# Artificial intelligence-enhanced electrocardiography for the identification of a sex-related cardiovascular risk continuum: a retrospective cohort study

*Arunashis Sau, Ewa Sieliwonczyk, Konstantinos Patlatzoglou, Libor Pastika, Kathryn A McGurk, Antônio H Ribeiro, Antonio Luiz P Ribeiro, Jennifer E Ho, Nicholas S Peters, James S Ware, Upasana Tayal, Daniel B Kramer, Jonathan W Waks, Fu Siong Ng*

## Summary

**Background** Females are typically underserved in cardiovascular medicine. The use of sex as a dichotomous variable for risk stratification fails to capture the heterogeneity of risk within each sex. We aimed to develop an artificial intelligence-enhanced electrocardiography (AI-ECG) model to investigate sex-specific cardiovascular risk.

**Methods** In this retrospective cohort study, we trained a convolutional neural network to classify sex using the 12-lead electrocardiogram (ECG). The Beth Israel Deaconess Medical Center (BIDMC) secondary care dataset, comprising data from individuals who had clinically indicated ECGs performed in a hospital setting in Boston, MA, USA collected between May, 2000, and March, 2023, was the derivation cohort (1 163 401 ECGs). 50% of this dataset was used for model training, 10% for validation, and 40% for testing. External validation was performed using the UK Biobank cohort, comprising data from volunteers aged 40–69 years at the time of enrolment in 2006–10 (42 386 ECGs). We examined the difference between AI-ECG-predicted sex (continuous) and biological sex (dichotomous), termed sex discordance score.

**Findings** AI-ECG accurately identified sex (area under the receiver operating characteristic 0·943 [95% CI 0·942–0·943] for BIDMC and 0·971 [0·969–0·972] for the UK Biobank). In BIDMC outpatients with normal ECGs, an increased sex discordance score was associated with covariate-adjusted increased risk of cardiovascular death in females (hazard ratio [HR] 1·78 [95% CI 1·18–2·70], p=0·006) but not males (1·00 [0·63–1·58], p=0·996). In the UK Biobank cohort, the same pattern was seen (HR 1·33 [95% CI 1·06–1·68] for females, p=0·015; 0·98 [0·80–1·20] for males, p=0·854). Females with a higher sex discordance score were more likely to have future heart failure or myocardial infarction in the BIDMC cohort and had more male cardiac (increased left ventricular mass and chamber volumes) and non-cardiac phenotypes (increased muscle mass and reduced body fat percentage) in both cohorts.

**Interpretation** Sex discordance score is a novel AI-ECG biomarker capable of identifying females with disproportionately elevated cardiovascular risk. AI-ECG has the potential to identify female patients who could benefit from enhanced risk factor modification or surveillance.

**Funding** British Heart Foundation.



*Lancet Digit Health* 2025; 7: e184–94

See **Comment** page e170

**National Heart and Lung Institute** (A Sau PhD, E Sieliwonczyk PhD, K Patlatzoglou PhD, L Pastika MBBS, K A McGurk PhD, Prof N S Peters MD, Prof J S Ware PhD, U Tayal PhD, D B Kramer MD, F S Ng PhD) **and MRC Laboratory of Medical Sciences** (E Sieliwonczyk, K A McGurk, Prof J S Ware), **Imperial College London, London, UK; Department of Cardiology, Imperial College Healthcare NHS Trust, London, UK** (A Sau, Prof N S Peters, F S Ng); **University of Antwerp and Antwerp University Hospital, Antwerp, Belgium** (E Sieliwonczyk); **Department of Information Technology, Uppsala University, Uppsala, Sweden** (A H Ribeiro PhD); **Department of Internal Medicine, Faculdade de Medicina, and Telehealth Center and Cardiology Service, Hospital das Clínicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil** (Prof A L P Ribeiro MD); **Cardiovascular Institute and Division of Cardiology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA** (J E Ho MD); **Department of Cardiology, Royal Brompton & Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London, UK** (U Tayal); **Richard A and Susan F Smith Center for Outcomes Research in Cardiology** (D B Kramer) **and Harvard-Thorndike Electrophysiology Institute** (J W Waks MD), **Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; Department of Cardiology, Chelsea and Westminster Hospital NHS Foundation Trust, London, UK** (F S Ng)

## Introduction

There are well documented sex differences in the prevalence, diagnosis, treatment, and outcomes for cardiovascular disease.[1] Females are often underserved, with increased rates of misdiagnosis and adverse events.[2] Notably, females are less likely to receive appropriate treatments, preventive care, or aggressive risk factor modification for cardiovascular risk factors.[3] This outcome might be in part due to lack of awareness of sex-specific risk factors in females, and an inappropriate bias in which clinicians perceive females to be at lower risk of cardiovascular events.[4]

There is growing awareness that biological sex is an important factor to be considered in risk stratification and management of cardiovascular health. Sex is often included in cardiovascular risk scores as a dichotomous variable, with females often considered at lower risk than males.[5] However, there is emerging evidence that the biological sex phenotype might be more accurately represented as a continuum.[6] Previous studies have suggested that there might be a sex continuum for both cardiometabolic and personality traits (such as degree of extraversion and assertiveness).[6,7] In the era of precision medicine, the dichotomisation of sex for purposes of risk stratification might prove a crude approach that fails to fully capture the continuum of inter-individual variability of sex-related cardiovascular risk.

Sex differences in electrocardiograms (ECGs) are well known.[8] Artificial intelligence-enhanced electrocardiography (AI-ECG) models have been shown to be able to identify sex [9,10] and sex misclassification by AI-ECG models appears to be associated with adverse prognosis.[10]

Correspondence to:
Dr Arunashis Sau, National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
**as7909@ic.ac.uk**
or
Dr Fu Siong Ng, National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
**f.ng@imperial.ac.uk**

## Research in context

### Evidence before this study

We searched PubMed for studies in English relating to sex differences in cardiovascular disease and the application of artificial intelligence-enhanced electrocardiography (AI-ECG) for sex identification published before Nov 1, 2023. We used the search terms ("artificial intelligence" OR "deep learning" OR "neural networks") AND ("ECG" OR "electrocardiogram") OR ("sex differences") AND ("cardiovascular disease"). We found previous studies have identified sex differences in the prevalence, diagnosis, treatment, and outcomes for cardiovascular disease. Females have been described to be at lower risk of cardiovascular events than males. AI-ECG has been shown to identify sex with very high accuracy, and sex misclassification has been described to be associated with adverse outcomes in males and females.

### Added value of this study

In this retrospective cohort study, using AI-ECG we developed a novel AI-ECG biomarker termed sex discordance score (ie, the difference between AI-ECG-predicted sex [continuous] and biological sex [dichotomous]). We found that a higher sex discordance score is associated with an increased risk of cardiovascular death in females, but not males, even after adjustment for clinical covariates. Sex discordance score could help to uncover potential underlying mechanisms of cardiovascular disease among females, and has the potential to identify female patients who might benefit from enhanced risk factor modification or surveillance.

### Implications of all the available evidence

The current dichotomous paradigm might in part be responsible for the incorrect generalisation that females are at low risk of cardiovascular disease, and in addressing this inequality in cardiovascular care, we propose a shift to an appreciation of sex-specific cardiovascular risk as a continuum rather than as a dichotomous trait when considering cardiovascular risk factors and risk stratification. AI-ECG sex discordance score could contribute to reducing female cardiovascular health inequalities.

