

1                   **Artificial intelligence–enabled electrocardiogram for mortality and**  
2                   **cardiovascular risk estimation: An actionable, explainable and biologically**  
3                   **plausible platform**

4   Arunashis Sau<sup>1,2</sup>, Libor Pastika<sup>1</sup>, Ewa Sieliwonczyk<sup>1,3,4</sup>, Konstantinos Patlatzoglou<sup>1</sup>,  
5   Antônio H. Ribeiro<sup>5</sup>, Kathryn A. McGurk<sup>1,3</sup>, Boroumand Zeidaabadi<sup>1</sup>, Henry Zhang<sup>1</sup>,  
6   Krzysztof Macierzanka<sup>1</sup>, Danilo Mandic<sup>6</sup>, Ester Sabino<sup>7</sup>, Luana Giatti<sup>7</sup>, Sandhi M  
7   Barreto<sup>8</sup>, Lidyane do Valle Camelo<sup>8</sup>, Ioanna Tzoulaki<sup>9,10</sup>, Declan P. O’Regan<sup>3</sup>,  
8   Nicholas S. Peters<sup>1,2</sup>, James S. Ware<sup>1,3</sup>, Antonio Luiz P. Ribeiro<sup>11</sup>, Daniel B.  
9   Kramer<sup>1,12</sup>, Jonathan W. Waks<sup>13</sup>, Fu Siong Ng<sup>1,2,14</sup>

10

11   **Affiliations**

- 12           1. National Heart and Lung Institute, Imperial College London, United Kingdom  
13           2. Department of Cardiology, Imperial College Healthcare NHS Trust, London,  
14           United Kingdom  
15           3. MRC Laboratory of Medical Sciences, Imperial College London, London, UK.  
16           4. University of Antwerp and Antwerp University Hospital, Antwerp, Belgium  
17           5. Department of Information Technology, Uppsala University, Uppsala, Sweden  
18           6. Department of Electrical and Electronic Engineering, Imperial College London,  
19           United Kingdom  
20           7. Department of Infectious Diseases, School of Medicine and Institute of  
21           Tropical Medicine, University of São Paulo, São Paulo, Brazil  
22           8. Department of Preventive Medicine, School of Medicine, and Hospital  
23           das Clínicas/EBSERH, Universidade Federal de Minas Gerais  
24           9. Systems Biology, Biomedical Research Foundation, Academy of Athens,  
25           Athens, Greece

1 10. Department of Biostatistics and Epidemiology, School of Public Health,  
2 Imperial College London, London, UK

3 11. Department of Internal Medicine, Faculdade de Medicina, and Telehealth  
4 Center and Cardiology Service, Hospital das Clínicas, Universidade Federal  
5 de Minas Gerais, Belo Horizonte, Brazil

6 12. Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology,  
7 Beth Israel Deaconess Medical Center, Harvard Medical School, Boston MA  
8 USA

9 13. Harvard-Thorndike Electrophysiology Institute, Beth Israel Deaconess Medical  
10 Center, Harvard Medical School, Boston, MA, USA.

11 14. Department of Cardiology, Chelsea and Westminster Hospital NHS  
12 Foundation Trust, London, United Kingdom

13

14 **Disclosures:** JWW was previously on the advisory board for Heartcor solutions LLC,  
15 the remaining authors have no conflicts to declare

16 **Correspondence:**

17 Fu Siong Ng

18 Reader in Cardiac Electrophysiology

19 National Heart and Lung Institute, Imperial College London

20 Hammersmith Campus

21 Du Cane Road

22 London W12 0NN

23 e-mail: [f.ng@imperial.ac.uk](mailto:f.ng@imperial.ac.uk)

24

## 1 **Abstract**

### 2 **Background and Aims**

3 Artificial intelligence-enhanced electrocardiograms (AI-ECG) can be used to predict  
4 risk of future disease and mortality but has not yet been adopted into clinical practice.  
5 Existing model predictions lack actionability at an individual patient level,  
6 explainability and biological plausibility. We sought to address these limitations of  
7 previous AI-ECG approaches by developing the AI-ECG risk estimator (AIRE)  
8 platform.

### 9 **Methods and Results**

10 The AIRE platform was developed in a secondary care dataset of 1,163,401 ECGs  
11 from 189,539 patients, using deep learning with a discrete-time survival model to  
12 create a subject-specific survival curve using a single ECG. Therefore, AIRE predicts  
13 not only risk of mortality, but *time-to-mortality*. AIRE was validated in five diverse,  
14 transnational cohorts from the USA, Brazil and the UK, including volunteers, primary  
15 care and secondary care subjects. AIRE accurately predicts risk of all-cause mortality  
16 (C-index 0.775 (0.773-0.776)), cardiovascular (CV) death 0.832 (0.831-0.834), non-  
17 CV death (0.749 (0.747-0.751)), future ventricular arrhythmia (0.760 (0.756-0.763)),  
18 future atherosclerotic cardiovascular disease (0.696 (0.694-0.698)) and future heart  
19 failure (0.787 (0.785-0.889))). Through phenome- and genome-wide association  
20 studies, we identified candidate biological pathways for the prediction of increased  
21 risk, including changes in cardiac structure and function, and genes associated with  
22 cardiac structure, biological aging and metabolic syndrome.

1 **Conclusion**

2 AIRE is an actionable, explainable and biologically plausible AI-ECG risk estimation  
3 platform that has the potential for use worldwide across a wide range of clinical  
4 contexts for short- and long-term risk estimation.

5

6

## 1 Introduction

2 The electrocardiogram (ECG) has been a fundamental tool in clinical medicine for  
3 over a century. With the recent advent of artificial intelligence (AI), and in particular  
4 deep learning, the potential applications of the ECG have significantly expanded,  
5 including both diagnostic and predictive capabilities (1-3). A major advantage of the  
6 deep learning approach is the ability to extract features relevant to the specific task,  
7 without anchoring on prior human beliefs. Recent studies have demonstrated the  
8 remarkable predictive capabilities of AI-ECG models, not only in terms of predicting  
9 mortality, but also future cardiac diseases (4-7). Novel AI-ECG biomarkers, such as  
10 AI-ECG predicted age and BMI (8, 9), have also been shown to be capable of  
11 capturing information on future risk.

12  
13 However, although many of these existing risk prediction models have good  
14 performance metrics, they have not been integrated into routine clinical care. Existing  
15 mortality prediction models are limited by prediction of survival at a small number of  
16 set time points, rather than providing an individualised survival prediction over time.  
17 Just as the Cox proportional hazards model is superior to logistic regression when  
18 time-to-event information is available (10), survival-based deep learning may  
19 outperform classification-based approaches.

20  
21 Another key limitation of existing models is the lack of information for clinicians on  
22 specific actionable pathways. A high-risk prediction is unhelpful to a clinician if there  
23 is no accompanying information on how to affect the survival trajectory of their  
24 patient. To make AI-ECG predictions more actionable, it is essential to consider not

1 only time-to-event predictions, but also specific predictions for diseases with  
2 established preventive and disease modifying treatments.

3

4 Furthermore, the adoption of AI into clinical practice is significantly limited by  
5 concerns regarding explainability and biological plausibility. Just as knowledge of the  
6 mechanisms of action of drugs are important for physicians to have confidence in  
7 their application, biological plausibility of AI predictions ensures their credibility and  
8 acceptance.

9

10 To address these limitations of existing risk prediction models, we aimed to develop  
11 and perform transnational validation on an AI-ECG risk prediction platform that is not  
12 only accurate, but also actionable, explainable, and biologically plausible.

13

1

## 2 **Methods**

3 In this study we first developed the AI-ECG risk estimation (**AIRE**) model for  
4 prediction of all-cause mortality. We subsequently developed seven additional  
5 submodels. The eight models together are referred to as the **AIRE platform**. A  
6 model development and validation flow chart is shown in **Figure 1**.

7

## 8 **Ethical approvals**

9 This study complies with all relevant ethical regulations, further details are provided  
10 in the **Supplementary Methods**.

11

## 12 **Cohorts**

13 We studied five cohorts, briefly the Beth Israel Deaconess Medical Center (BIDMC)  
14 cohort is a secondary care dataset comprised of routinely collected data from,  
15 Boston, USA. The São Paulo-Minas Gerais Tropical Medicine Research Center  
16 (SaMi-Trop) is a cohort of patients with chronic Chagas cardiomyopathy (11). The  
17 Clinical Outcomes in Digital Electrocardiography (CODE) cohort is a Brazilian  
18 database of ECGs recorded in primary care (8) containing ECGs of both 10s and 7s  
19 duration. The subset of this dataset with only 10s ECGs is referred to as CODE-10s.  
20 The Longitudinal Study of Adult Health (ELSA-Brasil) cohort consists of Brazilian  
21 public servants (12). The UK Biobank (UKB) is longitudinal study of volunteers (13).  
22 Further details are provided in the **Supplementary Methods**.

23

24

25

## 1 **ECG pre-processing**

2 12-lead ECGs were pre-processed with a bandpass filter 0.5 to 100Hz, a notch filter  
3 at 60Hz and re-sampling to 400Hz. Zero padding was added to make the input shape  
4 a power of 2. This resulted in 4096 samples for each lead for a 10s recording (4000  
5 samples + 48 zeros at the start and end), that was used as input to the neural  
6 network model. As leads III, aVL, aVR, aVF are linear combinations of leads I and II,  
7 these leads were not used for model development or evaluation. Therefore, the final  
8 input shape of a single ECG was 4096 x 8.

9

## 10 **AI-ECG risk estimation platform development**

11 We developed the **AIRE platform** using the BIDMC cohort as the derivation dataset.  
12 The data was split at a ratio of 50/10/40% for training, validation and internal test,  
13 respectively. For mortality end-points, ECGs without paired life status at 30 days  
14 were excluded. Data was split by patient ID stratified by presence of ECGs with  
15 paired 5-year life status. To prevent data leakage, a single subject could have ECGs  
16 assigned to only one of training/validation/testing datasets. We used a previously  
17 described convolutional neural network architecture based on residual blocks (14)  
18 and adapted the final layer to accommodate a discrete-time survival model (15). The  
19 discrete-time survival approach allows the model to account for both time to outcome  
20 (mortality) and censorship (i.e., loss to follow up). Unlike other models trained to  
21 predict mortality at one, or a small number of time points, our model predicts  
22 outcomes at numerous timepoints up to 10 years and accounts for right censored  
23 data during model training.

24



1 The lead I model, **AIRE-1L**, was developed using the same methodology above,  
2 using the same BIDMC data split but using just lead I as the model input. The  
3 Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or  
4 Diagnosis Checklist for Prediction Model Development and Validation was followed  
5 (16).

6

### 7 **Model fine-tuning for primary care population**

8 Using the CODE dataset, we finetuned the model to be more representative of a  
9 primary care population (**AIRE-primary care**) with a much lower risk of adverse  
10 events. We used 75% of the CODE dataset for finetuning, with 5 % as a validation  
11 set. The final 20% was used for internal validation of **AIRE-primary care**.

12

### 13 **Model fine-tuning for other endpoints**

14 We also developed five other subsequent models by fine-tuning the AIRE model  
15 separately for cardiovascular (CV) death, non-CV death, ventricular arrhythmia (VA),  
16 atherosclerotic cardiovascular disease (ASCVD) and heart failure (HF) in the BIDMC  
17 dataset. These models were named **AIRE-CV death**, **AIRE-NCV death**, **AIRE-VA**,  
18 **AIRE-ASCVD** and **AIRE-HF** respectively. The same splits were used as for training  
19 the original model. Fine-tuning was performed by loading the previous model and  
20 training using a low learning rate without freezing any layers. Internal validation and  
21 external validation datasets are shown in [Figure 1](#).

22

23 For non-mortality endpoints (ASCVD, VA, HF), subjects were coded as having  
24 prevalent disease, future disease or neither at the time of the ECG. Although the goal  
25 of these models was to predict future events, we hypothesised that information on

1 prevalent disease would be helpful for model training. In order to include prevalent  
2 disease in the discrete-time survival model, we encoded prevalent disease at the first  
3 timepoint in the discrete-time survival labels. When evaluating model performance,  
4 subjects with prevalent disease were excluded.

