SUPPLEMENTAL MATERIAL

Electrocardiogram-based Deep Learning and Clinical Risk Factors to Predict Incident Atrial Fibrillation

Index

Methods. Supplemental Methods

ECG-AI training

The input to ECG-AI was a 12-lead ECG in XML format, in which each lead is represented as a vector of relative voltage amplitude over 10 seconds sampled at either 250 or 500Hz. ECGs sampled at 250Hz were upsampled to 500Hz to allow for a uniform input shape. For individuals with multiple ECGs within the baseline window, the ECG used for training at each epoch was randomly selected among all ECGs for that individual.

Although incident AF can be modeled as a binary classification task,[10,11] such an approach does not account for censoring and may lead to poor sample-level calibration. To account for this, we expanded methods proposed by Gensheimer and Narasimhan[21] to develop and implement an encoding and loss function that are aware of both the time until the event (i.e., AF) and missingness introduced by right censoring. The encoding quantizes the total follow-up time into *n* bins, in our case *n* = 25, so each time bin spans ~72 days. Each individual is represented by two binary vectors of length *n*, one representing a binary mask for dates that are censored and the other represent AF event status via one-hot encoding. The censoring vector, $V_{\text{CENSOR,}}$ is 0 for every time bin in which the individual was censored and 1 for time bins during which they remained in the risk set. The event vector, $V_{\text{AF,}}$ is 1 if an AF diagnosis falls within that time bin, otherwise it is 0. ECG-AI emits a vector, $V_{\text{PREDICT,}}$ of length *n* representing AF survival probability at each time bin. Our loss function minimizes the negative log likelihood of ECG-AI's predictions. The likelihood is factored into contributions from time bins survived and time bins with events. Concretely, we minimize $\mathcal{L} = -\sum log(\mathcal{L}_{survival} +$

$\mathcal{L}_{event}$), where

$$\mathcal{L}_{survival} = 1 - (V_{AF} * V_{PREDICT})$$

$$\mathcal{L}_{event} = (V_{CENSOR} * V_{PREDICT}) + (1 - V_{CENSOR})$$

In this way, censored individuals do not contribute to the loss at time bins after censoring (e.g., earliest of death or last follow-up). The first time bin is reserved to encode events prior to the start of follow up. The model predicts an independent survival probability (a sigmoid activation) for each time bin, therefore making no assumptions of proportional hazards and facilitating modeling of discontinuous survival curves.

ECG-AI maps between two distinct time scales, the 10 second ECG and the 5-year predicted survival curve. This mapping is learned by a 1-dimensional convolutional neural network (CNN) with ECG input of shape (5000, 12) and output vector of 25 survival probabilities, one for each time bin. Architecture hyperparameters including width, depth, activation, normalization, and regularization were chosen via Bayesian hyperparameter optimization.[41] The best performing architecture contains over 18 million neurons, uses rectified linear unit (ReLU) activations, 16 layers of convolutions, each with a 21 timestep convolutional kernel, dense residual connections, Poincare normalization[42] on convolutional layers, max rather than average pooling and 208 neurons in the fully connected layer (see **Figure I** in the **Supplement**). ECG-AI was optimized using rectified Adam stochastic gradient descent,[43] with a learning rate of 2e-4, batch size of 48 and dropout[44] rates of 0.2 for convolutional layers and 0.5 for fully connected layers. Model training was complete after 4 hours using an Nvidia V100 (Santa Clara, CA) graphical processing unit.

Site-specific censoring in UK Biobank

In the UK Biobank, the date of last follow-up was dependent upon the availability of linked hospital data, and was therefore defined as February 28, 2021 for participants enrolled in England and Scotland, and February 28, 2018 for participants enrolled in Wales.

Saliency mapping

Saliency maps highlight the sections of the ECG where the smallest changes in input voltage lead to the greatest changes in AF prediction risk. Saliency is defined as the model output gradient with respect to an input ECG. Efficient computation is possible with the same backpropagation machinery used in model training except during training the gradient is of the loss function rather than the model output, and it is taken with respect to the model weights rather than the model input. Both cases rely on the chain rule and the automatic differentiation capabilities of the Python package “Tensorflow”. A median ECG waveform is overlaid on the ECG saliencies. QRS complex detection and bilinear interpolation map all ECGs to the same phase and heart rate. The median voltage is computed across these beats for each individual and for every time bin in the cardiac cycle. We used a similar approach to plot the median ECG waveform for 1,000 randomly selected individuals within binary strata of ECG-AI risk defined using our standard risk thresholds (i.e., low = <2.5% 5-year AF risk versus high = >5% 5-year AF risk).

Secondary analyses

We performed multiple secondary analyses. First, given increasingly availability of mobile devices capable of producing single-lead ECG tracings,[45] we assessed versions of ECG-AI and CH-AI provided only with lead I, and then separately lead II, of the standard 12-lead ECG. Second, since age is a very strong AF risk factor,[13,14,25] we compared the discrimination of CH-AI and CHARGE-AF within subgroups of age, with groups divided at approximate tertiles of the sample age distribution. Third, since AF risk estimation may be particularly important among individuals with heart failure and stroke, we assessed the performance of each model among individuals with prevalent heart failure, and separately prevalent stroke, at the time of ECG. We limited this analysis to the BWH test set given very limited event rates in the MGH and UK Biobank test sets. Fourth, to assess whether ECG-AI behavior varied according to clinical risk factor burden, we generated saliency maps across binary strata of CHARGE-AF risk and ECG-AI risk. Fifth, we assessed the Pearson correlation between the linear predictors of ECG-AI and CHARGE-AF, as a surrogate for whether ECG-AI may implicitly incorporate clinical AF risk factors information extracted from the ECG. Sixth, to quantify the relative contributions of specific clinical risk factors to discrimination using CH-AI, we fit Cox proportional hazards models in which covariates were added sequentially to ECG-AI. Covariates were added in groups, namely: a) age and sex, b) age, sex, systolic blood pressure, diastolic blood pressure, height, and weight, and c) age, sex, systolic blood pressure, diastolic blood pressure, height, weight, White race, smoking, anti-hypertensive use, diabetes, myocardial infarction, and heart failure (i.e., sex and each component of the CHARGE-AF score).

Table I. Performance of ECG-AI with versus without secondary tasks

Model Task(s)	**C-index in MGH test**	**C-index in BWH test**	**C-index in UK Biobank**
Incident AF	0.753	0.708	0.603
Incident AF, age, and sex	0.760	0.727	0.643
Incident AF, age, sex, and AF in the diagnostic statement	0.775	0.754	0.639
Incident AF, age, sex, and AF in the diagnostic statement with Bayesian hyperoptimization (final model)	0.788	0.742	0.690

Table II. Performance of ECG-AI in alternative training and test samples

Training Sample*	**5-year AUROC**
Sample 1	0.784 (0.753-0.815)
Sample 2	0.723 (0.690-0.757)
Sample 3	0.742 (0.707-0.779)
Sample 4	0.775 (0.746-0.805)
Sample 5	0.781 (0.751-0.811)
*Reflects ECG-AI training using a random resampling of the original MGH training and test sets, with the listed 5-year AUROC calculated in the respective test set independent of model training AUROC = area under the receiver operating characteristic curve	

Table III. Clinical factor definitions

Phenotype	Data fields	Field names	Data codes	Data code definitions
UK Biobank				
Atrial fibrillation	20002	Non-cancer illness code, self-reported	1471, 1483	Atrial fibrillation, Atrial flutter
Atrial fibrillation	20004	Operation code, self-reported	1524	Cardioversion
Atrial fibrillation	41202 41204 40001 40002	Diagnoses - main ICD10 Diagnoses – secondary ICD10 Underlying (primary) cause of death: ICD10 Contributory (secondary) cause of death: ICD10	I48, I48.0, I48.1, I48.2, I48.3, I48.4, I48.9	Atrial fibrillation and flutter, Paroxysmal atrial fibrillation, Persistent atrial fibrillation, Chronic atrial fibrillation, Typical atrial flutter, Atypical atrial flutter, Atrial fibrillation and flutter, unspecified
Atrial fibrillation	41203 41205	Diagnosis - main ICD9 Diagnoses - secondary ICD9	4273	Atrial fibrillation and flutter
Atrial fibrillation	41200 41210	Operative procedures – main OPCS Operative procedures – secondary OPCS	K57.1, K62.1, K62.2, K62.3, K62.4, X50.1, X50.2	Percutaneous transluminal ablation of atrioventricular node, Percutaneous transluminal ablation of pulmonary vein to left atrium conducting system, Percutaneous transluminal ablation of atrial wall for atrial flutter, Percutaneous transluminal ablation of conducting system of heart for atrial flutter NEC, Percutaneous transluminal internal cardioversion NEC, Direct current cardioversion, External cardioversion NEC
Diabetes	2443	Diabetes diagnosed by doctor	1	Yes
Diabetes	20002	Non-cancer illness code, self-reported	1220, 1221, 1222, 1223	Diabetes, Gestational diabetes, Type 1 diabetes, Type 2 diabetes
Diabetes	2986	Insulin use within one year	1	Started insulin within one year diagnosis of diabetes - Yes
Diabetes	6177	Medication for cholesterol, blood pressure or diabetes	3	Insulin
Diabetes	6153	Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones	3	Insulin
Diabetes	41202 41204 40001 40002	Diagnoses - main ICD10 Diagnoses – secondary ICD10 Underlying (primary) cause of death: ICD10 Contributory (secondary) cause of death: ICD10	E10, E10.0, E10.1, E10.2, E10.3, E10.4, E10.5, E10.6, E10.7, E10.8, E10.9, E11, E11.0, E11.1, E11.2, E11.3, E11.4, E11.5, E11.6, E11.7, E11.8, E11.9, E12, E12.1, E12.8, E12.9, E13, E13.1, E13.2, E13.3, E13.5, E13.6, E13.7,	Insulin-dependent diabetes mellitus, Insulin-dependent diabetes mellitus with coma, Insulin-dependent diabetes mellitus with ketoacidosis, Insulin-dependent diabetes mellitus with renal complications, Insulin-dependent diabetes mellitus with ophthalmic complications, Insulin-dependent diabetes mellitus with neurological complications, Insulin-dependent diabetes mellitus with peripheral circulatory complications, Insulin-dependent diabetes mellitus with other specified complications, Insulin-dependent diabetes mellitus with multiple complications, Insulin-dependent diabetes mellitus with unspecified complications, Insulin-dependent diabetes mellitus without complications, Non-insulin-dependent diabetes mellitus, Non-insulin-dependent diabetes mellitus - with coma, Non-insulin-dependent diabetes mellitus - with ketoacidosis, Non-insulin-dependent diabetes mellitus - with renal complications,

Phenotype	Data fields	Field names	Data codes	Data code definitions
			E13.8, E13.9, E14, E14.0, E14.1, E14.2, E14.3, E14.4, E14.5, E14.6, E14.7, E14.8, E14.9	Non-insulin-dependent diabetes mellitus - with ophthalmic complications, Non-insulin-dependent diabetes mellitus - with neurological complications, Non-insulin-dependent diabetes mellitus - with peripheral circulatory complications, Non-insulin-dependent diabetes mellitus - with other specified complications, Non-insulin-dependent diabetes mellitus - with multiple complications, Non-insulin-dependent diabetes mellitus - with unspecified complications, Non-insulin-dependent diabetes mellitus - without complications, Malnutrition-related diabetes mellitus, Malnutrition-related diabetes mellitus with ketoacidosis, Malnutrition-related diabetes mellitus with unspecified complications, Malnutrition-related diabetes mellitus without complications, Other specified diabetes mellitus, Other specified diabetes mellitus with ketoacidosis, Other specified diabetes mellitus with renal complications, Other specified diabetes mellitus with ophthalmic complications, Other specified diabetes mellitus with peripheral circulatory complications, Other specified diabetes mellitus with other specified complications, Other specified diabetes mellitus with multiple complications, Other specified diabetes mellitus with unspecified complications, Other specified diabetes mellitus without complications, Unspecified diabetes mellitus, Unspecified diabetes mellitus with coma, Unspecified diabetes mellitus with ketoacidosis, Unspecified diabetes mellitus with renal complications, Unspecified diabetes mellitus with ophthalmic complications, Unspecified diabetes mellitus with neurological complications, Unspecified diabetes mellitus with peripheral circulatory complications, Unspecified diabetes mellitus with other specified complications, Unspecified diabetes mellitus with multiple complications, Unspecified diabetes mellitus with unspecified complications, Unspecified diabetes mellitus without complications
Diabetes	41203 41205	Diagnosis - main ICD9 Diagnoses - secondary ICD9	2500, 25000, 25001, 25009, 2501, 25011, 25019, 2503, 2504, 2505, 25099	Diabetes mellitus without mention of complication, Diabetes mellitus without mention of complication (adult-onset type), Diabetes mellitus without mention of complication (juvenile type), Diabetes mellitus without mention of compl. (adult/juvenile unspec.), Diabetes with ketoacidosis, Diabetes with ketoacidosis (juvenile type), Diabetes with ketoacidosis (adult/juvenile unspec.), Diabetes with renal manifestations, Diabetes with ophthalmic manifestations, Diabetes with neurological manifestations, Diabetes with unspecified complications (unspecified onset)
Heart Failure	20002	Non-cancer illness code, self-reported	1076, 1079	Heart failure/pulmonary oedema, Cardiomyopathy
Heart Failure	41202 41204 40001 40002	Diagnoses - main ICD10 Diagnoses – secondary ICD10 Underlying (primary) cause of death: ICD10 Contributory (secondary) cause of death: ICD10	I11.0, I13.0, I13.2, I25.5, I42.0, I42.5, I42.8, I42.9, I50, I50.0, I50.1, I50.9	Hypertensive heart disease with (congestive) heart failure, Hypertensive heart and renal disease with (congestive) heart failure, Hypertensive heart and renal disease with both (congestive) heart failure and renal failure, Hypertensive heart disease with (congestive) heart failure, Hypertensive heart and renal disease with (congestive) heart failure, Hypertensive heart and renal disease with both (congestive) heart failure and renal failure, Ischaemic cardiomyopathy, Dilated cardiomyopathy, Other restrictive cardiomyopathy, Other cardiomyopathies, Cardiomyopathy, unspecified, Heart failure, Congestive heart failure, Left ventricular failure, Heart failure, unspecified