However, it remains unclear whether sex misclassification represents a failure of AI-ECG models to differentiate sex in the presence of gross or subtle ECG abnormalities secondary to pathology or if the misclassifications of sex by AI-ECG models reflect the presence of a true sex continuum of ECG phenotypes that is being detected by these models, which capture information on cardiovascular risk.

See Online for appendix

We hypothesised that the previously described[9] adverse prognosis associated with AI-ECG sex misclassification is due to higher risk individuals having more ECG abnormalities, which reduces the model's ability to pick out male-specific or female-specific ECG morphologies for accurate sex classification. We therefore investigated the significance of AI-ECG sex misclassification specifically in low-risk cohorts, where gross and subclinical ECG abnormalities are far less likely, and hypothesised that a more male ECG phenotype in females is associated with adverse cardiovascular prognosis.

## Methods

### Study design and participants

In this retrospective cohort study, we used two intentionally diverse cohorts. The Beth Israel Deaconess Medical Center (BIDMC) cohort is a secondary care dataset comprised of routinely collected data from Boston, MA, USA between May, 2000, and March, 2023; indivduals with an ECG recorded were included. 1 163 401 ECGs from 189 539 individuals (mean age of 57·68 years [±18·69]) were available.[11,12] The UK Biobank is a longitudinal study of more than 500 000 participants who enrolled voluntarily aged 40–69 years at the time of enrolment in 2006–10.[13] 42 386 individuals with digital ECGs taken at the second follow-up visit were available for analysis. As the BIDMC cohort was hospital based, the individuals had more comorbidities than the UK Biobank cohort, despite being of younger age on average. These two cohorts combined can be considered representative of a wide range of patient groups and volunteers. Sex and ethnicity were self-reported. Further details on the cohorts are provided in the appendix (p 1).

For the BIDMC cohort, ethics review and approval was provided by the Beth Israel Deaconess Medical Center Committee on Clinical Investigations (institutional review board protocol 2023P000042). Due to the retrospective nature of the BIDMC cohort, requirement for individual patient consent was waived. The UK Biobank received approval from the North West Multi-Centre Research Ethics Committee as a Research Tissue Bank (application identifications 48666 and 47602). Individual participant consent was obtained in the UK Biobank.

### Procedures

In addition to analyses of the whole cohort, we also aimed to investigate the significance of AI-ECG sex misclassification in subsets of cohorts with low cardiovascular risk. As there is no single definition of low cardiovascular risk, we defined several cohorts that could be considered low risk. These were individuals with normal ECGs and who were considered healthy or outpatients, and individuals aged 40 years and younger.

A subset of the BIDMC dataset had cardiologist reports for the ECGs (n=143 077). Normal ECGs in BIDMC were determined by searching for "normal ecg" in the free text reports; a whole word match was required to exclude

"abnormal ecg". ECGs with the phrase "otherwise" were also excluded from the normal definition. Additionally, we filtered by heart rate (60–100 beats per minute), PR interval (less than 200 milliseconds [ms]), QRS duration (less than 120 ms), and QTc interval (less than 470 ms). Clinician ECG reports were not available in the UK Biobank.

Healthy individuals were defined as those without hypertension, diabetes, hyperlipidaemia, a history of smoking, atherosclerotic cardiovascular disease, heart failure, and myocardial infarction, as defined by ICD-9 and ICD-10 codes. Higher risk cohorts individuals included those older than 40 years, inpatients, and patients with prevalent cardiovascular disease. A subgroup of patients with ECG abnormalities (broad QRS complex [QRS duration >120 ms] and tachycardia) were also analysed.

### Model development

12-lead ECGs were pre-processed with a bandpass filter 0·5–100 hz, a notch filter at 60 hz, and re-sampling to 400 hz (from 500 hz originally). Zero padding resulted in a signal, with 4096 samples for each lead for a 10 s recording. As leads III, aVL, aVR, and aVF are linear combinations of leads I and II, and therefore provide no independent data, these leads were not used for model development or evaluation.

With regard to model development, the BIDMC cohort was the derivation dataset. These data were split at a ratio of 50%:10%:40% for training, validation, and testing, respectively. Data were split by patient identification, stratified by presence of ECGs with paired 5-year life status. To prevent data leakage, a single individual could have ECGs assigned to only one of the training, validation, or testing datasets. All downstream BIDMC analysis was performed on the 40% test set. We used a previously described convolutional neural network architecture based on residual blocks[14] and trained for the binary task of self-reported sex classification. Further details of hyperparameters and model training are in the appendix (p 1).

In the BIDMC test set and UK Biobank cohort, we generated sex predictions (continuous) of all ECGs. AI-ECG sex discordance score (subsequently referred to as sex discordance score) was calculated by taking the absolute of AI-ECG sex prediction minus self-reported sex (encoded as 0 for male, 1 for female). Sex discordance score was analysed as a continuous variable for all analyses unless otherwise specified.

To understand the ECG morphologies associated with sex discordance score, we used three approaches. First, we trained a variational autoencoder using median ECG beats (appendix p 2). Median beats were extracted using the BRAVEHEART ECG analysis software as previously described.[15] Variational autoencoder latent features were input into a linear regression with AI-ECG-predicted sex as the output. The top three most important features as assessed by the *t* value were visualised by latent traversal. Second, using the median beats, we calculated the mean waveform from the 1000 ECGs with the lowest and highest AI-ECG sex prediction (ie, most confident male and female AI-ECG predictions, respectively). The mean and standard deviation of these waveforms were then plotted. Finally, in the UK Biobank cohort we performed univariate correlation between AI-ECG sex prediction and ECG parameters. Using linear regression, AI-ECG sex prediction was adjusted for age, age squared, height, weight, body surface area, and waist circumference.

### Outcomes

The main outcomes of the study were all-cause and cardiovascular death. Secondary outcomes were non-cardiovascular mortality, future heart failure, and myocardial infarction.

### Statistical analysis

Model performance was assessed using the area under the receiver operating characteristic (AUROC) curve. Sex discordance score quartiles were defined using the distributions in the test set, stratified by sex, for Kaplan–Meier curve visualisation. Statistical significance was assessed using the log-rank test. Multiple ECGs per individual were used during model training, to provide as much data as possible. However, Cox model analyses used the first ECG per individual as individuals often develop cardiovascular disease or ECG abnormalities with time, and therefore would no longer be consistent with the subgroup they started in. The sex discordance score variable was standardised, so hazard ratios (HRs) reflect a one standard deviation change in sex discordance score. Cox models were fit using the test dataset for BIDMC or UK Biobank datasets and adjusted for age, hypertension, hyperlipidaemia, diabetes, smoking history, heart failure, previous atherosclerotic cardiovascular disease, and previous myocardial infarction in BIDMC (as defined by hospital ICD-9 and ICD-10 codes in BIDMC and additionally primary care records and patient reported medical conditions in the UK Biobank) and additionally body-mass index, physical activity, and Index of Multiple Deprivation in the UK Biobank. Individuals were censored at the last in-person hospital contact (BIDMC) or the UK Biobank national censoring dates.