5

## 6 **Comparison of AIRE with other models**

7 We compared **AIRE and AIRE-primary care** to the previously described AI-ECG  
8 predicted age (8). AI-ECG predicted age values are publicly available for CODE15  
9 and SaMi-Trop. The remaining 85% of CODE was used to develop the AI-ECG age  
10 model, and therefore was not used. For ELSA-Brasil, the model weights were  
11 downloaded (17) and used to derive AI-ECG predicted age values.

12

13 We also compared **AIRE-primary care** and **AIRE-ASCVD** to the recently described  
14 SEER (Stanford Estimator of ECG Risk), which had a similar goal of predicting  
15 cardiovascular mortality and ASCVD (7). The model code and weights were  
16 downloaded (18) and performance evaluated in the UKB, as this was a dataset  
17 external to **AIRE-primary care, AIRE-ASCVD** and SEER, with cause of death and  
18 ASCVD event data available.

19

## 20 **Survival and statistical analyses**

21 In the test set, we generated predictions of all ECGs for the primary analyses of all-  
22 cause mortality. Sensitivity analyses including a single random ECG per subject were  
23 also performed. Model performance was reported using the C-index, and time-  
24 dependent AUROC. For analyses requiring a single predictor value (C-index and  
25 AUROC), the probability of survival at 5 years was used. Risk quartiles were defined

1 (low, intermediate-low, intermediate-high, high) based on values in the validation set.  
2 Given the diverse populations and event rates evaluated, risk quartiles were  
3 redefined in each dataset. In each case, where categorical risk levels were required,  
4 5% of the dataset was used to define the quartiles and evaluation was performed in  
5 the remaining 95%. Kaplan-Meier curves comparing the risk quartiles were plotted  
6 and statistical significance assessed using the log rank test.

7

8 Cox models were fit using the test dataset comparing demographics, clinical  
9 variables, imaging parameters and **AIRE platform** predictions. For the Cox models  
10 incorporating **AIRE platform** predictions, all model outputs (i.e., predicted  
11 probabilities of death at each timepoint) were used as inputs, as well as age, sex,  
12 heart rate, PR interval, QRS duration and QTc interval. These models are designated  
13 AIRE-Cox for the AIRE model and AIRE-CV Death-Cox, AIRE-ASCVD-Cox, AIRE-  
14 VA-Cox, AIRE-HF-Cox for the other models. Complete case analysis was used,  
15 therefore no variables were imputed. Recent work suggests virtually all real-world  
16 clinical datasets will violate the proportional hazards assumptions if sufficiently  
17 powered and that statistical tests for the proportional hazards assumption may be  
18 unnecessary (19). In line with these recommendations, the proportional hazards  
19 assumption was not evaluated and the hazard ratio from our Cox models should be  
20 interpreted as a weighted average of the true hazard ratios over the follow-up period.  
21 Nested Cox models were compared with the Likelihood Ratio test, while non-nested  
22 Cox models were compared with the Partial Likelihood Ratio test (20). Statistical  
23 analyses were performed with R 4.2.0 statistical package (R Core Team, Vienna,  
24 Austria) or Python (version 3.9).

25

## 1 **Diagnostic and imaging data**

2 ICD-9 and ICD-10 codes were used to define presence/absence of disease in the  
3 BIDMC and UKB cohorts. Cardiovascular death in the BIDMC cohort was defined as  
4 mortality occurring within 30 days of a diagnostic code for acute myocardial  
5 infarction, ischaemic stroke, intracranial haemorrhage, sudden cardiac death, or  
6 heart failure as previously described (7, 21). In the UKB, cause of death was  
7 ascertained based on the ICD10 code stated as the primary cause of death.  
8 Diagnostic codes were not available in the SaMi-Trop, CODE and ELSA-Brasil  
9 datasets. Echocardiograms within 60 days of an ECG were linked and used for  
10 analyses incorporating echocardiographic parameters. The pooled cohort equation  
11 was calculated using the PooledCohort R package. Medication usage was not  
12 available in the BIDMC cohort, therefore ICD9 and ICD10 codes consistent with a  
13 diagnosis of hypertension were used to code for antihypertensive medication use for  
14 calculation of the pooled cohort equation. Blood results and blood pressure (BP)  
15 readings taken within 180 days of the ECG were averaged. Sensitivity analyses were  
16 performed using 90 days and 30 days results.

17

18 We investigated the performance of AIRE in the high risk disease groups of severe  
19 aortic stenosis and primary pulmonary hypertension. Severe aortic stenosis was  
20 defined based on the reported overall severity on echocardiography reports (i.e. a  
21 subjective overall assessment by the clinician undertaking the echocardiogram).  
22 Primary pulmonary hypertension was defined using ICD9 and ICD10 codes. No fine-  
23 tuning was performed when evaluating the performance of **AIRE** in the severe aortic  
24 stenosis and primary pulmonary hypertension disease groups.

25

## 1 **Normal ECG definition**

2 We evaluated the performance of AIRE in clinician reported normal ECGs. In the  
3 BIDMC dataset a subset of ECGs had Cardiologist reports. Normal ECGs in BIDMC  
4 were determined by searching for 'normal ecg' in the free text reports, a whole word  
5 match was required in order to exclude 'abnormal ecg'. ECGs with the phrase  
6 'otherwise' were also excluded from the normal definition. Additionally we filtered by  
7 heart rate (60-100 bpm), PR interval (less than 200ms), QRS duration (less than  
8 120ms) and QTc interval (less than 470ms). Normal ECGs in ELSA-Brasil and CODE  
9 were defined as previously described (17).

10

11

12

## 13 **Explainability**

14 In order to understand the ECG morphologies associated with predicted survival we  
15 used two approaches. Median beats were extracted using the BRAVEHEART ECG  
16 analysis software as previously described (22).

17

18 First, we trained a variational autoencoder (VAE) as previously described (23) using  
19 median ECG beats. Further details in [Supplementary Methods](#). In preliminary  
20 analyses, models based on only the VAE latent features were found to be inferior to  
21 the supervised deep learning approach described above, therefore the VAE was  
22 used for explainability only, and not used for **AIRE** model training or any of the  
23 prediction models described in this manuscript. VAE latent features were input into a  
24 linear regression with predicted survival as the output. The top 3 most important  
25 features as assessed by the t-value were visualised by latent traversal (23).

1

2 Second, using the median beats we calculated the average waveform from the  
3 10,000 ECGs with the lowest and highest **AIRE** predicted mortality. The mean and  
4 standard deviation of these waveforms was then plotted.

5

## 6 **PheWAS**

7 To better understand the biology underlying **AIRE** and to explore the detailed  
8 phenogroup associations, we performed phenome-wide association studies  
9 (PheWAS). We performed PheWAS analysis in the UKB, which contains data from  
10 over 3000 phenotypes derived from patient measurements, surveys, and  
11 investigations. Univariate correlation was performed to investigate the association  
12 between ECG phenogroup and phenotypes, adjusted for age, sex and age<sup>2</sup>. We  
13 additionally investigated the association of predicted survival with continuous echo  
14 traits in the BIDMC dataset. Left ventricular trabeculation was calculated as  
15 previously described (24). Deep learning-derived brain age was calculated as  
16 previously described (25).

17

## 18 **GWAS**

19 To identify genetic associations with the ECG phenogroups, we performed a ge-  
20 nome-wide association study (GWAS) in the UKB. As the predicted survival trait was  
21 skewed, the data were normalized by rank-based inverse normal transform prior to  
22 the analysis. The GWAS analysis was adjusted for the following covariates: age at  
23 imaging visit, sex, height, body mass index (BMI), imaging assessment centre and  
24 the first 10 genetic principal components.

25

26 Further methods are described in the [Supplementary Methods](#).



1 **Results**

2

3 **AIRE accurately predicts mortality across diverse timepoints**

4 In the BIDMC cohort, 1,163,401 ECGs were available from 189,539 subjects. Mean  
5 follow-up period was  $5.46 \pm 5.81$  years on a per ECG basis, 3.41 (4.08) years taking a  
6 random ECG per subject. 34,938 (18.4%) subjects died during follow-up (**Table 1**).

7

8 **AIRE** produces subject-specific survival curves from only a single ECG and can  
9 predict *time-to-death*. **Figure 2A** and **2B** demonstrate two representative subject-  
10 specific survival curves in patients who died during follow up and two curves from  
11 subjects who survived through the follow up period, each generated by **AIRE** from a  
12 single ECG. **Figure 2C** demonstrates the evolution of AIRE-predicted survival based  
13 on multiple ECGs performed over several years of follow up. ECGs nearer to the  
14 subject's time of death show falling survival probabilities, particularly shortly before  
15 the subject's death.

16

17 In the hold out test set, **AIRE** predicted all-cause mortality with a concordance-index  
18 of 0.775 (0.773-0.776), further results are reported in the **Supplementary Results**.

19 Using the predicted probability of survival at 5 years in the validation set, quartiles of  
20 risk (low, intermediate-low, intermediate-high and high) were determined. **Figure 3A**  
21 shows the marked separation of survival curves of these quartiles in the test set.

22 **Table S1** shows age and sex adjusted hazard ratios for high risk vs low risk quartile  
23 for all cohorts.

24



1 A subset of ECGs had cardiologist interpretations available. These were more recent  
2 ECGs with shorter follow up durations compared to the whole cohort. When  
3 considering only ECGs labelled as normal by cardiologists, there remained a  
4 significant difference in mortality between high-risk and low-risk subjects (**Figure 3B**).  
5 Importantly, AIRE had similar performance in both men and women and in major  
6 ethnic groups (**Figure 3C** and **Table S2**).

7

### 8 **AIRE is superior to demographic data and traditional risk factors for mortality** 9 **prediction**

10 We compared the ability of **AIRE** to predict mortality, against using demographic  
11 data, risk factors and risk scores in the BIDMC test set (**Figure 3D**). **AIRE-Cox** had a  
12 significantly higher C-index than all other parameters combined (0.794 (0.792-0.795)  
13 vs 0.759 (0.758-0.761),  $p < 0.0001$ ). Numerical results for all Cox models are shown  
14 in **Table S3**.

15

16 **AIRE-CV Death** predicted CV death with a C-index of 0.832 (0.831-0.834). In the  
17 BIDMC test set, **AIRE-CV Death-Cox** had a significantly higher C-index for prediction  
18 of CV death than all other parameters combined (0.844 (0.839-0.849) vs 0.795  
19 (0.789-0.801),  $p < 0.0001$ ) (**Figure 3E**). Finally, **AIRE-Non-CV Death** predicted non-  
20 CV death with a C-index of 0.749 (0.747-0.751). **Figure S2** shows sensitivity  
21 analyses using a single random ECG per subject.

22

## 1 **AIRE predicts mortality in transnational external datasets**

2 First, we evaluated the performance of **AIRE** in the SaMi-Trop cohort of patients with  
3 chronic Chagas cardiomyopathy (11). Dataset demographics for all cohorts are  
4 shown in **Table 1**. The C-index was 0.773 (0.733-0.813, **Figure 4A**).

5

6 The CODE cohort is a Brazilian database of ECGs recorded in primary care (8). The  
7 dataset contains both 10s and 7s duration ECGs. We first evaluated the performance  
8 of **AIRE** without any fine-tuning. As **AIRE** was trained exclusively on 10s ECGs, we  
9 evaluated the model on the 10s subset (CODE-10s), **AIRE** had a C-index of 0.762  
10 (0.759-0.765) for all-cause mortality prediction. **AIRE-primary care** more accurately  
11 predicted mortality with an improved C-index of 0.802 (95% CI 0.799-0.805, **Figure**  
12 **4B**). When considering only the ECGs labelled as normal (26766 ECGs from 21897  
13 subjects), there remained a significant difference in mortality between high-risk and  
14 low-risk subjects based on model predictions (**Table S1**).

15

16 Further evaluation of **AIRE-primary care** was performed in another independent  
17 external dataset, ELSA-Brasil (n = 13739) a volunteer cohort of civil servants from  
18 Brazil (12)). The C-index was 0.713 (0.691-0.735 **Figure 4C**). Again, when  
19 considering normal ECGs only there was a significant difference in mortality between  
20 high and low risk subjects (**Table S1**).