Phenotype	Data fields	Field names	Data codes	Data code definitions
Myocardial infarction	20002	Non-cancer illness code, self-reported	1075	Heart attack/myocardial infarction
Myocardial infarction	41202 41204 40001 40002	Diagnoses - main ICD10 Diagnoses – secondary ICD10 Underlying (primary) cause of death: ICD10 Contributory (secondary) cause of death: ICD10	I21, I21.0, I21.1, I21.2, I21.3, I21.4, I21.9, I22, I22.0, I22.1, I22.8, I22.9, I23, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.8, I24.1, I25.2	Acute myocardial infarction, Acute transmural myocardial infarction of anterior wall, Acute transmural myocardial infarction of inferior wall, Acute transmural myocardial infarction of other sites, Acute transmural myocardial infarction of unspecified site, Acute subendocardial myocardial infarction, Acute myocardial infarction, unspecified, Subsequent myocardial infarction, Subsequent myocardial infarction of anterior wall, Subsequent myocardial infarction of inferior wall, Subsequent myocardial infarction of other sites, Subsequent myocardial infarction of unspecified site, Certain current complications following acute myocardial infarction, Hemopericardium as current complication following acute myocardial infarction, Atrial septal defect as current complication following acute myocardial infarction, Ventricular septal defect as current complication following acute myocardial infarction, Rupture of cardiac wall without hemopericardium as current complication following acute myocardial infarction, Rupture of chordae tendinae as current complication following acute myocardial infarction, Rupture of papillary muscle as current complication following acute myocardial infarction, Thrombosis of atrium, auricular appendage, and ventricle as current complication following acute myocardial infarction, Other current complications following acute myocardial infarction, Dressler's Syndrome, Old myocardial infarction
Myocardial infarction	41203 41205	Diagnosis - main ICD9 Diagnoses - secondary ICD9	410, 410.0, 410.1, 410.2, 410.3, 410.4, 410.5, 410.6, 410.7, 410.8, 410.9, 411, 411.0, 411.1, 411.8, 412, 412.9, 429.79	Myocardial infarction, Acute myocardial infarction of other anterior wall, episode of care unspecified, Acute myocardial infarction of inferolateral wall, episode of care unspecified, Acute myocardial infarction of inferoposterior wall, episode of care unspecified, Acute myocardial infarction of other inferior wall, episode of care unspecified, Acute myocardial infarction of other lateral wall, episode of care unspecified, True posterior wall infarction, episode of care unspecified, Subendocardial infarction, episode of care unspecified, Acute myocardial infarction of other specified sites, episode of care unspecified, Acute myocardial infarction of unspecified site, episode of care unspecified, Postmyocardial infarction syndrome, Intermediate coronary syndrome, Other acute and subacute complications of ischemic heart disease, Old myocardial infarction, Ill-defined descriptions and complications of heart disease, other
Mass General Brigham				
Diabetes	Diagnoses	ICD9	249, 249.01, 249.1, 249.11, 249.2, 249.21, 249.3, 249.31, 249.4, 249.41, 249.5, 249.51, 249.6, 249.61, 249.7, 249.71, 249.8, 249.81, 249.9,	Secondary diabetes mellitus without mention of complication, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus without mention of complication, uncontrolled, Secondary diabetes mellitus with ketoacidosis, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with ketoacidosis, uncontrolled, Secondary diabetes mellitus with hyperosmolarity, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with hyperosmolarity, uncontrolled, Secondary diabetes mellitus with other coma, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with other coma, uncontrolled, Secondary diabetes mellitus with renal manifestations, not stated as

Phenotype	Data fields	Field names	Data codes	Data code definitions
			249.91, 250, 250.01, 250.02, 250.03, 250.1, 250.11, 250.12, 250.13, 250.2, 250.21, 250.22, 250.23, 250.3, 250.31, 250.32, 250.33, 250.4, 250.41, 250.42, 250.43, 250.5, 250.51, 250.52, 250.53, 250.6, 250.61, 250.62, 250.63, 250.7, 250.71, 250.72, 250.73, 250.8, 250.81, 250.82, 250.83, 250.9, 250.91, 250.92, 250.93, 357.2, 362.01, 362.02, 362.03, 362.04, 362.05, 362.06, 362.07, 366.41, 791.6	uncontrolled, or unspecified, Secondary diabetes mellitus with renal manifestations, uncontrolled, Secondary diabetes mellitus with ophthalmic manifestations, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with ophthalmic manifestations, uncontrolled, Secondary diabetes mellitus with neurological manifestations, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with neurological manifestations, uncontrolled, Secondary diabetes mellitus with peripheral circulatory disorders, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with peripheral circulatory disorders, uncontrolled, Secondary diabetes mellitus with other specified manifestations, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with other specified manifestations, uncontrolled, Secondary diabetes mellitus with unspecified complication, not stated as uncontrolled, or unspecified, Secondary diabetes mellitus with unspecified complication, uncontrolled, Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled, Diabetes mellitus without mention of complication, type I [juvenile type], not stated as uncontrolled, Diabetes mellitus without mention of complication, type II or unspecified type, uncontrolled, Diabetes mellitus without mention of complication, type I [juvenile type], uncontrolled, Diabetes with ketoacidosis, type II or unspecified type, not stated as uncontrolled, Diabetes with ketoacidosis, type I [juvenile type], not stated as uncontrolled, Diabetes with ketoacidosis, type II or unspecified type, uncontrolled, Diabetes with ketoacidosis, type I [juvenile type], uncontrolled, Diabetes with hyperosmolarity, type II or unspecified type, not stated as uncontrolled, Diabetes with hyperosmolarity, type I [juvenile type], not stated as uncontrolled, Diabetes with hyperosmolarity, type II or unspecified type, uncontrolled, Diabetes with hyperosmolarity, type I [juvenile type], uncontrolled, Diabetes with other coma, type II or unspecified type, not stated as uncontrolled, Diabetes with other coma, type I [juvenile type], not stated as uncontrolled, Diabetes with other coma, type II or unspecified type, uncontrolled, Diabetes with other coma, type I [juvenile type], uncontrolled, Diabetes with renal manifestations, type II or unspecified type, not stated as uncontrolled, Diabetes with renal manifestations, type I [juvenile type], not stated as uncontrolled, Diabetes with renal manifestations, type II or unspecified type, uncontrolled, Diabetes with renal manifestations, type I [juvenile type], uncontrolled, Diabetes with ophthalmic manifestations, type II or unspecified type, not stated as uncontrolled, Diabetes with ophthalmic manifestations, type I [juvenile type], not stated as uncontrolled, Diabetes with ophthalmic manifestations, type II or unspecified type, uncontrolled, Diabetes with ophthalmic manifestations, type I [juvenile type], uncontrolled, Diabetes with neurological manifestations, type II or unspecified type, not stated as uncontrolled, Diabetes with neurological manifestations, type I [juvenile type], not stated as uncontrolled, Diabetes with neurological manifestations, type II or unspecified type, uncontrolled, Diabetes with neurological manifestations, type I [juvenile type], uncontrolled, Diabetes with peripheral circulatory disorders, type II or unspecified type, not stated as uncontrolled, Diabetes with peripheral circulatory disorders, type I [juvenile type], not

Phenotype	Data fields	Field names	Data codes	Data code definitions
				stated as uncontrolled, Diabetes with peripheral circulatory disorders, type II or unspecified type, uncontrolled, Diabetes with peripheral circulatory disorders, type I [juvenile type], uncontrolled, Diabetes with other specified manifestations, type II or unspecified type, not stated as uncontrolled, Diabetes with other specified manifestations, type I [juvenile type], not stated as uncontrolled, Diabetes with other specified manifestations, type II or unspecified type, uncontrolled, Diabetes with other specified manifestations, type I [juvenile type], uncontrolled, Diabetes with unspecified complication, type II or unspecified type, not stated as uncontrolled, Diabetes with unspecified complication, type I [juvenile type], not stated as uncontrolled, Diabetes with unspecified complication, type II or unspecified type, uncontrolled, Diabetes with unspecified complication, type I [juvenile type], uncontrolled, Polyneuropathy in diabetes, Background diabetic retinopathy, Proliferative diabetic retinopathy, Nonproliferative diabetic retinopathy NOS, Mild nonproliferative diabetic retinopathy, Moderate nonproliferative diabetic retinopathy, Severe nonproliferative diabetic retinopathy, Diabetic macular edema, Diabetic cataract, Acetonuria
Diabetes	Diagnoses	ICD10	E08.00, E08.01, E08.10, E08.11, E08.21, E08.22, E08.29, E08.311, E08.319, E08.321, E08.329, E08.331, E08.339, E08.341, E08.349, E08.351, E08.359, E08.36, E08.39, E08.40, E08.41, E08.42, E08.43, E08.44, E08.49, E08.51, E08.52, E08.610, E08.618, E08.620, E08.621, E08.622, E08.628, E08.630, E08.638, E08.641, E08.649, E08.65, E08.69, E08.8, E08.9, E08.90, E09.00, E09.01, E09.10, E09.11, E09.21, E09.22, E09.29, E09.311, E09.319, E09.321, E09.329, E09.331,	Diabetes Mellitus Due To Underlying Condition With Hyperosmolarity without nonketotic hyperglycemic-hyperosmolar coma (NKHHC), Diabetes Mellitus Due To Underlying Condition With Hyperosmolarity With Coma, Diabetes Mellitus Due To Underlying Condition With Ketoacidosis Without Coma, Diabetes Mellitus Due To Underlying Condition With Ketoacidosis With Coma, Diabetes Mellitus Due To Underlying Condition With Diabetic Nephropathy, Diabetes Mellitus Due To Underlying Condition With Diabetic chronic kidney disease, Diabetes Mellitus Due to underlying condition with other diabetic kidney complication, Diabetes Mellitus Due To Underlying Condition With Unspecified Diabetic Retinopathy With Macular Edema, Diabetes Mellitus Due To Underlying Condition With Unspecified Diabetic Retinopathy Without Macular Edema, Diabetes Mellitus due to underlying condition with mild nonproliferative diabetic retinopathy with macular edema, Diabetes mellitus due to underlying condition with mild nonproliferative diabetic retinopathy without macular edema, Diabetes mellitus due to underlying condition with moderate nonproliferative diabetic retinopathy with macular edema, Diabetes mellitus due to underlying condition with moderate nonproliferative diabetic retinopathy without macular edema, Diabetes mellitus due to underlying condition with severe nonproliferative diabetic retinopathy with macular edema, Diabetes mellitus due to underlying condition with severe nonproliferative diabetic retinopathy without macular edema, Diabetes mellitus due to underlying condition with proliferative diabetic retinopathy with macular edema, Diabetes mellitus due to underlying condition with proliferative diabetic retinopathy without macular edema, Diabetes Mellitus Due To Underlying Condition With Diabetic Cataract, Diabetes Mellitus Due To Underlying Condition With Other Diabetic Ophthalmic Complication, Diabetes Mellitus Due To Underlying Condition With Diabetic Neuropathy, Unspecified, Diabetes Mellitus Due To Underlying Condition With Diabetic Mononeuropathy, Diabetes Mellitus Due To Underlying Condition With Diabetic Polyneuropathy, Diabetes Mellitus Due To

Phenotype	Data fields	Field names	Data codes	Data code definitions
			E09.339, E09.341, E09.349, E09.351, E09.359, E09.36, E09.39, E09.40, E09.41, E09.42, E09.43, E09.44, E09.49, E09.51, E09.52, E09.59, E09.610, E09.618, E09.620, E09.621, E09.622, E09.628, E09.630, E09.638, E09.641, E09.649, E09.65, E09.69, E09.8, E09.9, E10.10, E10.11, E10.21, E10.22, E10.29, E10.311, E10.319, E10.321, E10.329, E10.331, E10.339, E10.341, E10.349, E10.351, E10.359, E10.36, E10.39, E10.40, E10.41, E10.42, E10.43, E10.44, E10.49, E10.51, E10.52, E10.59, E10.610, E10.618, E10.620, E10.621, E10.622, E10.628, E10.630, E10.638, E10.641, E10.649, E10.65, E10.69, E10.8, E10.9, E11.00, E11.01, E11.21, E11.22, E11.29, E11.311, E11.319, E11.321, E11.329, E11.331, E11.339, E11.341, E11.349, E11.351, E11.359, E11.36,	Underlying Condition With Diabetic Autonomic (Poly)Neuropathy, Diabetes Mellitus Due To Underlying Condition With Diabetic Amyotrophy, Diabetes Mellitus Due To Underlying Condition With Other Diabetic Neurological Complication, Diabetes Mellitus Due To Underlying Condition With Diabetic Peripheral Angiopathy Without Gangrene, Diabetes Mellitus Due To Underlying Condition With Diabetic Neuropathic Arthropathy, Diabetes Mellitus Due To Underlying Condition with diabetic neuropathic arthropathy, Diabetes Mellitus Due To Underlying Condition With Diabetic arthropathy, Diabetes Mellitus Due To Underlying Condition With diabetic dermatitis, Diabetes Mellitus Due To Underlying Condition With Foot Ulcer, Diabetes Mellitus Due To Underlying Condition With Other Skin ulcer, Diabetes Mellitus Due To Underlying Condition With other skin complications, Diabetes Mellitus Due To Underlying Condition With periodontal disease, Diabetes Mellitus Due To Underlying Condition With other oral complications, Diabetes Mellitus Due To Underlying Condition With hypoglycemia with coma, Diabetes mellitus due to underlying condition with hypoglycemia without coma, Diabetes Mellitus Due To Underlying Condition With Hyperglycemia, Diabetes Mellitus Due To Underlying Condition With Other Specified Complication, Diabetes Mellitus Due To Underlying Condition With Unspecified Complications, Diabetes mellitus due to underlying condition without Complications, Diabetes Mellitus Due To Underlying Condition Without Complications, Drug or Chemical induced diabetes mellitus with hyperosmolarity without nonketotic hyperglycemic-hyperosmolar coma (NKHHC), Drug Or Chemical Induced Diabetes Mellitus With Hyperosmolarity With Coma, Drug Or Chemical Induced Diabetes Mellitus With Ketoacidosis Without Coma, Drug Or Chemical Induced Diabetes Mellitus With Ketoacidosis With Coma, Drug Or Chemical Induced Diabetes Mellitus With Diabetic Nephropathy, Drug Or Chemical Induced Diabetes Mellitus With diabetic chronic kidney disease, Drug or chemical induced diabetes mellitus with other diabetic kidney complication, Drug Or Chemical Induced Diabetes Mellitus With Unspecified Diabetic Retinopathy With Macular Edema, Drug Or Chemical Induced Diabetes Mellitus With Unspecified Diabetic Retinopathy Without Macular Edema, Drug or chemical induced diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, Drug or chemical induced diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, Drug or chemical induced diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, Drug or chemical induced diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, Drug or chemical induced diabetes mellitus with severe nonproliferative diabetic retinopathy with macular edema, Drug or chemical induced diabetes mellitus with severe nonproliferative diabetic retinopathy without macular edema, Drug or chemical induced diabetes mellitus with proliferative diabetic retinopathy with macular edema, Drug or chemical induced diabetes mellitus with proliferative diabetic retinopathy without macular edema, Drug Or Chemical Induced Diabetes Mellitus With Diabetic Cataract, Drug Or Chemical Induced Diabetes Mellitus With Other Diabetic Ophthalmic Complication, Drug Or Chemical Induced Diabetes Mellitus With