In analyses of future disease (specifically heart failure and myocardial infarction) individuals with the prevalent disease being studied were excluded. Primary analysis was performed using a Cox model adjusted for the same covariates as above, whereas secondary analysis used a Fine and Gray model accounting for the competing risk of death.[16]

To better understand the biology underlying sex discordance score we performed phenome-wide association studies, to evaluate the association with sex discordance score with phenotypes of interest. We used

| | BIDMC cohort | UK Biobank cohort |
|---|---|---|
| Number of individuals | 189 539 | 42 386 |
| Age, years | 57·68 (18·69) | 64·14 (7·75) |
| Follow-up, years | 3·41 (4·08) | 4·83 (1·57) |
| Sex | | |
| Male | 90 792 (47·9%) | 20 538 (48·5%) |
| Female | 98 747 (52·1%) | 21 848 (51·6%) |
| Race and ethnicity | | |
| White | 122 344 (64·5%) | 40 926 (96·6%) |
| Black | 24 251 (12·8%) | 305 (0·7%) |
| Hispanic | 10 169 (5·4%) | NE* |
| Asian | 8924 (4·7%) | 603 (1·4%) |
| Other | 23 851 (12·6%) | 552 (1·3%) |
| Hypertension | 74 409 (39·3%) | 12 815 (30·2%) |
| Previous myocardial infarction | 11 788 (6·2%) | 961 (2·3%) |
| History of smoking | 23 343 (12·3%) | 1500 (3·5%) |
| Diabetes | 33 748 (17·8%) | 2324 (5·5%) |
| Hyperlipidaemia | 67 087 (35·4%) | 9513 (22·4%) |
| Atherosclerotic cardiovascular disease | 38 113 (20·1%) | 2872 (6·8%) |
| Heart failure | 15 666 (8·3%) | 272 (0·6%) |
| All-cause death | 34 938 (18·4%) | 809 (1·9) |
| Cardiovascular death | 6944 (3·7%) | 163 (0·4%)† |
| Non-cardiovascular death | 27 994 (14·8%) | 630 (1·4%) |

Data are n (%) or mean (SD). Data at the timepoint of a randomly selected ECG per individual is shown for the BIDMC and UK Biobank datasets. BIDMC=Beth Israel Deaconess Medical Center. ECG=electrocardiogram. NE=not estimable. *Hispanic ethnicity was not reported separately for the UK biobank. †Cause of death unavailable in 16 individuals.

***Table:*** **Dataset demographics**

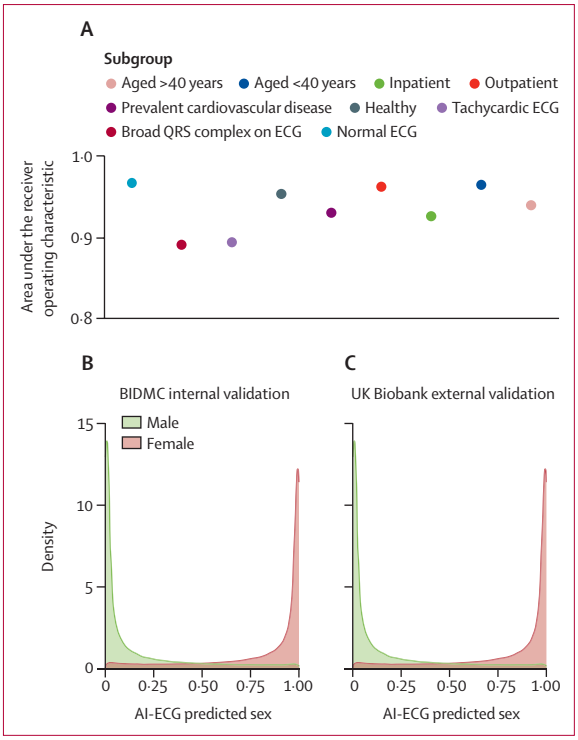



***Figure 1:*** **AI-ECG sex prediction performance metrics**
(A) Classification performance was reported according to area under the receiver operating characteristics curve for subgroups of the BIDMC dataset. Healthy individuals were defined as those without hypertension, diabetes, hyperlipidaemia, history of smoking, atherosclerotic cardiovascular disease, heart failure, and myocardial infarction. Prevalent cardiovascular disease was defined as previous atherosclerotic cardiovascular disease, heart failure, or myocardial infarction. Distribution of AI-ECG sex prediction according to the biological sex in the BIDMC (B) and UK Biobank (C) is shown. AI-ECG=artificial intelligence-enhanced electrocardiography. BIDMC=Beth Israel Deaconess Medical Center. ECG=electrocardiogram.

two approaches. First, we performed a phenome-wide association study of continuous traits using univariate correlation. We applied this approach to the individuals with ECGs within 60 days of the ECG in the BIDMC test set, in whom we investigated the association of sex discordance score with echocardiographic parameters. We additionally applied this approach to the UK Biobank that contains data from more than 3000 phenotypes derived from patient measurements, surveys, and investigations (including cardiac MRI). $p<0·05$ after Bonferroni correction was considered statistically significant. Further details on this phenome-wide association study are provided in the appendix (p 2). The other approach was a disease phenome-wide association study to explore the association of sex discordance score with both prevalent and incident diseases, as defined by ICD-9 and ICD-10 codes and converted into Phecodes in the BIDMC dataset. In this analysis we used logistic regression (appendix p 2).

To identify genetic associations with sex discordance score, we performed a genome-wide association study in the UK Biobank for males and females separately. This analysis was restricted to the autosomes. Significant single nucleotide polymorphisms were identified at a p value less than $5\times10^{-8}$ (further details are specified in the appendix [p 3]).

Statistical analyses were performed with R 4.2.0 statistical package or Python (version 3.9). $p<0·05$ was considered statistically significant, except in the case of the phenome-wide association studies and the genome-wide association study.


