ONLINE SUPPLEMENT

Artificial Intelligence Enabled Prediction of Heart Failure Risk from Single-lead Electrocardiograms

Lovedeep S Dhingra MBBS, Arya Aminorroaya MD, MPH, Aline F Pedroso PhD, Akshay Khunte BS, Veer Sangha BS, Daniel McIntyre MD, PhD, Clara K Chow MBBS, PhD, Folkert W Asselbergs MD, PhD, Luisa CC Brant MD, PhD, Sandhi M Barreto MD, PhD, Antonio Luiz P Ribeiro MD, PhD, Harlan M Krumholz MD, SM, Evangelos K Oikonomou MD, DPhil, Rohan Khera MD, MS

Table of Contents

eMethods 4

Online References 7

eFigure 1. Overview of Cohort Creation at the Yale New Haven Health System. 8

eFigure 2. Consort Diagram for Study Cohorts. 9

eFigure 3. Model Performance for Cross-sectional Detection of Left Ventricular Systolic Dysfunction in the Yale New Haven Hospital Held-out Test Set and Across External Validation Cohorts. 10

eFigure 4. Age- and Sex-adjusted Hazard for Heart Failure across Model Output Probability Bins. 11

eFigure 5. Net Benefit of AI-ECG Model Output Probability and Pooled Cohort Equations to Prevent Heart Failure) and Predicting Risk of Cardiovascular Disease Events Equations for Predicting Incident Heart Failure Across Thresholds at (A) Yale New Haven Health System (B) UK Biobank (C) Brazilian Longitudinal Study of Adult Health. 12

eFigure 6. Cumulative Hazard for Heart Failure Adjusted for Pooled Cohort Equations to Prevent Heart Failure (PCP-HF) Risk Score at (A) Yale New Haven Health System (B) UK Biobank (C) Brazilian Longitudinal Study of Adult Health. 13

eFigure 7. Cumulative Hazard for Heart Failure Adjusted for Pooled Cohort Equations to Prevent Heart Failure (PCP-HF) Risk Score at Yale New Haven Health System. 14

eFigure 8. Cumulative Hazard for Heart Failure Adjusted for the Predicting Risk of Cardiovascular Disease Events (PREVENT) Equations at (A) Yale New Haven Health System (B) UK Biobank (C) Brazilian Longitudinal Study of Adult Health. 15

eFigure 9. Cumulative Hazard for Heart Failure Adjusted for the Predicting Risk of Cardiovascular Disease Events (PREVENT) Equations at Yale New Haven Health System. 16

eTable 1. International Classification of Disease Tenth Revision Codes for the Identification of Comorbidities and Outcomes. 17

eTable 2. Model Performance Measures for Cross-sectional Detection of Left Ventricular Systolic Dysfunction in the Yale New Haven Hospital Held-out Test Set and Across External Validation Cohorts. 18

eTable 3. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis using Artificial Intelligence (TRIPOD + AI) Checklist. 19

eTable 4. Population Characteristics of the Yale New Haven Health System Sites. 21

eTable 5. Model Performance for Predicting Heart Failure Risk. 23

eTable 6. Model Performance for Prediction of Heart Failure Across Demographic Subgroups. 24

eTable 7. Age- and Sex- Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure-related Outcomes. 25

eTable 8. Age- and Sex- Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure-related Outcomes. 26

eTable 9. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure across Model Output Probabilities 27

eTable 10. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure across Model Output Probabilities at the Yale New Haven Health System Sites. 28

eTable 11. Comparison of Harrel's C-statistic for the Input Features of the Clinical Risk Scores with the Single-lead Artificial Intelligence-enhanced Electrocardiogram Model. 29

eTable 12. Categorical and Continuous Time-to-Event, Event and Non-event Net Reclassification Index of AI-ECG Model Output Probability over Pooled Cohort Equations to Prevent Heart Failure, and Predicting Risk of Cardiovascular Disease Events Equations. 30

eTable 13. Performance Measures of the AI-ECG Model Output Probability, Pooled Cohort Equations to Prevent Heart Failure, and Predicting Risk of Cardiovascular Disease Events Equations for Predicting Incident Heart Failure with Censoring the Observations at the Median Duration of Follow-up. 31

eTable 14. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Non-Heart Failure Clinical Outcomes. 32

eTable 15. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Non-Heart Failure Clinical Outcomes Across Yale New Haven Health System Sites. 33

eMethods

Data Sources

The Yale New Haven Health System (YNHHS) is the largest referral center in southern New England and serves a diverse patient population. The YNHHS includes five hospitals, Yale New Haven Hospital, Bridgeport Hospital, Greenwich Hospital, Lawrence and Memorial Hospital, and Westerly Hospital, and a large network of community outpatient clinics, the Northeast Medical Group. The electronic health records (EHR) data was acquired during patient care at YNHHS using Epic and was extracted from the Clarity database.[1,2]

UK Biobank (UKB) is a prospective cohort of 502,468 community-dwelling adults aged 40-69 years recruited during 2006-2010.[3] A group of these participants accepted to participate in the third or fourth UKB study visit during which the participants underwent 12-lead electrocardiograms (ECGs) in 2014-2021. The UKB dataset is linked with the national EHR from the UK National Health Service predating UKB enrollment, enabling access to EHR diagnosis codes.[4,5] We used data from UKB under research application #71033.

The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) study, a large multicenter prospective cohort study conducted in Brazil, enrolled,105 community-dwelling adults aged 35-74 years at their baseline visit during 2008-2010.[6,7] These participants represent active and retired civil servants from six higher education and research institutions in Brazilian state capitals in three geographical regions of the country: Southeast (Belo Horizonte, Rio de Janeiro, São Paulo and Vitória), South (Porto Alegre) and Northeast (Salvador).[8] The ELSA-Brasil study aimed to investigate the development and progression of chronic diseases and their determinants in the Brazilian adult population. Baseline data were collected using validated instruments, physical examinations, laboratory assessments, and imaging modalities.[6] Additionally, all participants underwent protocolized 12-lead ECG and echocardiogram.[6,7] To ascertain exposure status and to identify changes in baseline, ELSA-Brasil participants present for in-person follow-up visits every three to four years. Moreover, telephone interviews occur annually to obtain information on new diagnoses, hospitalization, and death with adjudicated clinical events based on expert medical record review.[6]

Study Population

In YNHHS, to identify patients with prevalent heart failure (HF) at the time of ECG, we identified the first recorded encounter for all patients within the EHR and followed for 1 year. Patients with prevalent HF based on either a diagnosis code for HF or an echocardiogram with left ventricular ejection fraction under 50% or left ventricular diastolic dysfunction (defined as “moderate” or “severe” left ventricular diastolic dysfunction) we excluded from the study. The baseline ECG for patients was defined as an outpatient ECG recorded after this 1-year blanking period to exclude prevalent HF (e**Figure 1**). The YNHHS cohort also excluded patients previously included in the development of the AI-ECG algorithm and a small proportion of individuals who opted out of research participation (<0.01% of all YNHHS patients).

Study Exposure

The model development population consisted of 503,516 ECGs from 110,228 unique patients (**Figure 1**). We used raw voltage data from 12-lead ECGs obtained as standard 10-second 12-lead ECGs at a sampling frequency of 500 Hz or 250 Hz and extracted the lead I waveforms. Median filtering was conducted by subtracting a one-second median filter from the acquired signals to eliminate baseline drift. To incorporate noise during the model development, we isolated four distinct noises from a 5-minute random Gaussian noise within four frequency ranges of 3-12 Hz, 12-50 Hz, 50-100 Hz, and 100-150 Hz, each corresponding to the frequency range of a specific type of real-world noise.[10,11] The noise with a frequency range of 3-12 Hz reflects the motion artifact noises attributable to tremors, 50-100 Hz accounts for the electrode contact noise, and 12-50 Hz and 100-150 Hz reflect the lower and higher-frequency muscle noises, respectively. Each ECG in the training set was included twice with different random noises signal-to-noise ratios. This augmentation involved a random type of noise and a random signal-to-noise ratio (SNR). For this purpose, we first randomly selected one of the four abovementioned distinct random Gaussian noises. Finally, the selected noise was introduced to the ECG waveform with a random SNR ranging from 0.5 to 1.25, representing a heavy and a light burden of noise in ECGs, respectively.

The employed convolutional neural network (CNN) architecture comprised an input layer with dimensions of (5000, 1, 1), representing a 10-second, 500 Hz, lead I ECG.1 The input layer was followed by seven 2-dimensional convolutional layers, progressively increasing the number of filters from 16 to 64 while incorporating varying kernel sizes (7x1, 5x1, and 3x1) to capture different levels of feature abstraction. A batch normalization layer, a rectified linear unit (ReLU) activation layer, and a 2-dimensional max-pooling layer with different pool sizes (2x1 and 4x1) followed each convolutional layer. Next, the output of the 7th convolutional layer was used as the input for a fully connected network that included two dense layers. Each dense layer was followed by a batch normalization layer, a ReLU activation layer, and a dropout layer with a rate of 0.5. Finally, the model output was a dense layer with a single class and a sigmoid activation to generate the output

probability of the label. The loss function was adjusted by calculating model weights using a class re-weighting approach to ensure that the learning is not impacted by the differential prevalence of positive and negative labels.

We defined a positive AI-ECG screen as a model output probability greater than 0.08, representing the probability threshold at which the model achieved a sensitivity of 90% for detecting LVSD during internal validation. We further defined graded thresholds based on AI-ECG probabilities of 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1 to evaluate the association of a higher risk score with HF. Notably, while the model was developed for detecting the cross-sectional signature of LVSD using data from the YNHH alone, it was applied across all YNHHS sites and the population-based cohorts without any further development or fine-tuning for prediction of HF risk.