Phenotype	Data fields	Field names	Data codes	Data code definitions
			E11.39, E11.40, E11.41, E11.42, E11.51, E11.52, E11.59, E11.610, E11.618, E11.620, E11.621, E11.622, E11.628, E11.630, E11.638, E11.641, E11.649, E11.65, E11.69, E11.8, E11.9, E13.00, E13.01, E13.10, E13.11, E13.21, E13.22, E13.29, E13.311, E13.319, E13.321, E13.329, E13.331, E13.339, E13.341, E13.349, E13.351, E13.359, E13.36, E13.39, E13.40, E13.41, E13.42, E13.43, E13.44, E13.49, E13.51, E13.52, E13.59, E13.610, E13.618, E13.620, E13.621, E13.622, E13.628, E13.630, E13.638, E13.641, E13.649, E13.65, E13.69, E13.8, E13.9, R82.4	Neurological Complications With Diabetic Neuropathy, Unspecified, Drug Or Chemical Induced Diabetes Mellitus With Neurological Complications With Diabetic Mononeuropathy, Drug Or Chemical Induced Diabetes Mellitus With Neurological Complications With Diabetic Polyneuropathy, Drug Or Chemical Induced Diabetes Mellitus With Neurological Complications With Diabetic Autonomic (Poly)Neuropathy, Drug Or Chemical Induced Diabetes Mellitus With Neurological Complications With Diabetic Amyotrophy, Drug Or Chemical Induced Diabetes Mellitus With Neurological Complications With Other Diabetic Neurological Complication, Drug Or Chemical Induced Diabetes Mellitus With Diabetic Peripheral Angiopathy Without Gangrene, Drug Or Chemical Induced Diabetes Mellitus With Diabetic peripheral angiopathy with gangrene, Drug or chemical induced diabetes mellitus with other circulatory complications, Drug Or Chemical Induced Diabetes Mellitus With Diabetic neuropathic Arthropathy, Drug or chemical induced diabetes mellitus with other diabetic arthropathy, Drug Or Chemical Induced Diabetes Mellitus With diabetic dermatitis, Drug Or Chemical Induced Diabetes Mellitus With foot ulcer, Drug Or Chemical Induced Diabetes Mellitus With Other Skin ulcer, Drug Or Chemical Induced Diabetes Mellitus With other skin complications, Drug Or Chemical Induced Diabetes Mellitus With periodontal disease, Drug Or Chemical Induced Diabetes Mellitus With other oral complications, Drug Or Chemical Induced Diabetes Mellitus With Hypoglycemia with coma, Drug Or Chemical Induced Diabetes Mellitus With hypoglycemia without coma, Drug Or Chemical Induced Diabetes Mellitus With Hyperglycemia, Drug Or Chemical Induced Diabetes Mellitus With Other Specified Complication, Drug Or Chemical Induced Diabetes Mellitus With Unspecified Complications, Drug Or Chemical Induced Diabetes Mellitus without complications, Type 1 Diabetes Mellitus With Ketoacidosis Without Coma, Type 1 Diabetes Mellitus With Ketoacidosis With Coma, Type 1 Diabetes Mellitus With Diabetic Nephropathy, Type 1 Diabetes Mellitus With diabetic chronic kidney disease, Type 1 Diabetes Mellitus With other diabetic kidney complication, Type 1 Diabetes Mellitus With Unspecified Diabetic Retinopathy With Macular Edema, Type 1 Diabetes Mellitus With Unspecified Diabetic Retinopathy Without Macular Edema, Type 1 Diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, Type 1 Diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, Type 1 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, Type 1 diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, Type 1 diabetes mellitus with severe nonproliferative diabetic retinopathy with macular edema, Type 1 diabetes mellitus with severe nonproliferative diabetic retinopathy without macular edema, Type 1 diabetes mellitus with proliferative diabetic retinopathy with macular edema, Type 1 diabetes mellitus with proliferative diabetic retinopathy without macular edema, Type 1 Diabetes Mellitus With Diabetic Cataract, Type 1 Diabetes Mellitus With Other Diabetic Ophthalmic Complication, Type 1 Diabetes Mellitus With Diabetic Neuropathy, Unspecified, Type 1 Diabetes Mellitus With diabetic mononeuropathy, Type 1 Diabetes Mellitus With Diabetic polyneuropathy, Type 1 Diabetes mellitus with

Phenotype	Data fields	Field names	Data codes	Data code definitions
				diabetic autonomic (poly)neuropathy, Type 1 diabetes mellitus with diabetic amyotrophy, Type 1 diabetes mellitus with other diabetic neurological complication, Type 1 Diabetes Mellitus With Diabetic Peripheral Angiopathy Without Gangrene, Type 1 diabetes mellitus with diabetic peripheral angiopathy with gangrene, Type 1 diabetes mellitus with other circulatory complications, Type 1 diabetes mellitus with diabetic neuropathic arthropathy, Type 1 Diabetes Mellitus With Diabetic arthropathy, Type 1 Diabetes Mellitus With diabetic dermatitis, Type 1 Diabetes Mellitus With Other foot ulcer, Type 1 Diabetes Mellitus With Other Skin ulcer, Type 1 Diabetes Mellitus With other skin complications, Type 1 Diabetes Mellitus With periodontal disease, Type 1 Diabetes Mellitus With other oral complications, Type 1 Diabetes Mellitus With Hypoglycemia With Coma, Type 1 Diabetes Mellitus With Hypoglycemia without coma, Type 1 Diabetes Mellitus With Hyperglycemia, Type 1 Diabetes Mellitus With Other Specified Complication, Type 1 Diabetes Mellitus With Unspecified Complications, Type 1 Diabetes Mellitus Without Complications, Type 2 Diabetes Mellitus With Hyperosmolarity Without Nonketotic Hyperglycemic-Hyperosmolar Coma (Nkhhc), Type 2 Diabetes Mellitus With Hyperosmolarity With Coma, Type 2 diabetes mellitus with diabetic nephropathy, Type 2 diabetes mellitus with diabetic chronic kidney disease, Type 2 diabetes mellitus with other diabetic kidney complication, Type 2 Diabetes Mellitus With Unspecified Diabetic Retinopathy With Macular Edema, Type 2 Diabetes Mellitus With Unspecified Diabetic Retinopathy Without Macular Edema, Type 2 Diabetes Mellitus With Mild Nonproliferative Diabetic Retinopathy With Macular Edema, Type 2 Diabetes Mellitus With Mild Nonproliferative Diabetic Retinopathy Without Macular Edema, Type 2 Diabetes Mellitus With Moderate Nonproliferative Diabetic Retinopathy With Macular Edema, Type 2 Diabetes Mellitus with moderate nonproliferative diabetic retinopathy Without Macular Edema, Type 2 Diabetes Mellitus With Severe Nonproliferative Diabetic Retinopathy With Macular Edema, Type 2 Diabetes Mellitus With severe nonproliferative Diabetic Retinopathy Without Macular Edema, Type 2 Diabetes Mellitus with proliferative diabetic retinopathy with macular edema, Type 2 Diabetes Mellitus With proliferative diabetic retinopathy without macular edema, Type 2 Diabetes Mellitus With Diabetic Cataract, Type 2 Diabetes Mellitus With Other Diabetic Ophthalmic Complication, Type 2 Diabetes Mellitus With Diabetic Neuropathy, Unspecified, Type 2 Diabetes Mellitus With Diabetic mononeuropathy, Type 2 Diabetes Melitus with diabetic polyneuropathy, Type 2 Diabetes Mellitus With Diabetic Peripheral Angiopathy Without Gangrene, Type 2 Diabetes Mellitus With diabetic peripheral Angiopathy With Gangrene, Type 2 Diabetes Mellitus with other circulatory complications, Type 2 diabetes mellitus with diabetic neuropathic arthropathy, Type 2 diabetes mellitus with other diabetic arthropathy, Type 2 Diabetes Mellitus with diabetic dermatitis, Type 2 Diabetes Mellitus With foot ulcer, Type 2 Diabetes Mellitus With Other Skin ulcer, Type 2 Diabetes Mellitus With other skin complications, Type 2 Diabetes Mellitus With periodontal disease, Type 2 Diabetes Mellitus With other oral complications, Type 2 Diabetes Mellitus With Hypoglycemia With Coma, Type 2 Diabetes Mellitus With Hypoglycemia without

Phenotype	Data fields	Field names	Data codes	Data code definitions
				coma, Type 2 Diabetes Mellitus With Hyperglycemia, Type 2 Diabetes Mellitus With Other Specified Complication, Type 2 Diabetes Mellitus With Unspecified Complications, Type 2 Diabetes Mellitus Without Complications, Other specified diabetes mellitus with hyperosmolarity without nonketotic hyperglycemic-hyperosmolar coma (NKHHC), Other specified diabetes mellitus with hyperosmolarity with coma, Other specified diabetes mellitus with ketoacidosis without coma, Other specified diabetes mellitus with ketoacidosis with coma, Other specified diabetes mellitus with diabetic nephropathy, Other specified diabetes mellitus with diabetic chronic kidney disease, Other specified diabetes mellitus with other diabetic kidney complication, Other specified diabetes mellitus with unspecified diabetic retinopathy with macular edema, Other specified diabetes mellitus with unspecified diabetic retinopathy without macular edema, Other specified diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, Other specified diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, Other specified diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, Other specified diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema, Other specified diabetes mellitus with severe nonproliferative diabetic retinopathy with macular edema, Other specified diabetes mellitus with severe nonproliferative diabetic retinopathy without macular edema, Other specified diabetes mellitus with proliferative diabetic retinopathy with macular edema, Other specified diabetes mellitus with proliferative diabetic retinopathy without macular edema, Other Specified Diabetes Mellitus With diabetic cataract, Other Specified Diabetes Mellitus With other diabetic ophthalmic complication, Other Specified Diabetes Mellitus With Diabetic neuropathy, unspecified, Other Specified Diabetes Mellitus With Diabetic mononeuropathy, Other Specified Diabetes Mellitus With Diabetic Polyneuropathy, Other Specified Diabetes Mellitus With Diabetic Autonomic (Poly)Neuropathy, Other Specified Diabetes Mellitus With Diabetic Amyotrophy, Other Specified Diabetes Mellitus With Other Diabetic Neurological Complication, Other specified diabetes mellitus with diabetic peripheral angiopathy without gangrene, Other specified diabetes mellitus with diabetic peripheral angiopathy with gangrene, Other Specified Diabetes Mellitus With Other Circulatory Complications, Other specified diabetes mellitus with diabetic neuropathic arthropathy, Other specified diabetes mellitus with other diabetic arthropathy, Other Specified Diabetes Mellitus With Diabetic Dermatitis, Other Specified Diabetes Mellitus With Foot Ulcer, Other Specified Diabetes Mellitus With Other Skin Ulcer, Other specified diabetes mellitus with other skin complications, Other specified diabetes mellitus with periodontal disease, Other specified diabetes mellitus with other oral complications, Other Specified Diabetes Mellitus With Hypoglycemia With Coma, Other Specified Diabetes Mellitus With Hypoglycemia Without Coma, Other Specified Diabetes Mellitus With Hyperglycemia, Other Specified Diabetes Mellitus With Other Specified Complication, Other Specified Diabetes Mellitus With Unspecified Complications, Other Specified Diabetes Mellitus Without Complications, Acetonuria

Phenotype	Data fields	Field names	Data codes	Data code definitions
Diabetes	Medications	Medications	metformin, glucophage, riomet, fortamet, glumetzia, chlorpropamide, glimepiride, glyburide, glipizide, tolazamide, tolbutamide, diabinese, amaryl, diabeta, micronase, glucotrol, glynase, tolinase, orinase, tolbutamide, repaglinide, nateglinide, prandin, starlix, pioglitazone, rosiglitazone, actos, avandia, sitagliptin, saxagliptin, linagliptin, alogliptin, januvia, onglyza, tradjenta, nesina, acarbose, miglitol, precose, glyset, pramlintide, symlin, liraglutide, exenatide, albiglutide, dulaglutide, victoza, bydureon, byetta, tanzeum, trulicity, canagliflozin, dapagliflozin, empagliflozin, invokana, farxiga, jardiance, actoplus, glucovance, metaglip, janumet, kombiglyze, prandimet, duetact, kazano, invokamet, xigduo, synjardy,	Biguanides, Sulfonylureas, Meglitinides, Thiazolidinediones, DPP-4 Inhibitors, Alpha-gluocosidase inhibitors, Amylin analogue, GLP-1 Agonist, SGLT2 Inhibitor, Combination pills, long-actin insulin

Phenotype	Data fields	Field names	Data codes	Data code definitions
			jentadueto, avandamet, oseni, glyxambi, avandaryl, juvisync, glargine, basaglar, lantus, toujeo, detemir, degludec, levemir, tresiba	
Heart failure	Diagnoses	ICD9	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428, 428.1, 428.2, 428.21, 428.22, 428.23, 428.3, 428.31, 428.32, 428.33, 428.4, 428.41, 428.42, 428.43, 428.9	Rheumatic heart failure (congestive), Malignant hypertensive heart disease with heart failure, Benign hypertensive heart disease with heart failure, Unspecified hypertensive heart disease with heart failure , Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified, Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease, Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified, Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease, Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified, Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease, Congestive heart failure, unspecified, Left heart failure, Systolic heart failure, unspecified, Acute systolic heart failure, Chronic systolic heart failure, Acute on chronic systolic heart failure, Diastolic heart failure, unspecified, Acute diastolic heart failure, Chronic diastolic heart failure, Acute on chronic diastolic heart failure, Combined systolic and diastolic heart failure, unspecified, Acute combined systolic and diastolic heart failure, Chronic combined systolic and diastolic heart failure, Acute on chronic combined systolic and diastolic heart failure, Heart failure, unspecified
Heart failure	Diagnoses	ICD10	I09.81, I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.9, I97.130, I97.131	Rheumatic Heart Failure, Hypertensive Heart Disease With Heart Failure, Hypertensive Heart And Chronic Kidney Disease With Heart Failure And Stage 1 Through Stage 4 Chronic Kidney Disease, Or Unspecified Chronic Kidney Disease, Hypertensive Heart And Chronic Kidney Disease With Heart Failure And With Stage 5 Chronic Kidney Disease, Or End Stage Renal Disease, Left Ventricular Failure, Unspecified Systolic (Congestive) Heart Failure, Acute Systolic (Congestive) Heart Failure, Chronic Systolic (Congestive) Heart Failure, Acute On Chronic Systolic (Congestive) Heart Failure, Unspecified Diastolic (Congestive) Heart Failure, Acute Diastolic (Congestive) Heart Failure, Chronic Diastolic (Congestive) Heart Failure, Acute On Chronic Diastolic (Congestive) Heart Failure, Unspecified Combined Systolic (Congestive) And Diastolic (Congestive) Heart Failure, Acute Combined Systolic (Congestive) And Diastolic (Congestive) Heart Failure, Chronic Combined Systolic (Congestive) And Diastolic (Congestive) Heart Failure, Acute On Chronic Combined Systolic (Congestive) And Diastolic (Congestive) Heart Failure, Heart