### Role of the funding source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

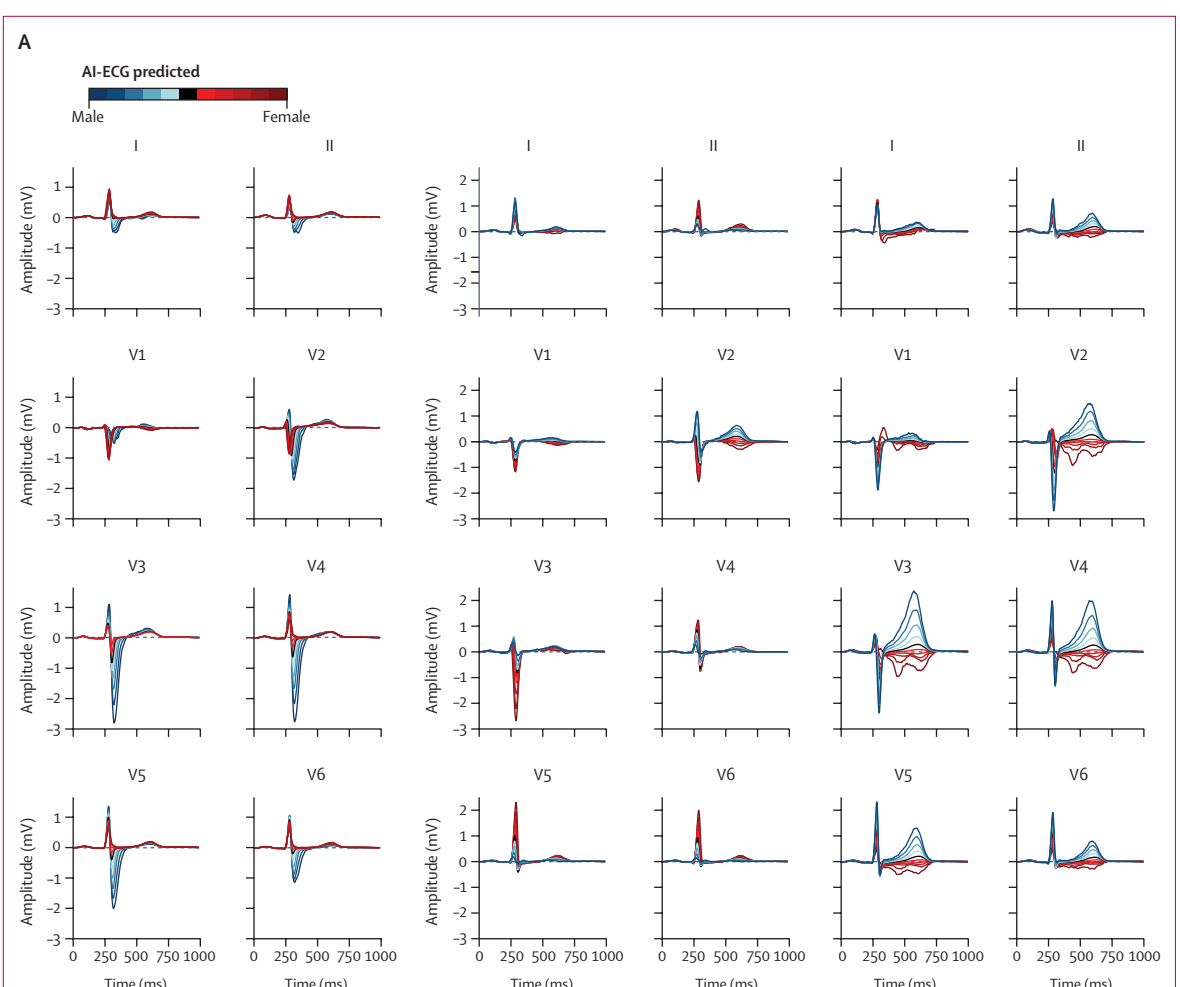

## Results
In the BIDMC cohort, 1 163 401 ECGs were available from 189 539 individuals (90 792 [47·9%] were male and 98 747 [52·1%] were female). Mean follow-up period was 3·41 years (IQR 4·08). 34 938 (18·4%) individuals died during follow-up (table). In the whole test set, AI-ECG accurately identified sex (AUROC 0·943 [95% CI 0·942–0·943]). Performance was reduced in higher risk

cohorts (older than 40 years, inpatients, and patients with prevalent cardiovascular disease) and in ECGs with abnormalities (broad QRS complex [QRS duration >120 ms] and tachycardia; figure 1A; appendix pp 4–5).

We externally validated the model in a cohort of volunteers from the UK Biobank, with 42 386 ECGs from the same number of individuals available (20 538 [48·5%] were male and 21 848 [51·6%] were female; table). Mean follow-up was 4·83 years (IQR 1·57). 784 (1·9%) individuals died during follow-up. AI-ECG sex prediction performance was maintained in the UK Biobank population (AUROC 0·971 [95% CI 0·969–0·972]). AI-ECG sex prediction distributions in both cohorts demonstrate excellent sex prediction accuracy, particularly at the extremes of the distributions (where sex prediction <0·2 or >0·8 accuracy was 93·1% in BIDMC and 95·5% in UK Biobank), and predictive accuracy did not differ between males and females (figure 1B, C).

We used a variational autoencoder to visualise the features most correlated with the AI-ECG sex predictions. The three latent features most highly correlated with AI-ECG sex predictions represented QRS morphology, amplitude, T-wave morphology, and QT interval (figure 2A). In a second approach, we created median beats from the 10-s ECGs in the BIDMC test set and created mean representations of groups of ECGs. Figure 2B shows mean representations of the 1000 ECGs with the most confident male and female AI-ECG sex predictions. These results similarly demonstrate the importance of QRS duration, amplitude, T-wave morphology, and QT interval in AI-ECG sex predictions. Finally, we assessed the correlation of established ECG parameters with AI-ECG sex prediction (figure 2C). These three methods highlight QRS duration, T-wave morphology, QTc interval, heart rate, and QRS voltage as the five most important factors in the derivation of AI-ECG sex prediction.

Using the BIDMC cohort test set, we investigated the association between discordant sex predictions and all-cause mortality. Separately for males and females, quartiles of sex discordance score were identified. When analysing ECGs from the entire cohort, which included a




(Figure 2 continues on next page)