21

22 Finally, we additionally evaluated the performance of **AIRE-primary care** in the UK  
23 Biobank, a relatively healthy volunteer population (n = 42386) with only 526 (1.2%)  
24 deaths during follow-up. The C-index was 0.638 (0.608-0.668 (**Figure 4D**)) for all-  
25 cause mortality. As cause of death was available in the UKB, we additionally

1 examined the ability of **AIRE-primary care** (which was trained for all-cause mortality)  
2 to predict CV death. C-index for CV death was 0.695 (0.636-0.754)

3

4 Hughes et al recently reported SEER (Stanford Estimator of ECG Risk) with the  
5 similar goal of predicting cardiovascular mortality (7). The model code and weights  
6 were downloaded, and performance evaluated in the UKB, as this was a dataset  
7 external to both **AIRE-primary care** and SEER, with cause of death available. **AIRE-**  
8 **primary care** was superior to SEER at predicting CV death in the UKB (SEER C-  
9 index 0.572 (0.514-0.630) p values for comparison to **AIRE-primary care** <0.001).

10

#### 11 **AIRE outperforms AI-ECG models of biological aging**

12 AI-ECG predicted age and the difference between AI-ECG predicted age and  
13 chronological age (“delta-age”) have been shown to be a better markers of biological  
14 aging than chronological age (8) . We compared **AIRE/AIRE-primary care**  
15 predictions to AI-ECG predicted age. For the subsequent analyses, **AIRE** was used  
16 for SaMi-Trop and **AIRE-primary care** was used for CODE and ELSA-Brasil. We  
17 found **AIRE/AIRE-primary care**-predicted survival inversely correlated with AI-ECG  
18 age (CODE:  $r = -0.595$ ,  $p < 0.0001$ , ELSA:  $r = -0.449$ ,  $p < 0.0001$ , SaMi-Trop  $-0.479$ ,  
19  $p < 0.0001$ ) and, to a lesser extent, chronological age (CODE:  $r = -0.499$ , ELSA: -  
20  $0.293$ , SaMi-Trop:  $-0.275$ ,  $p < 0.0001$  for all).

21

22 We compared the predictive capability of **AIRE/AIRE-primary care** with AI-ECG  
23 predicted age and delta-age in the CODE15 (a 15% subset of CODE), ELSA-Brasil  
24 and SaMi-Trop datasets. In age- and sex-adjusted models, **AIRE/AIRE-primary care**  
25 had a significantly higher C-index compared to AI-ECG predicted age or delta age in

1 all three cohorts (CODE: 0.823 (0.813-0.833) vs 0.806 (0.796-0.816), ELSA-Brasil:  
2 0.754 (0.734-0.774) vs 0.738 (0.718-0.758), SaMi-Trop 0.788 (0.749-0.827) vs 0.715  
3 (0.672-0.758);  $p < 0.001$  for all).

4

#### 5 **AIRE performance in high-risk disease groups**

6 For model predictions of mortality to be clinically useful, there needs to be specific  
7 interventions available to alter the trajectory of patients. Using the BIDMC test set,  
8 we investigated the performance of **AIRE** in two high-risk disease groups with  
9 existing risk stratification strategies and effective interventions. The data for high-risk  
10 disease group analysis was not available for the other cohorts.

11

12 For aortic stenosis, **AIRE** accurately predicted all-cause mortality, C-index: 0.701  
13 (0.681- 0.721, 1293 ECGs from 449 subjects). **AIRE-Cox** had a significantly higher  
14 C-index compared to all other parameters combined, **Figure 5A**, C-index 0.709  
15 (0.688-0.730) vs 0.683 (0.661-0.705),  $p < 0.001$ .

16

17 Similarly, for primary pulmonary hypertension, **AIRE** accurately predicted all-cause  
18 mortality, C-index 0.731 (0.724-0.738, 11741 ECGs from 789 subjects). In a subset  
19 with available recent echocardiograms (602 ECGs in 212 subjects, we evaluated  
20 **AIRE** in comparison and addition to other risk parameters. **AIRE-Cox** was superior  
21 all other parameters combined (C-index 0.753 (0.726-0.779) vs 0.700 (0.670-0.730),  
22  $p < 0.005$ , **Figure 5B**).

23

1 **Actionable predictions: (1) ASCVD**

2 **AIRE-ASCVD** was able to predict future ASCVD in subjects without known ASCVD  
3 (C-index 0.696 (0.694-0.698) n = 227588 ECGs from 56598 subjects). We externally  
4 validated these findings in the UKB Biobank (UKB); a healthy, volunteer population.  
5 **AIRE-ASCVD** had reduced performance in predicting ASCVD, C-index 0.643 (0.624-  
6 0.662) but was significantly better than the SEER model (7) (SEER C-index 0.547  
7 (0.527-0.567), p <0.0001 for comparison with AIRE).

8

9 We compared **AIRE-ASCVD** to other risk parameters, including the pooled cohort  
10 equation (PCE) and ASCVD risk factors, in a subset of outpatients (4580 ECGs from  
11 2926 subjects) in the BIDMC test set with appropriate available data. **AIRE-ASCVD-**  
12 **Cox** had a significantly higher C-index than all other factors combined (0.679 (0.651-  
13 0.708) vs 0.642 (0.613-0.672), **Figure 5C**, p < 0.00001). **Figure S3** shows sensitivity  
14 analyses using a single random ECG per subject for all three actionable prediction  
15 endpoints.

16

17 **Actionable predictions: (2) Ventricular arrhythmia**

18 **AIRE-VA** was able to accurately predict future VA (C-index 0.760 (0.756-0.763) n =  
19 393203 ECGs from 62443 subjects) in subjects without a previous history of VA.  
20 When compared to other conventional risk parameters, including ECG parameters  
21 and left ventricular ejection fraction, **AIRE-VA-Cox**, had a significantly higher C-index  
22 than all other factors combined (0.732 (0.723-0.741) vs 0.700 (0.692-0.709), **Figure**  
23 **5D**, p < 0.0001). The **Supplementary Results** describe performance in subgroups of  
24 LVEF <50% and dilated cardiomyopathy.

25

1 In the UKB, (n = 34400 with both ECG and cardiac magnetic resonance imaging  
2 (CMR) and without previous VA, 44 events), **AIRE-VA** had similar performance in  
3 predicting first occurrence of VA 0.719 (0.635-0.803) and performed at least  
4 equivalent to LVEF from cardiac magnetic resonance imaging (CMR) (C-index 0.595  
5 (0.494-0.697), p for comparison 0.11).

### 7 **Actionable predictions: (3) Future heart failure**

8 **AIRE-HF** was able to accurately predict future HF in subjects without a previous  
9 history of HF (C-index 0.787 (0.785-0.889) n = 310200, from 61747 unique subjects).  
10 In a subset of patients with the available data, we compared **AIRE-HF-Cox** to other  
11 conventional risk parameters in Cox models (36486 ECGs from 12288 subjects),  
12 including HF risk factors identified in the Atherosclerotic Risk in Communities (ARIC)  
13 study (26). **AIRE-HF-Cox**, had a significantly higher C-index than all other factors  
14 combined (0.761 (0.755-0.767) vs 0.716 (0.710-0.722), **Figure 5E**, p < 0.00001). In  
15 the external validation cohort, UKB, **AIRE-HF-Cox** had similar performance at  
16 predicting future HF (C-index 0.768 (0.733 -0.802).

### 18 **Single-lead ECG model performance**

19 To evaluate if **AIRE** could be potentially used in wearable devices or inpatient  
20 telemetry units, we trained a single-lead version of **AIRE** using lead I only (**AIRE-1L**).  
21 The performance of **AIRE-1L** (C-index 0.751 (0.750-752) **Figure S4**) was only  
22 slightly inferior in discrimination compared to the 8-lead **AIRE** model.

23

## 1 **Explainable ECG morphology associates with adverse prognosis**

2 Using a variational autoencoder, we found features of QRS morphology, particularly  
3 broader and more left bundle branch block morphologies, inverted and biphasic T  
4 waves as well as ST segment changes were identified as the most significant  
5 morphological features associated with high predicted mortality (**Figure 6A**).

6

7 In a second approach, using median beats the BIDMC test set we found poor  
8 precordial R wave progression, low QRS amplitude and T wave flattening/inversion  
9 as important features in **AIRE**-predicted survival (**Figure 6B**).

10

## 11 **Biological plausibility: Genetic associations of AIRE-predicted survival**

12 To investigate the underlying genetic associations with **AIRE**-predicted survival, we  
13 performed a genome-wide association study (**Figure 7A**). We found significant loci  
14 adjacent to *TBX3*, *VGLL2*, *CCDC91* and *KCNQ1*. 8% of the total phenotypic variance  
15 in predicted survival was caused by the additive effects of genetic variation. *TBX3*  
16 has been associated with blood pressure (27), ECG morphology (28), myocardial  
17 mass (29) and trabecular development (24). *VGLL2* has been associated with ECG  
18 parameters, blood pressure, atrial fibrillation and BMI (27, 28, 30, 31). *KCNQ1*  
19 (Potassium Voltage-Gated Channel Subfamily Q Member 1) encodes  $K_v7.1$  and is  
20 most well-known for its associated with Long Qt Syndrome 1 and Jervell And Lange-  
21 Nielsen Syndrome 1 and QT interval (32) but is also associated with metabolic  
22 syndrome phenotypes (33-35). Finally, *CCDC91* (Coiled-Coil Domain Containing 91)  
23 associates with BMI (33), a variant associated with the ECG has been previously  
24 described in this region (28).

25

## 1 **Biological plausibility: Phenotypic associations of AIRE-predicted survival**

2 In order to investigate the biological associations with **AIRE**-predicted mortality, we  
3 performed a PheWAS in the UK Biobank, **Figure 7B**. In particular, CMR associations  
4 included reduced left ventricular (LV) ejection fraction (EF), more positive, i.e.,  
5 abnormal, global longitudinal strain, increased LV mass and increased left atrial size,  
6 which were correlated with reduced **AIRE**-predicted survival. We additionally  
7 specifically examined the association between predicted survival and left ventricular  
8 trabeculation and found a significant negative correlation (**Table S5**). Important  
9 lifestyle risk factors included smoking status and alcohol intake which were both  
10 correlated with reduced predicted survival. Significant physical measures included  
11 systolic blood pressure, which was negatively correlated with predicted survival.

12

13 In BIDMC echocardiographic analyses (**Figure 7C**), we found LVEF was positively  
14 correlated with predicted survival while LA volume and surrogate measures of  
15 pulmonary pressure (TR velocity) and right ventricular diameter were negatively  
16 correlated.

17

18 Finally, investigation of non-cardiac imaging phenotypes identified associations with  
19 multiple multi-model brain imaging phenotypes including the total volume of white  
20 matter hyperintensities and deep learning-derived brain age (25) (**Table S6**).

21 Other non-cardiac imaging associations included predicted-survival inversely  
22 correlating with carotid intimal-media thickness, which is a marker of overall  
23 atherosclerotic burden (36).

24



## 1 **Discussion**

2 We describe, for the first time, an actionable, explainable, and biologically plausible  
3 mortality and risk prediction AI-ECG platform of eight AI-ECG models. Importantly,  
4 our platform was externally validated across ethnically and demographically diverse  
5 transnational cohorts.

6

### 7 **Importance of time-to-event for actionability**

8 This paper substantially extends the work of others on mortality protection using the  
9 ECG. Raghunath et al previously described the use of deep learning for mortality  
10 prediction (37), while Sun et al more recently built upon this work (6). Our study has  
11 several significant differences from these previous publications. Firstly, the use of a  
12 survival neural network architecture, provides our model with the ability to predict  
13 time of death without being constrained to a small number of time points.  
14 Additionally, this allows us to use training data from subjects that were censored, for  
15 whom the time for death was not known. Specifically, we created a model that  
16 provides an individualised survival curve based on a single ECG. The ability to  
17 identify patients at risk of short-term mortality versus those at risk of long-term  
18 mortality is an important distinction. For example, patients at risk of short-term  
19 mortality may benefit from early escalation to intensive care settings, while those at  
20 risk of long-term mortality may need more detailed outpatient, follow-up or aggressive  
21 medical or interventional treatment.

22

23 Furthermore, by comparing our model to existing clinical risk factors and  
24 demographic information we have demonstrated the significant additive value of our  
25 model beyond traditional approaches. Finally, we performed external validation

1 across diverse populations, demonstrating the wide applicability of our model  
2 platform.

