Integrating Clinical, Genetic, and Electrocardiogram-Based Artificial Intelligence to Estimate Risk of Incident Atrial Fibrillation

Shinwan Kany MD, MSc[1,2], Joel T. Rämö MD, PhD[1,3], Samuel F. Friedman PhD[4], Lu-Chen Weng PhD[1,5], Carolina Roselli PhD[1], Min Seo Kim, MD, MSc[1,5]; Akl C. Fahed, MD, MPH[1,5,6]; Steven A. Lubitz MD, MPH[1,5,7], Mahnaz Maddah PhD[4], Patrick T. Ellinor MD, PhD*[1,5,7], Shaan Khurshid MD, MPH*[1,5,7]

[1]Cardiovascular Disease Initiative, Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[2]Department of Cardiology, University Heart and Vascular Center Hamburg-Eppendorf, Hamburg, Germany
[3]Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki, Finland
[4]Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[5]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA
[6]Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts, USA
[7]Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, Massachusetts, USA

* Authors jointly supervised this work

Short Title: Clinical, genetic, and artificial intelligence AF risk

Journal subject terms: atrial fibrillation, genetics, risk factors, artificial intelligence, electrocardiogram

Word count: 3,068

Corresponding author:

Shaan Khurshid, MD, MPH: Demoulas Center for Cardiac Arrythmias and Cardiovascular Research Center, Massachusetts General Hospital, 55 Fruit Street, GRB 109, Boston, MA 02114; P:617-726-5036; F:617-726-7519; skhurshid@mgb.org

ABSTRACT

Background: AF risk estimation is feasible using clinical factors, inherited predisposition, and artificial intelligence (AI)-enabled electrocardiogram (ECG) analysis.

Objective: To test whether integrating these distinct risk signals improves AF risk estimation.

Methods: In the UK Biobank prospective cohort study, we estimated AF risk using three models derived from external populations: the well-validated Cohorts for Aging in Heart and Aging Research in Genomic Epidemiology AF (CHARGE-AF) clinical score, a 1,113,667-variant AF polygenic risk score (PRS), and a published AI-enabled ECG-based AF risk model (ECG-AI). We estimated discrimination of 5-year incident AF using time-dependent area under the receiver operating characteristic (AUROC) and average precision (AP).

Results: Among 49,293 individuals (mean age 65±8 years, 52% women), 825 (2.4%) developed AF within 5 years. Using single models, discrimination of 5-year incident AF was higher using ECG-AI (AUROC 0.705 [95%CI 0.686-0.724]; AP 0.085 [0.071-0.11]) and CHARGE-AF (AUROC 0.785 [0.769-0.801]; AP 0.053 [0.048-0.061]) versus the PRS (AUROC 0.618, [0.598-0.639]; AP 0.038 [0.028-0.045]). The inclusion of all components ("Predict-AF3") was the best performing model (AUROC 0.817 [0.802-0.832]; AP 0.11 [0.091-0.15], $p<0.01$ vs CHARGE-AF+ECG-AI), followed by the two component model of CHARGE-AF+ECG-AI (AUROC 0.802 [0.786-0.818]; AP 0.098 [0.081-0.13]). Using Predict-AF3, individuals at high AF risk (i.e., 5-year predicted AF risk >2.5%) had a 5-year cumulative incidence of AF of 5.83% (5.33-6.32). At the same threshold, the 5-year cumulative incidence of AF was progressively higher according to the number of models predicting high risk (zero: 0.67% [0.51-0.84], one: 1.48% [1.28-1.69], two: 4.48% [3.99-4.98]; three: 11.06% [9.48-12.61]), and Predict-AF3 achieved favorable net reclassification improvement compared to both CHARGE-AF+ECG-AI (0.039 [0.015-0.066]) and CHARGE-AF+PRS (0.033 [0.0082-0.059]).

Conclusions: Integration of clinical, genetic, and AI-derived risk signals improves discrimination of 5-year AF risk over individual components. Models such as Predict-AF3 have substantial potential to improve prioritization of individuals for AF screening and preventive interventions.

Key Words: atrial fibrillation, genetics, precision medicine, risk prediction, stroke, screening

Introduction

Atrial fibrillation (AF) is associated with increased risks of stroke, heart failure, and death.[1] AF screening can facilitate earlier diagnosis, and preventive treatment for AF-related morbidity, such as risk factor management or anticoagulation, can mitigate AF-related morbidity or even prevent AF altogether.[2] However, screening approaches to date have demonstrated modest yield of AF diagnosis and have failed to demonstrate improvements in hard outcomes such as stroke or mortality.[3–5]

Analogous to established screening approaches for selected conditions such as lung cancer[6] or osteoporosis,[7] the efficiency of AF screening may be improved by utilizing a risk-informed approach. AF risk can be predicted with reasonable accuracy on the basis of clinical risk factors,[8] inherited predisposition as assessed by polygenic risk scores [PRS]),[9,10] and most recently artificial intelligence-enabled analysis of the electrocardiogram (ECG-AI).[11,12] We have previously shown that the predictive power of a validated clinical risk score such as CHARGE-AF can be improved by the addition of either an AF PRS[9] or ECG-AI.[11] However, the degree to which each of these varied AF risk signals may overlap or complement one another within the context of a single comprehensive model remains unknown.

Here, we leveraged the UK Biobank – a unique resource of nearly 50,000 individuals with linkage to national health-related datasets, protocolized prospectively acquired 12-lead ECG, and genome-wide genotyping, to quantify the relative contributions of externally developed AF risk scores comprising a) clinical risk factors, b) common genetic variation, and c) AI-enabled ECG analysis. We hypothesized that by integrating complementary information, models incorporating each of the varying AF risk signals would achieve greater longitudinal AF discrimination.

Methods

The UK Biobank was approved by the UK Biobank Research Ethics Committee (reference 11/NW/0382). All UK Biobank participants provided written informed consent. Use of UK Biobank (application #7089) data was approved by the MGB institutional review board.