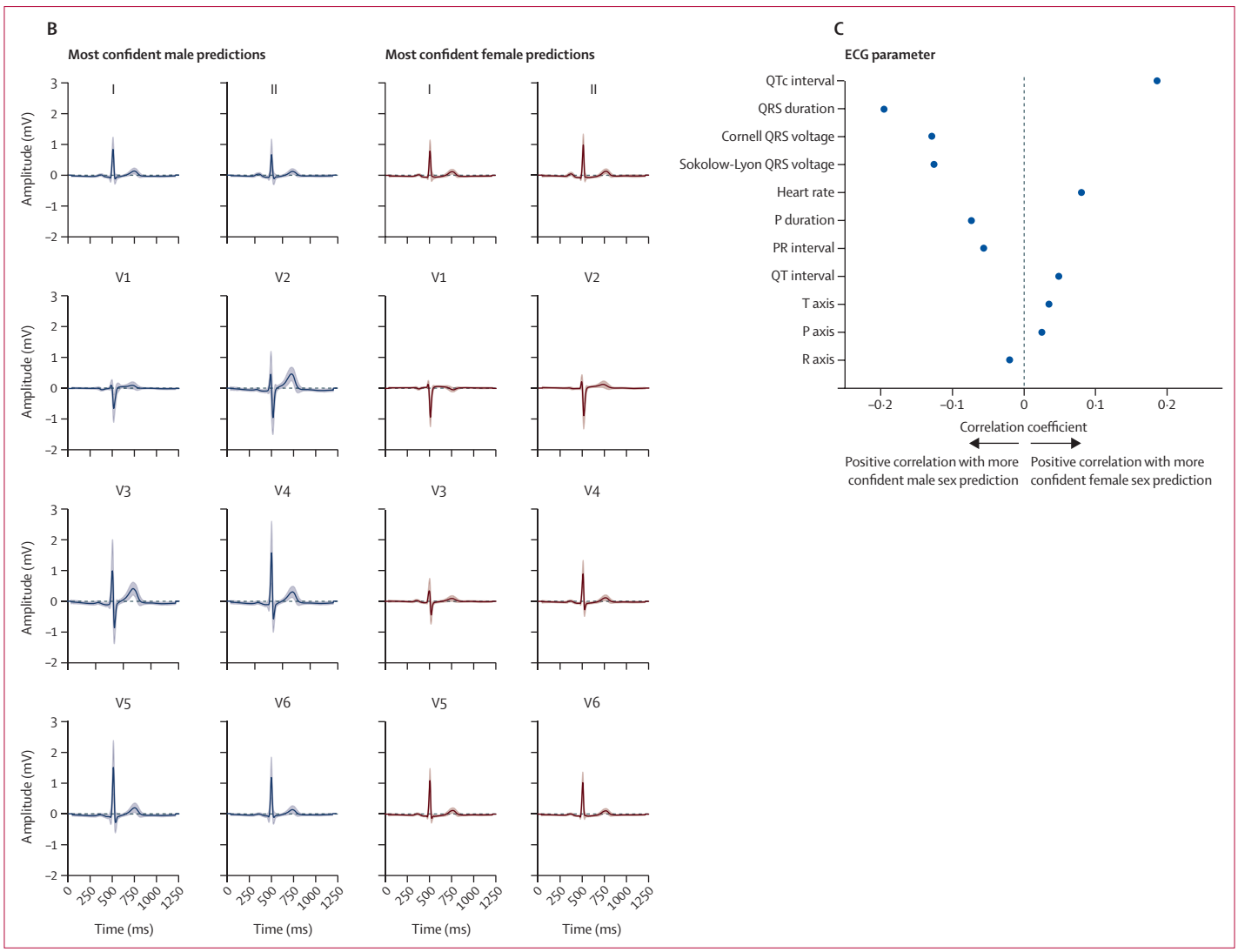



***Figure 2:*** **ECG morphologies associated with AI-ECG sex prediction**
(A) A variational autoencoder was used to identify the most important morphological features in AI-ECG sex prediction; each subpanel shows one of three latent features. (B) Mean (SD; shaded region) ECG waveforms for the 1000 highest and lowest AI-ECG sex predictions (ie, most confident male and female AI-ECG predictions) from the BIDMC test set. QRS duration, voltage, T-wave morphology, and QT interval were identified as the most important morphologies associated with AI-ECG sex prediction. (C) Univariate correlation was performed to evaluate the association of AI-ECG sex prediction with common ECG parameters. Positive correlation coefficients indicate positive correlation with more confident female sex prediction. For all correlations, p<0·001. AI-ECG=artificial intelligence-enhanced electrocardiography. BIDMC=Beth Israel Deaconess Medical Center. ECG=electrocardiogram. ms=millisecond.

mixture of individuals at high risk and low cardiovascular risk, increased sex discordance score was associated with increased all-cause mortality in both sexes (appendix p 11). We further investigated this association with covariate-adjusted Cox models. In the unselected BIDMC cohort, we found increased sex discordance score was associated with significantly greater all-cause mortality (HR 1·22 [95% CI 1·19–1·25] for males and HR 1·17 [1·14–1·20] for females), cardiovascular mortality (HR 1·20 [95% CI 1·14–1·25] for males and HR 1·25 [1·19–1·31] for females), and non-cardiovascular mortality (HR 1·23 [95% CI 1·20 –1·25] for males and HR 1·15 [1·12–1·18] for females) in males and females (figure 3A).

In line with our hypothesis, we proceeded to evaluate ECGs from low-risk populations only. Increased sex discordance score was associated with an age-adjusted increased risk of cardiovascular death only in females, but not males (figure 3B–D); we evaluated BIDMC outpatients with normal ECGs (HR 1·00 [95% CI 0·63–1·58] for males and 1·78 [1·18–2·70] for females), healthy individuals with normal ECGs (0·93 [0·49–1·74] for males and 1·87 [95% CI 1·11–3·16] for females), and patients aged 40 years or younger (1·05 [0·77–1·44] for males and 1·84 [1·38–2·46] for females). In evaluating the UK Biobank cohort, which is a low-risk volunteer cohort, the same pattern was seen (figure 3E).

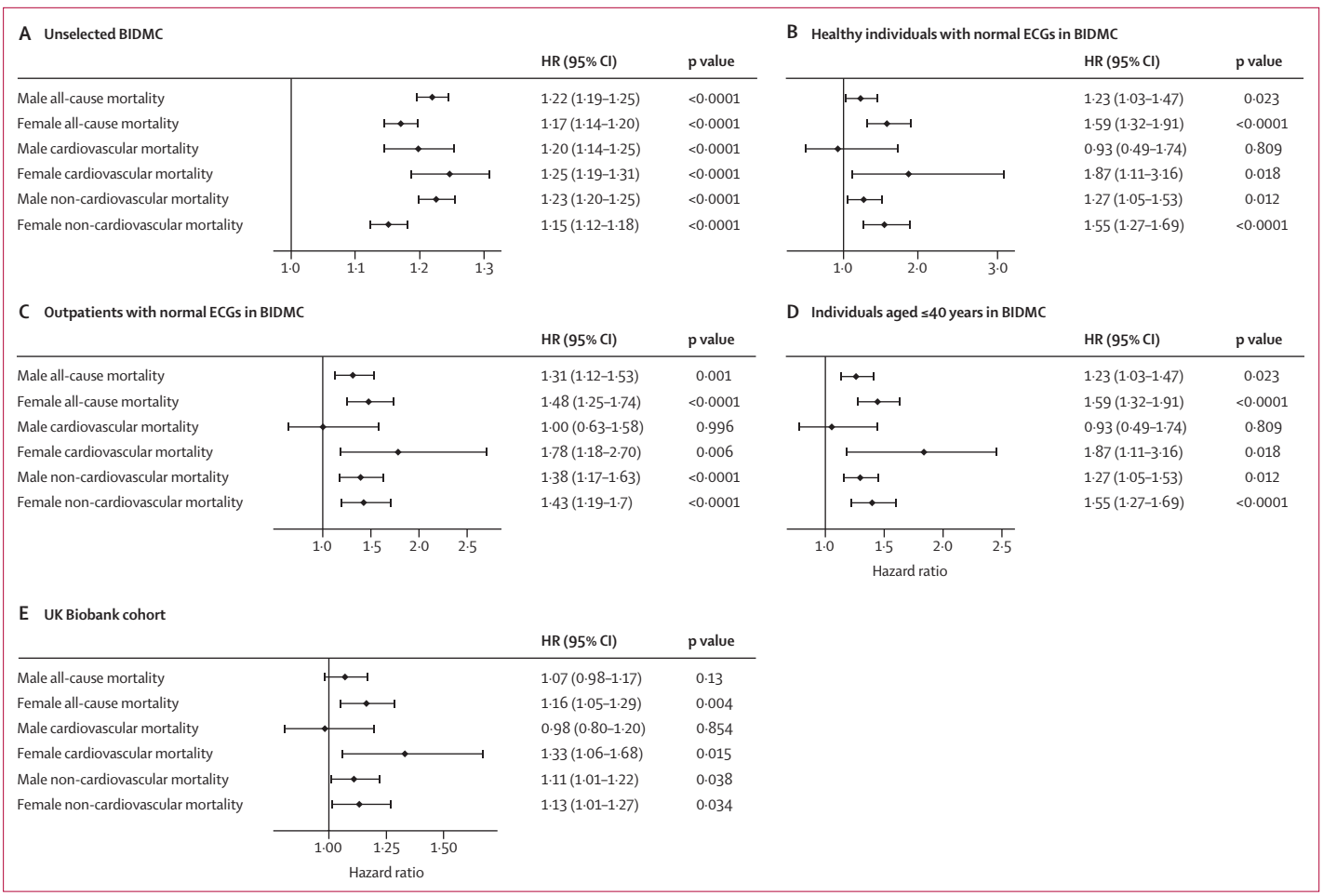



***Figure 3:*** **Association of sex discordance score with all-cause mortality, cardiovascular mortality, and non-cardiovascular mortality**
Covariate-adjusted, sex stratified Cox models were fit to evaluate the association of sex discordance score (continuous) with all-cause mortality, cardiovascular mortality, and non-cardiovascular mortality in the unselected BIDMC cohort (A), BIDMC healthy individuals with normal ECGs (B), BIDMC outpatients with normal ECGs (C), BIDMC individuals aged ≤40 years (D), and UK Biobank cohort (E). BIDMC=Beth Israel Deaconess Medical Center. ECG=electrocardiogram. HR=hazard ratio.