Study Outcomes and Covariates

We identified available demographic characteristics across cohorts, including age at the time of ECG, sex, and self-reported race and ethnicity. Comorbidities, including ischemic heart disease, hypertension, and type 2 diabetes mellitus, were defined using relevant EHR diagnosis codes in YNHHS and UKB (**eTable 1**). Obesity was defined as BMI ≥30 kg/m2.

In ELSA-Brasil, covariates were recorded at the baseline study visit.31 Race was self-classified based on Brazil's National Bureau of Statistics definition and classified as White, Black, "Pardo", Asian, or Others.31,32In ELSA-Brasil, HF was identified either by in-person interview or the annual telephonic surveillance and investigated by a designated committee that contacted health providers and requested copies of medical records for all hospitalizations. After investigation, the cardiovascular events were adjudicated by an independent review of two cardiologists. A third senior cardiologist defined the event in case of disagreement.[12,13] HF was identified from hospitalization records, based on the presence of a clinical diagnosis of HF, with the individual receiving pharmacological therapy for HF, in addition to any of the following: (1) pulmonary congestion on chest X-ray, (2) reduced ejection fraction or systolic dysfunction observed on cardiac imaging, or (3) preserved ejection fraction with evidence of moderate to severe diastolic dysfunction.

Information about all-cause death was available in the YNHHS EHR, with in-hospital mortality data supplemented from the Connecticut death index to improve capturing out-of-hospital patient mortality. Similarly, information about mortality was available in UKB via linkage to the EHR and the UK national death registries. Information about death in the ELSA-Brasil study was recorded via telephonic surveillance and confirmed using the national mortality database and death certificates.

Study Comparator

We employed two established clinical models to predict HF risk, the pooled cohort equations to prevent HF (PCP-HF) and the predicting risk of cardiovascular disease events (PREVENT) equations, as the baseline models.[11–13] The PCP-HF entails sex- and race-specific equations for estimating the 10-year risk of incident HF. However, the PREVENT predicts the 10-year HF risk independently of race. To align with the score development across cohorts, the PCP-HF score was calculated for White and Black individuals between 30 and 80 years of age, and the PREVENT score was computed for all individuals between 30 and 80 years of age with complete documentation of the score covariates. Of note, to conform with the design of PCP-HF and PREVENT, we have compared the performance of AI-ECG with these risk scores in individuals without a history of HF and ASCVD.[11,12] The calculated 10-year risk score was adjusted based on the length of follow-up for each individual to estimate the risk of HF over the study period.

In YNHHS, PCP-HF and PREVENT features were extracted from the EHR. Body mass index (BMI), systolic blood pressure, and laboratory measurements closest to and within two years of the ECG acquisition date were used for calculation. In UKB, the demographic features were identified from the baseline visit. Blood pressure measurement and smoking status assessment were conducted at the time of ECG acquisition. Laboratory values were measured in the first and second study visits, while ECGs were recorded in third and fourth visits.[14] We used the laboratory values closest to the ECG acquisition for the calculation of the PCP-HF and PREVENT scores. History of hypertension and diabetes were defined using ICD diagnosis codes from the linked EHR and self-reported use of anti-hypertensive and anti-hyperglycemic medications was recorded at the time of ECG acquisition.[15,16] In ELSA-Brasil, all features, including the ECG recording, were captured at the baseline visit using established study protocols.[17]

Statistical Analysis

Integrated discrimination improvement (IDI) was calculated as the difference between the improvements in the average predicted probabilities for those with and without the outcome for the AI-ECG model output vs. PCP-HF and PREVENT scores in each data source. Categorical net reclassification improvement (NRI) was calculated for the 0.08 threshold of the AI-ECG model. We also calculated event and non-event NRIs.

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of AI-ECG, PCP-HF, and PREVENT for predicting new-onset HF were also reported with censoring the observations at the median duration of follow-up. Net benefit evaluates true positives while accounting for the potential for increased false positives, ranging from 0-1, with higher values showing greater benefit. This was calculated using the following formula:

$$Net\ Benefit\ = \left(\frac{True\ Positive\ Rate - False\ Positive\ Rate}{1 - Probability\ Threshold}\right)$$

Categorical variables were reported as counts (percentages), and continuous variables as median (interquartile range [IQR]). All statistical tests were 2-sided with a level of significance set at 0.05. Harrell's C-statistic was calculated based on various Cox proportional hazard models with the AI-ECG model probability, age, sex, PCP-HF, and PREVENT scores as covariates. In Cox proportional hazard models, we treated death as a censoring event or, in sensitivity analyses, included death as part of a composite outcome in the dependent variable. compareC package in R was used for calculating and comparing Harrel's C-statistics.[17] Further, we evaluated the Harrel's C-statistics for input feature variables of the clinical risk scores. We also evaluated the association of AI-ECG with HF, overall and across racial groups, after adjusting for the clinical risk scores in Cox models. All analyses were conducted using a combination of Python 3.11.2 and R version 4.2.0. The Yale Institutional Review Board approved the study protocol and waived the need for informed consent as the study involves secondary analysis of pre-existing data.

Online References

1. Chishtie J, Sapiro N, Wiebe N, et al. Use of Epic Electronic health record system for health care research: Scoping review. *J Med Internet Res*. 2023;25(1):e51003.

2. Dhingra LS, Shen M, Mangla A, Khera R. Cardiovascular Care Innovation through Data-Driven Discoveries in the Electronic Health Record. *Am J Cardiol*. 2023;203:136-148.

3. Palmer LJ. UK Biobank: bank on it. *Lancet*. 2007;369(9578):1980-1982.

4. Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK Biobank: opportunities for cardiovascular research. *Eur Heart J*. 2019;40(14):1158-1166.

5. Raisi-Estabragh Z, Petersen SE. Cardiovascular research highlights from the UK Biobank: opportunities and challenges. *Cardiovasc Res*. 2020;116(1):e12-e15.

6. Aquino EML, Barreto SM, Bensenor IM, et al. Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): objectives and design. *Am J Epidemiol*. 2012;175(4):315-324.

7. Schmidt MI, Duncan BB, Mill JG, et al. Cohort profile: Longitudinal study of adult health (ELSA-brasil). *Int J Epidemiol*. 2015;44(1):68-75.

8. Aquino EML, Araujo MJ, Almeida M da CC, et al. Recrutamento de participantes no Estudo Longitudinal de Saúde do Adulto. *Rev Saúde Pública*. 2013;47:10-18.

9. Khunte A, Sangha V, Oikonomou EK, et al. Detection of left ventricular systolic dysfunction from single-lead electrocardiography adapted for portable and wearable devices. *NPJ Digit Med*. 2023;6(1):124.

10. Barreto SM, Ladeira RM, Bastos M do SCB de O, et al. ELSA-Brasil strategies for outcome identification, investigation and ascertainment. *Rev Saude Publica*. 2013;47 Suppl 2:79-86.

11. Khan SS, Ning H, Shah SJ, et al. 10-year risk equations for incident heart failure in the general population. *J Am Coll Cardiol*. 2019;73(19):2388-2397.

12. Khan SS, Matsushita K, Sang Y, et al. Development and validation of the American heart association's PREVENT equations. *Circulation*. 2024;149(6):430-449.

13. Khan SS, Coresh J, Pencina MJ, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: A scientific statement from the American Heart Association. *Circulation*. 2023;148(24):1982-2004.

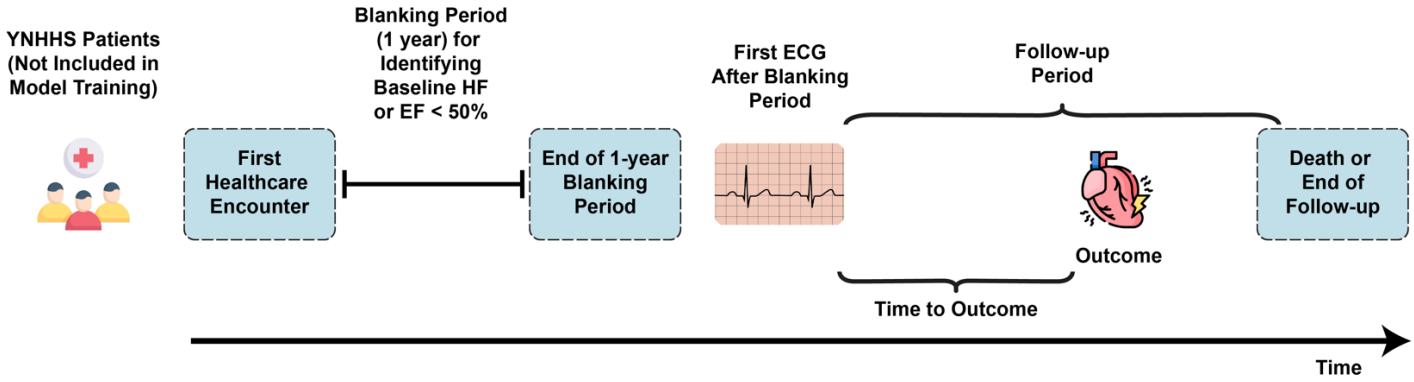
14. Elliott P, Peakman TC, UK Biobank. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol*. 2008;37(2):234-244.

15. Mill JG, Pinto K, Griep RH, et al. Medical assessments and measurements in ELSA-Brasil. *Rev Saude Publica*. 2013;47 Suppl 2:54-62.