Data availability

UK Biobank data are accessible to researchers by application (www.ukbiobank.ac.uk). The PRS used in this study was obtained from the most recent AFGen consortium genome-wide association study by Roselli et al., which utilized a version of the analysis excluding UK Biobank participants.[13] The published version of the ECG-AI model[11] used to generate the AI-based predictions evaluated in the current analysis is available at https://github.com/broadinstitute/ml4h/tree/master/model_zoo/ECG2AF. Data processing scripts underlying the current analysis are available at https://github.com/shaankhurshid/af_prediction3.

UK Biobank cohort

We analyzed the UK Biobank, a prospective cohort of 502,629 participants aged 40-69 that were recruited between 2006-2010.[14] Participants underwent an extensive assessment in various stages ("instances"), including questionnaire data, anthropometric measures, and laboratory values. Instance 2 included a structured imaging visit for a subset of participants that included whole-body magnetic resonance imaging as well as protocolized resting 12-lead ECG. We analyzed individuals without prevalent AF who underwent resting 12-lead ECG at the instance 2 study visit.

AF polygenic risk score

Details of genotyping, imputation, and quality control have been published previously and are provided in **Supplementary Methods**.[14,15] The SNP-level weights used to calculate the AF PRS used in the current study were derived from a recent AF GWAS meta-analysis including

154,330 AF cases and 999,609 controls.[13] The weights were derived using the PRS-CS approach.[16] The UK Biobank cohort was not included in the version of the meta-analysis used as input to the PRS-CS algorithm. We calculated AF genetic risk in UK Biobank individuals using Plink2, summing up the weighted effect allele dosage of 1,113,667 variants included in the score with good quality (imputation info ≥0.4).[17]

ECG-AI

To obtain ECG-based AI-enabled AF risk estimates, we applied a contemporary version of ECG-AI, a previously published convolutional neural network-based deep learning model designed to estimate 5-year risk of AF using a single 12-lead ECG.[11] The current version of ECG-AI utilized a larger training set (450,000 standard 12-lead ECGs representing over 100,000 primary care and cardiology patients at Massachusetts General Hospital) and achieved higher discrimination of 5-year incident AF (c-index 0.761 vs. 0.716), but otherwise had identical architecture to the published model described in detail previously.[11] UK Biobank participants were not included in any aspect of ECG-AI development.

CHARGE-AF

To estimate the predictive utility of clinical risk factors, we calculated the CHARGE-AF score, a widely validated risk factor-based AF prediction tool.[8,18] UK Biobank participants were not included in the original derivation of CHARGE-AF. To calculate the score, age, sex, race, height, weight, and blood pressure values were obtained from structured assessments using the value closest to the ECG date. Anti-hypertensive use was determined using self-reported medication data. Tobacco use was categorized as present or absent. Race was classified as White or non-White, as performed previously using CHARGE-AF. In cases of multiple available

values, values from the assessment most closely preceding the ECG were used preferentially. The presence of heart failure, diabetes, and myocardial infarction were ascertained using previously published diagnostic and procedural codes.[9,11] A full summary of clinical factor definitions is provided in **Supplementary Table 1**.

Outcomes

The prediction target for each model was 5-year incident AF. AF events were identified using a previously published combination of self-reported illness codes, OPCS Classification of Interventions and Procedures version 4 codes for cardioversion or catheter ablation, and International Classification of Diseases, Ninth- or Tenth Revision (ICD-10) codes for AF or atrial flutter (**Supplementary Table 1**).[19]

Statistical analysis

Discrimination of 5-year incident AF was measured using the area under the time-dependent receiver operating characteristic curve (AUROC) and time-dependent average precision (AP).[20,21] Similar to how AUROC provides a composite measure of test sensitivity and specificity across a range of predictor thresholds, AP provides a composite measure of test precision (i.e., positive predictive value) and recall (i.e., sensitivity) across a range of predictor thresholds. Both AP and AUROC were calculated using inverse probability of censoring weights, which more accurately account for the potential bias introduced by censoring when compared to unweighted measures such as Harrell's c-index.[22]

Discrimination of individual models was assessed as raw scores (CHARGE-AF, PRS) or probabilities (ECG-AI). Combination models were then developed by fitting Cox proportional hazards models with terms for each respective component (e.g., CHARGE-AF + PRS comprises a Cox model including a term for CHARGE-AF and a term for the AF PRS). As performed with the two-component Predict-AF model previously,[23] prior to inclusion in Cox

models the ECG-AI AF probabilities were logit-transformed to achieve an approximately linear relationship with log hazard. A model with a coefficient for each of CHARGE-AF, the PRS, and ECG-AI was termed “Predict-AF3”. Discrimination of the combination models was assessed using their linear predictors (i.e., score values). Discrimination was assessed overall and across subgroups of age (age<60 years, age 60-70 years, age>70 years, approximating tertiles of the distribution) and sex.

We then assessed model calibration. To allow fair comparison across each component, including components with no intrinsic translation to a longitudinal risk estimate (i.e., AF PRS), we fit univariable Cox proportional hazards models with a single term for each individual component. Linear predictors of the individual fitted models as well as the combination models outlined above were then converted into 5-year absolute risk estimates using the equation $1-S_0^{\exp(\Sigma X-\Sigma Y)}$, where S_0 is the average 5-year AF-free survival of the sample, ΣX is the individual’s linear predictor or score value, and ΣY is the average score of the sample. Calibration of absolute risk estimates was then assessed by plotting smoothed curves of absolute versus predicted risk using adaptive hazard regression, and calculating the integrated calibration index (ICI), a measure of the average prediction error weighted by the empirical risk distribution and where an ICI of zero indicates perfect absolute risk estimates.[24] For AP, AUROC, and ICI estimates, 95% confidence intervals were calculated using bootstrapping (500 iterations for AP and AUROC, and 200 iterations for ICI), which also provided estimates of standard error to perform pairwise *z* testing. To estimate the clinical impact of implementing one AF risk model over another, we calculated net reclassification improvement, both at the >2.5% 5-year AF risk threshold and as a continuous measure.[8,25] We considered 2-sided $p<0.05$ statistically significant. Analyses were performed using Python v3.10 and R v4.3.

Results

Sample *characteristics*

We analyzed 49,293 individuals without prevalent AF (age 65 ± 8 years, 52% women, 96% White) at the time of the UK Biobank instance 2 visit. Over 5 years of follow-up, 825 participants (2.4%) developed incident AF. An overview of the study is provided in **Figure 1** and baseline characteristics are provided in **Table 1**.