Phenotype	Data fields	Field names	Data codes	Data code definitions
				Failure, Unspecified, Postprocedural heart failure following cardiac surgery, Postprocedural heart failure following other surgery
Myocardial infarction	Diagnoses	ICD9	410, 410.01, 410.02, 410.1, 410.11, 410.12, 410.2, 410.21, 410.22, 410.3, 410.31, 410.32, 410.4, 410.41, 410.42, 410.5, 410.51, 410.52, 410.6, 410.61, 410.62, 410.7, 410.71, 410.72, 410.8, 410.81, 410.82, 410.9, 410.91, 410.92, 412, 429.79	Acute myocardial infarction of anterolateral wall, episode of care unspecified, Acute myocardial infarction of anterolateral wall, initial episode of care, Acute myocardial infarction of anterolateral wall, subsequent episode of care , Acute myocardial infarction of other anterior wall, episode of care unspecified, Acute myocardial infarction of other anterior wall, initial episode of care, Acute myocardial infarction of other anterior wall, subsequent episode of care, Acute myocardial infarction of inferolateral wall, episode of care unspecified, Acute myocardial infarction of inferolateral wall, initial episode of care, Acute myocardial infarction of inferolateral wall, subsequent episode of care, Acute myocardial infarction of inferoposterior wall, episode of care unspecified, Acute myocardial infarction of inferoposterior wall, initial episode of care, Acute myocardial infarction of inferoposterior wall, subsequent episode of care, Acute myocardial infarction of other inferior wall, episode of care unspecified, Acute myocardial infarction of other inferior wall, initial episode of care, Acute myocardial infarction of other inferior wall, subsequent episode of care, Acute myocardial infarction of other lateral wall, episode of care unspecified, Acute myocardial infarction of other lateral wall, initial episode of care, Acute myocardial infarction of other lateral wall, subsequent episode of care, True posterior wall infarction, episode of care unspecified, True posterior wall infarction, initial episode of care, True posterior wall infarction, subsequent episode of care, Subendocardial infarction, episode of care unspecified, Subendocardial infarction, initial episode of care, Subendocardial infarction, subsequent episode of care, Acute myocardial infarction of other specified sites, episode of care unspecified, Acute myocardial infarction of other specified sites, initial episode of care, Acute myocardial infarction of other specified sites, subsequent episode of care, Acute myocardial infarction of unspecified site, episode of care unspecified, Acute myocardial infarction of unspecified site, initial episode of care, Acute myocardial infarction of unspecified site, subsequent episode of care, Old myocardial infarction , Certain sequelae of myocardial infarction, not elsewhere classified, other
Myocardial infarction	Diagnoses	ICD10	I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I22.0, I22.1, I22.2, I22.8, I22.9, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.7, I23.8, I24.1, I25.2	ST elevation (STEMI) myocardial infarction involving left main coronary artery, ST elevation (STEMI) myocardial infarction involving left anterior descending coronary artery, ST elevation (STEMI) myocardial infarction involving other coronary artery of anterior wall, ST elevation (STEMI) myocardial infarction involving right coronary artery, ST elevation (STEMI) myocardial infarction involving other coronary artery of inferior wall, ST elevation (STEMI) myocardial infarction involving left circumflex coronary artery, ST elevation (STEMI) myocardial infarction involving other sites, ST elevation (STEMI) myocardial infarction of unspecified site , Non-ST elevation (NSTEMI) myocardial infarction , Subsequent ST elevation (STEMI) myocardial infarction of anterior wall, Subsequent ST elevation (STEMI) myocardial infarction of inferior wall, Subsequent non-ST elevation (NSTEMI) myocardial infarction , Subsequent ST elevation (STEMI) myocardial infarction of other sites, Subsequent ST elevation (STEMI) myocardial infarction of unspecified site, Hemopericardium as

Phenotype	Data fields	Field names	Data codes	Data code definitions
				current complication following acute myocardial infarction, Atrial septal defect as current complication following acute myocardial infarction, Ventriclar septal defect as current complication following acute myocardial infarction, Rupture of cardiac wall without hemopericardium as current complication following acute myocardial infarction, Rupture of chordae tendineae as current complication following acute myocardial infarction, Rupture of papillary muscle as current complication following acute myocardial infarction, Thrombosis of atrium, auricular appendage, and ventricle as current complications following acute myocardial infarction, Postinfarction angina, Other current complications following acute myocardial infarction, Dressler's syndrome, Old myocardial infarction
Atrial fibrillation (AF) in Mass General Brigham defined using a previously validated algorithm (≥1 inpatient diagnosis code, ≥2 outpatient diagnosis code, ≥1 AF-related procedure, or ≥1 electrocardiogram with AF). See above for individual codes. This algorithm has a reported positive predictive value 92%[24] AF in UK Biobank defined as ≥1 inpatient diagnosis or procedural code, or self-reported diagnosis. When validated in an external dataset, this definition had a positive predictive value of 92%.[25] Myocardial infarction defined as ≥2 codes from any setting[23] Heart failure defined as ≥1 inpatient code				

Table IV. Cox model parameters from MGH training set

Model	**β**	**S_0**	**ΣβY**
5-Year Atrial Fibrillation			
Age & Sex	Age (years): 0.060413 Sex (male): 0.466141	0.9536967	4.44×10^{-14}
ECG-AI	0.61270	0.9517232	-4.22×10^{-14}
CH-AI	CHARGE-AF: 0.35655 ECG-AI: 0.44266	0.9607289	-5.87×10^{-14}
CHARGE-AF*	-	0.9718413	12.58156
2-Year Atrial Fibrillation			
Age & Sex	Age (years): 0.058810 Sex (male): 0.499409	0.9818978	2.19×10^{-14}
ECG-AI	0.75530	0.9831352	6.95×10^{-15}
CH-AI	CHARGE-AF: 0.30351 ECG-AI: 0.53105	0.9851827	8.47×10^{-16}
A predicted AF risk estimate can be obtained for each respective window using the equation $1 - s_0^{\exp(\sum \beta X - \sum \beta Y)}$ where s_0 is the average AF-free survival probability at the window of interest in the MGH training set, $\sum \beta X$ is the individual's score, and $\sum \beta Y$ is the average score in the MGH training set			

Table V. Model discrimination over time

	Massachusetts General Hospital (N=4,166)					***Brigham and Women's Hospital (N=37,963)***					***UK Biobank (N=41,033)***	
Model	**1 year**	**2 years**	**3 years**	**4 years**	**5 years**	**1 year**	**2 years**	**3 years**	**4 years**	**5 years**	**1 year**	**2 years**
Average Precision												
Age and sex	0.04 (0.02-0.07)	0.07 (0.05-0.09)	0.09 (0.07-0.11)	0.13 (0.10-0.73)	0.16 (0.13-0.20)	0.04 (0.03-0.04)	0.07 (0.06-0.08)	0.10 (0.09-0.11)	0.12 (0.11-0.13)	0.14 (0.13-0.15)	0.010 (0.007-0.021)	0.018 (0.015-0.024)
CHARGE-AF	0.04 (0.03-0.07)	0.10 (0.07-0.16)	0.12 (0.09-0.17)	0.18 (0.14-0.24)	0.21 (0.17-0.27)	0.04 (0.03-0.05)	0.08 (0.07-0.09)	0.11 (0.10-0.13)	0.14 (0.13-0.15)	0.17 (0.15-0.18)	0.011 (0.008-0.017)	0.020 (0.016-0.027)
ECG-AI	0.10 (0.06-0.16)	0.13 (0.09-0.19)	0.17 (0.13-0.24)	0.22 (0.18-0.29)	0.27 (0.22-0.34)	0.05 (0.04-0.06)	0.09 (0.08-0.10)	0.13 (0.12-0.14)	0.16 (0.14-0.17)	0.19 (0.17-0.20)	0.032 (0.018-0.052)	0.060 (0.043-0.087)
CH-AI	0.10 (0.06-0.18)	0.14 (0.10-0.21)	0.18 (0.14-0.24)	0.24 (0.18-0.31)	0.30 (0.24-0.38)	0.06 (0.05-0.07)	0.10 (0.09-0.12)	0.14 (0.13-0.16)	0.17 (0.16-0.19)	0.21 (0.19-0.22)	0.032 (0.019-0.056)	0.059 (0.041-0.082)
Area Under the Receiver Operating Characteristic Curve												
Age and sex	0.739 (0.668-0.795)	0.752 (0.696-0.798)	0.754 (0.707-0.796)	0.762 (0.723-0.800)	0.768 (0.732-0.805)	0.704 (0.679-0.730)	0.719 (0.700-0.738)	0.726 (0.710-0.741)	0.731 (0.717-0.743)	0.730 (0.717-0.743)	0.711 (0.674-0.749)	0.728 (0.702-0.755)
CHARGE-AF	0.754 (0.686-0.817)	0.785 (0.741-0.829)	0.793 (0.759-0.829)	0.796 (0.761-0.830)	0.802 (0.767-0.836)	0.717 (0.691-0.740)	0.733 (0.715-0.751)	0.744 (0.728-0.757)	0.751 (0.737-0.764)	0.752 (0.741-0.763)	0.716 (0.682-0.750)	0.732 (0.704-0.759)
ECG-AI	0.834 (0.778-0.884)	0.837 (0.798-0.868)	0.831 (0.793-0.861)	0.829 (0.793-0.856)	0.823 (0.790-0.856)	0.745 (0.721-0.765)	0.747 (0.732-0.763)	0.751 (0.738-0.764)	0.746 (0.734-0.758)	0.747 (0.736-0.759)	0.696 (0.658-0.735)	0.705 (0.673-0.737)
CH-AI	0.820 (0.756-0.875)	0.837 (0.795-0.875)	0.841 (0.805-0.873)	0.838 (0.802-0.869)	0.838 (0.807-0.869)	0.753 (0.730-0.775)	0.764 (0.747-0.780)	0.773 (0.760-0.787)	0.774 (0.763-0.786)	0.777 (0.766-0.788)	0.736 (0.697-0.775)	0.746 (0.716-0.776)

Table VI. Performance of CH-AI versus CHARGE-AF at fixed operating points

	Massachusetts General Hospital (N=4,166)				*Brigham and Women's Hospital (N=37,963)*				*UK Biobank (N=41,033)*			
Model	**Sensitivity**	**Specificity**	**Precision/PPV**	**NPV**	**Sensitivity**	**Specificity**	**Precision/PPV**	**NPV**	**Sensitivity**	**Specificity**	**Precision/PPV**	**NPV**
85% Specificity												
CH-AI	63.8	85	9.74	98.9	51.6	85	8.58	98.5	49.3	85	2.70	99.5
ECG-AI	59.2	85	9.15	98.8	47.5	85	7.99	98.3	44.6	85	2.45	99.5
CHARGE-AF	53.6	85	8.37	98.6	48.0	85	8.08	98.4	41.1	85	2.26	99.4
Age and sex	51.1	85	7.86	98.6	45.1	85	7.63	98.3	40.0	85	2.21	99.4
90% Specificity												
CH-AI	53.0	90	11.9	98.7	42.3	90	10.4	98.3	42.3	90	3.43	99.5
ECG-AI	51.8	90	11.7	98.7	39.7	90	9.85	98.2	37.6	90	3.11	99.4
CHARGE-AF	42.7	90	9.69	98.4	37.6	90	9.33	98.1	31.0	90	2.55	99.4
Age and sex	37.2	90	8.53	98.3	35.2	90	8.80	98.1	28.0	90	2.31	99.3
95% Specificity												
CH-AI	42.3	95	17.9	98.5	31.3	95	14.6	98.1	30.4	95	4.85	99.4
ECG-AI	41.2	95	17.3	98.5	26.6	95	12.6	97.9	25.4	95	4.12	99.3
CHARGE-AF	24.6	95	11.0	98.0	25.2	95	12.0	97.9	20.0	95	3.28	99.3
Age and sex	23.4	95	10.7	98.0	22.7	95	11.0	97.8	17.9	95	2.94	99.3
85% Sensitivity												
CH-AI	85	64.6	5.76	99.4	85	47.7	4.26	99.2	85	43.4	1.25	99.7
ECG-AI	85	66.5	5.84	99.3	85	45.3	4.07	99.1	85	31.5	1.03	99.6
CHARGE-AF	85	51.2	4.24	99.3	85	41.1	3.79	99.0	85	46.0	1.31	99.7
Age and sex	85	44.4	3.72	99.2	85	39.8	3.72	99.0	85	45.7	1.30	99.7
90% Sensitivity												
CH-AI	90	49.7	4.32	99.5	90	37.3	3.78	99.3	90	35.4	1.16	99.8
ECG-AI	90	55.6	4.86	99.5	90	34.8	3.64	99.2	90	20.7	0.95	99.6
CHARGE-AF	90	48.0	4.18	99.5	90	29.3	3.36	99.1	90	38.2	1.21	99.8
Age and sex	90	35.4	3.39	99.3	90	28.3	3.32	99.0	90	34.8	1.15	99.8
95% Sensitivity												
CH-AI	95	40.1	3.85	99.7	95	21.9	3.22	99.4	95	15.8	0.94	99.7
ECG-AI	95	37.2	3.68	99.7	95	20.7	3.17	99.3	95	9.6	0.88	99.6
CHARGE-AF	95	30.4	3.33	99.6	95	17.2	3.04	99.2	95	22.7	1.03	99.8
Age and sex	95	24.3	3.07	99.5	95	12.5	2.89	98.9	95	25.1	1.06	99.9
PPV = positive predictive value; NPV = negative predictive value												

Table VII. Net reclassification using standard risk thresholds

	MGH				BWH				UK Biobank			
		CH-AI				CH-AI				CH-AI		
	CHARGE AF	<2.5%	2.5-5%	≥5%	**CHARGE AF**	<2.5%	2.5-5%	≥5%	**CHARGE AF**	<0.5%	0.5-1%	≥1%
AF events	<2.5%	10 (4.9%)	8 (3.9%)	1 (0.5%)	<2.5%	103 (5.4%)	42 (2.2%)	20 (1.1%)	<0.5%	21 (6.5%)	9 (2.8%)	9 (2.8%)
	2.5-5%	0 (0%)	7 (3.4%)	11 (5.4%)	2.5-5%	26 (1.4%)	133 (7.0%)	105 (5.5%)	0.5-1%	20 (6.2%)	31 (9.7%)	35 (10.9%)
	≥5%	0 (0%)	6 (3.0%)	160 (78.8%)	≥5%	1 (0.05%)	60 (3.2%)	1408 (74.2%)	≥1%	17 (5.3%)	46 (14.3%)	133 (41.4%)
AF non events	<2.5%	1319 (33.3%)	244 (6.2%)	30 (0.8%)	<2.5%	9675 (26.8%)	1363 (3.8%)	178 (0.5%)	<0.5%	12670 (31.1%)	2724 (6.7%)	664 (1.6%)
	2.5-5%	113 (2.9%)	498 (12.6%)	202 (5.1%)	2.5-5%	1927 (5.3%)	5205 (14.4%)	1974 (5.5%)	0.5-1%	5097 (12.5%)	5264 (12.9%)	1974 (4.8%)
	≥5%	3 (0.8%)	216 (5.5%)	1338 (33.8%)	≥5%	71 (0.2%)	2483 (6.9%)	13189 (36.6%)	≥1%	1425 (3.5%)	4825 (11.9%)	6069 (14.9%)
	NRI+	5.9% (0.7 to 10.6)			NRI+	3.6% (1.9 to 5.1)			NRI+	-10.2% (-16.7 to -3.2)		
	NRI-	-3.7% (-5.2 to -2.1)			NRI-	2.7% (2.3 to 3.2)			NRI-	14.7% (14.1 to 15.3)		
	NRI	0.022 (-0.031 to 0.073)			NRI	0.063 (0.046-0.079)			NRI	0.045 (-0.022 to 0.11)		
		MGH				**BWH**				**UK Biobank**		
		ECG-AI				**ECG-AI**				**ECG-AI**		
	CHARGE AF	<2.5%	2.5-5%	≥5%	**CHARGE AF**	<2.5%	2.5-5%	≥5%	**CHARGE AF**	<0.5%	0.5-1%	≥1%
AF events	<2.5%	6 (3.0%)	6 (3.0%)	7 (3.4%)	<2.5%	34 (1.8%)	56 (3.0%)	75 (4.0%)	<0.5%	17 (5.3%)	10 (3.1%)	12 (3.7%)
	2.5-5%	1 (0.5%)	5 (2.5%)	12 (5.9%)	2.5-5%	17 (9.0%)	82 (4.3%)	165 (8.7%)	0.5-1%	20 (6.2%)	28 (8.7%)	38 (11.8%)
	≥5%	1 (0.5%)	17 (8.4%)	148 (72.9%)	≥5%	13 (0.7%)	133 (7.0%)	1323 (69.7%)	≥1%	26 (8.1%)	50 (15.6%)	120 (37.4%)
AF non events	<2.5%	936 (23.6%)	487 (12.2%)	170 (4.3%)	<2.5%	4652 (12.9%)	4538 (12.6%)	2026 (5.6%)	<0.5%	9519 (23.4%)	5044 (12.4%)	1495 (3.7%)
	2.5-5%	183 (4.6%)	376 (9.5%)	254 (6.4%)	2.5-5%	1100 (3.1%)	4049 (11.2%)	3957 (11.0%)	0.5-1%	4838 (11.9%)	5265 (12.9%)	2232 (5.5%)
	≥5%	117 (3.0%)	470 (11.9%)	970 (24.5%)	≥5%	406 (1.1%)	3774 (10.5%)	11563 (32.1%)	≥1%	2683 (6.6%)	5158 (12.7%)	4478 (11.0%)
	NRI+	2.2% (-3.5% to 8.3%)			NRI+	6.1% (4.0 to 8.3)			NRI+	-12.0% (-19.2 to -4.4)		
	NRI-	-3.6% (-5.4% to -0.02%)			NRI-	-14.7% (-15.3 to -13.9)			NRI-	9.6% (8.8 to 10.3)		
	NRI	-0.014 (-0.078 to 0.057)			NRI	-0.086 (-10.9 to -6.4)			NRI	-0.024 (-0.095 to 0.057)		