3

#### 4 **Actionable risk prediction in high-risk disease groups**

5 By predicting risk of death in high-risk disease groups, we were able to provide  
6 potentially actionable information that could guide treatment decisions. We described  
7 how **AIRE** can be used to predict the risk of death in severe aortic stenosis and  
8 primary pulmonary hypertension. Severe aortic stenosis has a high mortality if left  
9 untreated. Current clinical guidelines in general advocate intervention based on  
10 echocardiographic measures and the patient's symptoms (38, 39). However, there is  
11 increasing evidence that earlier intervention in asymptomatic individuals at high-risk  
12 of death may be of benefit, and there are randomised control trials currently  
13 investigating this approach (40, 41). Primary pulmonary hypertension also has a  
14 significant mortality rate (42). Medical therapy is available, and transplantation may  
15 be undertaken (43), therefore accurate risk stratification is paramount in guiding  
16 treatment selection for pulmonary hypertension patients. Mortality prediction using  
17 **AIRE** may be helpful in guiding treatments in both these conditions.

18

#### 19 **Event-specific risk prediction**

20 Using disease specific models, we showed the **AIRE platform** can also predict future  
21 cardiovascular events such as ASCVD, HF and VA, in addition to predicting mortality.  
22 ASCVD prediction is currently used extensively in international guidelines for  
23 decision-making around lipid lowering therapies. A 10-year ASCVD is calculated and  
24 can be used to consider initiation of primary prevention statin therapy (44). In order to  
25 assess ASCVD risk, various scoring systems are used, including the PCE, SCORE2

1 and QRISK3 (44-46) . The ECG, despite being a cheap and readily available  
2 investigation, does not feature in any of these scoring systems. In this study, we have  
3 shown that **AIRE-ASCVD** can provide additional information that could improve  
4 ASCVD risk prediction and may be superior or at least additive to the existing PCE.

5  
6 Our analyses are additive to those of Hughes et al (7), who recently describe SEER,  
7 an AI-ECG model for the prediction of CV death. **AIRE-primary care** and **AIRE-**  
8 **ASCVD** was superior to SEER for CV death and ASCVD prediction in the UKB.  
9 Additionally, the time-to-event nature of our platform provides additive actionable  
10 information as discussed above (7).

11  
12 Similarly, predicting future VA is a particularly important endpoint, as there are clear  
13 preventative, and therapeutic options. For example, in-hospital predictions of short-  
14 term high risk of VA could be used to guide the need for cardiac monitoring and  
15 potentially even prophylactic use of antiarrhythmics. Medium-term risk could be  
16 managed with a wearable cardioverter defibrillator (47), while patients at longer-term  
17 risk of ventricular arrhythmias may benefit from primary prevention implantable  
18 cardioverter defibrillator implantation. Current guidelines advocate LVEF as the  
19 primary factor in determining eligibility for a primary prevention ICD. In our study, we  
20 demonstrated that **AIRE-VA** is a better predictor of future VT/VF than LVEF and  
21 could therefore potentially be incorporated into this decision-making paradigm.

22  
23 Lastly, predicting future heart failure is important given the high number of unplanned  
24 hospital admissions due to undiagnosed heart failure (48). Early clinical assessment,

1 echocardiography and institution of appropriate therapies may reduce adverse  
2 events, particularly in HF with reduced ejection fraction (HFrEF) (49).

3

#### 4 **Single-lead ECG applications**

5 Extending risk prediction to the single-lead ECG is particularly important given the  
6 rapidly increasing number of single lead devices, including consumer products (50) .

7 Others have shown the applicability of AI-ECG algorithms to consumer single-lead  
8 products (51). Our study highlights the excellent performance of **AIRE** at mortality

9 prediction on only a single lead. This could be particularly applicable for inpatient

10 cardiac monitoring, or for remote monitoring in the outpatient setting using wearable

11 devices, for example in patients with chronic diseases such as heart failure, where

12 high-risk predictions could trigger pre-emptive treatments to prevent hospital

13 admissions.

14

#### 15 **Explainability and biological plausibility**

16 A major advantage of the deep learning approach is the ability to extract features

17 relevant to the specific task, without anchoring on prior beliefs. However, a significant

18 challenge is explainability of the model predictions. A significant barrier to the

19 adoption of AI tools in clinical practice is physician reluctance to adopt technologies

20 that are often “black boxes” (52). Explainability and biological plausibility is therefore

21 key to improving clinician trust in AI. In this study using multiple approaches, we have

22 explored the underlying biology behind **AIRE** risk predictions. Our ECG morphology

23 findings are in line with prior studies that highlight these features as being

24 prognostically importantly (53-55). Importantly, the significant predictive value of

1 **AIRE**, even in normal ECGs, shows how deep learning can additionally make use of  
2 ECG morphological features that clinicians deem to be normal.

3

4 The accuracy of **AIRE** at mortality prediction raises the important question as to  
5 which biological pathways are being identified in these predictions. Using cardiac  
6 imaging data in two distinct datasets, we identified associations with cardiac chamber  
7 structure, function and measures of pulmonary pressure, that may be reflected in the  
8 ECG (56). We also identified multiple established factors, whose effects maybe  
9 reflected directly or indirectly in the ECG, including blood pressure, smoking and  
10 alcohol intake.

11

12 Previous work has described the importance of ischaemic heart disease on brain  
13 aging and outcomes (57). Our work has identified significant associations between  
14 ECG-predicted survival and brain imaging phenotypes that may reflect the combined  
15 effect of cardiovascular risk factors on cardiac and brain outcomes.

16

### 17 **AIRE-predicted survival is a biomarker of overall health**

18 Through GWAS, we have identified plausible biological pathways, including  
19 myocardial mass and trabecular development, that could be a significant factors  
20 influencing **AIRE**-predicted survival. *TBX3* and *VGLL2* have been previously  
21 described in relation to AI-ECG derived delta-age, which is a marker of accelerated  
22 biological aging (8). Indeed, there was a moderate correlation between **AIRE**-  
23 predicted survival and AI-ECG predicted age, which was stronger than the correlation  
24 with chronological age. We also identified **AIRE**-predicted survival as inversely  
25 correlated with deep learning-derived brain-age. Additionally, we identified variants in

1 *KCNQ1* and *CCDC91* that suggest **AIRE** may capture metabolic risk as an additional  
2 mechanism. These findings suggest **AIRE**-predicted survival is a biomarker of overall  
3 health, including biological age and the presence of clinical and subclinical disease.

4

#### 5 **Reduced performance in volunteer populations**

6 In our analyses, in general, **AIRE-primary care** had reduced performance metrics in  
7 the volunteer populations (UK Biobank, ELSA-Brasil). Our findings are consistent  
8 with other studies in that AI-ECG model performance is generally reduced in these  
9 volunteer populations (8, 58). A similar phenomenon is seen in other risk prediction  
10 models applied to the UK Biobank, such as QRISK3 (59). This may be due to  
11 differences in population characteristics, for example the UK Biobank population is  
12 older but healthier than the general population (60), and low event rates.

13

## 1 **Limitations**

2 There are limitations to the accuracy and granularity of ICD diagnostic codes that are  
3 used in this study to ascertain disease status. In particular, ventricular arrhythmias as  
4 reported by ICD codes are not necessarily sustained or haemodynamically significant  
5 and therefore patients predicted to have these events would not necessarily benefit  
6 from an implantable cardioverter defibrillator. Further evaluation in an implantable  
7 cardioverter defibrillator cohort is needed. The GWAS findings should be considered  
8 hypothesis generating given the absence of a replication dataset. The GWAS and  
9 UKB PheWAS results were drawn from a population of predominantly European  
10 ancestry and may not apply to other populations. Despite these limitations, we have  
11 demonstrated that the **AIRE platform** is highly effective at predicting the timing of  
12 multiple outcomes in multiple, diverse, transnational datasets.

13

## 1 **Conclusion**

2 In conclusion, we describe the **AIRE platform**, an actionable, explainable and  
3 biologically plausible AI-ECG risk estimation platform that has the potential for use  
4 worldwide across a wide range of clinical contexts, including primary and secondary  
5 care, for short- and long-term risk prediction at a population and disease-specific  
6 levels.



**Funding:** AS is funded by a British Heart Foundation (BHF) clinical research training fellowship (FS/CRTF/21/24183). FSN and NSP are supported by the BHF (RG/F/22/110078 and RE/18/4/34215). KAM is support by a BHF fellowship (FS/IPBSRF/22/27059). FSN is supported by the National Institute for Health Research Imperial Biomedical Research Centre. ES is supported by a EJP RD Research Mobility Fellowship (European Reference Networks).

DO'R is supported by the Medical Research Council (MC\_UP\_1605/13); National Institute for Health Research (NIHR) Imperial College Biomedical Research Centre; and the British Heart Foundation (RG/19/6/34387, RE/18/4/34215). For the purpose of open access, the authors have applied a creative commons attribution (CC BY) licence to any author accepted manuscript version arising.

**Acknowledgments:** This research has been conducted using the UK Biobank Resource under Application Numbers 48666, 40616 and 47602. The authors would also like to thank InSIGHT Core in the Center for Healthcare Delivery Science at Beth Israel Deaconess Medical Center for assistance in obtaining primary data.

**Data availability:**

SaMi-Trop cohort was made openly available (<https://doi.org/10.5281/zenodo.4905618>). The CODE-15% cohort was also made openly available (<https://doi.org/10.5281/zenodo.4916206>). Restrictions apply to additional clinical information on the CODE-15% and SaMi-Trop cohorts; to the full CODE cohort, the ELSA-Brasil cohort. UK Biobank data are available upon

application (<http://www.ukbiobank.ac.uk/>). The BIDMC dataset is restricted due to ethical limitations. Researchers affiliated to educational, or research institutions may make requests to access the datasets. Requests should be made to the corresponding author of this paper. They will be forwarded to the relevant steering committee.

### **Code availability**

The programming code relating to these analyses will be made available under GNU General Public License version 3 upon request to the corresponding author.

**Table 1**  
**Dataset demographics**

Data at the timepoint of a randomly selected ECG per subject is shown for the BIDMC and CODE datasets. Categorical variables n (%), continuous variables mean (SD)

	BIDMC	SaMi-TROP	CODE	ELSA-Brasil	UK Biobank
N subjects	189539	1631	1558421	13739	42386
Age	57.68 (18.69)	59.40 (12.79)	51.66 (17.59)	52.26 (9.10)	64.14 (7.75)
Follow up (years)	3.41 (4.08)	2.08 (0.39)	3.68 (1.87)	9.35 (1.28)	3.73 (1.57)
Sex (M)	90792 (47.9)	534 (32.7)	627042 (40.2)	7483 (54.5)	20538 (48.5)
Hypertension	74409 (39.3)	-	492640 (31.6)	4924 (35.9)	6256 (14.8)
Previous MI	11788 (6.2)	-	11604 (0.7)	251 (1.8)	633 (1.5)
Smoker	23343 (12.3)	-	108815 (7.0)	1801 (13.1)	1111 (2.6)
Diabetes Mellitus	33748 (17.8)	-	101470 (6.5)	2728 (19.9)	1334 (3.1)
Hyperlipidaemia	67087 (35.4)	-	60590 (3.9)	6811 (49.6)	2985 (7.0)
Mortality	34938 (18.4)	104 (6.4)	52127 (3.3)	599 (4.4)	526 (1.2)

**Table 2**

**Mortality prediction results summary table**

Model performance as assessed by C-index (95% CI) is shown

**AIRE:** Artificial-intelligence enhanced ECG risk estimator

	<b>BIDMC Test set</b>	<b>SaMi-Trop</b>	<b>CODE-10s</b>	<b>CODE Test set</b>	<b>UK Biobank</b>	<b>ELSA-Brasil</b>
<b>N ECGs</b>	434,629	1631	958,954	464,429	42,386	13,739
<b>N Subjects</b>	64,105	1631	645,372	311,684	42,386	13,739
<b>Internal/external validation</b>	Internal	External	External	Internal	External	External
<b>AIRE</b>	0.775 (0.773-0.776)	0.773 (0.733-0.813)	0.762 (0.759-0.765)	-	-	-
<b>AIRE-primary care</b>	-	-	-	0.802 (0.799-0.805)	0.638 (0.608-0.668)	0.713 (0.691-0.735)

## Figure Legends

### Figure 1

Schematic depicting all eight models in the **AIRE platform**, training datasets and validation datasets. **AIRE** and **AIRE-primary care** were trained for all-cause mortality, the remaining models were trained for the outcomes they are named after. CODE-10s denotes the subset of CODE with ECGs of 10s duration. CV: cardiovascular, NCV: non-cardiovascular, ASCVD: atherosclerotic cardiovascular disease, HF: heart failure, VA: ventricular arrhythmia. BIDMC: Beth Israel Deaconess Medical Center, CODE: Clinical Outcomes in Digital Electrocardiography, SaMi-TROP: São Paulo-Minas Gerais Tropical Medicine Research Center, ELSA-Brasil: The Brazilian Longitudinal Study of Adult Health.

