, Curtis LH, Sotoodehnia N, Correa A. Pericardial fat volume and incident atrial fibrillation in the Multi-Ethnic Study of Atherosclerosis and Jackson Heart Study. *Obesity* 2017;25:1115–1121. [PubMed: 28452401]
17. Tibshirani R The LASSO method for variable selection in the Cox model. *Stat Med* 1997;16:385–395. [PubMed: 9044528]
18. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the Performance of Prediction Models. *Epidemiology* 2010;21:128–138. [PubMed: 20010215]
19. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Stat Med* 2004;23:2109–2123. [PubMed: 15211606]
20. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000;56:337–344. [PubMed: 10877287]
21. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Stat Med* 2013;32:1467–1482. [PubMed: 23296397]
22. Demler O V, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med* 2015;34:1659–1680. [PubMed: 25684707]
23. Wang TJ, Larson MG, Levy D, Benjamin EJ, Leip EP, Omland T, Wolf PA, Vasan RS. Plasma Natriuretic Peptide Levels and the Risk of Cardiovascular Events and Death. *N Engl J Med* 2004;350:655–663. [PubMed: 14960742]
24. Patton KK, Ellinor PT, Heckbert SR, Christenson RH, Defilippi C, Gottdiener JS, Kronmal RA. N-Terminal pro-b-type natriuretic peptide is a major predictor of the development of atrial fibrillation: The cardiovascular health study. *Circulation* 2009;120:1768–1774. [PubMed: 19841297]
25. Schnabel RB, Larson MG, Yamamoto JF, Sullivan LM, Pencina MJ, Meigs JB, Tofler GH, Selhub J, Jacques PF, Wolf PA, Magnani JW, Ellinor PT, Wang TJ, Levy D, Vasan RS, Benjamin EJ. Relations of biomarkers of distinct pathophysiological pathways and atrial fibrillation incidence in the community. *Circulation* 2010;121:200–207. [PubMed: 20048208]
26. Smith JG, Newton-Cheh C, Almgren P, Struck J, Morgenthaler NG, Bergmann A, Platonov PG, Hedblad B, Engström G, Wang TJ, Melander O. Assessment of conventional cardiovascular risk factors and multiple biomarkers for the prediction of incident heart failure and atrial fibrillation. *J Am Coll Cardiol* 2010;56:1712–1719. [PubMed: 21070922]
27. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, O'Leary DH, Tracy R, Watson K, Wong ND, Kronmal RA. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med* 2008;358:1338–1345.
28. O'Neal WT, Efird JT, Dawood FZ, Yeboah J, Alonso A, Heckbert SR, Soliman EZ. Coronary artery calcium and risk of atrial fibrillation (from the Multi-Ethnic Study of Atherosclerosis). *Am J Cardiol* 2014;114:1707–1712. [PubMed: 25282316]
29. O'Neal WT, Efird JT, Qureshi WT, Yeboah J, Alonso A, Heckbert SR, Nazarian S, Soliman EZ. Coronary Artery Calcium Progression and Atrial Fibrillation: The Multi-Ethnic Study of Atherosclerosis. *Circ Cardiovasc Imaging* 2015;8:e003786.
30. Vinter N, Christesen AMS, Mortensen LS, Urbonaviciene G, Lindholt J, Johnsen SP, Frost L. Coronary artery calcium score and the long-term risk of atrial fibrillation in patients undergoing non-contrast cardiac computed tomography for suspected coronary artery disease: a Danish registry-based cohort study. *Eur Heart J Cardiovasc Imaging* 2018;19:926–932. [PubMed: 28977363]

Figure 1. Discrimination of Atrial Fibrillation 5-year Risk Scores in MESA

ROC curves for 4 AF risk prediction models based on 23 candidate predictor variables included in previous risk prediction models of AF and a machine learning analysis in MESA. Higher values of the c-statistic indicate better performance.

MESA = Multi-Ethnic Study of Atherosclerosis; ROC = receiver operating characteristic

Figure 2. Observed vs Predicted Probability of Atrial Fibrillation at 5 Years

The predicted and observed event probability estimates represent the mean predicted probability from the Cox regression model and the mean observed probability from the population divided into categories of predicted probability. Panel A, Calibration of the CHARGE-AF Simple model; Panel B, Calibration of the CHARGE-AF Enriched model; Panel C, Calibration of the All Predictors Model; Panel D, Calibration of the Novel MESA Model.

CHARGE = Cohorts for Heart and Aging Research in Genomic Epidemiology; MESA = Multi-Ethnic Study of Atherosclerosis

Table 1.