A further analysis was performed, in which each cohort was divided into groups by sex discordance score and the impact of sex in these groups was explored. Our findings were consistent with those aforementioned (appendix pp 7–8). To further evaluate the phenotypic differences associated with sex discordance score, we investigated correlations with diverse phenotypes. We found differential phenotypic associations between the sexes. In females, greater sex discordance score was associated with reduced left ventricular ejection fraction, and increased left ventricular, right ventricular, left atrial, and aortic size (ie, volume or dimension), and increased left ventricular wall thickness, whereas in males greater sex discordance score was associated with the opposite findings, in particular reduced aortic dimensions (figure 4).

We additionally performed a phenome-wide association study using the UK Biobank cohort. Cardiac MRI findings were similar to the echocardiography findings described previously, with greater sex discordance score in females being associated with reduced left ventricular ejection fraction, and increased ventricular, atrial, and aortic size (appendix p 12). Non-cardiac associations of greater sex discordance score in females included reduced body impedance (suggesting increased muscular mass) and body fat percentage, increased handgrip strength, and increased bone area. In men these included reduced brain volumes and bone area, increased fat mass, and increased body impedance (appendix p 13).

To investigate the diseases associated with sex discordance score, we performed a phenome-wide association study of sex discordance score against 2304 Phecodes in the BIDMC test set. Again, we found differential disease associations between the sexes. The most significant associations in females with high sex discordance score were broadly cardiovascular and respiratory diseases, including heart failure, myocardial infarction, cardiomyopathy, and respiratory failure

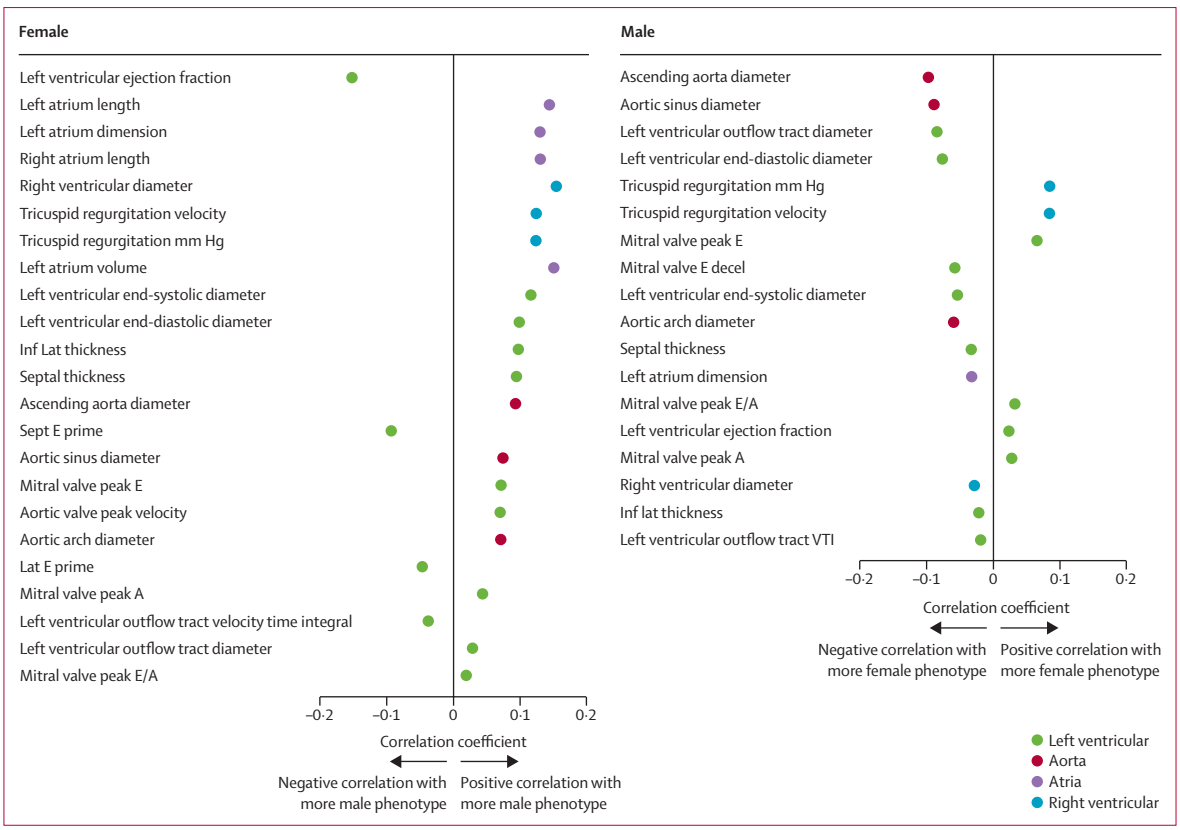



***Figure 4:*** **Association of sex discordance score with echocardiographic parameters**
Sex-stratified, univariate correlation of sex discordance score (continuous) and echocardiographic parameters was performed in the Beth Israel Deaconess Medical Center test set. A higher sex discordance score is indicative of an ECG phenotype closer to the other sex; therefore, positive correlations indicate the echocardiographic parameter is positively associated with an ECG phenotype of the opposite sex. Comparisons meeting significance after Bonferroni correction are shown. ECG=electrocardiogram.

(appendix p 14). By contrast, the strongest associations in males with high sex discordance score were broadly non-cardiac conditions, including liver pathologies.

We also investigated the association of sex discordance score with future disease in BIDMC individuals with normal ECGs, finding different associations between the sexes. In covariate-adjusted Cox models, sex discordance score was significantly associated with future heart failure in females but not males (appendix p 9). The same pattern was seen for myocardial infarction (appendix p 9). In the UK Biobank cohort, there were no differences in future heart failure or myocardial infarction (appendix p 9). Secondary analyses using a Fine and Gray model adjusted for the competing risk of death yielded similar results (appendix p 10).

To investigate whether there were underlying genetic associations with sex discordance score, we performed a genome-wide association study in the UK Biobank cohort (appendix p 15). In females, we found significant loci adjacent to insulin-like growth factor 1 receptor (*IGF1R*) and N-myc downregulated gene family *member 4*, which have previously been associated with QRS duration, left ventricular mass, and left ventricular septal thickness, and sex hormone binding globulin, testosterone, and QT interval, respectively.[17–20] In males, sex discordance score was associated with variants adjacent to single-stranded DNA binding protein 3, protein tyrosine phosphatase receptor type N2, *IGF1R*, and lipopolysaccharide induced TNF factor. The associations were broadly similar to those for females, with the addition of lipid levels and bone density.[21–23] In males, 11·0% of phenotypic variance was explained by genetic variation, whereas for females this was 9·8%.