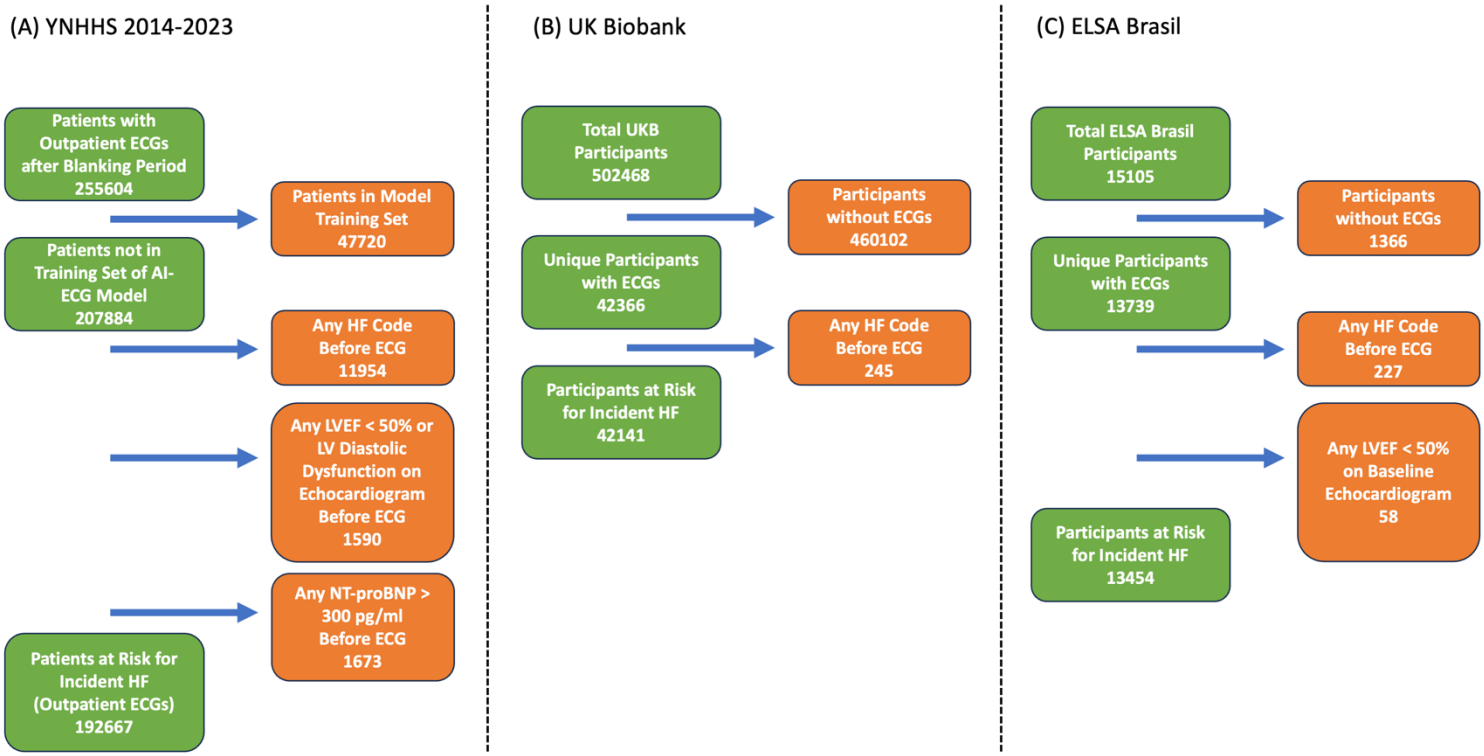
16. Bensenor IM, Griep RH, Pinto KA, et al. Routines of organization of clinical tests and interviews in the ELSA-Brasil investigation center. *Rev Saude Publica*. 2013;47 Suppl 2:37-47.

17. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med*. 2015;34(4):685-703.

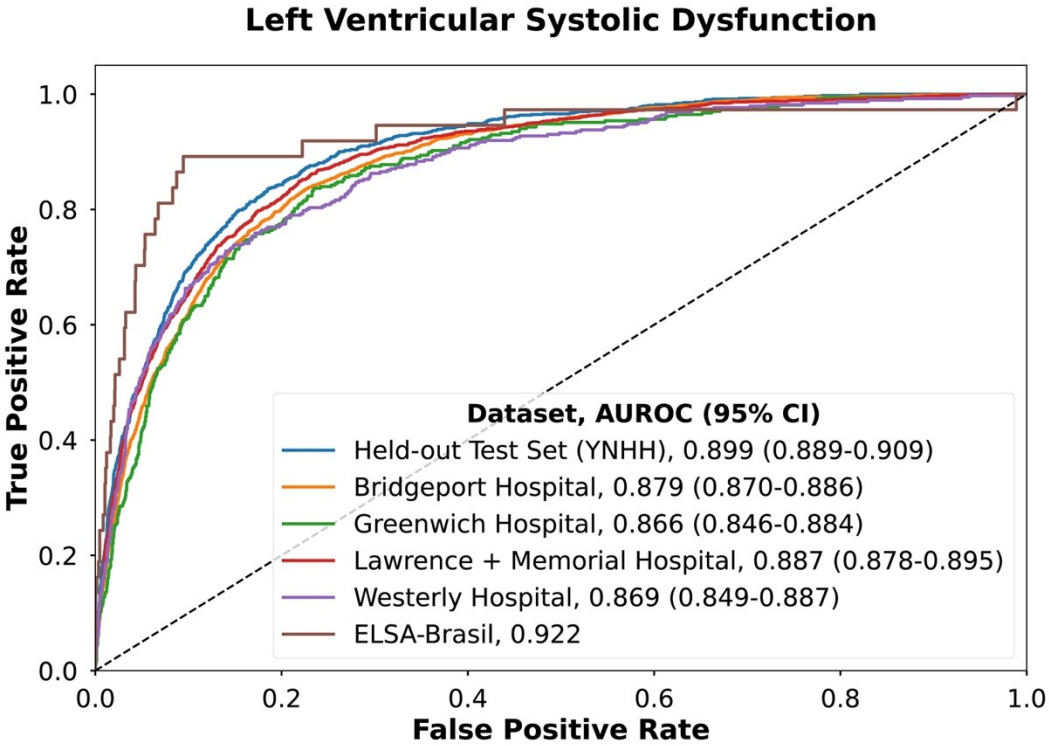
eFigure 1. Overview of Cohort Creation at the Yale New Haven Health System. Abbreviations: ECG, Electrocardiograms; EF, Ejection Fraction; HF, Heart Failure; YNHHS, Yale New Haven Health System.



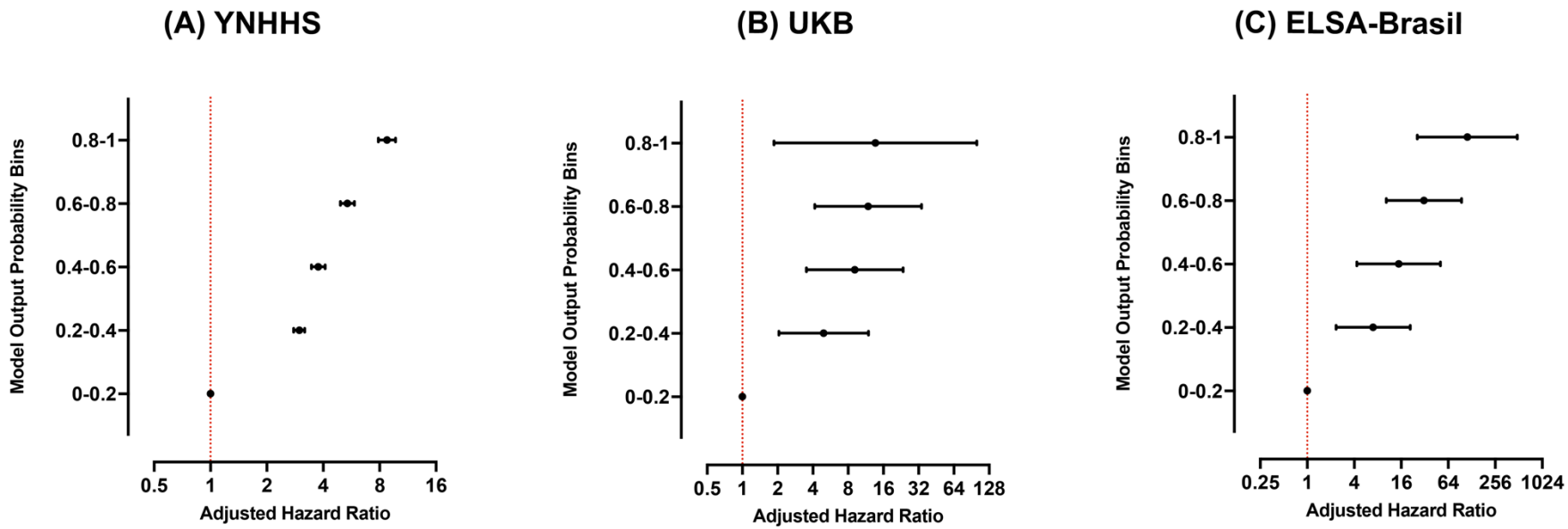
eFigure 2. Consort Diagram for Study Cohorts. Abbreviations: ECG, Electrocardiogram; ELSA Brasil, Brazilian Longitudinal Study of Adult Health; HF, Heart Failure; YNHHS, Yale New Haven Health System