### Figure 2

#### Example subject-specific survival predictions

**AIRE** outputs subject-specific survival curves. Two examples are shown for subjects who died during follow up (A), and two for subjects who survived through the follow up period (B). Dashed red lines indicate the date of death and dashed black lines indicate **AIRE**-predicted date of death. (C) Examples of subjects with many ECGs during the study period, each blue dot is a survival prediction from a single ECG. **AIRE**-predicted survival trends down over time and predicted probability of survival is particularly low prior to actual time of death (red dashed line).

### Figure 3

#### Mortality prediction performance – BIDMC Test set

Kaplan-Meier curves of **AIRE**-predicted all-cause mortality by risk quartile in the whole BIDMC test set (A) and a subset of normal ECGs (B). (C) Comparison of **AIRE** performance across sex and major ethnic groups, **AIRE** performs well across all demographic groups. Using Cox models, **AIRE** was compared with existing risk factors and ECG parameters. In these Cox models, age, sex and ECG parameters were incorporated with **AIRE** to create **AIRE-Cox**. In all comparisons, **AIRE-Cox** had a significantly higher C-index than all comparators for both all-cause mortality (D), and cardiovascular mortality (E). ECG parameters: heart rate, PR interval, QRS duration, QTc interval, CV risk factors: diabetes mellitus, hypertension, smoking history, hyperlipidaemia, ethnicity. ASCVD risk factors: systolic blood pressure, total cholesterol, HDL cholesterol, hypertension, smoking history, diabetes mellitus, ethnicity. 10-year ASCVD risk assessed using the pooled cohort equation.

#### **Figure 4**

##### **Survival analysis in four diverse, transnational, external validation datasets**

In all cohorts, **AIRE/AIRE-primary care** successfully identified groups at higher risk of all-cause mortality. (A) São Paulo-Minas Gerais Tropical Medicine Research Center (SaMi-TROP) cohort of subjects with Chagas disease, **AIRE** predictions are shown, (B) CODE cohort of primary care subjects in Brazil, (C) The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) volunteer cohort, (D) UK Biobank volunteer cohort. **AIRE**-primary care results are shown for panels B-D. Sub panels show truncated y-axes for the populations with low event rates.

#### **Figure 5**

## **Mortality prediction in high-risk disease groups and prediction of actionable end-points - BIDMC Test set**

Using Cox models, **AIRE** was compared with existing risk factors, ECG and imaging parameters in subgroups. In these Cox models, age, sex and ECG parameters were incorporated with **AIRE** to create **AIRE-Cox**. Severe aortic stenosis (A), and primary pulmonary hypertension (B). We also evaluated disease specific models, **AIRE-ASCVD** (atherosclerotic cardiovascular disease) (C), **AIRE-VA** (ventricular arrhythmia) (D) and **AIRE-HF** (heart failure) (E). Echocardiographic parameters for severe aortic stenosis: left ventricular ejection fraction (LVEF), aortic valve area, peak gradient and mean gradient. Echocardiographic parameters for primary pulmonary hypertension: LVEF, LV end diastolic diameter, tricuspid regurgitation (TR) pressure gradient, TR severity, right ventricular (RV) function, RV diameter. ECG parameters: heart rate, PR interval, QRS duration, QTc interval. ASCVD risk factors: systolic blood pressure, total cholesterol, HDL cholesterol, hypertension, smoking history, diabetes mellitus, ethnicity. 10-year ASCVD risk assessed using the pooled cohort equation. ARIC-HF risk factors: body mass index, systolic blood pressure, prevalent ASCVD, diabetes mellitus, smoking history, previous myocardial infarction, hypertension, ethnicity.

### **Figure 6**

**AIRE model explainability - BIDMC Test set:** (A) A variational auto-encoder was used to identify the most important morphological features in **AIRE**-predicted mortality, each subpanel shows one of three latent features, identifying the importance of a broad QRS complex in a left bundle morphology as well as biphasic and inverted T waves. (B) Average  $\pm$  standard deviation (shaded region) ECG

waveforms for the 10,000 highest and lowest predicted survival ECGs from the BIDMC test set. This analysis identified poor R wave progression, low QRS amplitude and T wave flattening/inversion as important features in **AIRE**-predicted survival.

## Figure 7

### Exploration of underlying biology through Phenome and Genome-wide association studies.

(A) Genome-wide association study (GWAS) in the UK Biobank. Manhattan plots of genomic loci associated with predicted survival. Nearest genes to significant single nucleotide polymorphisms are shown. The red line depicts the genome-wide significant threshold ( $P < 5 \times 10^{-8}$ ). (B) Phenome-wide association study in the UK Biobank. Cardiac associations include left ventricular ejection fraction (LVEF), atrial and right ventricular phenotypes. Non-cardiac associations included brain phenotypes such as total volume of white matter hyperintensities and pack years of smoking. (C) Association of **AIRE**-predicted survival with echocardiographic parameters in the BIDMC test set. LA: left atrium, LAEF: LA ejection fraction, TR: tricuspid regurgitation, MV: mitral valve, LVESD: LV end-systolic diameter, LVEDD: LV end-diastolic diameter, RA: right atrium, AV: aortic valve.



## Graphical Abstract

### Artificial intelligence-enhanced ECG risk estimator (AIRE)

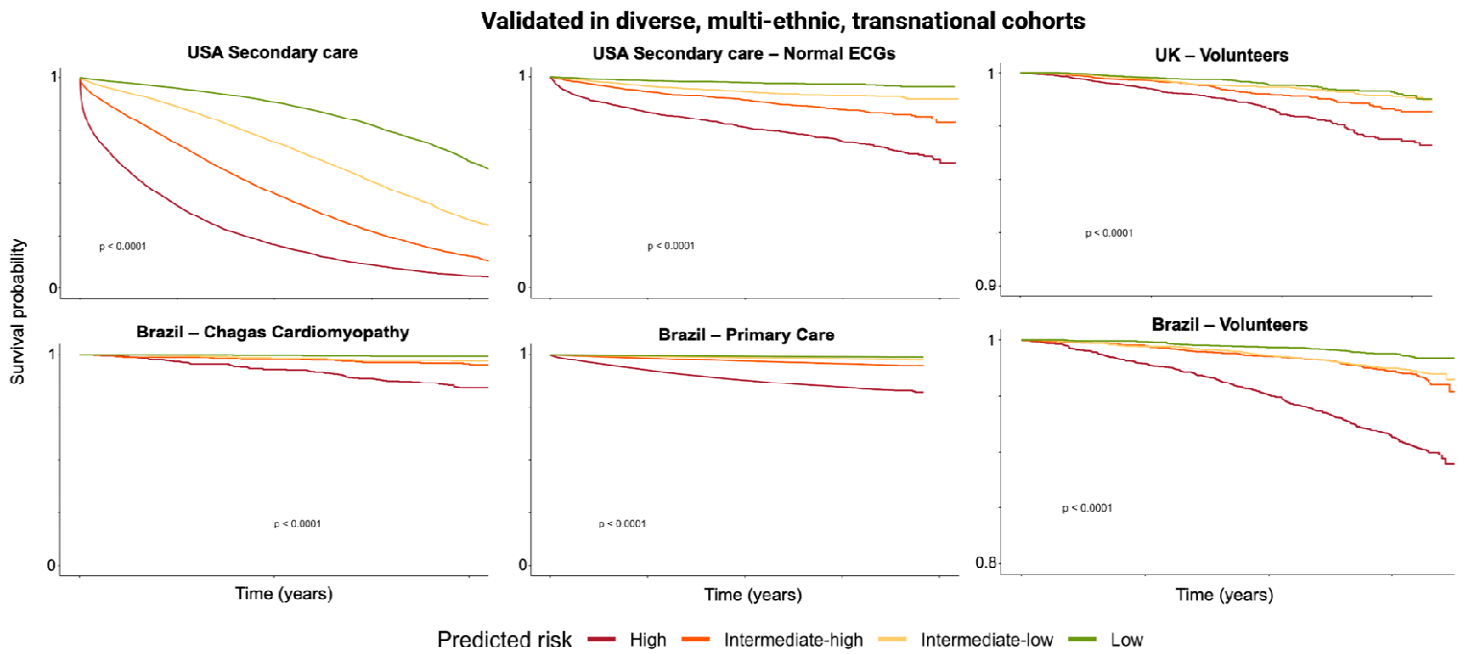
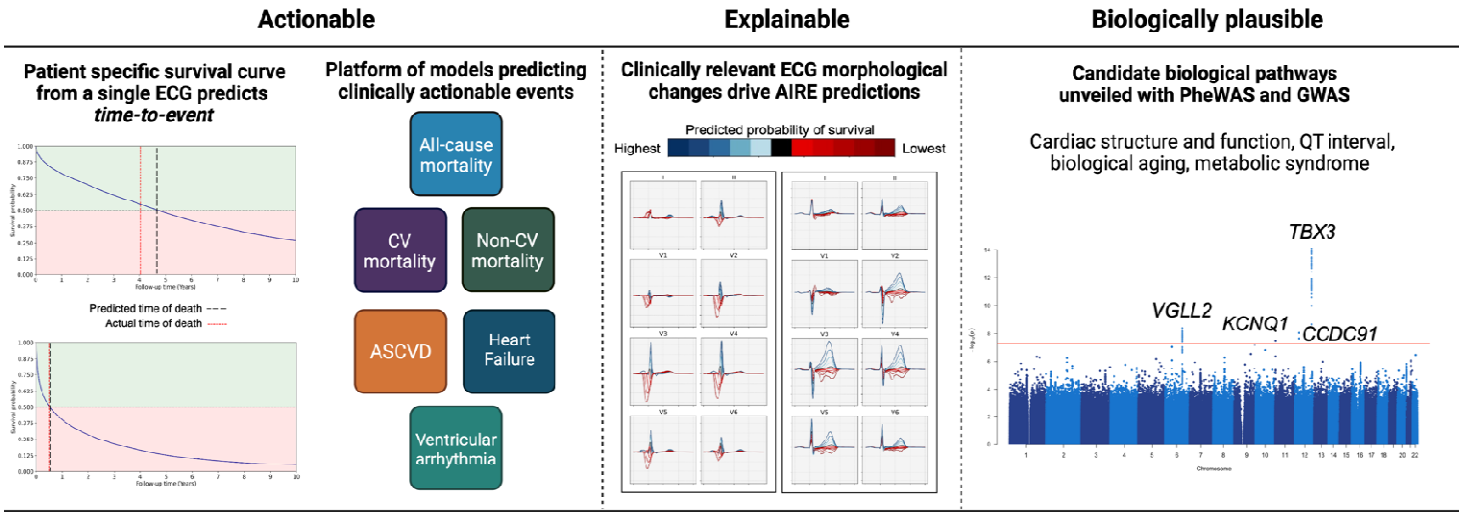
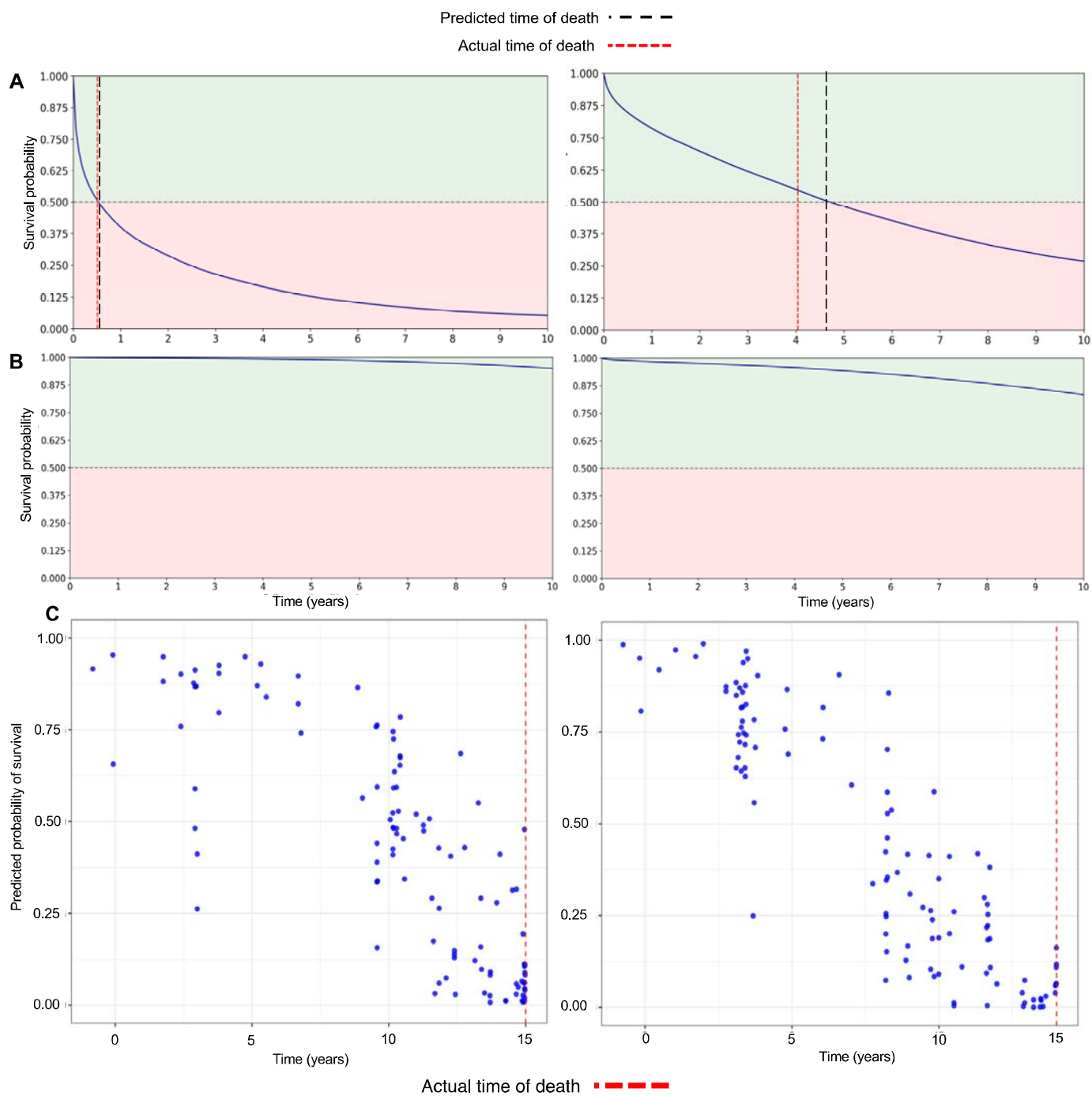