## Discussion

In this retrospective cohort study, we describe an AI-ECG model for sex classification, validated in two diverse populations, that identifies a novel digital biomarker of AI-ECG sex discordance score. Importantly, in lower risk populations, discordance in AI-ECG-predicted sex and biological sex identifies higher cardiovascular risk in females but not in males.

We showed that AI-ECG can identify sex highly accurately, as previously described.[9,10] The AI model performed well across both high cardiovascular risk and low cardiovascular risk groups, but the model

performance was greatest in relatively low cardiovascular risk populations, in which there are fewer ECG changes due to pathology. This finding was particularly notable in the UK Biobank cohort, where the model had greater performance, despite this being the external dataset, likely because the UK Biobank is a volunteer cohort that is healthier than the BIDMC derivation cohort. We specifically showed improved model performance in normal ECGs compared with ECGs with abnormalities. Any abnormalities on an ECG, however subtle, might obscure sex-specific ECG morphologies, thus making it more challenging for the model to identify sex, which would explain the superior model performance in the lower risk and more healthy subgroups. This is a key finding that informed subsequent analyses.

Through explainability analyses, we demonstrated with three methods that the five ECG features most correlated with AI-ECG sex prediction were QRS duration, QRS voltage, T-wave morphology, QT interval, and heart rate, which are broadly in line with previous descriptions.[8]

By use of the continuous sex classification model predictions, we developed the AI-ECG sex discordance score. A previous study described binary sex misclassification as being associated with higher mortality for both males and females.[10] Their key finding was derived in a hospital population. Considering the whole cohort, increased sex discordance score in our study was also associated with worse outcomes in the sexes. As discussed in the section above, the likely explanation for this observation is because individuals at high cardiovascular risk are more likely to have subtle or overt ECG abnormalities, which reduce the model's ability to detect sex-specific ECG morphologies, leading to more sex misclassifications in this group. Therefore, in higher risk groups, sex misclassification and sex score discordance are manifestations of the presence of ECG abnormalities, which, as expected, result in a higher risk of adverse outcomes for males and females.

We aimed to explore the prognostic implications of sex misclassification and sex score discordance in ECGs without abnormalities; therefore, we chose to focus additional analyses on subsets of individuals at low cardiovascular risk (ie, with mostly normal ECGs). We report a novel finding of sex differential effects, with females with higher sex discordance scores having greater cardiovascular mortality, without a similar pattern observed in males. This female-specific pattern was also validated in a second large cohort (UK Biobank). Our findings support our hypothesis that in populations at low cardiovascular risk, increased sex discordance score is associated with adverse cardiovascular prognosis in females, but not in males.

Through phenome-wide and genome-wide association analyses in both cohorts, we found the AI-ECG sex discordance score to reflect physiological differences between sexes. Even after adjusting for body size, females with greater sex discordance score had a more male phenotype, with greater heart size, reduced fat mass, and increased bone density. The opposite was true for males. These findings are consistent with previously described phenotypic differences between males and females.[24,25]

As males in general have an adverse cardiovascular prognosis compared with females,[25] these findings support our observation that females with greater sex discordance score (and therefore a more male ECG phenotype) have a comparatively worse cardiovascular prognosis. As hypothesised, the opposite was not observed in males.

Interestingly, our genome-wide association study suggested that there are genetic determinants of sex discordance score that broadly reflect the phenotypes identified in the phenome-wide association studies. We, therefore, showed that genetic modifiers on autosomal chromosomes could contribute to sex discordance score, which might relate to the increased risk seen in females with a more male ECG phenotype.

We identified heart failure and myocardial infarction as potential contributors to the increased cardiovascular mortality seen in females with greater sex discordance scores. The associations with liver disease shown in males might be due to the established bidirectional association of liver disease with sex hormones.[26]

Overall, our findings support our hypothesis that there is a continuum of sex-related ECG morphologies, and the ECG contains information on sex-specific cardiovascular risk and biological differences beyond the physiological changes due to body shape and size.

Concerns have been expressed over AI having the potential to exacerbate existing societal, racial, and sex biases.[27] However, AI models might, instead of exacerbating biases, have the potential to aid in overcoming them. Although cardiovascular disease rates are generally higher in males,[25] cardiovascular disease is also the leading cause of death in females.[28] Under the current paradigm, females receive more misdiagnoses and less preventive care and aggressive risk factor modification.[2,3,29] This might be in part due to the perceived lower cardiovascular risk in females compared with males.[4] The current dichotomous paradigm could in part be responsible for the incorrect generalisation that females are at low risk of cardiovascular disease, and in addressing this inequality in cardiovascular care, we propose a shift to an appreciation of sex-specific cardiovascular risk as a continuum rather than as a dichotomous trait when considering cardiovascular risk factors and risk stratification.

A potential way to address the disadvantage to females would be to integrate sex discordance score into the electronic health records, and flag females for whom an adverse (ie, more male) ECG phenotype is detected. This personalised approach could help focus preventive treatments to those females at highest risk. The ECG is a ubiquitous and inexpensive investigation, already performed millions of times a day.[30] Application of AI to

one of the most commonly performed cardiovascular investigations is a natural extension of the current clinical paradigm and could be used to help overcome the widespread sex bias that exists in health care.[4]

This study did have some limitations. Our analysis focused on self-identified sex and, therefore, could not capture any complexities between self-identified sex and biological sex. We were unable to effectively study the associations between sex discordance score and sex hormone concentrations as the blood results were drawn several years in advance of the ECG recordings in the UK Biobank. Furthermore, we were unable to assess for the impact of significant female life events (menarche, pregnancy, or menopause) on our findings. With respect to menopause specifically, the consistent findings in both predominantly postmenopausal (UK Biobank) and premenopausal (BIDMC subgroup aged <40 years) groups suggest that our key findings are applicable regardless of menopausal status. Validation of our genome-wide association study findings through replication studies in additional populations is needed.

In conclusion, we describe sex discordance score as a novel AI-ECG biomarker capable of identifying females with disproportionately elevated cardiovascular risk. Sex discordance score could help uncover potential underlying mechanisms of cardiovascular disease among females, and future clinical studies are needed to determine its potential utility in identification of females at higher risk of cardiovascular events. Rather than exacerbating biases, AI-ECG has the potential to reduce female cardiovascular health inequalities.

For the **UK Biobank data** see http://www.ukbiobank.ac.uk

For the **AI-ECG code** see https://github.com/antonior92/automatic-ecg-diagnosis

#### Contributors

AS and FSN conceptualised the study; AS, LP, and JWW accessed and verified the data. AS, LP, FSN, KP, ES, KAM, AHR, ALPR, and JSW developed the methodology and performed data analysis; NSP, DBK, JWW, and AS collected the data; AS wrote the first draft; AS, LP, FSN, NSP, KAM, and ES acquired funding. AS, ES, KAM, JEH, NSP, JSW, UT, DBK, JWW, AHR, ALPR, and FSN interpreted the data. All authors critically reviewed and commented on the manuscript. All authors had access to the data in the study and had final responsibility for the decision to submit for publication.