Discrimination of incident atrial fibrillation using single models

Using single prediction models, discrimination of 5-year incident AF as measured by AUROC was highest using CHARGE-AF (0.785 [95% CI: 0.769-0.801]), followed by ECG-AI 0.705 [0.686-0.724]) and the PRS (0.618 [0.598-0.639]). Discrimination using CHARGE-AF tended to increase with longer prediction windows, while discrimination using ECG-AI and AF PRS remained largely stable over time (**Figure 2**). While CHARGE-AF had highest discrimination among individuals aged <60 years (AUROC 0.739 [0.678-0.798]), discrimination was more similar across models in those aged >70 years (e.g. CHARGE-AF AUROC 0.688 vs ECG-AI AUROC 0.687) (**Supplementary Table 2**). Models performed similarly among men versus women (**Supplementary Table 3**).

Measured using average precision, discrimination was highest using ECG-AI (0.085 [0.071-0.11]), followed by CHARGE-AF (0.053 [0.048-0.061]) and the PRS (0.038 [0.028-0.045]), indicative of relatively good performance using ECG-AI for the detection of the highest risk individuals. Consistent with an increasing cumulative event rate (e.g., 0.4% at 1 year to 2.4% at 5 years), the AP for each model increased over time (**Figure 2**) and among older subgroups of age (**Supplementary Table 2**). Trends in performance were similar among men versus women (**Supplementary Table 3**).

As expected, there was no substantive correlation between PRS and CHARGE-AF (r=-0.014 [-0.022 to -0.0048]). There was very weak correlation between PRS and ECG-AI (r=0.041

[0.032-0.050]). There was moderate correlation between ECG-AI and CHARGE-AF (r=0.34 [0.33-0.35])

Discrimination of incident atrial fibrillation using combined models

We then evaluated discrimination of the combined models, namely a) CHARGE-AF + PRS, b) ECG-AI + PRS, c) CHARGE-AF + ECG-AI, and CHARGE-AF + ECG-AI + PRS ("Predict-AF3"). The best performing two-component models were CHARGE-AF + ECG-AI (AUROC 0.802 [0.786-0.818]; AP 0.098 [0.081-0.13]), followed by CHARGE-AF + PRS (AUROC 0.802 [0.786-0.818]; AP 0.053 [0.048-0.061]). The best performing model overall was Predict-AF3 (AUROC 0.817 [0.802-0.832]; AP 0.11 [0.091-0.15], p<0.01 vs CHARGE-AF + ECG-AI and CHARGE-AF + PRS) (**Figure 3**). Addition of the PRS to each component contributed relatively modest but consistently detectable improvements in discrimination (improvement in AUROC with addition of PRS to CHARGE-AF: 0.017 [0.0096-0.025], addition of PRS to ECG-AI: 0.023 [0.012-0.034], addition of PRS to CHARGE-AF + ECG-AI: 0.015 [0.0089-0.021]) (**Supplementary Table 4**). Improvements were similar according to AP (**Supplementary Table 4**). Relative performance patterns were similar when assessing discrimination of AF at 1 year (e.g., Predict-AF3 AUROC 0.791 [0.758-0.823]; AP 0.030 [0.020-0.050], **Figure 3**).

Calibration and net reclassification of atrial fibrillation risk estimates

Absolute risk estimates for the single and combined models were well-calibrated (**Supplementary Figure 1**). Absolute error rates were consistently very low (ICI range 4.16×10^{-5} for CHARGE-AF to 0.0018 for Predict-AF3). At the >2.5% 5-year AF risk threshold, Predict-AF3 resulted in favorable net reclassification improvement versus CHARGE-AF + ECG-AI (NRI 0.039 [0.015-0.066]), which was driven by both favorable case reclassification (NRI+ 2.59% [0.18-5.30]) and favorable non-case reclassification (NRI- 1.35% [1.06-1.67]) (**Supplementary Table 5**). Predict-AF3 also provided similar reclassification improvement over CHARGE-AF +

PRS, although it was driven solely by favorable non-case reclassification (NRI 0.033 [0.0082-0.059]; NRI+ 0.52% [-2.06 to 3.00]; NRI- 2.81% [2.54-3.08]) (**Supplementary Table 5**). Continuous reclassification improvement was also favorable using Predict-AF3 (0.34 [0.26-0.41] vs. CHARGE-AF + ECG-AI; 0.23 [0.16-0.31] vs. CHARGE-AF + PRS).

Stratification of longitudinal atrial fibrillation incidence using Predict-AF3

Use of Predict-AF3 effectively stratified longitudinal AF risk, with individuals at high AF risk (5-year predicted AF risk >2.5%) having a 5-year cumulative incidence of AF of 5.83% (5.33-6.32), those with a predicted risk of 1-2.5% having a cumulative incidence of 1.51% (1.26-1.75) and those with a predicted risk of ≤1% having a cumulative incidence of 0.56% (0.43-0.69) (**Figure 4**).

Using the >2.5% 5-year AF risk threshold, the 5-year cumulative incidence of AF was progressively higher according to the number of individual models predicting high risk (zero: 0.67% [0.51-0.84], one: 1.48% [1.28-1.69], two 4.48% [3.99-4.98], three: 11.06% [9.48-12.61]) (**Figure 4**). Among the 825 participants who developed AF during the 5-year period, 208 (25.2%) had high estimated AF risk according to all three models, 340 (41.2%) according to two models, 207 (25.1%) according to one model, and 70 (8.5%) according to zero models.

Discussion

In this study, we leveraged a unique resource of nearly 50,000 prospective cohort study participants with detailed clinical data, genome-wide genotyping, and protocolized 12-lead ECG to compare the relative predictive utility of varied forms of AF risk information. Specifically, we applied externally derived and contemporary clinical, genetic and ECG-based AI-enabled AF risk models separately and in combination. Our findings demonstrate that clinical, genetic, and AI-based AF risk signals are complementary. There was a graded increase in AF incidence as individuals were identified as high-risk by a greater number of risk signals. The Predict-AF3

score, which combines all three elements, achieved higher predictive utility than any single or two-model combination. Overall, our work establishes the value of integrating varying data types to achieve increasingly accurate AF risk estimates, providing a foundation for efforts to better prioritize individuals for AF screening and related preventive efforts. Our findings yield several key implications.