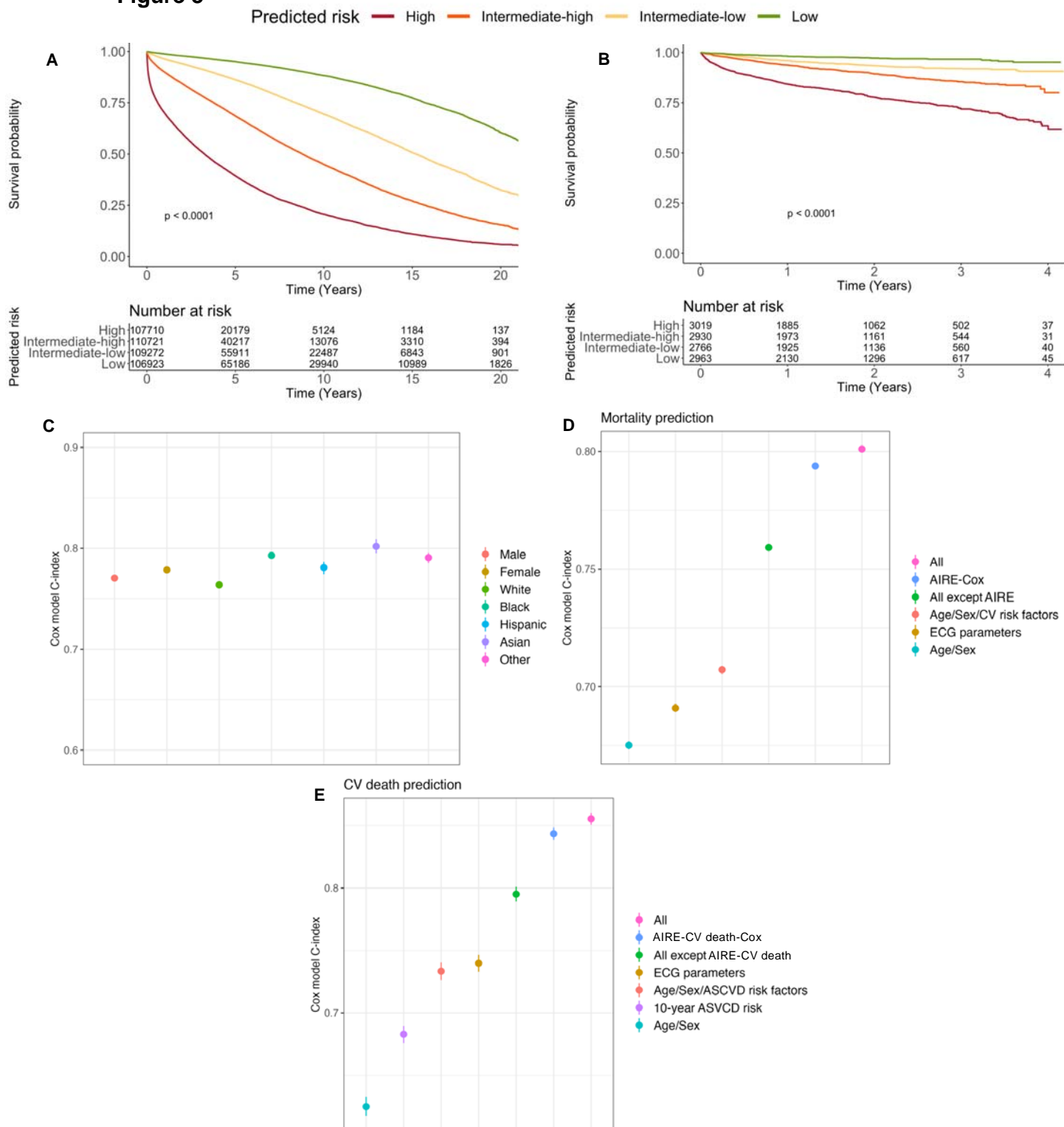
Figure 1

Model		Training and Internal validation dataset	External validation dataset
1. Mortality prediction	AIRE	BIDMC	CODE-10s SaMi-Trop
<b>AIRE submodels finetuned from AIRE</b>			
2. Mortality prediction in lower risk primary care cohort	AIRE-Primary care	CODE	ELSA-Brasil UK Biobank
3. Cardiovascular death prediction	AIRE-CV Death	BIDMC	————
4. Non-cardiovascular death prediction	AIRE-NCV Death	BIDMC	————
5. Future atherosclerotic cardiovascular disease prediction	AIRE-ASCVD	BIDMC	UK Biobank
6. Future heart failure prediction	AIRE-HF	BIDMC	UK Biobank
7. Future ventricular arrhythmia prediction	AIRE-VA	BIDMC	UK Biobank
<b>Single lead model</b>			
8. Mortality prediction with lead I only for wearables	AIRE-1L	BIDMC	————