Baseline Participant Characteristics in the Multi-Ethnic Study of Atherosclerosis by Incident Atrial Fibrillation within Five Years

Variables	Atrial Fibrillation		P Value
	No (n = 3410)	Yes (n = 124)	
Age (years)	60.9 (10.0)	70.9 (7.3)	<0.001
Men	1646 (48%)	70 (56%)	0.09
Non-Hispanic white	1311 (38%)	59 (48%)	0.21
Chinese American	524 (15%)	14 (11%)	
Non-Hispanic black	762 (22%)	25 (20%)	
Non-Hispanic black	762 (22%)	25 (20%)	
Hispanic	813 (24%)	26 (21%)	0.72
Current smoker	416 (12%)	17 (14%)	
Body mass index (kg/m ²)	27.5 (4.9)	28.0 (4.8)	
Systolic blood pressure (mm Hg)	124.8 (21.1)	133.3 (21.1)	
Diastolic blood pressure (mm Hg)	71.8 (10.2)	71.2 (10.4)	0.51
Pulse pressure (mm Hg)	53.0 (16.6)	62.1 (17.4)	<0.001
Antihypertensive medication use	1118 (33%)	64 (52%)	<0.001
Diabetes mellitus	380 (11%)	13 (10%)	0.93
Coronary artery calcium (Agatston units)			<0.001
0	1812 (53%)	32 (26%)	
1–99	885 (26%)	39 (31%)	
100–399	438 (13%)	25 (20%)	
400+	275 (8%)	28 (23%)	
Ankle-brachial index	1.12 (0.11)	1.09 (0.16)	0.002
Common carotid intima-media thickness (mm)	0.85 (0.18)	0.95 (0.21)	<0.001
Internal carotid intima-media thickness (mm)	1.02 (0.56)	1.27 (0.75)	<0.001
Serum creatinine (mg/dL)	0.96 (0.24)	1.04 (0.29)	<0.001
N-terminal pro-B-type natriuretic peptide (pg/mL)	50.1 [22.0, 97.6]	106.9 [51.3, 233.5]	<0.001
Detectable cardiac troponin-T	34 (1%)	5 (4%)	0.006
R amplitude in lead V4 (uV)	1429.8 (545.5)	1478.5 (660.3)	0.33
STJ amplitude in lead V5 (uV)	18.5 (35.3)	8.37 (42.56)	0.002
Heart rate (beats per minute)	62.7 (9.1)	62.3 (10.7)	0.64
Heart rate variability (ms)	22.6 (16.0)	20.2 (15.2)	0.10
QRS axis (degrees)	21.9 (31.4)	12.1 (31.2)	0.001
End-systolic basal superior lateral wall thickness (mm)	15.6 (3.0)	16.4 (3.4)	0.003
End-systolic mid-ventricular anterior wall thickness (mm)	15.9 (3.1)	17.3 (3.7)	<0.001

Values are mean (standard deviation), median [interquartile range]

Table 2.

Discrimination and Calibration for Sequential Five-year Atrial Fibrillation Risk Prediction Models

Model		Discrimination		Calibration	
		C-statistic (95% CI)	P Value	GND X ² value	P Value
1	CHARGE-AF Simple *	0.795 (0.764–0.827)	-	6.2	0.28
2	CHARGE-AF Enriched †	0.804 (0.771–0.837)	<0.001	2.4	0.80
3	+ Biomarkers ‡	0.804 (0.771–0.837)	0.71	2.4	0.79
4	+ MRI §	0.803 (0.770–0.837)	0.13	4.2	0.53
5	+ ECG ¶	0.805 (0.773–0.837)	0.86	4.0	0.55
6	+ Subclinical CVD #	0.805 (0.772–0.837)	0.82	3.2	0.67
7	All Predictors **	0.806 (0.774–0.839)	0.81	4.3	0.51
8	Novel MESA ††	0.802 (0.769–0.835)	0.93	3.9	0.57

Higher values for C-statistic indicate better models. Discrimination P values are for a likelihood ratio test comparing each model with Model 2, except for 1) Model 2, which is compared with Model 1; and 2) Model 8, which is compared with Model 7. Calibration P values >0.05 indicate adequate fit.