First, it is feasible to integrate varying AF risk signals to achieve more accurate AF risk estimates. In our analysis, the Predict-AF3 score consistently exhibited favorable discrimination compared to the individual components of the score, implying that clinical, genetic, and ECG-based AI risk estimates provide complementary information. By incorporating all three components, Predict-AF3 achieves the highest discrimination of 5-year incident AF risk (AUROC 0.82, AP 0.11) of any AF prediction model previously applied in a prospective community cohort.[8,9,11,18,26,27] Our findings are consistent with prior observations from our group suggesting that the addition of ECG-derived AI risk to clinical risk improves AF risk estimation.[11] Despite some evidence that ECG-based AI may encode some aspects of inherited predisposition to AF,[28] we consistently observed improvements in AF discrimination with the addition of PRS to any individual component or two-component combination, suggesting that genetic risk remains orthogonal to current ECG-based AI models. However, as with prior work, the degree of discrimination improvement observed with the addition of PRS was generally modest,[9,29] although we did note favorable net reclassification improvement when adding the PRS to the best performing two-component model (CHARGE-AF + ECG-AI) to create Predict-AF3. Importantly, we observed that AF incidence increased progressively with the number of distinct elements portending high AF risk, suggesting that individuals with risk conferred by the combination of clinical factors, genetics, and AI-enabled ECG signals appear particularly vulnerable to developing AF.

Second, the optimal approach to AF risk estimation to guide AF screening and related preventive interventions may be a function of available data and the specific setting in which risk stratification is intended. Although Predict-AF3 achieved the highest discrimination of longitudinal AF among all approaches tested, genome-wide genotyping data is not commonly available in many populations in which AF screening is considered (e.g., routine primary care). To this end, we observed good discrimination using CHARGE-AF + ECG-AI, a model requiring only routine clinical factors and a single 12-lead ECG, an inexpensive diagnostic test available within minutes in most clinic settings. Future work is warranted to assess whether the predictive utility of ECG-AI may extend to single-lead ECGs, which are increasingly available using mobile and consumer devices and may increase the reach of AF risk estimation further.[30] Conversely, the combination of CHARGE-AF + PRS also achieved similar performance to CHARGE-AF + ECG-AI. Such a model may have particular value in population health interventions targeting healthcare-related biobanks, where risk estimation can be run on previously acquired samples, and leveraging linked electronic health record data.

Third, the integration of an increasing variety of clinically relevant risk markers has potential to further the goal of personalized medicine with regard to AF. While opportunistic screening is recommended in the most recent European Society of Cardiology guidelines for people of older age,[31] recent randomized trials of screening guided only by age have resulted in little to no increase in AF diagnosis, and absence of meaningful improvements in AF-related complications such as stroke or mortality.[3,4] Subgroup analyses have suggested higher AF screening yield among individuals with higher AF incidence,[3] suggesting that screening based on AF risk may be more efficient. To this end, recent work has identified the ability to stratify AF risk based on a large breadth of potential markers, including not only clinical factors, polygenic risk, and AI-based signals, but additionally imaging-based features and blood-based biomarkers.[32,33] Therefore, although Predict-AF3 provides an important demonstration of the potential value of

incorporating varying contributors to AF risk to achieve more precise AF risk estimates, future work is warranted to integrate even more markers, and potentially leverage emerging statistical learning techniques capable of incorporating high-level interactions between varied data types. Ultimately, randomized trials are needed to assess whether AF risk estimation using progressively more accurate models leads to more efficient AF screening interventions.

Limitations

Our study should be interpreted in the context of design. The UK Biobank is primarily of European ancestry and our results might not be generalizable to people of other ancestries. Second, the UK Biobank is not reflective of the general population and comprises healthier individuals.[34] Due to survivorship bias, the instance 2 cohort that received protocolized 12-lead ECG is even healthier than the UK Biobank at large. Third, our models were assessed against the clinical diagnosis of AF, and therefore their performance for detection of undiagnosed AF in a screening setting may vary. Fourth, although the clinical risk score, PRS, and ECG-based AI model assessed in the current study were all developed externally to the UK Biobank, the relative contributions of each component in the combination models were weighted within-sample. The relative importance of AF risk components may differ in other populations, and therefore further validation of the Predict-AF3 multi-component model is warranted.

Conclusions

Integration of clinical, genetic, and ECG-based AI risk signals for AF into a single model (Predict-AF3) leads to greater predictive utility of 5-year incident AF compared to the use of individual elements. Consistent with the presence of complementary information, AF incidence increased progressively with the total number of distinct elements portending high AF risk. Scores such as Predict-AF3 may pave the way for integrated and personalized prioritization of individuals for AF screening and related preventive interventions.

1

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 17488**.** The authors would like to thank study staff and investigators from the UK Biobank.

Funding

SKany is supported by the Walter Benjamin Fellowship from the Deutsche Forschungsgemeinschaft (521832260). JTR is supported by a research fellowship from the Sigrid Jusélius Foundation. ACF is supported by grants from the National Institutes of Health (K08 HL161448 and R01 HL16462). PTE is supported by grants from the National Institutes of Health (RO1HL092577, RO1HL157635), from the American Heart Association (18SFRN34230127, 961045), and from the European Union (MAESTRIA 965286). SAL previously received support from NIH grants R01HL139731 and R01HL157635, and American Heart Association 18SFRN34250007. SKhurshid is supported by the NIH (K23HL169839-01) and the American Heart Association (2023CDA1050571).

Disclosures

ACF reports being co-founder of Goodpath, serving as scientific advisor to MyOme and HeartFlow, and receiving a research grant from Foresite Labs. PTE receives sponsored research support from Bayer AG, IBM Research, Bristol Myers Squibb, Pfizer and Novo Nordisk; he has also served on advisory boards or consulted for Bayer AG. SAL is an employee of Novartis as of July 2022. SAL was previously supported by NIH grants R01HL139731 and R01HL157635, and American Heart Association 18SFRN34250007. SAL received sponsored research support from Bristol Myers Squibb, Pfizer, Boehringer Ingelheim, Fitbit, Medtronic, Premier, and IBM, and has consulted for Bristol Myers Squibb, Pfizer, Blackstone Life Sciences, and Invitae. The remaining authors have nothing to disclose.