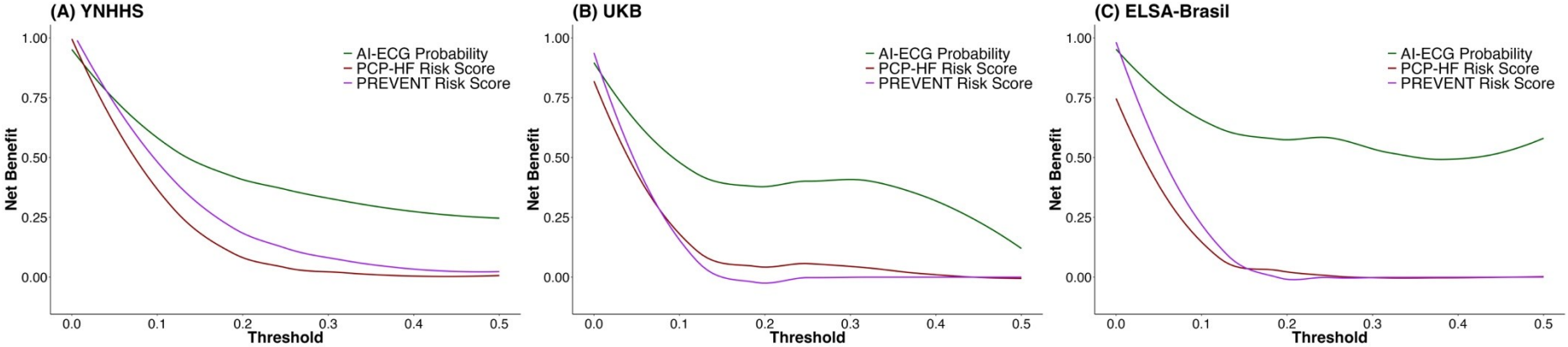
eFigure 3. Model Performance for Cross-sectional Detection of Left Ventricular Systolic Dysfunction in the Yale New Haven Hospital Held-out Test Set and Across External Validation Cohorts. Abbreviations: AUROC, area under the receiver operating characteristic curve; CI, confidence interval; YNHH, Yale New Haven Hospital.



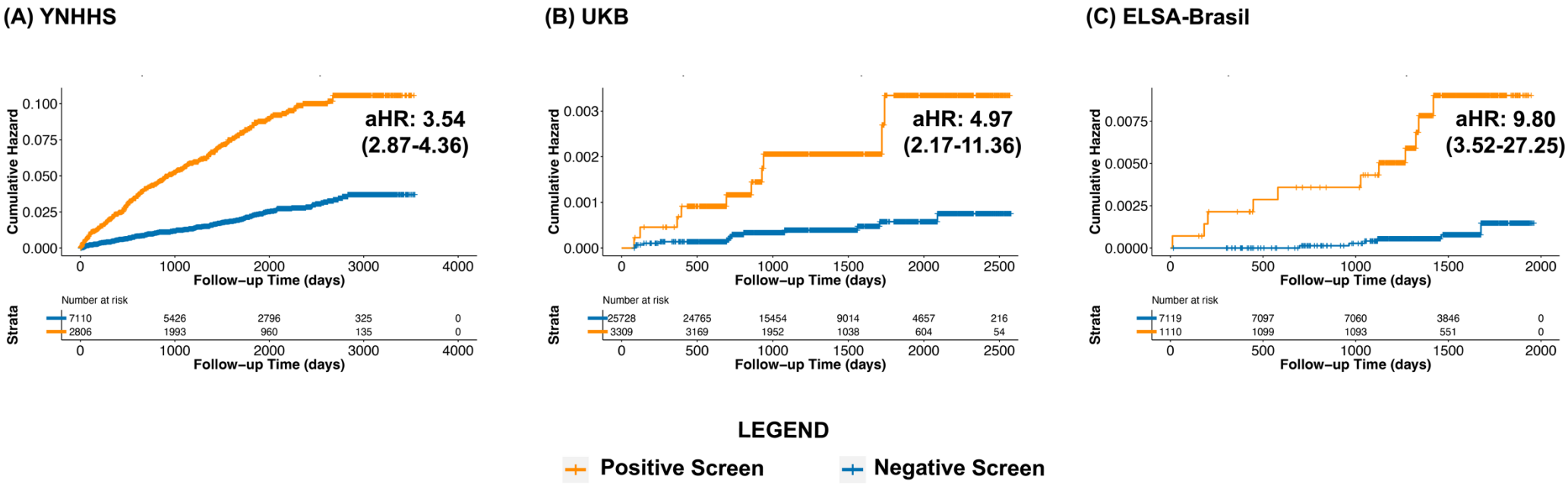
eFigure 4. Age- and Sex-adjusted Hazard for Heart Failure across Model Output Probability Bins. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System.



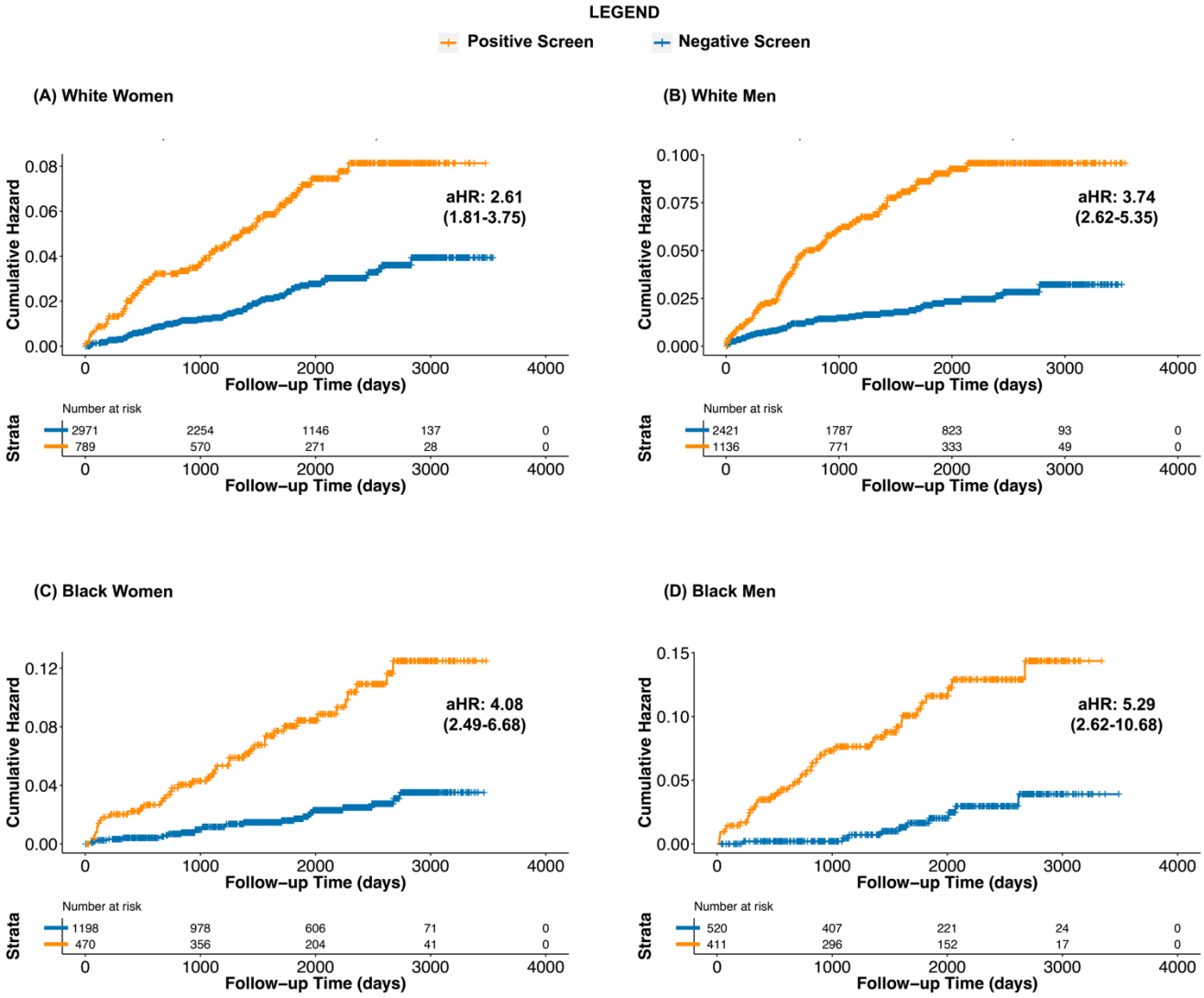
eFigure 5. Net Benefit of AI-ECG Model Output Probability and Pooled Cohort Equations to Prevent Heart Failure) and Predicting Risk of Cardiovascular Disease Events Equations for Predicting Incident Heart Failure Across Thresholds at (A) Yale New Haven Health System (B) UK Biobank (C) Brazilian Longitudinal Study of Adult Health. Abbreviations: AI-ECG, Artificial Intelligence-enhanced Electrocardiography; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; PCP-HF, Pooled Cohort Equations to Prevent Heart Failure; PREVENT, Predicting Risk of Cardiovascular Disease Events; UKB, UK Biobank; YNHHS, Yale New Haven Health System