Green cells denote appropriate reclassification
Red cells denote inappropriate reclassification
NRI=Net reclassification improvement; NRI+=Event reclassification improvement; NRI-=Non-event reclassification improvement
NRI values calculated using the Kaplan-Meier estimator to account for censored survival data[33]

Table VIII. Net reclassification using high risk thresholds

		MGH		BWH			UK Biobank		
		CH-AI			CH-AI			CH-AI	
	CHARGE AF	<20%	≥20%	CHARGE AF	<20%	≥20%	CHARGE AF	<2%	≥2%
AF events	<20%	112 (34.9%)	42 (13.1%)	<20%	1250 (65.9%)	343 (18.1%)	<2%	194 (60.4%)	59 (18.4%)
	≥20%	12 (3.7%)	37 (11.5%)	≥20%	46 (2.4%)	259 (13.6%)	≥2%	26 (8.1%)	42 (13.1%)
AF non events	<20%	3644 (92.0%)	87 (2.2%)	<20%	33571 (93.1%)	1297 (3.6%)	<2%	36289 (89.1%)	1412 (3.5%)
	≥20%	91 (2.3%)	141 (3.6%)	≥20%	354 (1.0%)	843 (2.3%)	≥2%	2005 (4.9%)	1006 (2.5%)
	NRI+	13.5% (6.3 to 20.4)		NRI+	15.3% (13.4 to 17.4)		NRI+	9.7% (3.5 to 15.1)	
	NRI-	0.23% (-0.51 to 0.88)		NRI-	-2.5% (-2.7 to -2.3)		NRI-	1.5% (1.2 to 1.7)	
	NRI	0.14 (0.063 to 0.21)		NRI	0.13 (0.11 to 0.15)		NRI	0.11 (0.049 to 0.17)	
		MGH		BWH			UK Biobank		
		ECG-AI			ECG-AI			ECG-AI	
	CHARGE AF	<20%	≥20%	CHARGE AF	<20%	≥20%	CHARGE AF	<2%	≥2%
AF events	<20%	121 (59.6%)	33 (16.3%)	<20%	1285 (67.7%)	308 (16.2%)	<2%	197 (61.4%)	56 (17.4%)
	≥20%	23 (11.3%)	26 (12.8%)	≥20%	151 (8.0%)	154 (8.1%)	≥2%	45 (14.0%)	23 (7.2%)
AF non events	<20%	3627 (91.5%)	104 (2.6%)	<20%	33618 (83.2%)	1250 (3.5%)	<2%	36245 (89.0%)	1456 (3.6%)
	≥20%	182 (4.6%)	50 (1.3%)	≥20%	799 (2.2%)	398 (1.1%)	≥2%	2506 (6.2%)	505 (1.2%)
	NRI+	1.4% (-6.3 to 9.7)		NRI+	6.9% (4.4 to 9.2)		NRI+	2.0% (-5.1 to 8.3)	
	NRI-	1.8% (1.0 to 2.7)		NRI-	-1.3% (-1.5 to -1.0)		NRI-	2.6% (2.3 to 2.9)	
	NRI	0.032 (-0.046 to 0.12)		NRI	0.056 (0.033 to 0.080)		NRI	0.045 (-0.026 to 0.11)	

Green cells denote appropriate reclassification
Red cells denote inappropriate reclassification
NRI=Net reclassification improvement; NRI+=Event reclassification improvement; NRI-=Non-event reclassification improvement
NRI values calculated using the Kaplan-Meier estimator to account for censored survival data[33]

Table IX. Continuous net reclassification improvement

`	**NRI**	**NRI+**	**NRI-**
MGH			
ECG-AI	0.0042 (-0.13 to 0.15)	13% (0.1% to 26%)	-13% (-16% to 9.6%)
CH-AI	0.18 (0.023 to 0.31)	36% (21% to 50%)	-18% (-21% to -15%)
BWH			
ECG-AI	-0.084 (-0.13 to -0.035)	14% (9.6% to 19%)	-22% (-23% to -21%)
CH-AI	0.36 (0.32 to 0.40)	31% (28% to 36%)	4.4% (3.3% to 5.4%)
UK Biobank			
ECG-AI	-0.045 (-0.15 to 0.065)	-11.0% (-22% to -0.44%)	6.5% (5.7% to 7.5%)
CH-AI	0.18 (0.076 to 0.31)	-1.8% (-13% to 10%)	20% (19% to 21%)
All values reflect continuous net reclassification improvement (NRI) values for the indicated model versus the CHARGE-AF score			

Table X. Model performance among individuals with prevalent heart failure and stroke in BWH test set

*Brigham and Women's Hospital**					
Model	**Hazard ratio (per 1-SD)**	**5-year AUROC**	**5-year average precision**	**Calibration slope**	**ICI‡**
Prevalent heart failure (n=1,388)					
Age and sex	1.45 (1.26-1.67)	0.627 (0.590-0.664)	0.25 (0.22-0.29)	0.40 (0.25-0.56)	0.1212
CHARGE-AF	1.54 (1.33-1.79)	0.645† (0.600-0.690)	0.27 (0.23-0.31)	0.26 (0.17-0.35)	0.1247
ECG-AI	1.54 (1.39-1.72)	0.673 (0.631-0.715)	0.32†‡ (0.27-0.37)	0.46 (0.34-0.57)	0.0827
CH-AI	1.74 (1.53-1.98)	0.695†‡ (0.657-0.733)	0.31†‡ (0.27-0.36)	0.53 (0.41-0.65)	0.0561
Prevalent stroke (n=984)					
Age and sex	1.71 (1.35-2.16)	0.666 (0.609-0.724)	0.18 (0.14-0.24)	0.59 (0.33-0.85)	0.0311
CHARGE-AF	1.97 (1.55-2.51)	0.696† (0.633-0.759)	0.22† (0.17-0.30)	0.40 (0.26-0.54)	0.0698
ECG-AI	1.62 (1.37-1.93)	0.648 (0.579-0.718)	0.24† (0.17-0.35)	0.54 (0.35-0.73)	0.0386
CH-AI	1.95 (1.58-2.40)	0.690 (0.631-0.748)	0.26† (0.18-0.35)	0.60 (0.41-0.78)	0.0373

*Only Brigham and Women's test set included given insufficient event rates in MGH and UK Biobank test sets
†p<0.05 for comparison against age and sex
‡p<0.05 for comparison against CHARGE-AF
§Integrated calibration index (ICI), a quantitative measure of the average difference between predicted event risk and observed event incidence, weighted by the empirical distribution of event risk.[30] Smaller values indicate better calibration.
Difference in c-index for CH-AI vs ECG-AI: heart failure AUROC p<0.05, average precision p=NS; stroke AUROC p<0.05, average precision p=NS
AUROC = area under the receiver operating characteristic curve; ICI = integrated calibration index; SD = standard deviation

Table XI. Performance of single-lead ECG models

	Massachusetts General Hospital (N=4,166)					*Brigham and Women's Hospital (N=37,963)*					*UK Biobank (N=41,033)*				
Model	**Hazard ratio (per 1-SD)**	**5-year AUROC**	**5-year average precision**	**Calibration slope**	**ICI‡**	**Hazard ratio (per 1-SD)**	**5-year AUROC**	**5-year average precision**	**Calibration slope**	**ICI**	**Hazard ratio (per 1-SD)**	**2-year AUROC**	**2-year average precision**	**Calibration slope**	**ICI**
Referent models															
Age and sex	2.91 (2.44-3.47)	0.768 (0.732-0.805)	0.16 (0.13-0.20)	1.05 (0.88-1.23)	0.0074	2.48 (2.35-2.62)	0.730 (0.717-0.743)	0.14 (0.013-0.015)	0.94 (0.88-1.00)	0.0072	2.21 (1.96-2.50)	0.728 (0.702-0.755)	0.018 (0.015-0.024)	1.48 (1.25-1.71)	0.0019§
CHARGE-AF	3.36 (2.98-4.30)	0.802* (0.767-0.836)	0.21* (0.17-0.26)	0.68 (0.58-0.77)	0.0320	2.78 (2.63-2.94)	0.752* (0.741-0.763)	0.17* (0.15-0.18)	0.57 (0.53-0.60)	0.0344	2.26 (2.00-2.55)	0.732 (0.704-0.759)	0.020 (0.016-0.026)	0.87 (0.75-1.00)	0.0011§
12-lead ECG models															
ECG-AI	2.45 (2.23-2.69)	0.823* (0.790-0.856)	0.27* (0.21-0.34)	1.06 (0.95-1.17)	0.0212	2.05 (1.98-2.11)	0.747* (0.736-0.759)	0.19*† (0.17-0.20)	0.81 (0.77-0.84)	0.0129	2.01 (1.88-2.14)	0.705 (0.673-0.737)	0.060*† (0.044-0.090)	0.75 (0.68-0.82)	0.0035§
CH-AI	3.74 (3.24-4.33)	0.838*† (0.807-0.869)	0.30*† (0.24-0.38)	1.13 (1.01-1.25)	0.0120	2.76 (2.64-2.88)	0.777*† (0.766-0.788)	0.21*† (0.19-0.23)	0.77 (0.74-0.81)	0.0108	2.27 (2.11-2.44)	0.746 (0.716-0.776)	0.059*† (0.042-0.083)	1.01 (0.92-1.10)	0.0001§
Lead I only models															
ECG-AI	2.12 (1.92-2.33)	0.759† (0.720-0.798)	0.22* (0.17-0.28)	1.08 (0.94-1.21)	0.0142	1.89 (1.83-1.95)	0.715*† (0.703-0.728)	0.16* (0.15-0.17)	0.98 (0.93-1.03)	0.0079	2.00 (1.84-2.18)	0.689*† (0.660-0.718)	0.039*† (0.028-0.062)	0.95 (0.84-1.07)	0.0003§
CH-AI	3.60 (3.08-4.20)	0.814* (0.780-0.849)	0.28*† (0.22-0.37)	1.14 (1.00-1.28)	0.0078	2.80 (2.67-2.94)	0.767*† (0.755-0.779)	0.19*† (0.18-0.21)	0.97 (0.92-1.01)	0.0049	2.42 (2.20-2.67)	0.750 (0.725-0.775)	0.035*† (0.027-0.050)	1.26 (1.12-1.40)	0.0016§
Lead II only models															
ECG-AI	2.19 (1.98-2.42)	0.753† (0.712-0.793)	0.23* (0.18-0.30)	1.08 (0.94-1.22)	0.0072	1.92 (1.86-1.98)	0.717*† (0.704-0.730)	0.16* (0.14-0.17)	1.02 (0.96-1.07)	0.0096	1.98 (1.82-2.15)	0.690*† (0.659-0.720)	0.043*† (0.029-0.067)	1.14 (1.00-1.28)	0.0008§
CH-AI	3.62 (3.10-4.24)	0.811* (0.777-0.846)	0.29*† (0.23-0.37)	1.14 (1.00-1.27)	0.0080	2.84 (2.70-2.98)	0.765*† (0.753-0.777)	0.19*† (0.18-0.21)	0.98 (0.93-1.03)	0.0056	2.49 (2.25-2.75)	0.759*† (0.729-0.789)	0.036*† (0.026-0.051)	1.49 (1.33-1.66)	0.0024§

*p<0.05 for comparison against age and sex

†p<0.05 for comparison against CHARGE-AF

‡Integrated calibration index (ICI), a quantitative measure of the average difference between predicted event risk and observed event incidence, weighted by the empirical distribution of event risk.[30] Smaller values indicate better calibration.