Authors' Contributions

SKany – Study conception, design, data analysis, data interpretation, drafting

SKhurshid – Study conception, design, data analysis, data interpretation, drafting, critical review and editing

CR, LCW, SFF – Data analysis, data interpretation, critical review

PTE – Study conception, design, data interpretation, drafting, critical review

JTR, MSK, ACF, SAL AAP, MM – Data interpretation, critical review

References

1. Corley SD, Epstein AE, DiMarco JP, et al. Relationships between sinus rhythm, treatment, and survival in the Atrial Fibrillation Follow-Up Investigation of Rhythm Management (AFFIRM) Study. *Circulation*. 2004;109:1509–1513.

2. Khurshid S, Healey JS, McIntyre WF, Lubitz SA. Population-Based Screening for Atrial Fibrillation. *Circ Res*. 2020;127:143–154.

3. Lubitz SA, Atlas SJ, Ashburner JM, et al. Screening for atrial fibrillation in older adults at primary care visits: VITAL-AF randomized controlled trial. *Circulation*. 2022;145:946–954.

4. Uittenbogaart SB, Verbiest-van Gurp N, Lucassen WAM, et al. Opportunistic screening versus usual care for detection of atrial fibrillation in primary care: cluster randomised controlled trial. *BMJ*. 2020;370:m3208.

5. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass Screening for Untreated Atrial Fibrillation: The STROKESTOP Study. *Circulation*. 2015;131:2176–2184.

6. Jaklitsch MT, Jacobson FL, Austin JHM, et al. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. *J Thorac Cardiovasc Surg*. 2012;144:33–38.

7. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX™ and the assessment of fracture probability in men and women from the UK. *Osteoporos Int*. 2008;19:385–397.

8. Alonso A, Krijthe BP, Aspelund T, et al. 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. Khurshid S, Mars N, Haggerty CM, et al. Predictive Accuracy of a Clinical and Genetic Risk

Model for Atrial Fibrillation. *Circ Genom Precis Med*. 2021;14:e003355.

10. Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*. 2020;26:549–557.

11. Khurshid S, Friedman S, Reeder C, et al. ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. *Circulation*. 2022;145:122–133.

12. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, et al. Deep Neural Networks Can Predict New-Onset Atrial Fibrillation From the 12-Lead ECG and Help Identify Those at Risk of Atrial Fibrillation–Related Stroke. *Circulation*. 2021;143:1287–1298.

13. Roselli C, Surakka I, Olesen MS, Sveinbjornsson G, Marston NA, Choi SH, Holm H, Chaffin M, Gudbjartsson D, Hill MC, Aegisdottir H, Albert CM, Alonso A, Anderson CD, Arking DE, Arnar DO, Barnard J, Benjamin EJ, Braunwald E, Brumpton B, Campbell A, Chami N, Chasman DI, Cho K, Choi EK, Christophersen IE, Chung MK, Conen D, Crijns HJ, Cutler MJ, Czuba T, Damrauer SM, Martin D, Dorr M, Dudink E, Duong T, Eriskstrup C, Esko T, Fatkin D, Faul JD, Ferreira M, Freitag DF, Ganesh SK, Gaziano M, Geelhoed B, Ghouse J, Gieger C, Giulianini F, Graham SE, Gudnason V, Guo X, Haggerty C, Jukema JW, Jurgens SJ, Kaab S, Kane JP, Kany S, Kardia SLR, Kavousi M, Khurshid S, Kamanu FK, Kirchhof P, Kleber ME, Knight S, Komuro I, Krieger JE, Launer LJ, Li D, Lin H, Lin HJ, Loos RJF, Lotta L, Lubitz SA, Lunetta KL, Macfarlane PW, Magnusson PKE, Malik R, Mantineo H, Marcus GM, Marz W, McManus DD, Melander O, Melloni GEM, Meyre P, Miyazawa K, Mohanty S, Monfort LM, Muller-Nurasyid M, Nafissi NA, Natale A, Nazarian S, Ostrowski SR, Pak HN, Pang S, Pedersen OB, Pedersen NL, Pereira AC, Pirruccello JP, Preuss M, Psaty BM, Pullinger CR, Rader DJ, Ramo JT, Ridker PM, Rienstra M, Risch L, Roden DM, Rotter JI, Sabatine MS, Schunkert H, Shah SH, Shim J, Shoemaker MB, Simonson B, Sinner MF, Smit RAJ, Smith JA, Smith NL, Smith JG, Soliman

EZ, Sorensen E, Sotoodehnia N, Strbian D, Stricker BH, Teder-Laving M, Sun YV, Theriault S, Thorolfsdottir RB, Thorsteinsdottir U, Tveit A, Van der Harst P, van Meurs J, Wang B, Weiss S, Wells QS, Weng LC, Wilson PW, Xiao L, Yang PS, Yao J, Yoneda ZT, Zeller T, Zeng L, Zhao W, Zhou W, Zollner S, The Biobank Japan Project, Regeneron Genetics Center, DBDS Genomic Consortium, Ruff CT, Bundgaard H, Willer C, Stefansson K, Ellinor PT. Meta-analysis of Genome-wide associations and polygenic risk prediction for atrial fibrillation in more than 180,000 cases. *Nat Genet*. 2024. Published online2024.

14. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–209.

15. Roselli C, Chaffin MD, Weng L-C, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet*. 2018;50:1225–1233.

16. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10:1776.

17. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4: 7. 2015.

18. Christophersen IE, Yin X, Larson MG, et al. A comparison of the CHARGE-AF and the CHA2DS2-VASc risk scores for prediction of atrial fibrillation in the Framingham Heart Study. *Am Heart J*. 2016;178:45–54.

19. Khurshid S, Choi SH, Weng L-C, et al. Frequency of Cardiac Rhythm Abnormalities in a Half Million Adults. *Circ Arrhythm Electrophysiol*. 2018;11:e006273.

20. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med*. 2013;32:2430–2442.

21. Yuan Y, Zhou QM, Li B, Cai H, Chow EJ, Armstrong GT. A threshold-free summary index of prediction accuracy for censored time to event data. *Stat Med*. 2018;37:1671–1681.

22. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56:779–788.

23. Christopoulos G, Graff-Radford J, Lopez CL, et al. Artificial Intelligence–Electrocardiography to Predict Incident Atrial Fibrillation. *Circ Arrhythm Electrophysiol*. 2020;13:e009355.

24. Austin PC, Harrell FE Jr, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. *Stat Med*. 2020;39:2714–2742.

25. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30:11–21.

26. Wu J, Nadarajah R, Nakao YM, et al. Risk calculator for incident atrial fibrillation across a range of prediction horizons. *Am Heart J*. 2024;272:1–10.

27. Khurshid S, Kartoun U, Ashburner JM, et al. Performance of Atrial Fibrillation Risk Prediction Models in Over 4 Million Individuals. *Circ Arrhythm Electrophysiol*. 2021;14:e008997.

28. Wang X, Khurshid S, Choi SH, et al. Genetic Susceptibility to Atrial Fibrillation Identified via Deep Learning of 12-Lead Electrocardiograms. *Circ Genom Precis Med*. 2023;16:340–349.

29. Marston NA, Garfinkel AC, Kamanu FK, et al. A polygenic risk score predicts atrial fibrillation in cardiovascular disease. *Eur Heart J*. 2023;44:221–231.

30. Pipilas DC, Khurshid S, Al-Alusi MA, et al. Automated interpretations of single-lead electrocardiograms predict incident atrial fibrillation: The VITAL-AF trial. *Heart Rhythm*. 2024. Published onlineApril 29, 2024. https://doi.org/10.1016/j.hrthm.2024.04.085.

31. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J*. 2021;42:373–498.

32. Staerk L, Preis SR, Lin H, et al. Protein Biomarkers and Risk of Atrial Fibrillation: The FHS. *Circ Arrhythm Electrophysiol*. 2020;13:e007607.

33. Schrage B, Geelhoed B, Niiranen TJ, et al. Comparison of Cardiovascular Risk Factors in European Population Cohorts for Predicting Atrial Fibrillation and Heart Failure, Their Subsequent Onset, and Death. *J Am Heart Assoc*. 2020;9:e015218.

34. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017;186:1026–1034.

Tables

Table 1. Baseline Characteristics

	Female	Male	Overall
N	25742	23551	49293
Age at ECG, years	64.48 (7.7)	65.73 (7.9)	65.08 (7.8)
Height, cm	162.6 (6.3)	175.9 (6.6)	169.0 (9.2)
Weight, kg	69.1 (13.4)	83.7 (13.6)	76.1 (15.3)
Current smoker	762 (3.0)	971 (4.1)	1733 (3.5)
Systolic blood pressure, mmHg	136.70 (19.7)	142.53 (17.6)	139.49 (18.9)
Diastolic blood pressure, mmHg	79.25 (10.1)	77.66 (10.0)	80.99 (9.9)
Diabetes mellitus	1136 (4.4)	1928 (8.2)	3064 (6.2)
Heart failure	80 (0.3)	173 (0.7)	253 (0.5)
Myocardial infarction	235 (0.9)	952 (4.0)	1187 (2.4)
Values are shown as mean (standard deviation) for continuous measures and N (%) for categorical measures.			