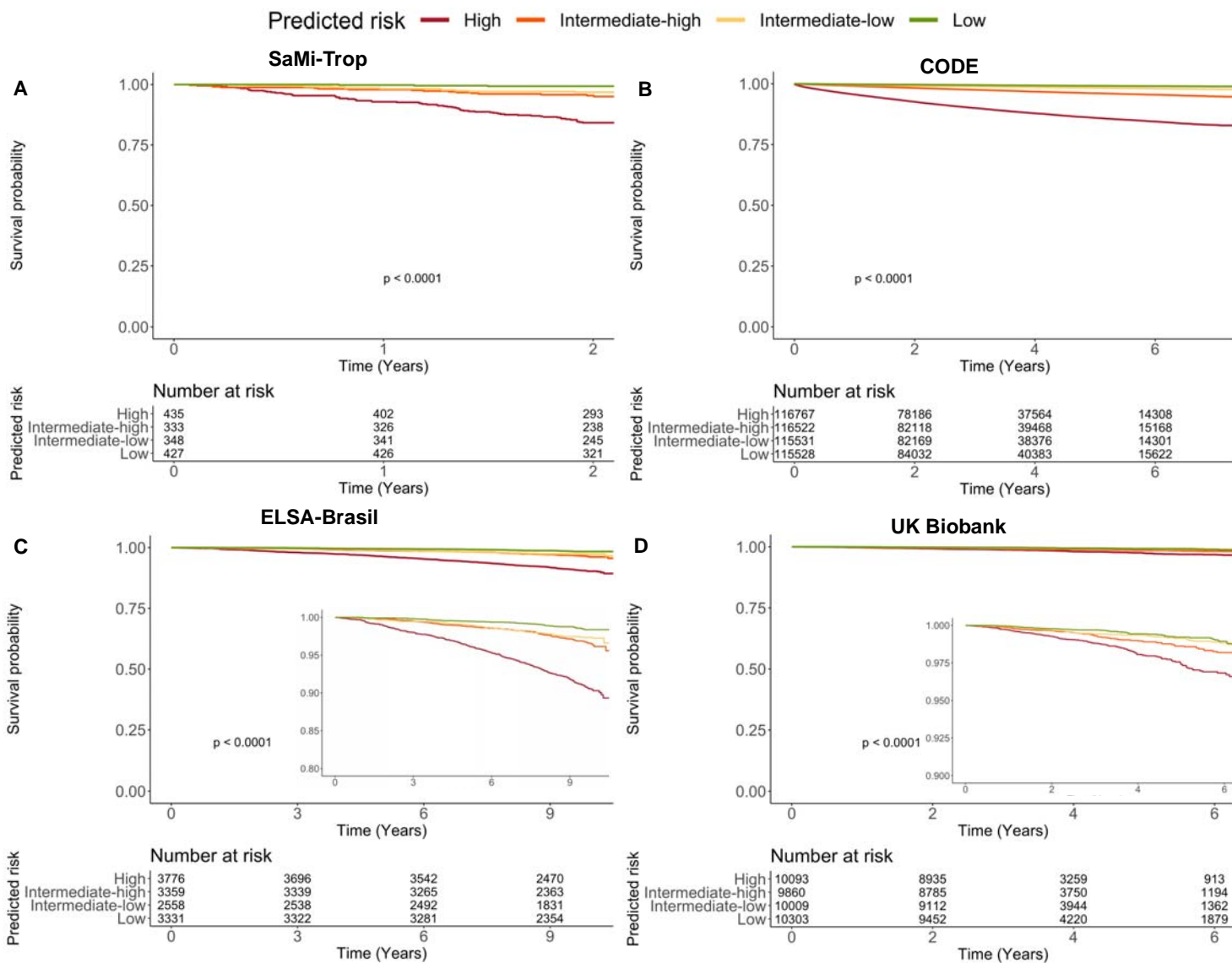
**Figure 2**



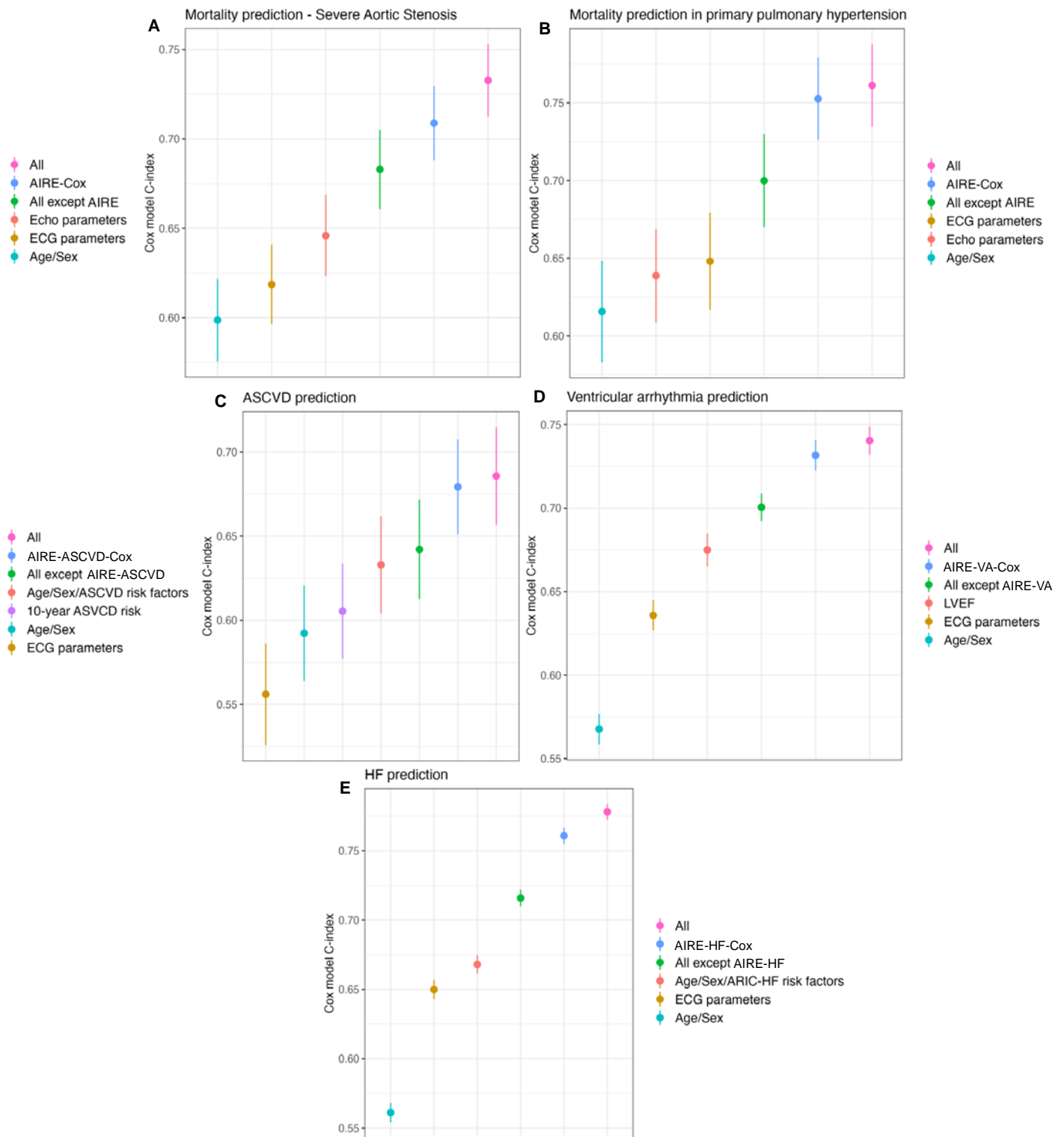
**Figure 3**



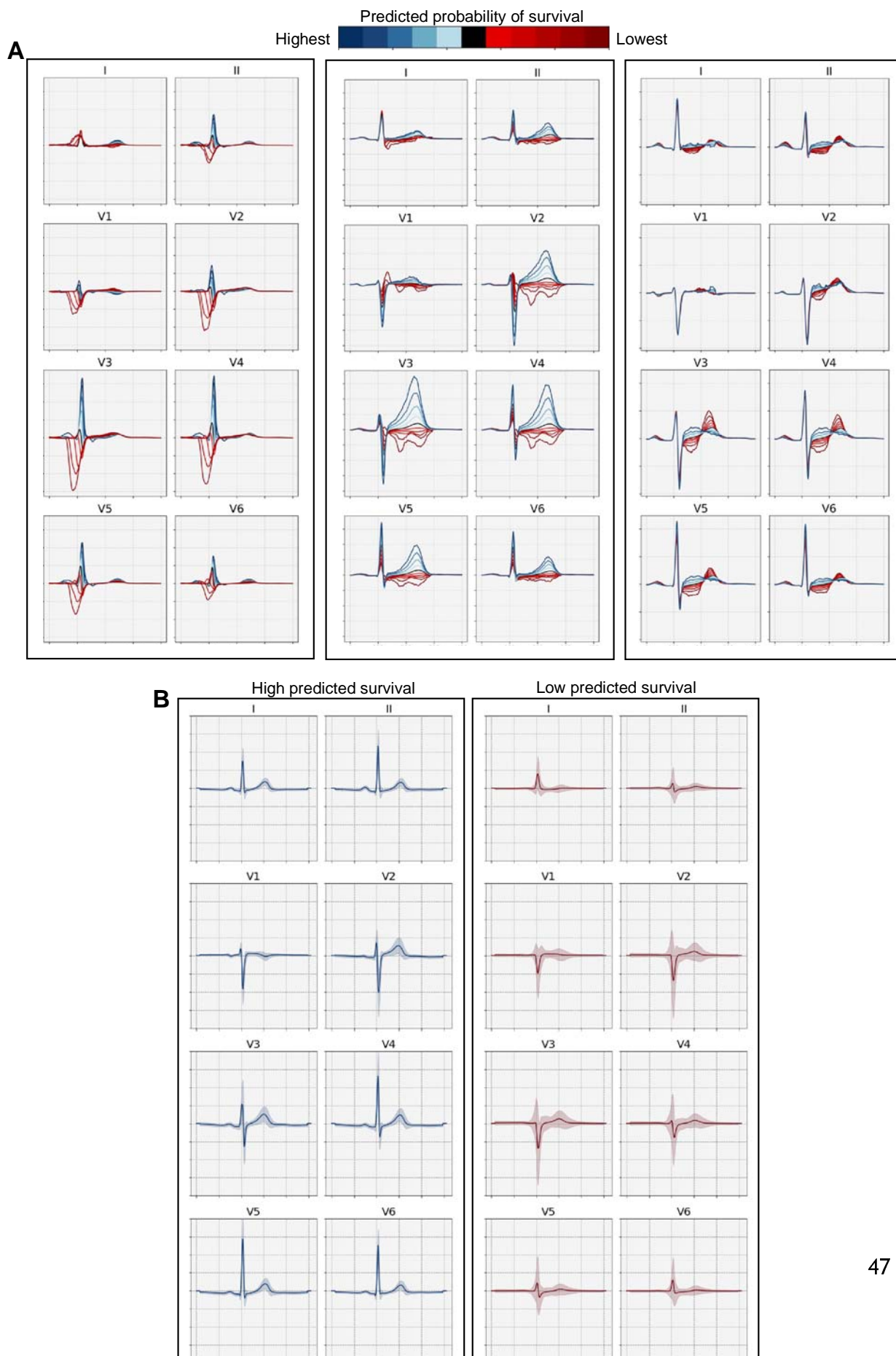
**Figure 4**

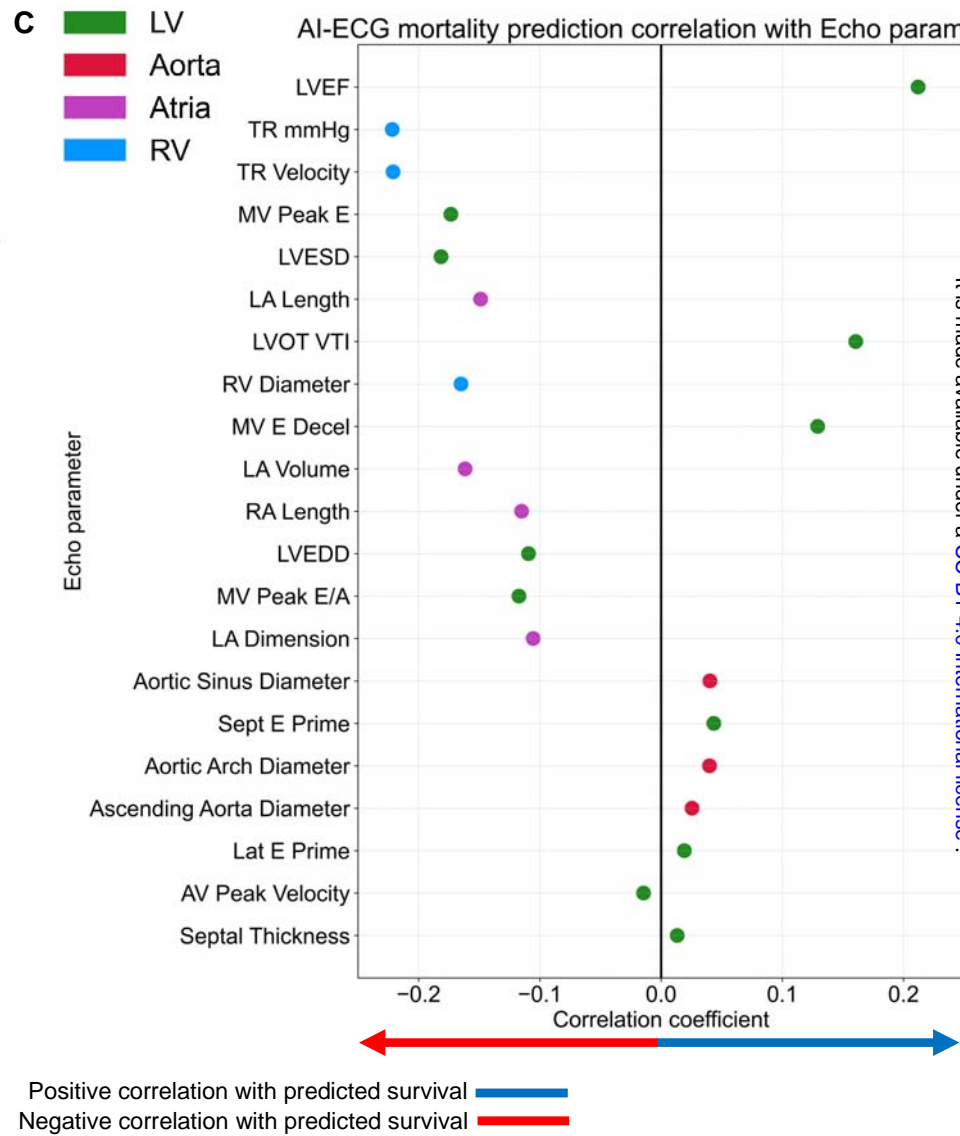
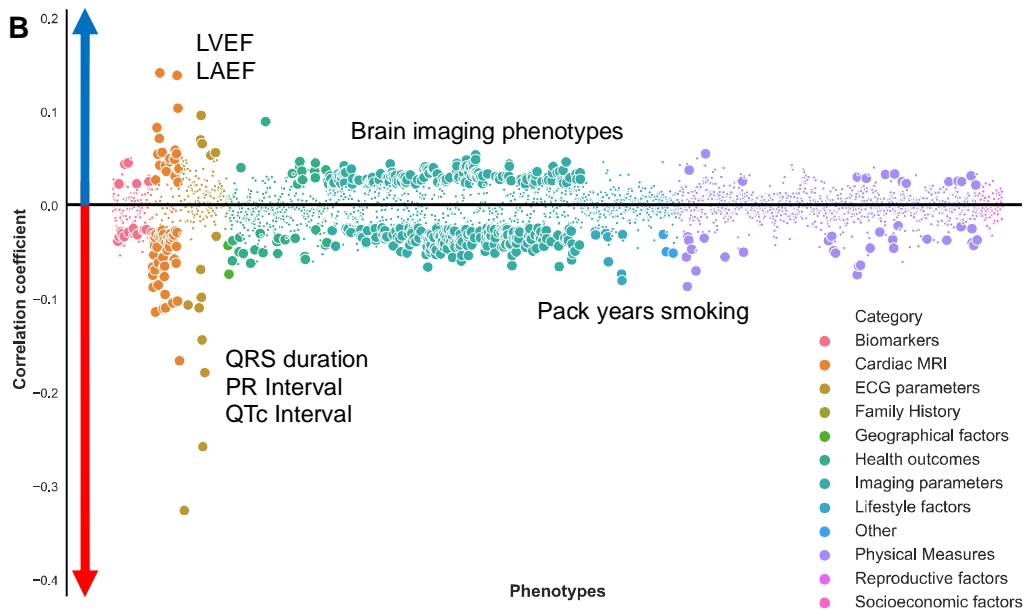
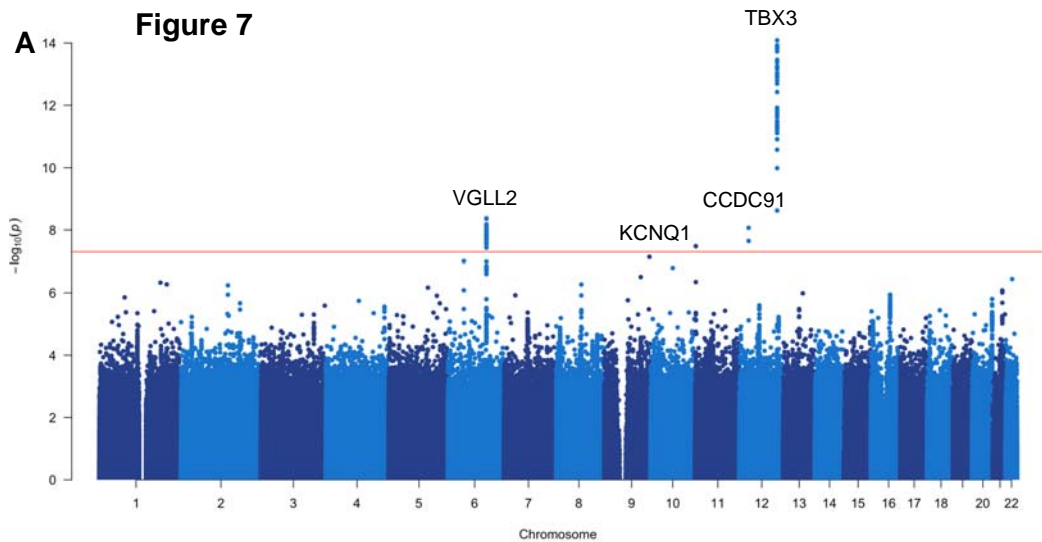