§Values reflect ICI after recalibration to the baseline 2-year AF risk in the UK Biobank

Lead I models: Difference in AUROC for CH-AI vs ECG-AI: MGH p<0.05, BWH p<0.05, UK Biobank p<0.05; average precision MGH p<0.05, BWH p<0.05, UK Biobank p=NS

Lead II models: Difference in AUROC for CH-AI vs ECG-AI: MGH p<0.05, BWH p<0.05, UK Biobank p<0.05; average precision MGH p<0.05, BWH p<0.05, UK Biobank p=NS

AUROC = area under the receiver operating characteristic curve; SD = standard deviation

Table XII. Model performance for 2-year incident AF in MGH and BWH test sets

Massachusetts General Hospital (N=4,166)					***Brigham and Women's Hospital (N=37,963)***			
Model	**Hazard ratio (per 1-SD)**	**5-year AUC**	**Calibration slope**	**ICI**	**Hazard ratio (per 1-SD)**	**5-year AUC**	**Calibration slope**	**ICI**
Deep learning architectures								
ECG-AI	-	0.793 (0.744-0.841)	-	0.0066	-	0.793 (0.744-0.841)	-	0.0161
Cox proportional hazards models								
Age & Sex	2.79 (2.17-3.59)	0.752 (0.706-0.800)	1.02 (0.77-1.28)	0.0028	2.52 (2.33-2.73)	0.719 (0.702-0.736)	0.92 (0.84-1.00)	0.0019
CHARGE-AF	3.35 (2.56-4.34)	0.785* (0.740-0.830)	0.64 (0.50-0.78)	0.0363	2.62 (2.42-2.83)	0.733* (0.716-0.750)	0.53 (0.49-0.58)	0.0343
ECG-AI	3.19 (2.60-3.91)	0.793 (0.744-0.841)	1.11 (0.92-1.31)	0.0035	2.21 (2.09-2.35)	0.793 (0.744-0.841)	0.75 (0.70-0.81)	0.0068
CH-AI	4.04 (3.14-5.20)	0.810* (0.769-0.851)	1.15 (0.94-1.36)	0.0038	2.63 (2.46-2.82)	0.747*† (0.731-0.764)	0.83 (0.77-0.89)	0.0043

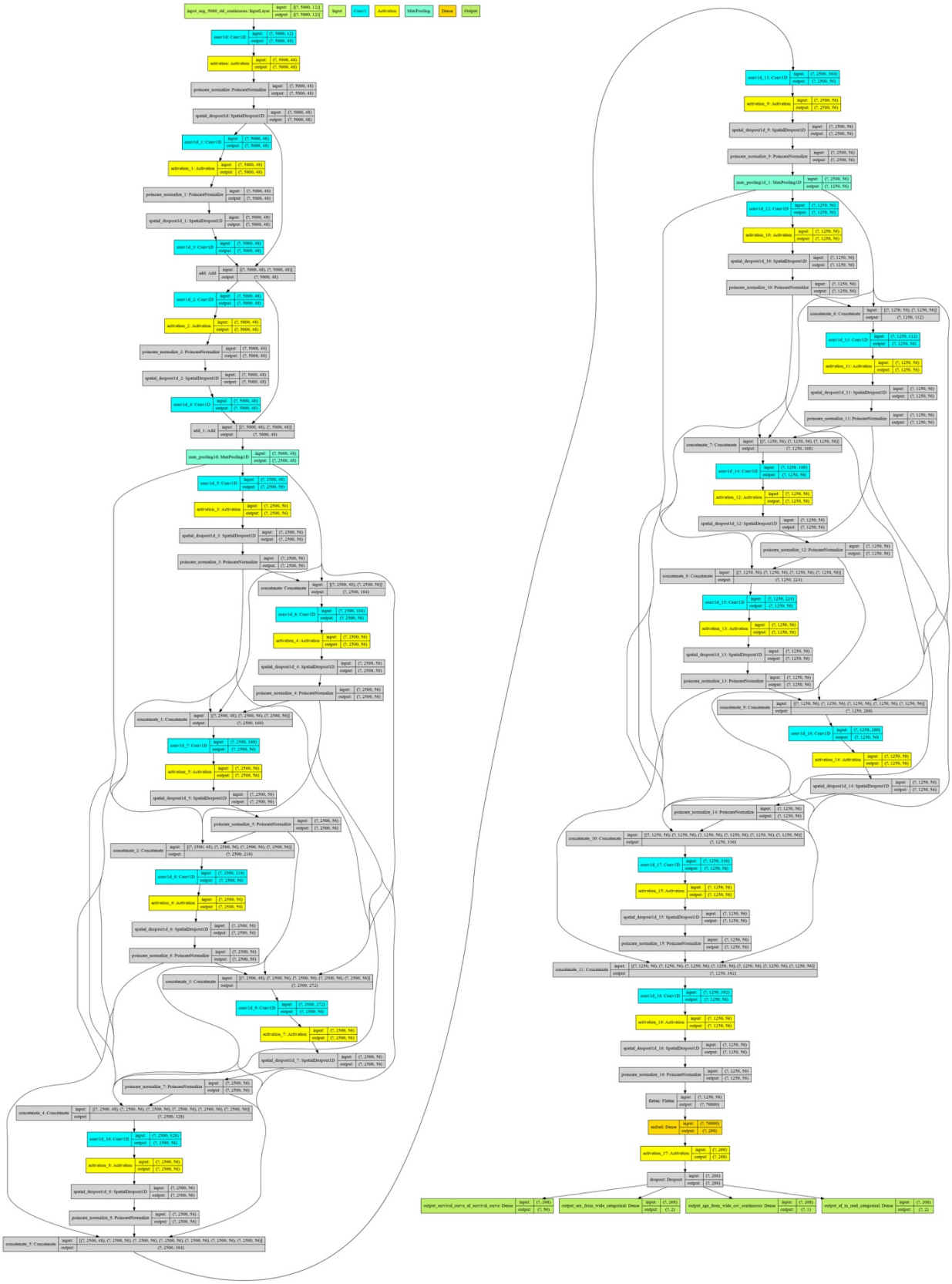
*p<0.05 for comparison against age and sex
†p<0.05 for comparison against CHARGE-AF
Difference in c-index for CH-AI vs ECG-AI: MGH p=0.088, BWH p<0.05

Table XIII. Sequential ECG-AI and clinical risk factor models

	Massachusetts General Hospital (N=4,166)					*Brigham and Women's Hospital (N=37,963)*					*UK Biobank (N=41,033)*				
Model	**Hazard ratio (per 1-SD)**	**5-year AUROC**	**5-year average precision**	**Calibration slope**	**ICI***	**Hazard ratio (per 1-SD)**	**5-year AUROC**	**5-year average precision**	**Calibration slope**	**ICI***	**Hazard ratio (per 1-SD)**	**2-year AUROC**	**2-year average precision**	**Calibration slope**	**ICI***
Primary analysis models															
ECG-AI	2.45 (2.23-2.69)	0.823 (0.790-0.856)	0.27 (0.21-0.34)	1.06 (0.95-1.17)	0.0212	2.05 (1.98-2.11)	0.747 (0.736-0.759)	0.19 (0.17-0.20)	0.81 (0.77-0.84)	0.0129	2.01 (1.88-2.14)	0.705 (0.673-0.737)	0.060*† (0.044-0.090)	0.75 (0.68-0.82)	0.0035†
CH-AI	3.74 (3.24-4.33)	0.838 (0.807-0.869)	0.30 (0.24-0.38)	1.13 (1.01-1.25)	0.0120	2.76 (2.64-2.88)	0.777 (0.766-0.788)	0.21 (0.19-0.23)	0.77 (0.74-0.81)	0.0108	2.27 (2.11-2.44)	0.746 (0.716-0.776)	0.059*† (0.042-0.083)	1.01 (0.92-1.10)	0.0001†
Sequentially adjusted models															
ECG-AI + Age + Sex	3.29 (2.88-3.75)	0.827 (0.793-0.860)	0.28 (0.23-0.36)	1.09 (0.97-1.21)	0.0143	2.56 (2.46-2.67)	0.768 (0.756-0.780)	0.20 (0.18-0.21)	0.89 (0.85-0.93)	0.0108	2.25 (2.10-2.42)	0.743 (0.708-0.777)	0.060 (0.043-0.091)	1.07 (0.98-1.17)	0.0126†
ECG-AI + Age + Sex + Vital Signs	3.47 (3.03-3.98)	0.831 (0.799-0.863)	0.30 (0.23-0.39)	1.12 (1.00-1.24)	0.0082	2.63 (2.52-2.74)	0.771 0.760-0.783)	0.20 (0.19-0.22)	0.89 (0.85-0.93)	0.0121	2.31 (2.15-2.49)	0.746 (0.715-0.776)	0.060 (0.043-0.091)	1.10 (1.00-1.20)	0.0128†
ECG-AI + Age + Sex + Vital Signs + Clinical Risk Factors	3.73 (3.24-4.30)	0.841 (0.811-0.872)	0.31 (0.24-0.39)	1.14 (1.01-1.26)	0.0089	2.76 (2.65-2.89)	0.780 (0.770-0.791)	0.21 (0.19-0.23)	0.91 (0.87-0.95)	0.0106	2.30 (2.14-2.49)	0.745 (0.717-0.773)	0.057 (0.040-0.085)	1.19 (1.08-1.30)	0.0132†

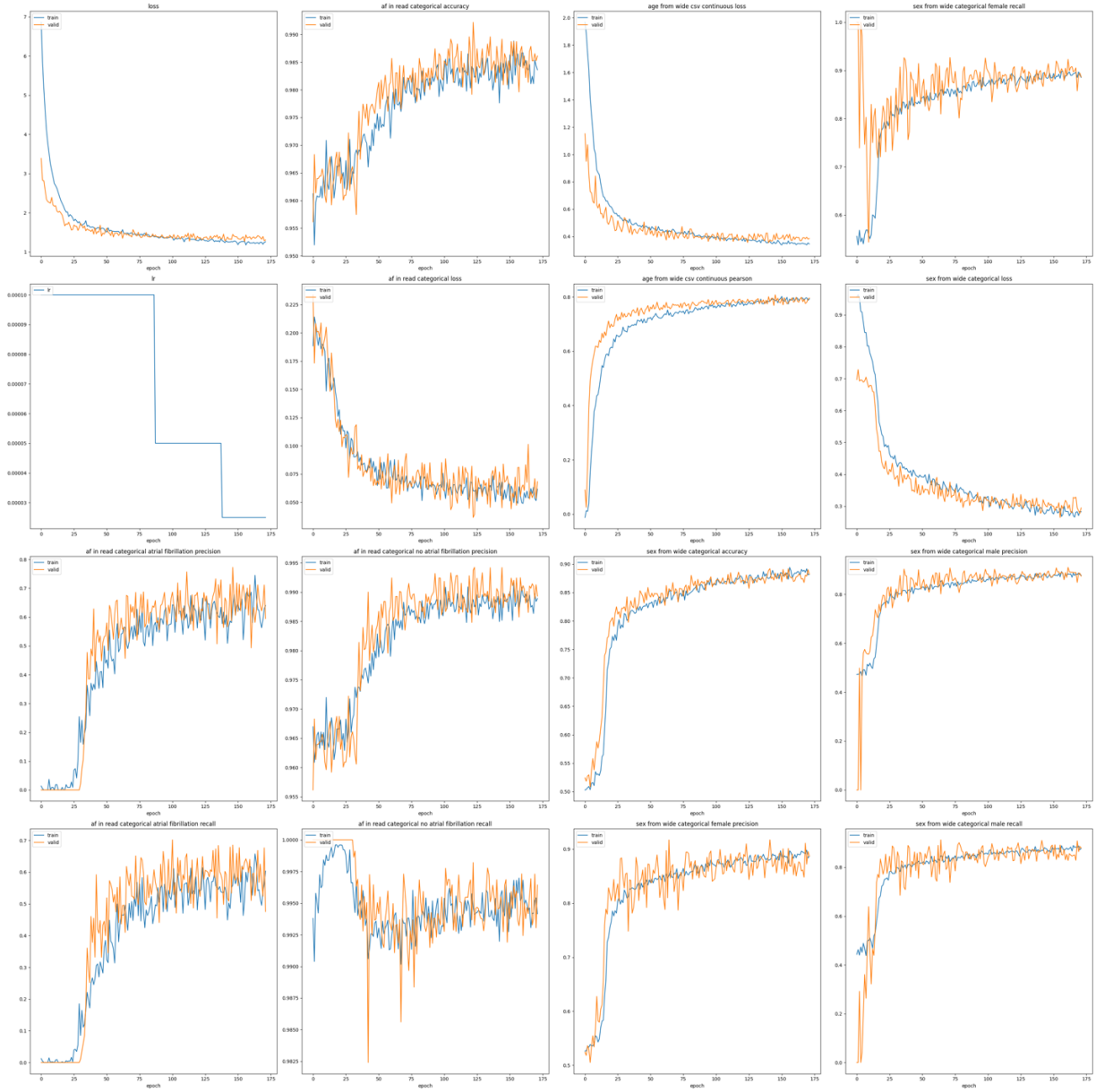
*Integrated calibration index (ICI), a quantitative measure of the average difference between predicted event risk and observed event incidence, weighted by the empirical distribution of event risk.[30] Smaller values indicate better calibration.
†Values reflect ICI after recalibration to the baseline 2-year AF risk in the UK Biobank
Vital Signs = height, weight, diastolic blood pressure, systolic blood pressure
Clinical Risk Factors = race, smoking, anti-hypertensive use, diabetes, myocardial infarction, heart failure
AUROC = area under the receiver operating characteristic curve; ICI = integrated calibration index; SD = standard deviation

Figure I. ECG-AI model architecture



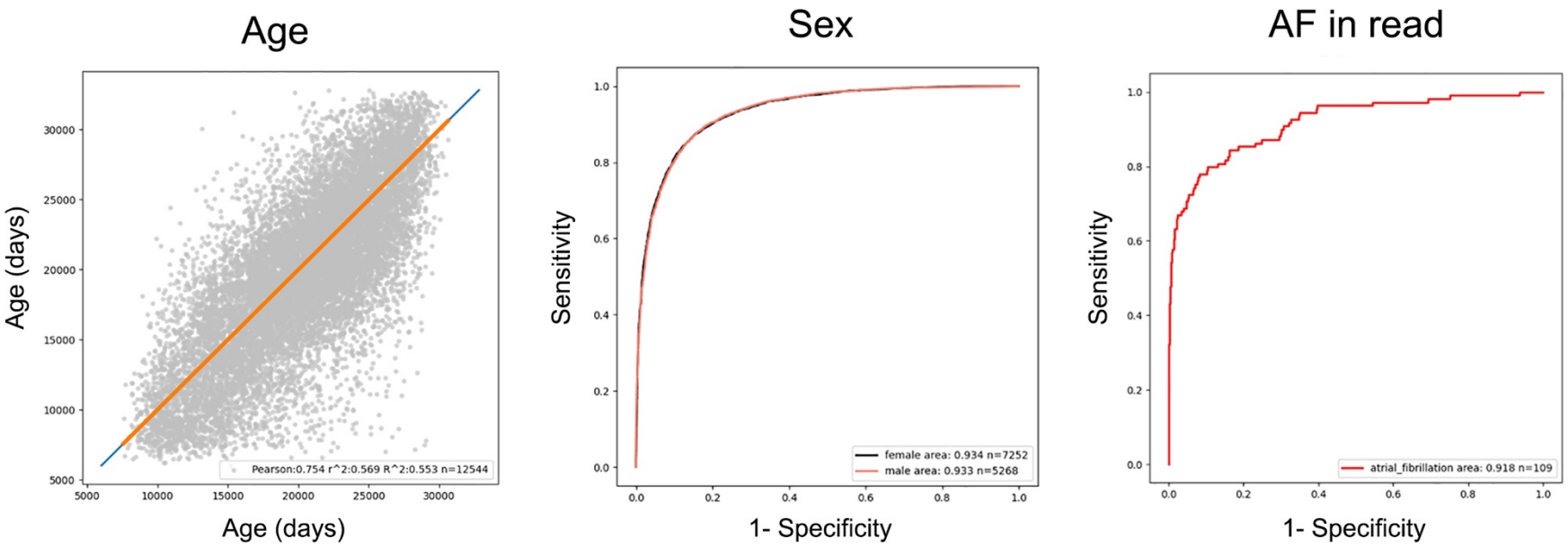
Depicted is a summary of the architecture of ECG-AI. Briefly, the model takes 10 seconds of continuous 12-lead ECG waveform data as input into the first convolutional layer. The fully-connected layers take convolved ECG waveform data only to produce an estimate of time to AF (primary) as well as predictions of age, sex, and presence of AF in the diagnostic statement (secondary). Arrows depict connections between layers. Conv1D = one-dimensional convolution, MaxPooling1D = one-dimensional maximum pooling.