Figures

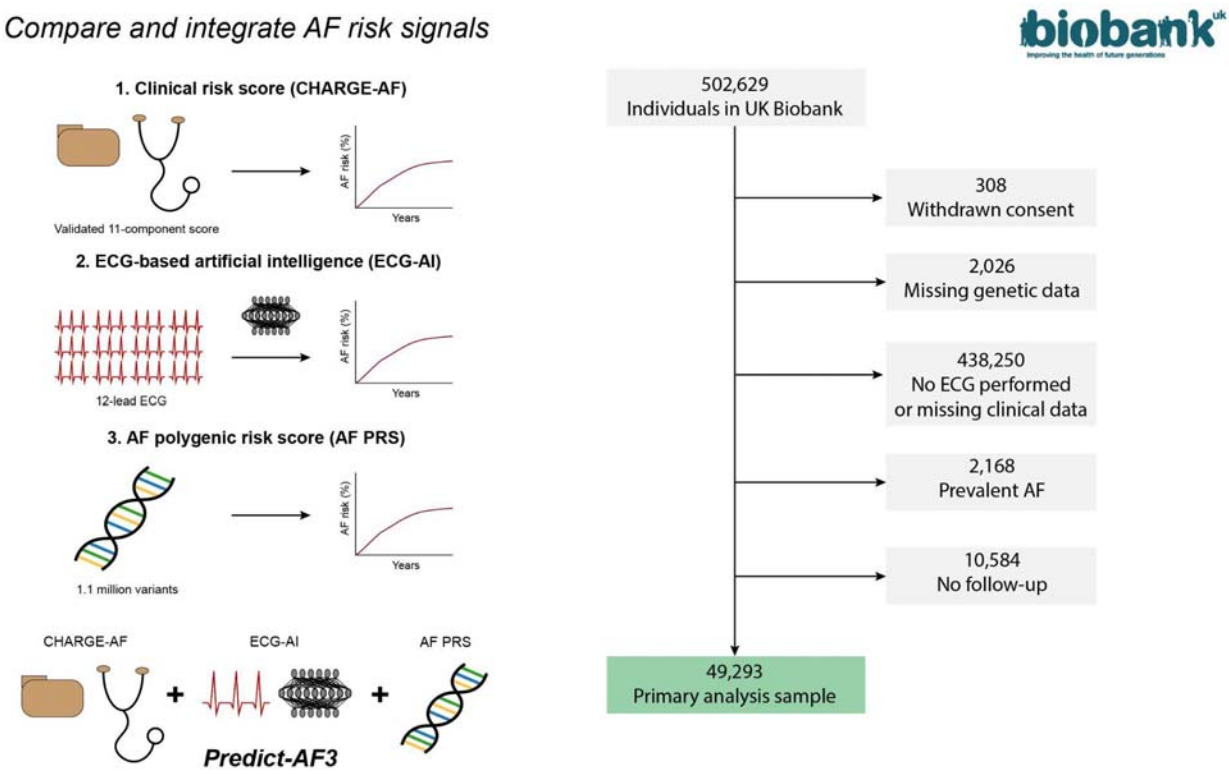
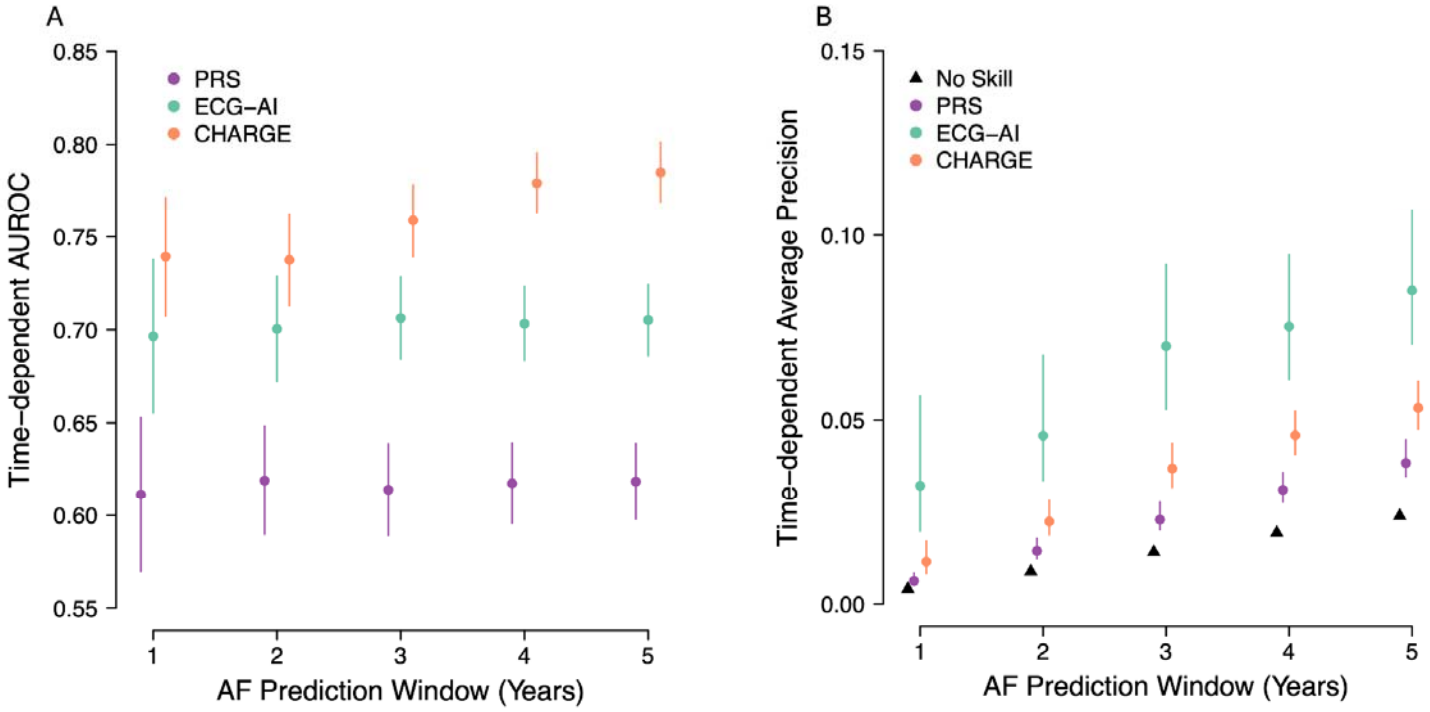


Figure 1. Flow chart of study cohort

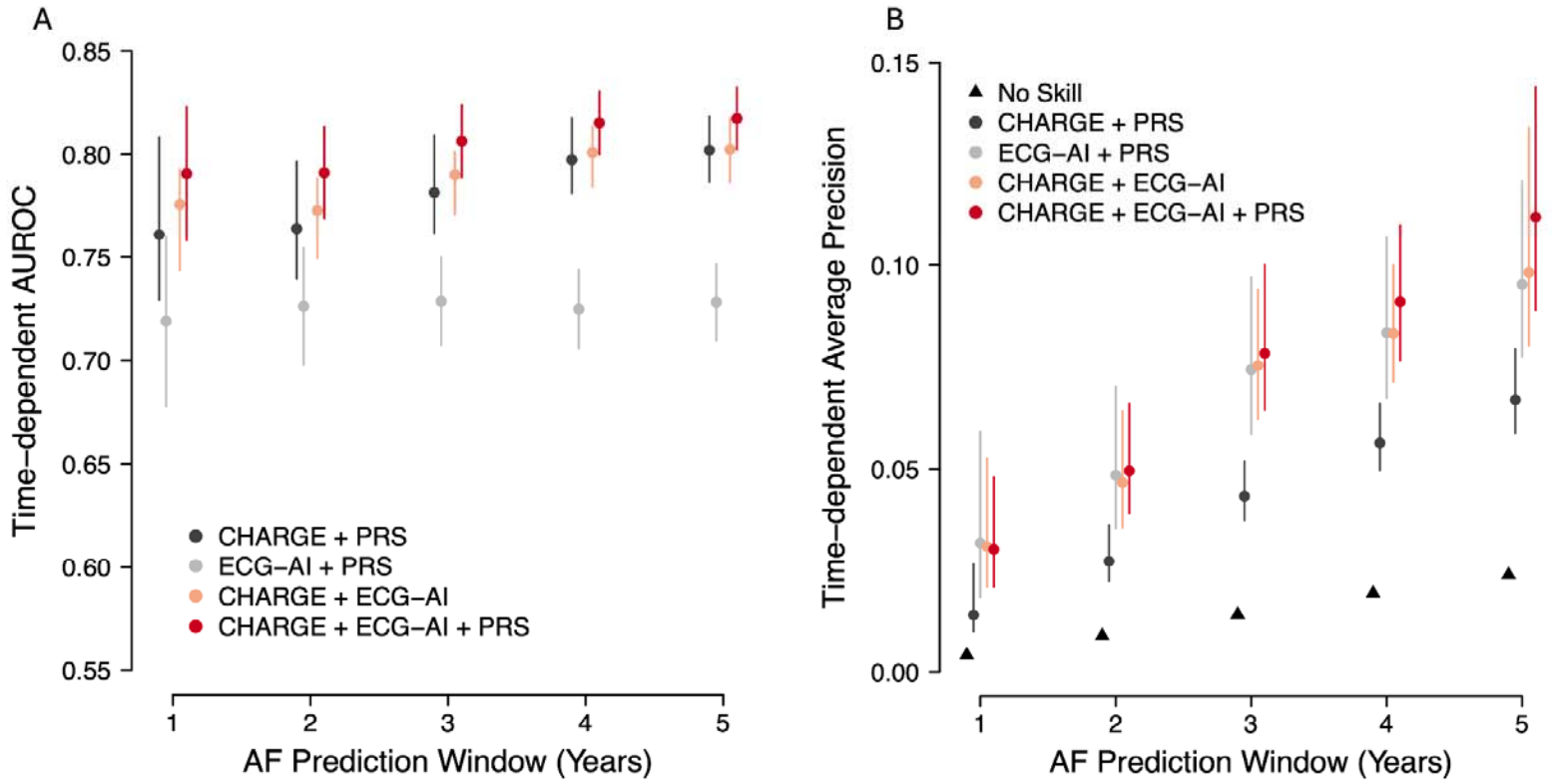
Depicted is an overview of the cohort creation. CHARGE-AF is a clinical risk model for incident atrial fibrillation (AF) risk prediction. ECG-AI is a deep-learning model trained to predict 5-year risk of AF and the polygenic risk score (PRS) is a genetic risk score estimating the risk based on common variant risk from genome-wide association studies of incident AF. Predict-AF3 combines all three risk prediction models for a comprehensive AF risk prediction.