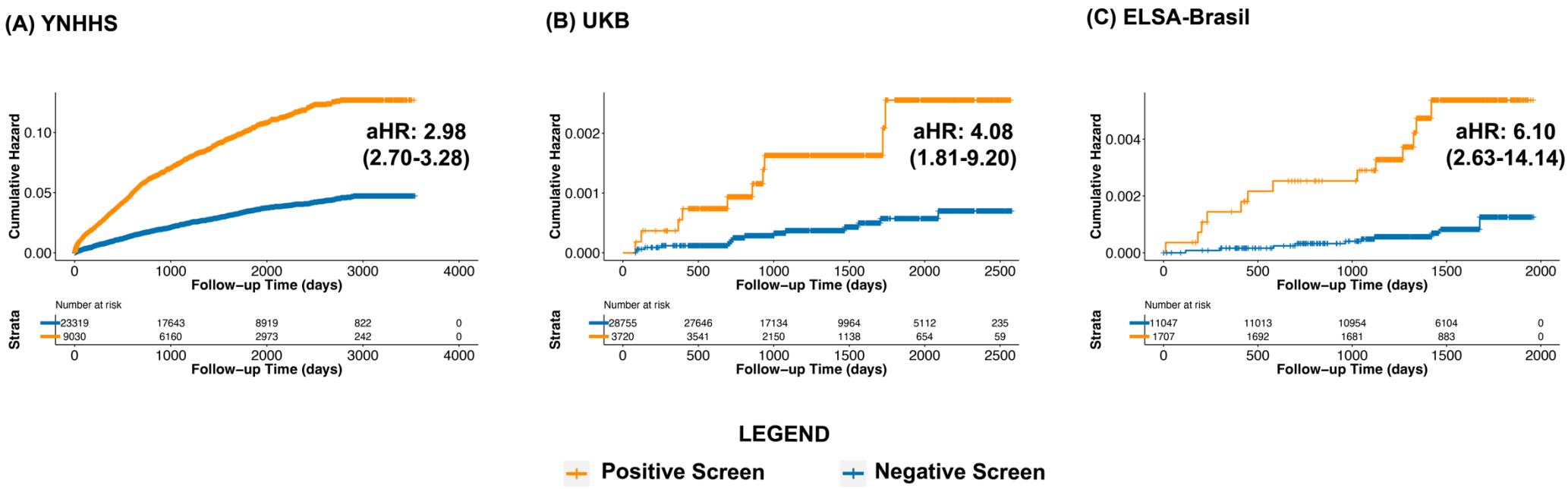
eFigure 6. Cumulative Hazard for Heart Failure Adjusted for Pooled Cohort Equations to Prevent Heart Failure (PCP-HF) Risk Score at (A) Yale New Haven Health System (B) UK Biobank (C) Brazilian Longitudinal Study of Adult Health. Abbreviations: aHR, Adjusted Hazard Ratio; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System.



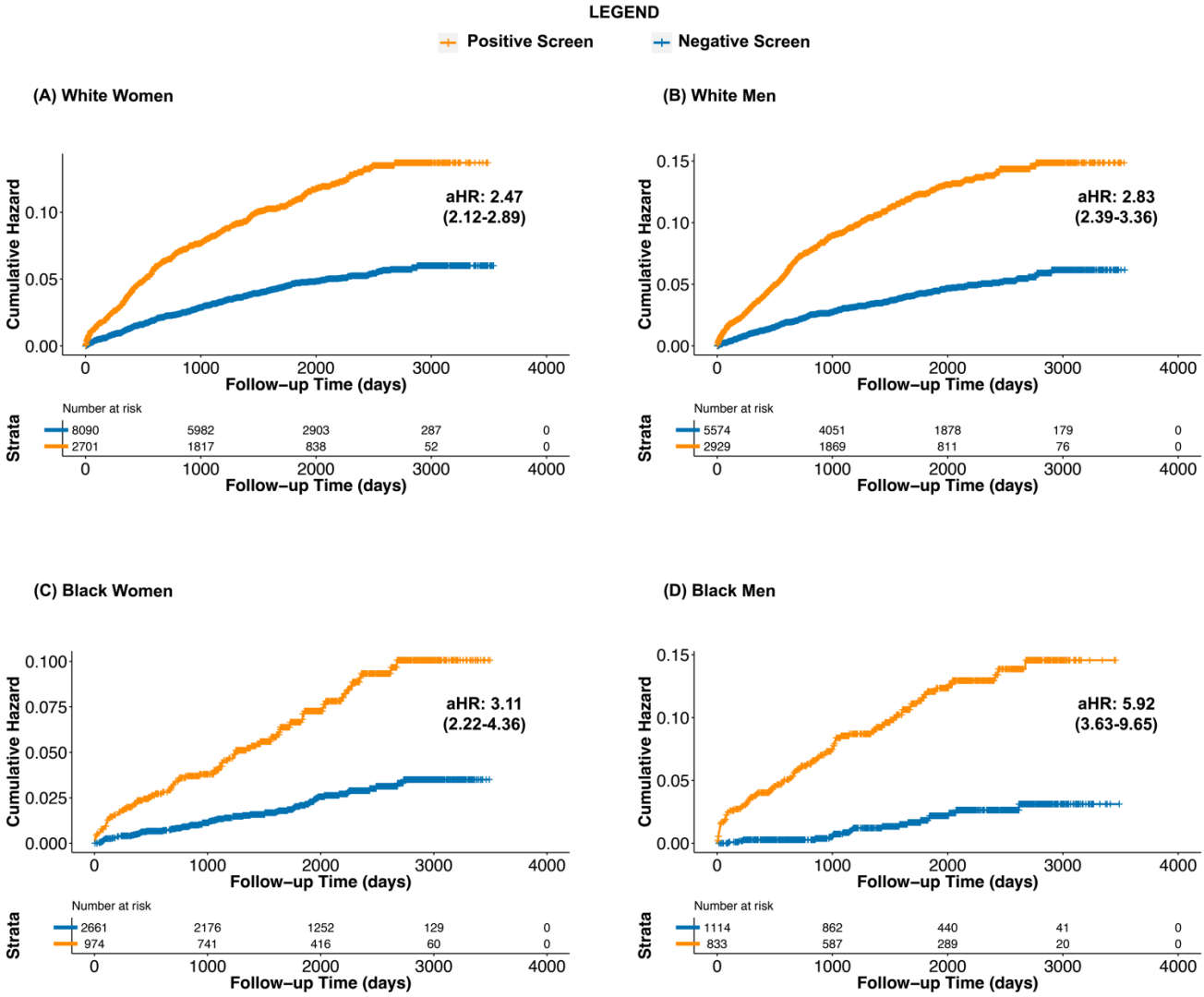
eFigure 7. Cumulative Hazard for Heart Failure Adjusted for Pooled Cohort Equations to Prevent Heart Failure (PCP-HF) Risk Score at Yale New Haven Health System. Abbreviations: aHR, Adjusted Hazard Ratio



eFigure 8. Cumulative Hazard for Heart Failure Adjusted for the Predicting Risk of Cardiovascular Disease Events (PREVENT) Equations at (A) Yale New Haven Health System (B) UK Biobank (C) Brazilian Longitudinal Study of Adult Health. Abbreviations: aHR, Adjusted Hazard Ratio; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System.



eFigure 9. Cumulative Hazard for Heart Failure Adjusted for the Predicting Risk of Cardiovascular Disease Events (PREVENT) Equations at Yale New Haven Health System. Abbreviations: aHR, Adjusted Hazard Ratio



eTable 1. International Classification of Disease Tenth Revision Codes for the Identification of Comorbidities and Outcomes. Abbreviations: ICD-10-CM, International Classification of Disease Tenth Revision Clinical Modification Codes.

Condition	ICD-10-CM codes
Heart Failure	'I11.0','I13.0','I13.2','I50','I50.0','I50.1','I50.9','Z95.81','I09.81'
Acute Myocardial Infarction	'I21', 'I22', 'I23', 'I24.0', 'I24.8', 'I24.9'
Stroke	'G45','G45.0','G45.1','G45.2','G45.3','G45.4','G45.8','G45.9', 'I63','I63.0','I63.1','I63.2','I63.3','I63.4','I63.5','I63.8','I63.9','I64', 'I65','I65.0','I65.1','I65.2','I65.3','I65.8','I65.9','I66','I66.0','I66.1', 'I66.2','I66.3','I66.4','I66.8','I66.9','I67.2','I69.3','I69.4'
Type 2 Diabetes Mellitus	'E11','E11.0','E11.1','E11.2','E11.3','E11.4','E11.5','E11.6', 'E11.7','E11.8','E11.9','O24.1'
Hypertension	'I10','I11','I11.0','I11.9','I12','I12.0','I12.9', 'I13','I13.0','I13.1','I13.2','I13.9','I67.4', 'O10','O10.0','O10.1','O10.2','O10.3','O10.9','O11'
Ischemic Heart Disease	'I20', 'I20.0', 'I20.8', 'I20.9', 'I21', 'I21.0', 'I21.1', 'I21.2', 'I21.3', 'I21.4', 'I21.9', 'I21.X', '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', 'I24.0', 'I24.1', 'I24.8', 'I24.9', 'I25', 'I25.0', 'I25.1', 'I25.2', 'I25.5', 'I25.6', 'I25.8', 'I25.9', 'Z95.1', 'Z95.5'

eTable 2. Model Performance Measures for Cross-sectional Detection of Left Ventricular Systolic Dysfunction in the Yale New Haven Hospital Held-out Test Set and Across External Validation Cohorts. Abbreviations: AUROC, area under the receiver operating characteristic curve; NPV, negative predictive value; OR, odds ratio; PPV, positive predictive value.