Figure II. ECG-AI learning curves



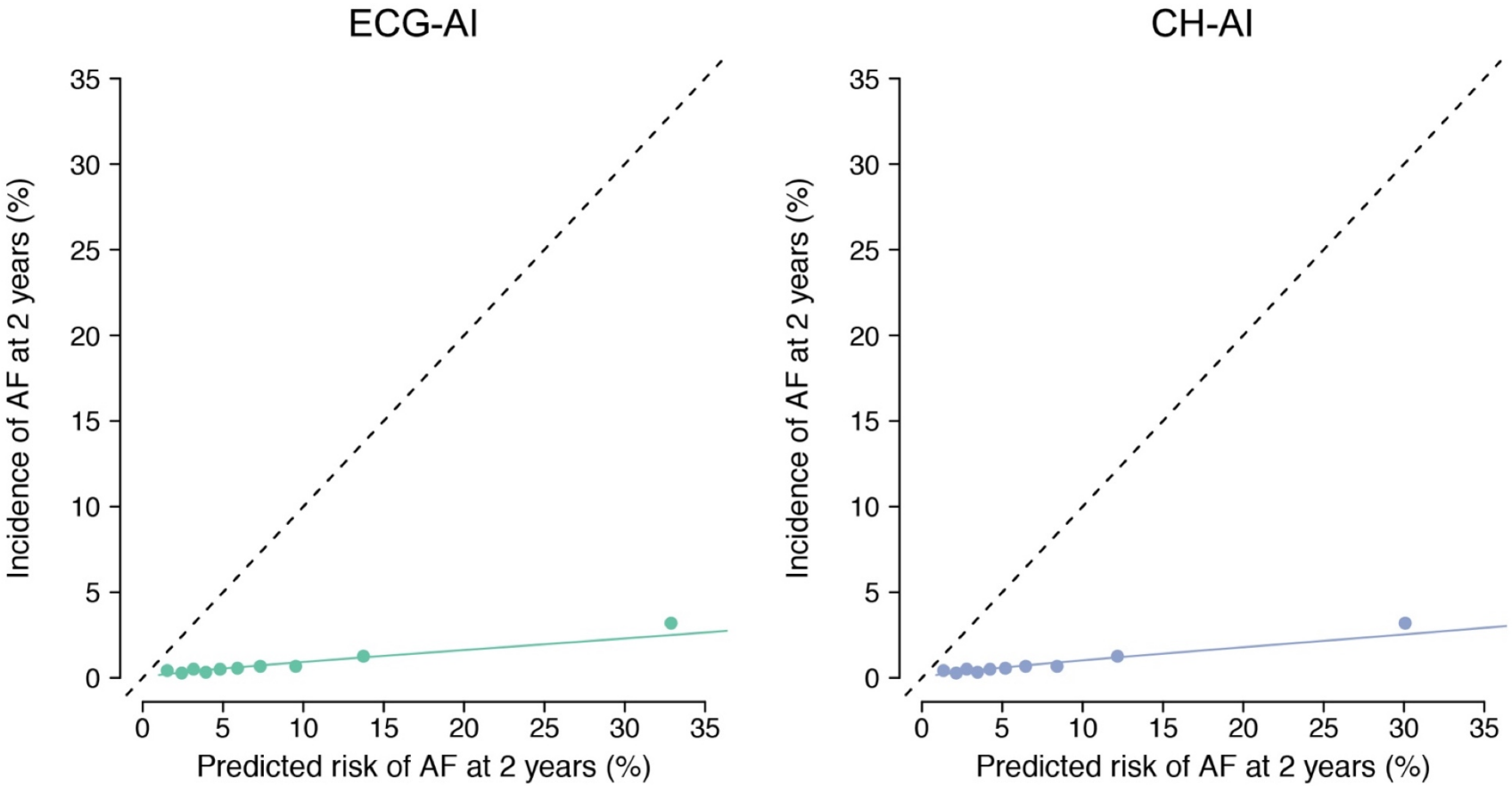
Depicted are learning curves for ECG-AI. The blue curve depicts the training set, and the orange curve depicts the internal validation set across sequential epochs (x-axis).

Figure III. Performance of ECG-AI for secondary tasks



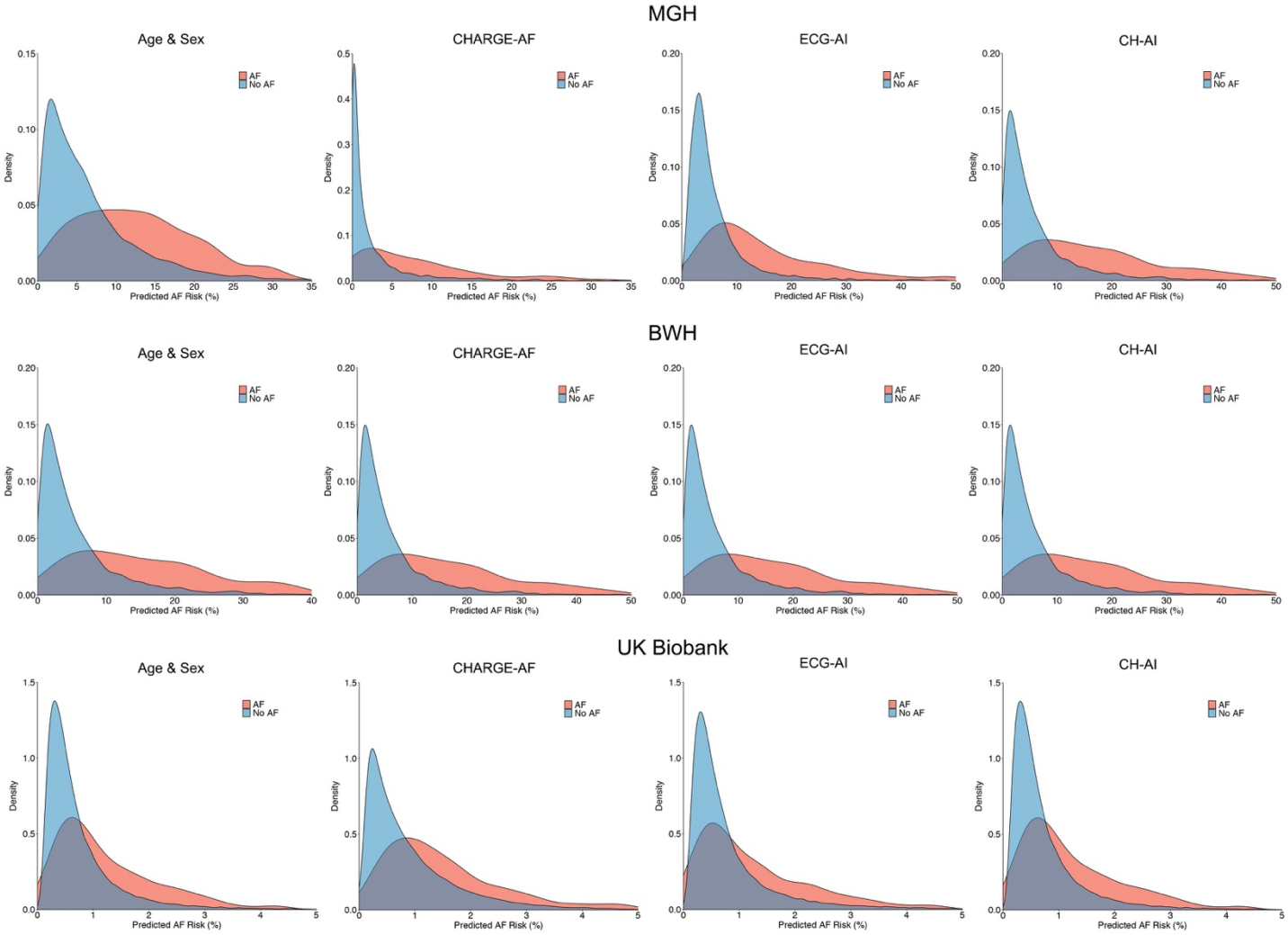
Depicted is the performance of ECG-AI on three secondary tasks: 1) prediction of age, 2) prediction of sex, and 3) classification of the presence of AF in the diagnostic statement. The left plot shows the correlation between ECG-AI predicted age (x-axis) and true age (y-axis), with the Pearson correlation depicted on the bottom right. The middle and right plots show receiver operating characteristic curves for the labeled classification task, with the area under the curve depicted on the bottom right.

Figure IV. Calibration of ECG-AI and CH-AI in UK Biobank prior to recalibration



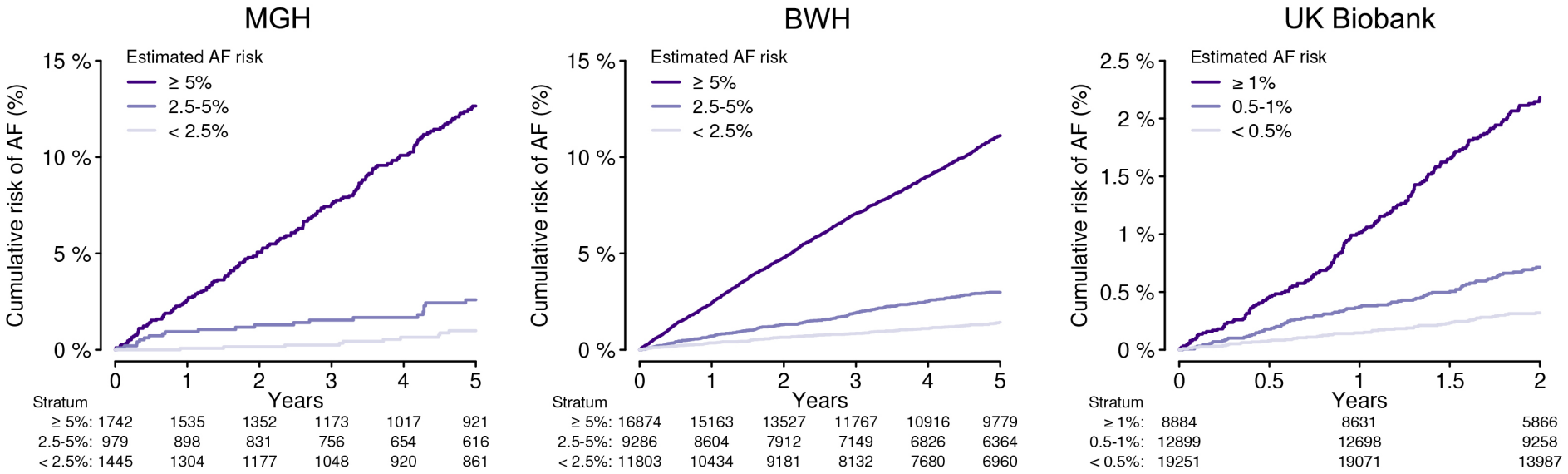
Depicted is the relationship between predicted event risk (x-axis) and observed cumulative event incidence (y-axis) for ECG-AI (left) and CH-AI (right) in the UK Biobank prior to recalibration. Each point represents increasing decile of predicted risk. Perfect calibration is indicated by the hashed diagonal line, denoting perfect correspondence between predicted and observed risk. A fitted curve was obtained using adaptive hazard regression[30] relating predicted risk and observed event risk.

Figure V. Distributions of AF risk in each test set



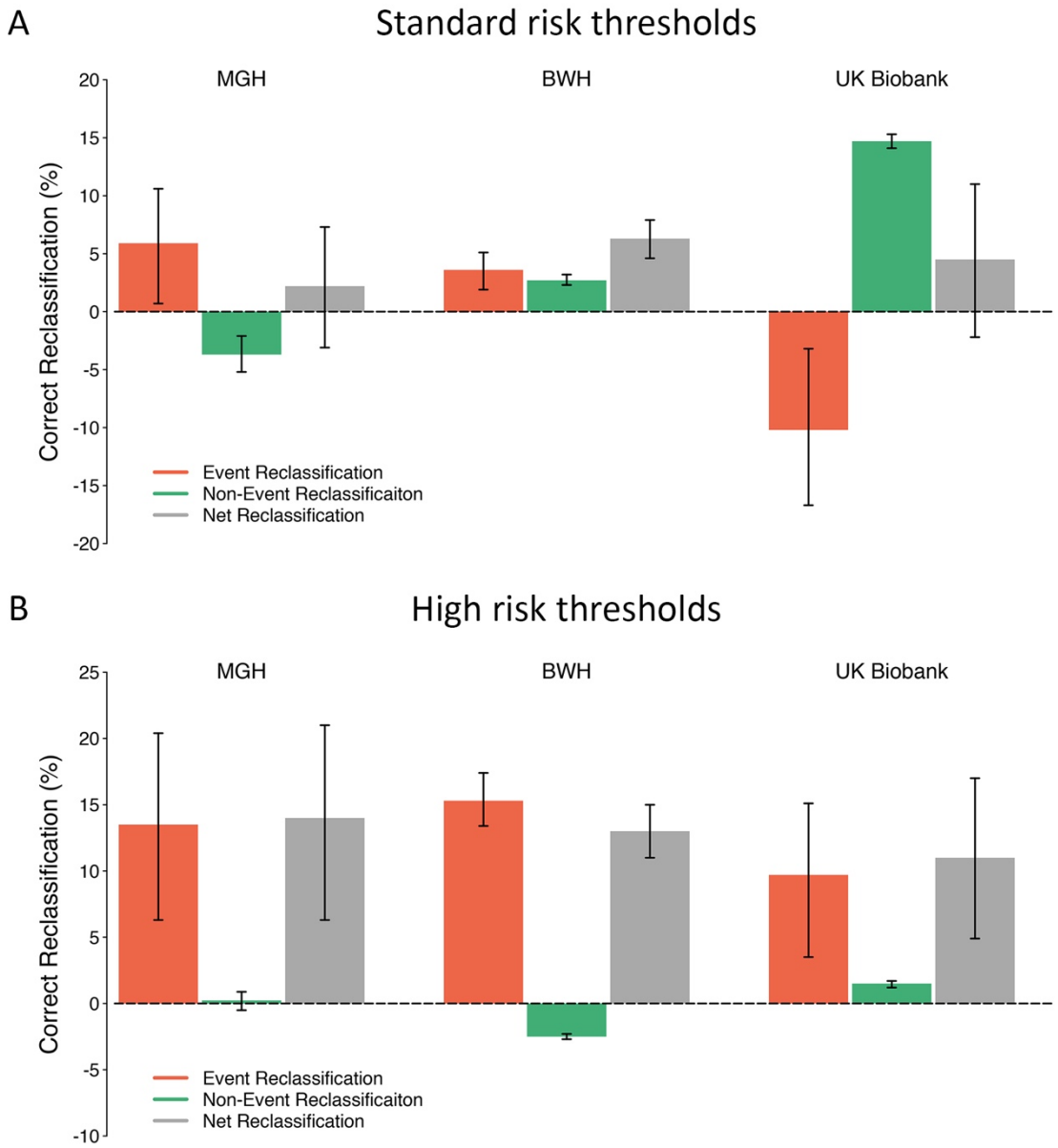
Depicted are density plots showing the distribution of predicted AF risk in each test set (at 5 years for MGH and BWH, and 2 years for UK Biobank). Individuals who developed AF within the respective window are depicted in red, while individuals who did not develop AF are depicted in blue. All models were recalibrated to the sample-level 2-year AF hazard in the UK Biobank (see main text).