#### Declaration of interests

AS is funded by a British Heart Foundation (BHF) clinical research training fellowship (FS/CRTF/21/24183) and National Institute for Health Research (NIHR) Academic Clinical Lectureship. FSN and NSP are supported by the BHF (RG/F/22/110078). LP is funded by a Medical Research Council (MRC) clinical research training fellowship (MR/Y000803/1). KAM is supported by a BHF fellowship (FS/IPBSRF/22/27059). ES is supported by an EJP RD Research Mobility Fellowship (European Reference Networks), the Sir Jules Thorn Charitable Trust, and was previously supported by an FWO PhD Fellowship. JEH is supported by the National Institutes for Health (R01 HL160003, R01 HL168889, and K24 HL153669). ALPR is supported in part by CNPq (465518/2014-1, 310790/2021-2, and 409604/2022-4) and by FAPEMIG (PPM-00428-17, RED-00081-16, and PPE-00030-21). UT is supported by the MRC (MR/W023830/1). AS, LP, FSN, AHR, and ALPR are supported by the Academy of Medical Sciences (NGR1\1746). JWW was previously on the advisory board for Heartcor solutions and reports research funding from Anumana. JSW reports research grants from the Sir Jules Thorn Charitable Trust, MRC, BHF, Bristol Myers Squibb, and Pfizer; consulting fees from Bristol Myers Squibb, Pfizer, Foresite Labs, Health Lumen, and Tenaya; honoraria from Global Heart Hub; and is on the clinical advisory group for Cardiomyopathy UK. FSN reports speaker fees from GE healthcare and is on the advisory board for AstraZeneca. UT has received fees for educational content from Chiesi and has roles within the British Cardiovascular Society, Royal Society, and DCM SHaRe registry. The authors acknowledge support from Imperial's BHF Centre for Excellence Award (RE/18/4/34215 and RE/24/130023) and NIHR Imperial Biomedical Research Centre. KP and DBK declare no competing interests.

#### Data sharing

The full UK Biobank data used in this study are available to researchers upon application. The Beth Israel Deaconess Medical Center (BIDMC) dataset is restricted due to ethical limitations. The code underlying the artificial intelligence-enhanced electrocardiography (AI-ECG) model architecture used in this study is available. The remaining programming code will be made available upon reasonable request to the corresponding author.

#### Acknowledgments

This research has been conducted using the UK Biobank Resource under application numbers 48666 and 47602. The authors would also like to thank InSIGHT Core in the Center for Healthcare Delivery Science at BIDMC for assistance in obtaining primary data. For the purpose of open access, the authors have applied a creative commons attribution (CC BY) licence to any author accepted manuscript version arising.

#### References

1 Peters SAE, Muntner P, Woodward M. Sex differences in the prevalence of, and trends in, cardiovascular risk factors, treatment, and control in the United States, 2001 to 2016. *Circulation* 2019; **139:** 1025–35.

2 Pope JH, Aufderheide TP, Ruthazer R, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. *N Engl J Med* 2000; **342:** 1163–70.

3 Agarwala A, Michos ED, Samad Z, Ballantyne CM, Virani SS. The use of sex-specific factors in the assessment of women's cardiovascular risk. *Circulation* 2020; **141:** 592–99.

4 Leifheit-Limson EC, D'Onofrio G, Daneshvar M, et al. Sex differences in cardiac risk factors, perceived risk, and health care provider discussion of risk and risk modification among young patients with acute myocardial infarction: the VIRGO study. *J Am Coll Cardiol* 2015; **66:** 1949–57.

5 Goff DC Jr, Lloyd-Jones DM, Bennett G, 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.

6 Vosberg DE, Syme C, Parker N, Richer L, Pausova Z, Paus T. Sex continuum in the brain and body during adolescence and psychological traits. *Nat Hum Behav* 2021; **5:** 265–72.

7 Vosberg DE, Pausova Z, Paus T. The genetics of a "femaleness/maleness" score in cardiometabolic traits in the UK Biobank. *Sci Rep* 2023; **13:** 9109.

8 Moss AJ. Gender differences in ECG parameters and their clinical implications. *Ann Noninvasive Electrocardiol* 2010; **15:** 1–2.

9 Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol* 2019; **12:** e007284.

10 Siegersma KR, van de Leur RR, Onland-Moret NC, et al. Deep neural networks reveal novel sex-specific electrocardiographic features relevant for mortality risk. *Eur Heart J Digit Health* 2022; **3:** 245–54.

11 Sau A, Ribeiro AH, McGurk KA, et al. Prognostic significance and associations of neural network-derived electrocardiographic features. *Circ Cardiovasc Qual Outcomes* 2024; **17:** e010602.

12 Sau A, Pastika L, Sieliwonczyk E, et al. Artificial intelligence-enabled electrocardiogram for mortality and cardiovascular risk estimation: an actionable, explainable and biologically plausible platform. *Lancet Digit Health* 2024; **6:** e791–802.

13 Sudlow C, Gallacher J, Allen N, 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:** e1001779.

14 Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 2020; **11:** 1760.

15 Stabenau HF, Waks JW. BRAVEHEART: open-source software for automated electrocardiographic and vectorcardiographic analysis. *Comput Methods Programs Biomed* 2023; **242:** 107798.

16 Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999; **94:** 496–509.

17 Prins BP, Mead TJ, Brody JA, et al. Exome-chip meta-analysis identifies novel loci associated with cardiac conduction, including ADAMTS6. *Genome Biol* 2018; **19:** 87.

18 Ruth KS, Day FR, Tyrrell J, et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med* 2020; **26:** 252–58.

19 van Duijvenboden S, Ramírez J, Young WJ, et al. Genomic and pleiotropic analyses of resting QT interval identifies novel loci and overlap with atrial electrical disorders. *Hum Mol Genet* 2021; **30:** 2513–23.

20 Tadros R, Francis C, Xu X, et al. Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat Genet* 2021; **53:** 128–34.

21 Graham SE, Clarke SL, Wu KH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* 2021; **600:** 675–79.

22 Kim SK. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PLoS One* 2018; **13:** e0200785.

23 Nielsen JB, Rom O, Surakka I, et al. Loss-of-function genomic variants highlight potential therapeutic targets for cardiovascular disease. *Nat Commun* 2020; **11:** 6417.

24 Devereux RB, Lutas EM, Casale PN, et al. Standardization of M-mode echocardiographic left ventricular anatomic measurements. *J Am Coll Cardiol* 1984; **4:** 1222–30.

25 Zhu K, Briffa K, Smith A, et al. Gender differences in the relationships between lean body mass, fat mass and peak bone mass in young adults. *Osteoporos Int* 2014; **25:** 1563–70.

26 Bannister P, Losowsky MS. Sex hormones and chronic liver disease. *J Hepatol* 1988; **6:** 258–62.

27 Howard A, Borenstein J. The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. *Sci Eng Ethics* 2018; **24:** 1521–36.

28 Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation* 2015; **131:** e29–322.

29 Mosca L, Linfante AH, Benjamin EJ, et al. National study of physician awareness and adherence to cardiovascular disease prevention guidelines. *Circulation* 2005; **111:** 499–510.

30 Reichlin T, Abächerli R, Twerenbold R, et al. Advanced ECG in 2016: is there more than just a tracing? *Swiss Med Wkly* 2016; **146:** w14303.