Figure 2. Discrimination of incident atrial fibrillation



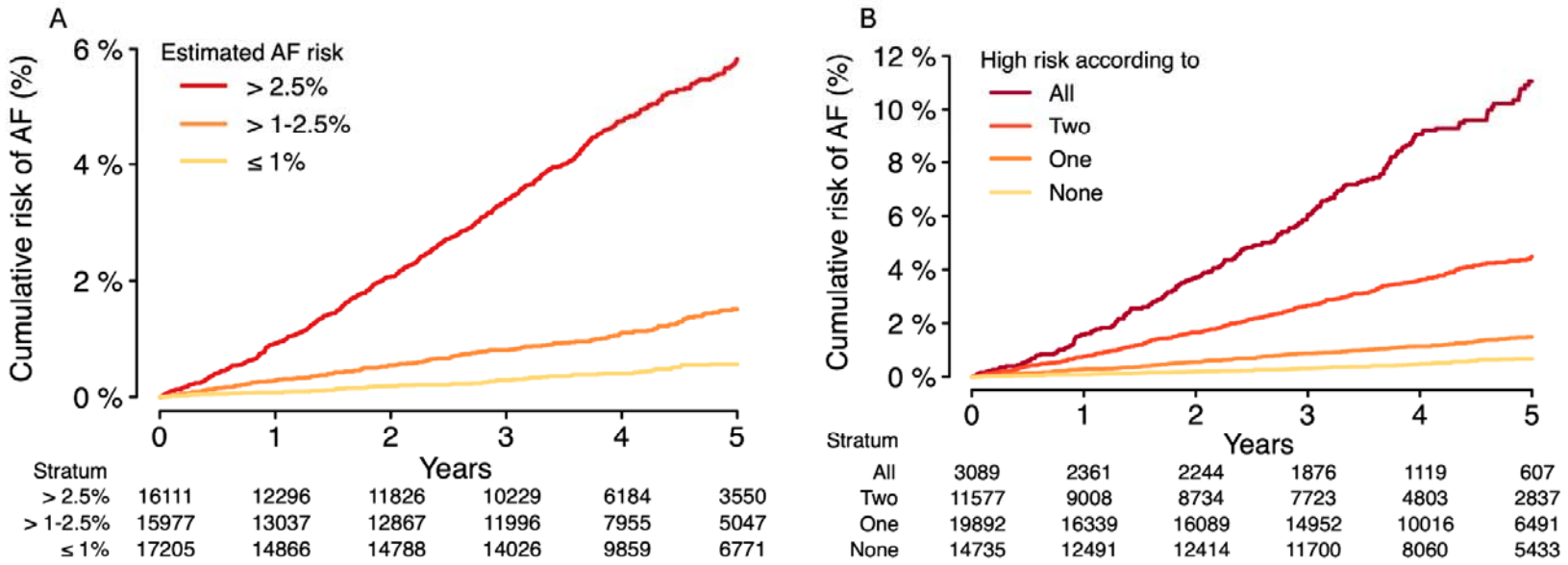
Depicted is model discrimination for a 5-year window of incident atrial fibrillation (AF) using a polygenic risk score (PRS) model (teal), an ECG-derived artificial intelligence (AI) prediction (turquoise) and a model based on the CHARGE-AF score (orange). Panel A depicts discrimination measured using a time-dependent area under the receiver operating characteristic curve (AUROC) while Panel B depicts the time-dependent average precision. A model with no discriminative power is depicted by black triangles.

Figure 3. Discrimination of incident atrial fibrillation in combined models



Depicted is model discrimination for a 5-year window of incident atrial fibrillation (AF) using models a) combining polygenic risk score (PRS) and CHARGE-AF (black), a model combining an ECG-derived artificial intelligence (ECG-AI) prediction and the PRS (grey), a model combining ECG-AI and the CHARGE-AF (orange), and a model combining ECG-AI, CHARGE-AF and the PRS (red). Panel A depicts discrimination measured using a time-dependent area under the receiver operating characteristic curve (AUROC) while Panel B depicts the time-dependent average precision. A model with no discriminative power is depicted by black triangles.

Figure 4. Cumulative risk of incident atrial fibrillation stratified by the combined models



Depicted is the cumulative risk of AF across strata of predicted risk using Predict-AF3. Panel A plots cumulative risk across categories of Predict-AF3 estimated risk (thresholds chosen to approximate tertiles of the risk distribution), and Panel B plots cumulative risk across strata of high risk (i.e., 2-year AF risk >2.5%) by each model component. The number at risk across each stratum over time is depicted below each plot.