Figure VI. Cumulative risk of AF according to CH-AI in the test sets



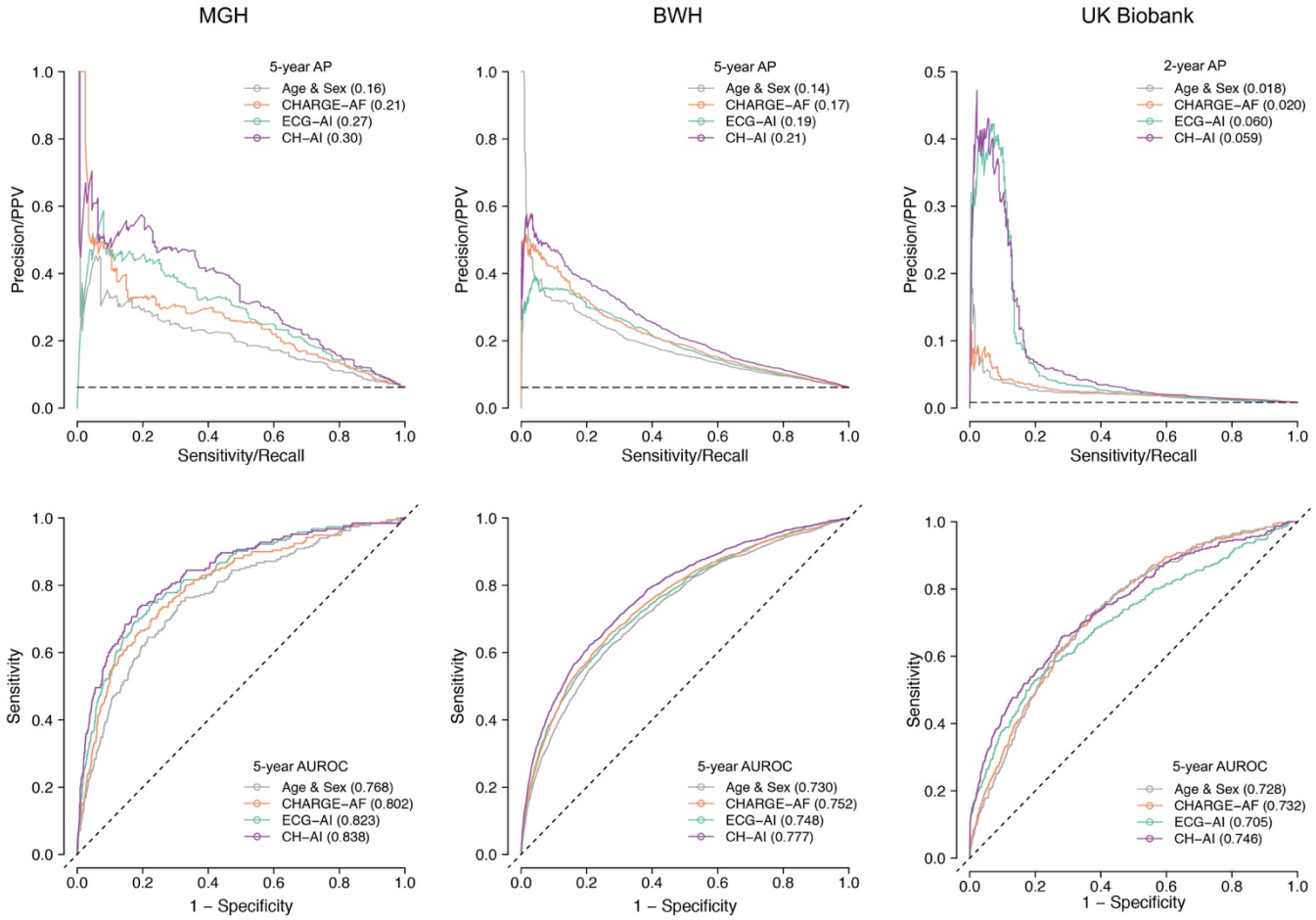
Depicted is the cumulative risk of AF at 5 years in the MGH (left) and BWH (middle) test sets, and the cumulative risk of AF at 2 years in the UK Biobank test set. In MGH and BWH, risk is stratified using the CH-AI score at predicted risk levels previously utilized to denote low, medium, and high risk using the CHARGE-AF score (i.e., <2.5%, 2.5-5%, and ≥5%).[13] In the UK Biobank test set, strata of CH-AI were defined at 2-year AF probability thresholds approximating tertiles (i.e., <0.5%, 0.5-1%, and ≥1%). CH-AI was recalibrated to the sample-level 2-year AF hazard in the UK Biobank (see main text). The number at risk within each stratum depicted below each plot.

Figure VII. Net reclassification improvement using CH-AI versus CHARGE-AF



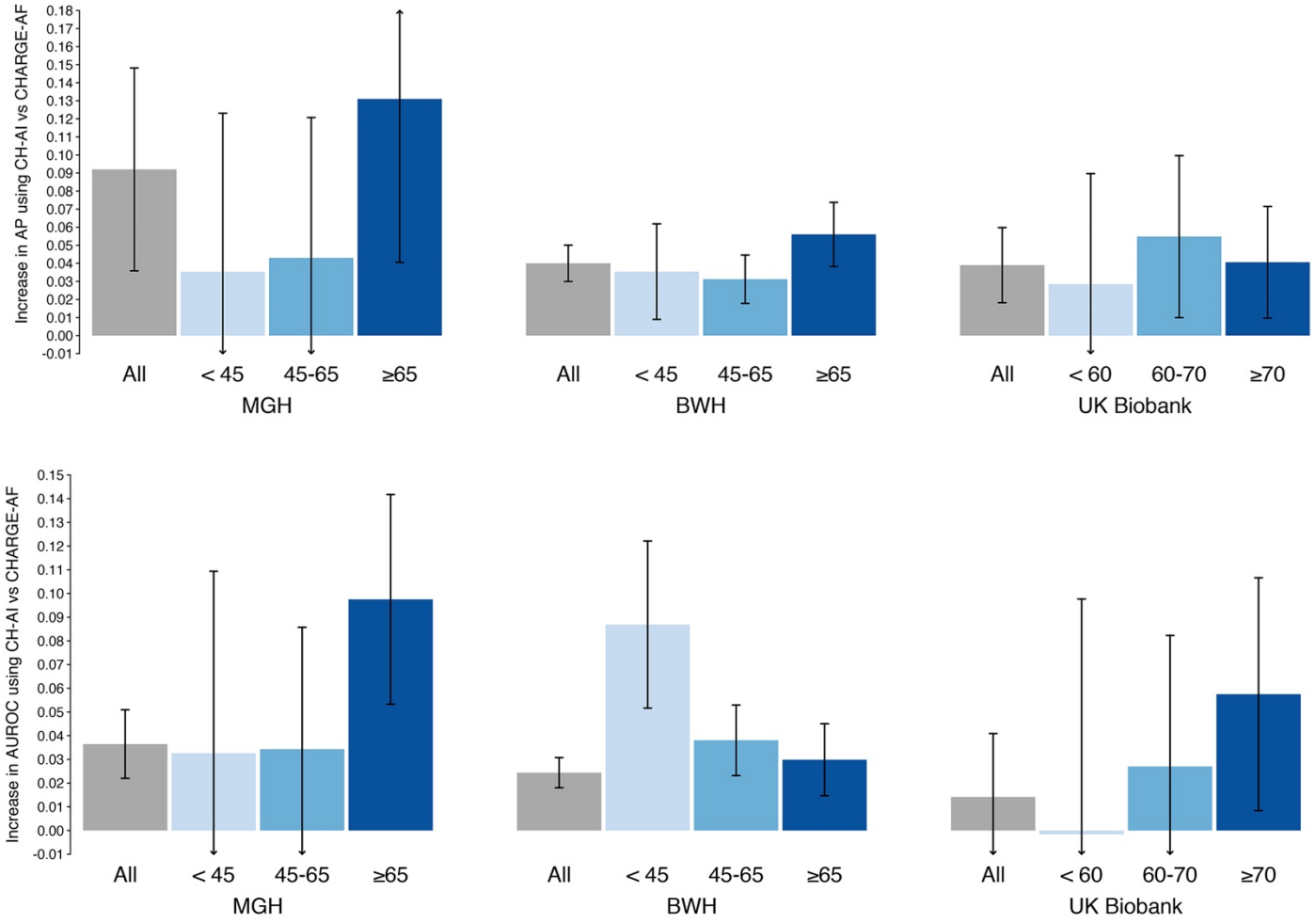
Depicted is reclassification performance using CH-AI versus CHARGE-AF. Each plot depicts time-dependent categorical reclassification, with **Panel A** utilizing the risk thresholds used in the original CHARGE-AF derivation study (i.e., 5-year AF risk <2.5%, 2.5-5%, and ≥5%), and **Panel B** utilizing higher risk thresholds (i.e, 5-year AF risk <20% and ≥20%). Given our 2-year window in UK Biobank, we utilized analogous thresholds (standard: 2-year AF risk <0.5%, 0.5-1%, and ≥1%; high: <2% and ≥2%). In each plot, event reclassification is depicted in red, non-event reclassification is depicted in green, and net reclassification is depicted in gray. Error bars depict 95% confidence intervals and the horizontal dashed line depicts no reclassification

Figure 8. Precision-recall and receiver operating characteristic curves



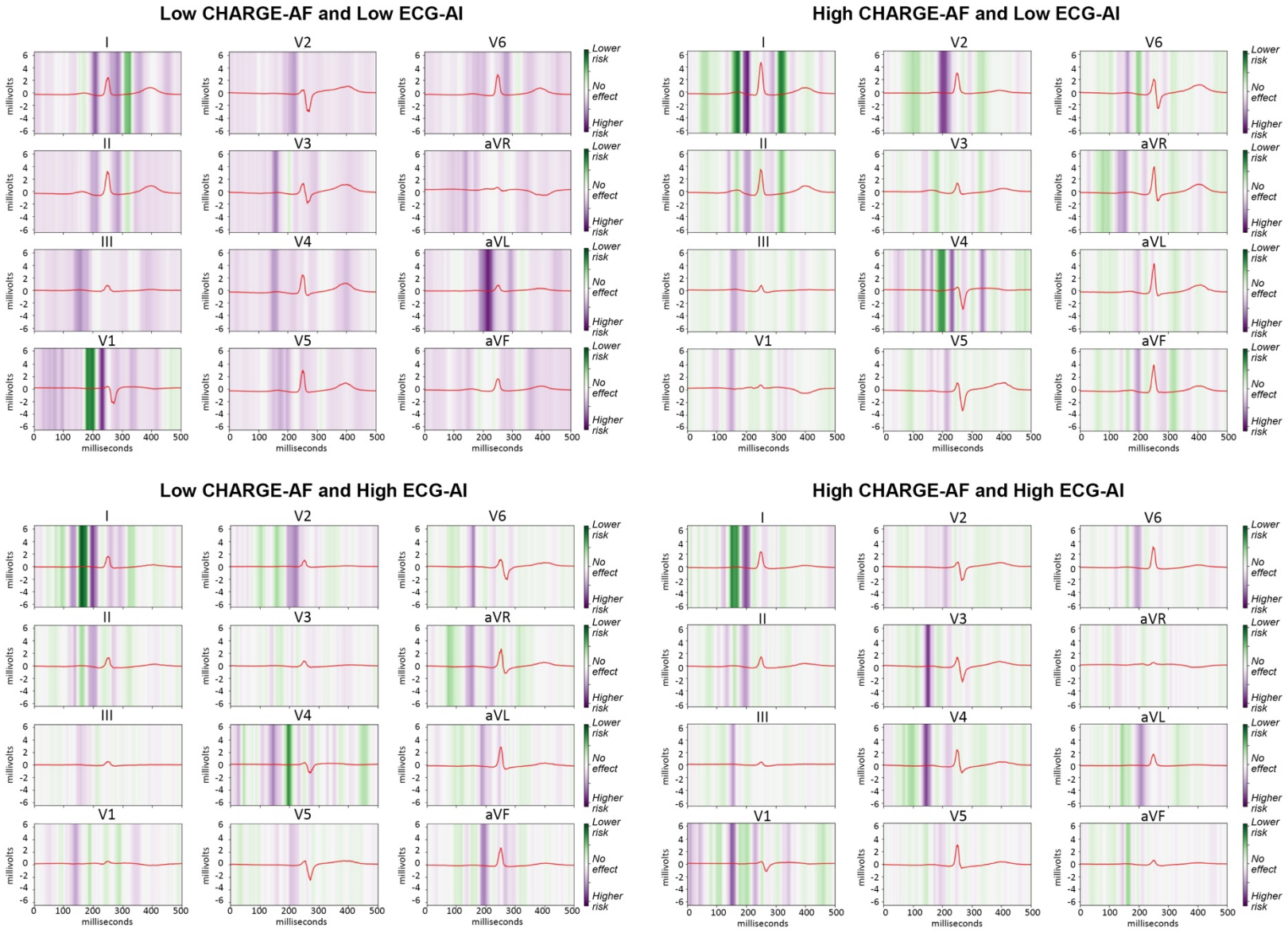
Depicted are precision-recall (upper panels) and receiver operating characteristic (bottom panels) curves for CH-AI (blue), ECG-AI (green), CHARGE-AF (orange) and age and sex (gray). Each plot displayed time-dependent test characteristics at the 5-year window (MGH and BWH) and 2-year window (UK Biobank). The area under each curve is listed in the legend next to the corresponding model. The hashed horizontal (upper panels) and diagonal (lower panels) lines depict the performance of randomly guessing model. PPV = positive predictive value; AP = average precision; AUROC = area under the receiver operating characteristic curve.

Figure 9. Improvement in discrimination using CH-AI versus CHARGE-AF across subgroups of age



Depicted is the improvement in average precision (AP, upper panels) and receiver operating characteristic curve (AUROC, lower panels) using CH-AI versus CHARGE-AF overall and across subgroups of age. Thresholds of age were chosen to approximate tertiles of the age distribution for each sample. Error bars depict 95% confidence intervals obtained using bootstrapping, with arrows denoting confidence intervals that have been trimmed for graphical purposes.

Figure 10. ECG-AI saliency map across strata of ECG-AI and CHARGE-AF risk



Depicted is a saliency map of ECG-AI across binary strata of predicted AF risk according to CHARGE-AF and ECG-AI. High AF risk was defined as 5-year predicted AF risk ≥5%. Each quadrant shows saliency maps of all 12 leads for the labeled combination of CHARGE-AF and ECG-AI risk. Each map demarcates regions of the ECG waveform having the greatest influence on AF risk predictions. Specifically, green shades depict areas of the ECG where greater amplitude is associated with lower predicted AF risk, purple shades depict areas of the ECG where greater amplitude is associated with higher predicted AF risk, and white depicts areas of the ECG where amplitude changes have no effect on predicted AF risk. Saliency was averaged over a random sample of 256 individuals in the BWH test set, and the red waveform depicts the median waveform in each lead among the 256 individuals.