* Model 1: age, white race, height, weight, systolic BP, diastolic BP, current smoking, antihypertensive medication use, diabetes

† Model 2: Model 1 + ln(NT-proBNP)

‡ Model 3: Model 2 + serum creatinine, cardiac troponin-T

§ Model 4: Model 2 + end-systolic basal superior lateral wall thickness, end-systolic mid-ventricular anterior wall thickness

¶ Model 5: Model 2 + heart rate, heart rate variability, R amplitude in lead V4, STJ amplitude in lead V5, QRS axis

Model 6: Model 2 + ln(CAC + 1), ankle-brachial index, common cIMT, internal cIMT

** Model 7: all candidate variables

†† Model 8: age, weight, current smoking, ln(NT-proBNP), ln(CAC+1), and cardiac troponin-T

Abbreviations: BP = blood pressure; CAC = coronary artery calcium; CI = confidence interval; cIMT = carotid intima-media thickness; CVD = cardiovascular disease; ECG = electrocardiography; GND = Greenwood-Nam-D'Agostino; LRT = likelihood ratio test; MRI = magnetic resonance imaging; NT-proBNP = N-terminal pro-B-type natriuretic peptide

Table 3.

Hazard Ratios and Predicted Risk Calculation for CHARGE-AF Enriched, All Predictors, and Novel MESA
Five-year Atrial Fibrillation Risk Prediction Models

Variables	CHARGE-AF Enriched		All Predictors		Novel MESA	
	HR (95% CI)	Beta *	HR (95% CI)	Beta *	HR (95% CI)	Beta *
Age (per 5 years)	1.64 (1.45–1.85)	0.0984	1.56 (1.36–1.79)	0.0886	1.58 (1.40–1.80)	0.0917
Non-Hispanic white	1.09 (0.75–1.59)	0.0875	1.18 (0.79–1.77)	0.1665		
Non-Hispanic white	1.03 (0.82–1.30)	0.0032	1.02 (0.79–1.32)	0.0023		
Weight (per 15 kg)	1.37 (1.11–1.69)	0.0208	1.32 (1.05–1.66)	0.0186	1.37 (1.15–1.63)	0.0208
Systolic BP (per 20 mm Hg)	0.93 (0.74–1.18)	-	0.93 (0.72–1.20)	-		
Diastolic BP (per 10 mm Hg)	1.03 (0.80–1.32)	0.0030	1.00 (0.76–1.31)	-		
Current smoker	2.05 (1.21–3.48)	0.7184	2.06 (1.19–3.58)	0.7231	1.97 (1.17–3.32)	0.6774
Antihypertensive medication use	1.23 (0.84–1.78)	0.2036	1.21 (0.83–1.77)	0.1932		
Diabetes mellitus	0.76 (0.42–1.38)	-	0.68 (0.37–1.26)	-		
Ln(NT-proBNP) (per 1 SD)	1.54 (1.27–1.88)	0.4343	1.55 (1.26–1.90)	0.4355	1.51 (1.26–1.82)	0.4148
Serum creatinine (per 0.1 mg/dL)			0.99 (0.93–1.06)	-		
Detectable cardiac troponin-T			1.75 (0.61–4.99)	0.0545	0.5596 (0.54–3.52)	1.38 0.3221
Basal superior lateral wall thickness (per 5 mm) [†]			0.87 (0.59–1.26)	0.0290		
Mid-ventricular anterior wall thickness (per 5 mm) [†]			1.42 (1.02–1.97)	0.0696		
Heart rate (per 5 beats per minute)			1.04 (0.94–1.15)	0.0074		
Heart rate variability (per 10 ms)			0.95 (0.85–1.07)	-		
R amplitude in lead V4 (per 100 uV)			1.01	0.0001		
STJ amplitude in lead V5 (per 10 uV)			1.02 (0.97–1.07)	0.0019		

Variables	CHARGE-AF Enriched		All Predictors		Novel MESA	
	HR (95% CI)	Beta *	HR (95% CI)	Beta *	HR (95% CI)	Beta *
QRS Axis (per 10 degrees)			0.99 (0.93–1.05)	- 0.0010		
Ln(CAC + 1) (per 1 SD)			1.04 (0.96–1.13)	0.0432	1.05 (0.97–1.13)	0.0448
Ankle-brachial index (per 0.05)			0.99 (0.93–1.06)	- 0.1251		
Common cIMT (per 0.5 mm)			1.07 (0.65–1.77)	0.1424		
Internal cIMT (per 0.5 mm)			0.96 (0.83–1.11)	0.0810		

The 5-year risk for the CHARGE-AF Enriched model can be calculated as $1-0.9819\exp(\sum bX-9.7729)$ where b is the regression coefficient (beta) and X is the level for each risk factor; the risk for the All Predictors model can be calculated as $1-0.9823\exp(\sum bX-9.6722)$; the risk for the Novel MESA model can be calculated as $1-0.9815\exp(\sum bX-8.9785)$.

* Per 1-unit increase

[†] Measured during end-s stole

Abbreviations: BP = blood pressure; CAC = coronary artery calcium; CI = confidence interval; cIMT = carotid int. media thickness; NT-proBNP = N-terminal pro-B-type natriuretic peptide