Dataset	Total Number	Diagnostic OR	AUROC	F1 Score	Prevalence	Sensitivity	Specificity	PPV	NPV
Held-out Test Set									
Yale New Haven Hospital	10860	24.9 (19.7-31.5)	0.899 (0.889-0.909)	0.346	7.6%	90.3% (89.7-90.8)	72.9% (72.1-73.8)	21.4% (20.6-22.2)	98.9% (98.7-99.1)
External Validation Sites									
Bridgeport Hospital	17915	19.2 (16.2-22.8)	0.879 (0.870-0.886)	0.353	9.9%	91.5% (91.1-91.9)	64.2% (63.4-64.9)	21.9% (21.3-22.5)	98.6% (98.4-98.7)
Greenwich Hospital	4306	15.6 (11.3-21.4)	0.866 (0.846-0.884)	0.334	8.5%	87.8% (86.8-88.8)	68.4% (67.0-69.8)	20.6% (19.4-21.8)	98.4% (98.0-98.7)
Lawrence + Memorial Hospital	17730	21.4 (17.9-25.6)	0.887 (0.878-0.895)	0.338	8.3%	90.7% (90.2-91.1)	68.8% (68.1-69.5)	20.8% (20.2-21.4)	98.8% (98.6-99.0)
Westerly Hospital	3614	13.9 (10.2-19.0)	0.869 (0.849-0.887)	0.371	10.7%	87.8% (86.8-88.9)	65.8% (64.3-67.4)	23.5% (22.1-24.9)	97.8% (97.4-98.3)
ELSA-Brasil	3012	51.1 (22.2-117.7)	0.922	0.201	1.2%	81.1% (79.7-82.5)	92.3% (91.3-93.2)	11.5% (10.4-12.7)	99.7% (99.6-99.9)

eTable 3. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis using Artificial Intelligence (TRIPOD + AI) Checklist.

Section/Topic	Item	Development / evaluation[1]	Checklist item	Reported on page
TITLE				
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Pg 1
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	Pg 3-4
INTRODUCTION				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Pg 5
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	Pg 5-6
	3c	D;E	Describe any known health inequalities between sociodemographic groups	Pg 5-6
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Pg 6
METHODS				
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	Pg 6-7 and suppl.
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Pg 6-7 and suppl.
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	Pg 7-9
	6b	D;E	Describe the eligibility criteria for study participants	Pg 7-9
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	N/A
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Pg 8-10
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	Pg 8
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	N/A
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	N/A
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	Pg 8-10
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	Pg 8-10, and suppl.
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	Pg 9-10, 13-14, and suppl.
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	Pg 7
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	Pg 8-9 and suppl.
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	Pg 10, and suppl.
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	Pg 10, and suppl.
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	Pg 10, and suppl.
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations[3]	N/A
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	Pg 10, and suppl.
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	N/A
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	Pg 10, and suppl.
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	N/A
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	Pg 8-10
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Pg 8-9, and suppl.

[1] D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

[2] Separately for all model building approaches.

[3] TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	N/A
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	Pg 21
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	Pg 21
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	Pg 20-21
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	Not prepared
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	N/A
Data sharing	18e	D;E	Provide details of the availability of the study data	Pg 21
Code sharing	18f	D;E	Provide details of the availability of the analytical code[4]	Pg 21
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	N/A
RESULTS				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Suppl.
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	Suppl.
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	Suppl.
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	Suppl.
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)[5]	Pg 8-9, and suppl.
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	Pg 11-14
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	N/A
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	N/A
DISCUSSION				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Pg 14-15
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	Pg 17-18
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	Pg 17-18
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	N/A
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	Pg 18

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

[4] This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.
[5] This relates to the code to implement the model to get estimates of risk for a new individual.

eTable 4. Population Characteristics of the Yale New Haven Health System Sites. Abbreviations: AMI, acute myocardial infarction; ECG, Electrocardiogram; HF, heart failure; IQR, Interquartile Range; LVEF, Left Ventricular Ejection Fraction; NEMG, Northeast Medical Group; L&M, Lawrence and Memorial Hospital; YNHH, Yale New Haven Hospital

Characteristic		YNHH	Bridgeport	Greenwich	L&M	Westerly	NEMG
Number		96317	32377	17746	19080	4529	22618
Age at ECG, Median [IQR]		54.3 [38.6,67.1]	51.4 [37.3,64.0]	56.0 [42.6,70.2]	58.3 [43.5,71.0]	65.5 [53.8,76.0]	64.1 [54.3,73.7]
Female Sex, N (%)		55580 (57.7)	19669 (60.7)	10406 (58.6)	11235 (58.9)	2464 (54.4)	11827 (52.3)
Race/Ethnicity, N (%)	**White**	57231 (59.4)	12469 (38.5)	12081 (68.1)	13579 (71.2)	4185 (92.4)	18312 (81.0)
	Black	18293 (19.0)	8232 (25.4)	969 (5.5)	1726 (9.0)	81 (1.8)	1322 (5.8)
	Hispanic	14942 (15.5)	10291 (31.8)	3580 (20.2)	2723 (14.3)	121 (2.7)	1599 (7.1)
	Asian	2033 (2.1)	348 (1.1)	509 (2.9)	266 (1.4)	34 (0.8)	363 (1.6)
	Other	1138 (1.2)	309 (1.0)	137 (0.8)	376 (2.0)	57 (1.3)	142 (0.6)
	Missing	2680 (2.8)	728 (2.2)	470 (2.6)	410 (2.1)	51 (1.1)	880 (3.9)
Death, N (%)		5082 (5.3)	1587 (4.9)	811 (4.6)	1161 (6.1)	381 (8.4)	1359 (6.0)
Follow-up Time, Years; Median [IQR]		4.9 [2.7,6.9]	4.8 [3.2,6.6]	5.3 [3.6,6.8]	3.4 [2.0,4.7]	3.3 [1.5,4.6]	4.7 [3.0,6.7]
Positive Screens, N (%)		20670 (21.5)	6930 (21.4)	3227 (18.2)	4597 (24.1)	1354 (29.9)	5997 (26.5)
Hypertension at baseline, N (%)		42576 (44.2)	13411 (41.4)	6235 (35.1)	9390 (49.2)	2722 (60.1)	13881 (61.4)
Type-2 diabetes mellitus at baseline, N (%)		16901 (17.5)	6356 (19.6)	2345 (13.2)	3846 (20.2)	1402 (31.0)	4672 (20.7)
Obesity at baseline, N (%)		16543 (17.2)	4950 (15.3)	1262 (7.1)	3316 (17.4)	1178 (26.0)	3244 (14.3)
Atrial fibrillation at baseline, N (%)		2036 (2.1)	479 (1.5)	407 (2.3)	621 (3.3)	229 (5.1)	974 (4.3)
Left bundle branch block at baseline, N (%)		1212 (1.3)	227 (0.7)	151 (0.9)	326 (1.7)	98 (2.2)	383 (1.7)
Use of antihypertensive drugs at baseline, N (%)		22806 (23.7)	7549 (23.3)	2560 (14.4)	4815 (25.2)	1828 (40.4)	8053 (35.6)
Use of antihyperglycemic drugs at baseline, N (%)		14584 (15.1)	5650 (17.5)	1652 (9.3)	3355 (17.6)	1089 (24.0)	4190 (18.5)
End-stage renal disease, N (%)		329 (0.3)	79 (0.2)	24 (0.1)	60 (0.3)	15 (0.3)	40 (0.2)
Primary HF hospitalization during follow-up, N (%)		1454 (1.5)	594 (1.8)	305 (1.7)	526 (2.8)	231 (5.1)	587 (2.6)
Primary HF hospitalization or an echocardiogram with LVEF < 50% during follow-up, N (%)		2904 (3.0)	1185 (3.7)	460 (2.6)	990 (5.2)	336 (7.4)	1639 (7.2)

Any HF hospitalization during follow-up, N (%)	6313 (6.6)	2079 (6.4)	917 (5.2)	1443 (7.6)	602 (13.3)	2351 (10.4)
Any HF hospitalization or an echocardiogram with LVEF < 50% during follow-up, N (%)	7047 (7.3)	2374 (7.3)	995 (5.6)	1689 (8.9)	656 (14.5)	2944 (13.0)
Primary AMI hospitalization during follow-up, N (%)	50 (0.1)	47 (0.1)	16 (0.1)	37 (0.2)	22 (0.5)	194 (0.9)
Primary Stroke hospitalization during follow-up, N (%)	1210 (1.3)	539 (1.7)	370 (2.1)	529 (2.8)	160 (3.5)	473 (2.1)
Major Adverse Cardiovascular Events during follow-up, N (%)	7162 (7.4)	2513 (7.8)	1367 (7.7)	1999 (10.5)	682 (15.1)	2316 (10.2)

eTable 5. Model Performance for Predicting Heart Failure Risk. Abbreviations: IHD, Ischemic Heart Disease; HTN, hypertension; NEMG, Northeast Medical Group; L&M, Lawrence and Memorial Hospital; T2DM, type-2 diabetes mellitus; YNHH, Yale New Haven Hospital