**Figure 5**



**Figure 6**







## References

1. Sau A, Ibrahim S, Ahmed A, Handa B, Kramer DB, Waks JW, et al. Artificial intelligence-enabled electrocardiogram to distinguish cavotricuspid isthmus dependence from other atrial tachycardia mechanisms. *European Heart Journal - Digital Health*. 2022:ztac042.
2. Sau A, Ibrahim S, Kramer DB, Waks JW, Qureshi N, Koa-Wing M, et al. Artificial intelligence-enabled electrocardiogram to distinguish atrioventricular re-entrant tachycardia from atrioventricular nodal re-entrant tachycardia. *Cardiovascular Digital Health Journal*. 2023;4(2):60-7.
3. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25(1):70-4.
4. Sau A, Ribeiro AH, McGurk KA, Pastika L, Bajaj N, Ardissino M, et al. Neural network-derived electrocardiographic features have prognostic significance and important phenotypic and genotypic associations. *medRxiv*. 2023:2023.06.15.23291428.
5. Raghunath S, Ulloa Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med*. 2020;26(6):886-91.
6. Sun W, Kalmady SV, Sepehrvand N, Salimi A, Nademi Y, Baine K, et al. Towards artificial intelligence-based learning health system for population-level mortality prediction using electrocardiograms. *npj Digit Med*. 2023;6(1):21.
7. Hughes JW, Tooley J, Torres Soto J, Ostropolets A, Poterucha T, Christensen MK, et al. A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease. *npj Digit Med*. 2023;6(1):169.
8. Lima EM, Ribeiro AH, Paixão GMM, Ribeiro MH, Pinto-Filho MM, Gomes PR, et al. Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nature Communications*. 2021;12(1):5117.
9. Li X, Patel KHK, Sun L, Peters NS, Ng FS. Neural networks applied to 12-lead electrocardiograms predict body mass index, visceral adiposity and concurrent cardiometabolic ill-health. *Cardiovascular Digital Health Journal*. 2021.
10. van der Net JB, Janssens ACJW, Eijkemans MJC, Kastelein JJP, Sijbrands EJG, Steyerberg EW. Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *European Journal of Human Genetics*. 2008;16(9):1111-6.
11. Cardoso CS, Sabino EC, Oliveira CDL, de Oliveira LC, Ferreira AM, Cunha-Neto E, et al. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. *BMJ Open*. 2016;6(5):e011181.
12. Schmidt MI, Duncan BB, Mill JG, Lotufo PA, Chor D, Barreto SM, et al. Cohort Profile: Longitudinal Study of Adult Health (ELSA-Brasil). *Int J Epidemiol*. 2015;44(1):68-75.
13. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.

14. Ribeiro AH, Ribeiro MH, Paixao GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun.* 2020;11(1):1760.
15. Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks. *PeerJ.* 2019;7:e6257.
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine.* 2015;13(1):1.
17. Ribeiro AH. Predicting age from the electrocardiogram and its usage as a mortality predictor 2021 [
18. Hughes JW. A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease 2023 [
19. Stensrud MJ, Hernan MA. Why Test for Proportional Hazards? *JAMA.* 2020;323(14):1401-2.
20. Fine JP. Comparing nonnested Cox models. *Biometrika.* 2002;89(3):635-48.
21. Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet.* 2019;394(10211):1816-26.
22. Stabenau HF, Waks JW. BRAVEHEART: Open-source software for automated electrocardiographic and vectorcardiographic analysis. *Computer Methods and Programs in Biomedicine.* 2023;242:107798.
23. van de Leur RR, Bos MN, Taha K, Sammani A, Yeung MW, van Duijvenboden S, et al. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders *European Heart Journal - Digital Health.* 2022;3(3):390-404.
24. Meyer HV, Dawes TJW, Serrani M, Bai W, Tokarczuk P, Cai J, et al. Genetic and functional insights into the fractal structure of the heart. *Nature.* 2020;584(7822):589-94.
25. Jonsson BA, Bjornsdottir G, Thorgeirsson TE, Ellingsen LM, Walters GB, Gudbjartsson DF, et al. Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications.* 2019;10(1):5409.
26. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, et al. Prediction of incident heart failure in general practice: the Atherosclerosis Risk in Communities (ARIC) Study. *Circ Heart Fail.* 2012;5(4):422-9.
27. Zhu X, Zhu L, Wang H, Cooper RS, Chakravarti A. Genome-wide pleiotropy analysis identifies novel blood pressure variants and improves its polygenic risk scores. *Genet Epidemiol.* 2022;46(2):105-21.
28. Verweij N, Benjamins JW, Morley MP, van de Vegte YJ, Teumer A, Trenkwalder T, et al. The Genetic Makeup of the Electrocardiogram. *Cell Syst.* 2020;11(3):229-38.e5.
29. van der Harst P, van Setten J, Verweij N, Vogler G, Franke L, Maurano MT, et al. 52 Genetic Loci Influencing Myocardial Mass. *J Am Coll Cardiol.* 2016;68(13):1435-48.
30. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshihara S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet.* 2021;53(10):1415-24.

31. Christakoudi S, Evangelou E, Riboli E, Tsilidis KK. GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. *Sci Rep*. 2021;11(1):10688.
32. Jespersen T, Grunnet M, Olesen S-P. The KCNQ1 potassium channel: from gene to physiological function. *Physiology*. 2005;20(6):408-16.
33. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS Genet*. 2017;13(4):e1006528.
34. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet*. 2018;50(3):390-400.
35. Zhao W, Rasheed A, Tikkanen E, Lee JJ, Butterworth AS, Howson JMM, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet*. 2017;49(10):1450-7.
36. Sillesen H, Muntendam P, Adourian A, Entrekin R, Garcia M, Falk E, et al. Carotid plaque burden as a measure of subclinical atherosclerosis: comparison with other tests for subclinical arterial disease in the High Risk Plaque BiImage study. *JACC Cardiovasc Imaging*. 2012;5(7):681-9.
37. Raghunath S, Ulloa Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nature Medicine*. 2020;26(6):886-91.
38. Otto CM, Nishimura RA, Bonow RO, Carabello BA, Erwin JP, 3rd, Gentile F, et al. 2020 ACC/AHA Guideline for the Management of Patients With Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2021;143(5):e72-e227.
39. Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J, et al. 2021 ESC/EACTS Guidelines for the management of valvular heart disease: Developed by the Task Force for the management of valvular heart disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). *European Heart Journal*. 2022;43(7):561-632.
40. Génèreux P. EARLY TAVR: Evaluation of TAVR Compared to Surveillance for Patients With Asymptomatic Severe Aortic Stenosis (EARLY TAVR) [Available from: <https://clinicaltrials.gov/study/NCT03042104>].
41. McCann G. The Early Valve Replacement in Severe ASymptomatic Aortic Stenosis Study (EASY-AS).
42. Chang KY, Duval S, Badesch DB, Bull TM, Chakinala MM, De Marco T, et al. Mortality in Pulmonary Arterial Hypertension in the Modern Era: Early Insights From the Pulmonary Hypertension Association Registry. *J Am Heart Assoc*. 2022;11(9):e024969.
43. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J*. 2022;43(38):3618-731.
44. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. 2021;42(34):3227-337.

45. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr., Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(25 Pt B):2935-59.
46. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ*. 2017;357:j2099.
47. Olgin JE, Pletcher MJ, Vittinghoff E, Wranicz J, Malik R, Morin DP, et al. Wearable Cardioverter–Defibrillator after Myocardial Infarction. *New England Journal of Medicine*. 2018;379(13):1205-15.
48. Bachtiger P, Kelshiker MA, Petri CF, Gandhi M, Shah M, Kamalati T, et al. Survival and health economic outcomes in heart failure diagnosed at hospital admission versus community settings: a propensity-matched analysis. *BMJ Health Care Inform*. 2023;30(1).
49. Yang H, Negishi K, Otahal P, Marwick TH. Clinical prediction of incident heart failure risk: a systematic review and meta-analysis. *Open Heart*. 2015;2(1):e000222.
50. Bachtiger P, Petri CF, Scott FE, Ri Park S, Kelshiker MA, Sahemey HK, et al. Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: a prospective, observational, multicentre study. *The Lancet Digital Health*. 2022:S2589750021002569.
51. Attia ZI, Harmon DM, Dugan J, Manka L, Lopez-Jimenez F, Lerman A, et al. Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. *Nat Med*. 2022;28(12):2497-503.
52. Rosenberg MA. Trusting Magic. *Circulation*. 2021;143(13):1299-301.
53. Istolahti T, Lyytikäinen LP, Huhtala H, Nieminen T, Kähönen M, Lehtimäki T, et al. The prognostic significance of T-wave inversion according to ECG lead group during long-term follow-up in the general population. *Ann Noninvasive Electrocardiol*. 2021;26(1):e12799.
54. Schröder LC, Holkeri A, Eranti A, Haukilahti MAE, Kerola T, Kenttä TV, et al. Poor R-wave progression as a predictor of sudden cardiac death in the general population and subjects with coronary artery disease. *Heart Rhythm*. 2022;19(6):952-9.
55. Imanishi R, Seto S, Ichimaru S, Nakashima E, Yano K, Akahoshi M. Prognostic Significance of Incident Complete Left Bundle Branch Block Observed Over a 40-Year Period. *The American Journal of Cardiology*. 2006;98(5):644-8.
56. Seyyedi SR, Sharif-Kashani B, Sadr M, Chitsazan M, Malekmohammad M, Abedini A, et al. The Relationship between Electrocardiographic Changes and Prognostic Factors in Severely Symptomatic Pulmonary Hypertension. *Tanaffos*. 2019;18(1):34-40.
57. Rauseo E, Salih A, Raisi-Estabragh Z, Aung N, Khanderia N, Slabaugh GG, et al. Ischemic Heart Disease and Vascular Risk Factors Are Associated With Accelerated Brain Aging. *JACC Cardiovasc Imaging*. 2023;16(7):905-15.
58. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. *Circulation*. 2022;145(2):122-33.

59. Ruth EP, Xiaonan L, Jennifer AC, David AC, Benjamin JC, Lei C. Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank. *Heart*. 2023;109(22):1690.
60. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017;186(9):1026-34.