Model	Covariates	YNHH		Bridgeport		Greenwich		L&M		Westerly		NEMG	
		Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment
Cox Proportional Hazard Model	Model Probability	4.98 (4.49-5.52)	1.47 (1.44-1.50)	5.17 (4.39-6.08)	1.46 (1.42-1.50)	5.52 (4.41-6.92)	1.54 (1.48-1.61)	4.43 (3.72-5.27)	1.43 (1.39-1.48)	4.17 (3.19-5.44)	1.34 (1.28-1.41)	4.45 (3.76-5.25)	1.37 (1.33-1.41)
Cox Proportional Hazard Model	Model Probability + Age + Sex	3.47 (3.12-3.87)	1.34 (1.31-1.37)	3.30 (2.79-3.91)	1.33 (1.28-1.37)	2.82 (2.23-3.57)	1.33 (1.26-1.39)	2.92 (2.44-3.49)	1.31 (1.27-1.36)	3.09 (2.36-4.05)	1.24 (1.18-1.31)	3.08 (2.59-3.65)	1.27 (1.23-1.31)
Cox Proportional Hazard Model	Model Probability + Age + Sex + IHD + HTN + T2DM + Obesity	2.97 (2.67-3.31)	1.30 (1.28-1.33)	2.73 (2.31-3.24)	1.28 (1.24-1.32)	2.46 (1.95-3.10)	1.27 (1.21-1.34)	2.46 (2.05-2.94)	1.28 (1.24-1.33)	2.80 (2.13-3.68)	1.21 (1.15-1.28)	2.69 (2.27-3.20)	1.25 (1.21-1.29)
Fine-Gray Subdistribution Hazard Model	Model Probability + Age + Sex + Competing Risk of Death	3.38 (3.09-3.69)	1.32 (1.30-1.34)	3.30 (2.92-3.73)	1.31 (1.28-1.34)	2.88 (2.45-3.37)	1.28 (1.25-1.32)	3.01 (2.63-3.45)	1.30 (1.27-1.33)	3.04 (2.47-3.74)	1.23 (1.18-1.27)	2.96 (2.49-3.52)	1.26 (1.22-1.30)
Fine-Gray Subdistribution Hazard Model	Model Probability Age + Sex + IHD + HTN + T2DM + Obesity + Competing Risk of Death	2.93 (2.62-3.27)	1.30 (1.27-1.32)	2.71 (2.28-3.21)	1.27 (1.23-1.31)	2.39 (1.90-3.00)	1.24 (1.18-1.30)	2.32 (1.94-2.78)	1.26 (1.22-1.31)	2.86 (2.19-3.72)	1.20 (1.14-1.27)	2.57 (2.16-3.06)	1.23 (1.19-1.27)

eTable 6. Model Performance for Prediction of Heart Failure Across Demographic Subgroups. Abbreviations: CI, Confidence Interval; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; HF, Heart Failure; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Subgroup		YNHHS			UKB			ELSA-Brasil		
		Total Number of Individuals at Risk	Number of HF Events	Age- and Sex-Adjusted Cox Proportional Hazard Ratios (95% CI)	Total Number of Individuals at Risk	Number of HF Events	Age- and Sex-Adjusted Cox Proportional Hazard Ratios (95% CI)	Total Number of Individuals at Risk	Number of HF Events	Age- and Sex-Adjusted Cox Proportional Hazard Ratios (95% CI)
Age < 65		131335 (68.2)	1061	4.84 (4.28-5.47)	20802 (49.4)	9	6.30 (1.68-23.57)	12038 (89.4)	21	11.41 (4.55-28.64)
Age ≥ 65		61332 (31.8)	2636	2.80 (2.59-3.03)	21345 (50.6)	37	5.86 (3.06-11.23)	1416 (10.6)	10	4.69 (1.31-16.71)
Female		111181 (57.7)	1885	2.95 (2.69-3.24)	21795 (51.7)	11	3.76 (1.10-12.88)	7348 (54.6)	11	10.21 (3.88-26.91)
Male		81486 (42.3)	1812	3.75 (3.40-4.15)	20346 (48.3)	35	6.92 (3.52-13.60)	6106 (45.4)	20	6.76 (2.03-22.47)
Race/ Ethnicity	**White**	117857 (61.2)	2769	3.02 (2.80-3.27)	40691 (96.6)	46	5.96 (3.32-10.68)	6920 (51.4)	15	10.17 (3.40-30.46)
	Black	30623 (15.9)	470	4.36 (3.48-5.44)	304 (0.7)	0	-	2130 (15.8)	9	11.80 (2.38-58.47)
	Hispanic	33256 (17.3)	338	4.40 (3.54-5.47)	0	-	-	-	-	-
	Asian	3553 (1.8)	32	5.21 (2.50-10.87)	600 (1.4)	0	-	-	-	-
	Other	2159 (1.1)	28	2.98 (1.39-6.39)	546 (1.3)	0	-	637 (4.7)	0	-
	Brazilian "Pardo"	-	-	-	-	-	-	3767 (28.0)	7	3.69 (0.81-16.8)
	Missing	5219 (2.7)	60	2.34 (1.38-3.97)	0	-	-	-	-	-

eTable 7. Age- and Sex- Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure-related Outcomes. Abbreviations: AMI, acute myocardial infarction; ELSA Brasil, Brazilian Longitudinal Study of Adult Health; HF, heart failure; LVEF, Left Ventricular Ejection Fraction; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Outcome	YNHHS	UKB	ELSA-Brasil
Primary HF Hospitalization	3.31 (3.10-3.54)	5.96 (3.32-10.68)	8.74 (4.13-18.48)
Primary HF Hospitalization or an Echocardiogram with LVEF < 50%	3.87 (3.69-4.06)	-	-
Primary HF Hospitalization or Death	1.97 (1.90-2.04)	1.91 (1.51-2.41)	2.59 (1.99-3.35)
Any Hospitalization with HF	2.43 (2.34-2.51)	3.55 (2.72-4.64)	-
Any Hospitalization with HF or an Echocardiogram with LVEF < 50%	2.62 (2.54-2.70)	-	-
Any HF Hospitalization or Death	2.06 (2.01-2.12)	2.18 (1.81-2.63)	-

eTable 8. Age- and Sex- Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure-related Outcomes. Abbreviations: AMI, acute myocardial infarction; HF, heart failure; NEMG, Northeast Medical Group; LVEF, Left Ventricular Ejection Fraction; L&M, Lawrence and Memorial Hospital; YNHH, Yale New Haven Hospital

Outcome	YNHH	Bridgeport	Greenwich	L&M	Westerly	NEMG
Primary HF Hospitalization	4.98 (4.49-5.52)	3.30 (2.79-3.91)	2.82 (2.23-3.57)	2.92 (2.44-3.49)	3.09 (2.36-4.05)	3.08 (2.59-3.65)
Primary HF Hospitalization or an Echocardiogram with LVEF < 50%	3.76 (3.49-4.06)	3.61 (3.20-4.07)	2.76 (2.28-3.34)	3.46 (3.04-3.95)	2.71 (2.17-3.37)	4.86 (4.37-5.41)
Primary HF Hospitalization or Death	1.89 (1.79-1.99)	1.81 (1.65-1.98)	1.97 (1.74-2.23)	2.20 (1.99-2.44)	1.75 (1.48-2.08)	2.09 (1.9-2.30)
Any Hospitalization with HF	2.39 (2.28-2.52)	2.47 (2.26-2.70)	2.17 (1.89-2.48)	2.69 (2.42-2.99)	2.00 (1.70-2.35)	2.25 (2.07-2.45)
Any Hospitalization with HF or an Echocardiogram with LVEF < 50%	2.51 (2.39-2.63)	2.56 (2.36-2.78)	2.19 (1.93-2.50)	2.83 (2.57-3.13)	1.97 (1.68-2.30)	2.87 (2.67-3.09)
Any HF Hospitalization or Death	2.01 (1.93-2.10)	1.99 (1.85-2.14)	1.96 (1.76-2.18)	2.31 (2.12-2.51)	1.74 (1.52-2.00)	2.05 (1.91-2.21)

eTable 9. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure across Model Output Probabilities. Abbreviations: ELSA Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Model output probability bins	YNHHS	UKB	ELSA-Brasil
0-0.2	Reference	Reference	Reference
0.2-0.4	2.91 (2.65-3.19)	4.94 (2.05-11.93)	6.98 (2.34-20.82)
0.4-0.6	3.55 (3.15-3.99)	9.11 (3.52-23.57)	14.83 (4.32-50.95)
0.6-0.8	4.88 (4.29-5.56)	11.87 (4.15-33.9)	31.17 (10.27-94.66)
0.8-1.0	7.49 (6.33-8.87)	13.68 (1.86-100.45)	112.33 (25.75-489.96)

eTable 10. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Heart Failure across Model Output Probabilities at the Yale New Haven Health System Sites. Abbreviations: NEMG, Northeast Medical Group; L&M, Lawrence and Memorial Hospital; YNHH, Yale New Haven Hospital

Model output probability bins	YNHH	Bridgeport	Greenwich	L&M	Westerly	NEMG
0-0.2	Reference	Reference	Reference	Reference	Reference	Reference
0.2-0.4	3.58 (3.10-4.14)	2.20 (1.72-2.83)	2.63 (1.90-3.64)	2.34 (1.81-3.02)	3.12 (2.25-4.33)	2.34 (1.86-2.94)
0.4-0.6	3.67 (3.03-4.45)	3.97 (2.95-5.35)	3.85 (2.55-5.82)	3.55 (2.65-4.76)	2.12 (1.22-3.69)	3.08 (2.30-4.13)
0.6-0.8	5.65 (4.63-6.91)	5.08 (3.60-7.15)	4.56 (2.80-7.43)	4.34 (3.10-6.07)	3.80 (2.30-6.26)	3.79 (2.73-5.24)
0.8-1.0	9.14 (7.05-11.85)	6.44 (4.17-9.94)	8.35 (4.10-17.0)	9.59 (6.40-14.36)	3.12 (1.27-7.63)	5.15 (3.36-7.90)

eTable **11**. Comparison of Harrel's C-statistic for the **I**nput **Features** of the **C**linical **R**isk **S**cores with the **S**ingle-lead **A**rtificial **I**ntelligence-**e**nhanced **E**lectrocardiogram **M**odel. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; PCP-HF, Pooled Cohort Equations to Prevent Heart Failure; PREVENT, Predicting Risk of Cardiovascular Disease Events; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Clinical Risk Score	Covariates	YNHHS			UKB			ELSA-Brasil		
		Harrel's C-statistic	Marginal difference over Harrel's C-statistic for clinical risk score input variables	P-value	Harrel's C-statistic	Marginal difference over Harrel's C-statistic for clinical risk score input variables	P-value	Harrel's C-statistic	Marginal difference over Harrel's C-statistic for clinical risk score input variables	P-value
PCP-HF	**PCP-HF Input Variables**	0.696 (0.669 - 0.723)	-	-	0.794 (0.699 - 0.889)	-	-	0.923 (0.894 - 0.952)	-	-
	AI-ECG Model Output Probability	0.723 (0.694 - 0.752)	0.026 (-0.011 - 0.063)	0.16	0.736 (0.606 - 0.867)	-0.057 (-0.191 - 0.076)	0.39	0.828 (0.692 - 0.964)	-0.095 (-0.233 - 0.043)	0.18
	AI-ECG Model Output Probability + Age + Sex	0.720 (0.692 - 0.748)	0.023 (-0.001 - 0.049)	0.06	0.800 (0.707 - 0.894)	0.005 (-0.035 - 0.046)	0.77	0.897 (0.820 - 0.975)	-0.026 (-0.112 - 0.061)	0.56
	AI-ECG Model Output Probability + PCP-HF Input Variables	0.760 (0.734 - 0.785)	0.063 (0.044 - 0.082)	< 0.001	0.829 (0.743 - 0.915)	0.035 (-0.009 - 0.079)	0.12	0.945 (0.912 - 0.979)	0.022 (-0.053 - 0.009)	0.16
PREVENT	**PREVENT Input Variables**	0.701 (0.673 - 0.723)	-	-	0.778 (0.667 - 0.890)	-	-	0.894 (0.861 - 0.926)	-	-
	AI-ECG Model Output Probability	0.723 (0.694 - 0.752)	0.022 (-0.017 - 0.061)	0.27	0.736 (0.606 - 0.867)	-0.041 (-0.200 - 0.117)	0.60	0.828 (0.692 - 0.964)	-0.066 (-0.207 - 0.076)	0.36
	AI-ECG Model Output Probability + Age + Sex	0.720 (0.692 - 0.748)	0.194 (-0.008 - 0.046)	0.16	0.800 (0.707 - 0.894)	0.022 (-0.048 - 0.092)	0.54	0.897 (0.820 - 0.975)	0.003 (-0.081 - 0.089)	0.93
	AI-ECG Model Output Probability + PREVENT Input Variables	0.765 (0.739 - 0.790)	0.064 (0.046 - 0.081)	< 0.001	0.828 (0.743 - 0.912)	0.049 (-0.010 - 0.108)	0.10	0.932 (0.896 - 0.968)	0.038 (0.003 - 0.073)	0.03

eTable 12. Categorical and Continuous Time-to-Event, Event and Non-event Net Reclassification Index of AI-ECG Model Output Probability over Pooled Cohort Equations to Prevent Heart Failure, and Predicting Risk of Cardiovascular Disease Events Equations. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; NRI, Net Reclassification Index; PCP-HF, Pooled Cohort equations to Prevent Heart Failure; PREVENT, Predicting Risk of Cardiovascular Disease Events; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Metric	**YNHHS**		**UKB**		**ELSA-Brasil**	
Comparator	**PCP-HF**	**PREVENT**	**PCP-HF**	**PREVENT**	**PCP-HF**	**PREVENT**
Categorical NRI						
NRI	0.182 (0.100 to 0.263)	0.118 (0.034 to 0.199)	0.198 (-0.076 to 0.465)	0.289 (0.017 to 0.537)	0.472 (0.131 to 0.749)	0.475 (0.173 to 0.809)
NRI+	0.261 (0.180 to 0.338)	0.163 (0.081 to 0.242)	0.301 (0.029 to 0.568)	0.395 (0.124 to 0.643)	0.585 (0.251 to 0.864)	0.583 (0.282 to 0.917)
NRI-	-0.078 (-0.090 to -0.066)	-0.045 (-0.057 to -0.032)	-0.102 (-0.106 to -0.098)	-0.106 (-0.110 to -0.102)	-0.113 (-0.120 to -0.104)	-0.107 (-0.115 to -0.099)
Continuous NRI						
NRI	0.210 (0.094 to 0.325)	0.207 (0.094 to 0.323)	0.096 (-0.347 to 0.506)	0.309 (-0.140 to 0.724)	0.095 (-0.242 to 0.324)	0.188 (-0.268 to 0.531)
NRI+	0.360 (0.249 to 0.471)	0.226 (0.114 to 0.337)	0.405 (-0.037 to 0.815)	0.418 (-0.035 to 0.832)	0.775 (0.432 to 0.999)	0.663 (0.209 to 0.999)
NRI-	-0.150 (-0.171 to -0.130)	-0.019 (-0.038 to 0.002)	-0.309 (-0.320 to -0.298)	-0.109 (-0.120 to -0.098)	-0.680 (-0.698 to -0.664)	-0.475 (-0.490 to -0.455)

eTable 13. Performance Measures of the AI-ECG Model Output Probability, Pooled Cohort Equations to Prevent Heart Failure, and Predicting Risk of Cardiovascular Disease Events Equations for Predicting Incident Heart Failure with Censoring the Observations at the Median Duration of Follow-up.
Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; NPV, negative predictive value; PCP-HF, Pooled Cohort Equations to Prevent Heart Failure; PPV, positive predictive value; PREVENT, Predicting Risk of Cardiovascular Disease Events; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Metric	YNHHS			UKB			ELSA-Brasil		
	PCP-HF	PREVENT	AI-ECG	PCP-HF	PREVENT	AI-ECG	PCP-HF	PREVENT	AI-ECG
Sensitivity	41.9% (36.3-47.5)	53.6% (47.9-59.2)	62.1% (56.6-67.6)	14.7% (0-30.3)	5.4% (0-15.6)	45.0% (21.6-68.4)	11.1% (0-25.8)	11.1% (0-25.8)	68.3% (44.9-91.7)
Specificity	72.3% (71-73.5)	65.0% (63.7-66.4)	73.9% (72.7-75.1)	1.9% (1.7-2.2)	1.5% (1.3-1.7)	10.0% (9.5-10.5)	98.3% (97.9-98.7)	97.8% (97.4-98.3)	87.6% (86.6-88.6)
PPV	5.7% (4.7-6.7)	5.8% (4.9-6.6)	8.7% (7.5-9.9)	0.6% (0-1.2)	0.3% (0-0.8)	0.3% (0.1-0.6)	1.4% (0-3.3)	1.1% (0-2.7)	1.2% (0.5-1.9)
NPV	96.9% (96.4-97.3)	97.2% (96.8-97.7)	98% (97.6-98.4)	99.9% (99.9-100)	99.9% (99.9-100)	100% (99.9-100)	99.8% (99.7-99.9)	99.8% (99.7-99.9)	99.9% (99.8-100)

eTable 14. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Non-Heart Failure Clinical Outcomes. Abbreviations: AMI, acute myocardial infarction; ELSA Brasil, Brazilian Longitudinal Study of Adult Health; HF, heart failure; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Outcome	YNHHS	UKB	ELSA-Brasil
Primary AMI Hospitalization	1.12 (0.89-1.40)	1.23 (0.85-1.76)	3.00 (1.78-5.08)
Primary Stroke Hospitalization	1.18 (1.09-1.27)	1.54 (1.10-2.15)	3.86 (2.28-6.51)
All-cause Death	0.99 (0.95-1.03)	1.58 (1.22-2.05)	2.18 (1.65-2.88)
Major Adverse Cardiovascular Events	1.76 (1.70-1.82)	1.62 (1.37-1.93)	2.68 (2.14-3.36)

eTable 15. Age- and Sex-Adjusted Cox Proportional Hazard Models for the Prediction of Non-Heart Failure Clinical Outcomes Across Yale New Haven Health System Sites. Abbreviations: AMI, acute myocardial infarction; HF, heart failure; NEMG, Northeast Medical Group; L&M, Lawrence and Memorial Hospital; YNHH, Yale New Haven Hospital

Outcome	YNHH	Bridgeport	Greenwich	L&M	Westerly	NEMG
Primary AMI Hospitalization	1.33 (0.73-2.42)	1.31 (0.70-2.44)	2.05 (0.71-5.88)	1.68 (0.85-3.29)	1.93 (0.82-4.55)	0.76 (0.55-1.05)
Primary Stroke Hospitalization	1.11 (0.98-1.26)	1.20 (0.99-1.45)	1.34 (1.07-1.67)	1.24 (1.04-1.49)	1.12 (0.81-1.55)	1.05 (0.86-1.28)
All-cause Death	0.95 (0.90-1.01)	1.01 (0.91-1.12)	1.24 (1.08-1.43)	0.99 (0.88-1.12)	0.95 (0.77-1.17)	1.00 (0.90-1.11)
Major Adverse Cardiovascular Events	1.73 (1.65-1.81)	1.66 (1.53-1.80)	1.77 (1.58-1.98)	1.93 (1.76-2.12)	1.57 (1.35-1.83)	1.70 (1.56-1.85)