SUPPLEMENTAL MATERIALS

Figure S1: Primary model architecture
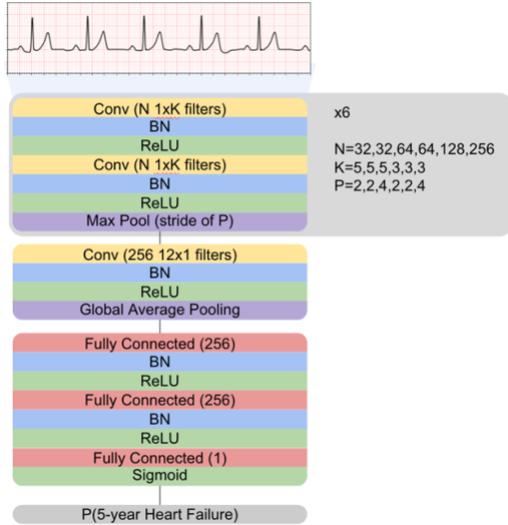
Figure S2: Breakdown of electrocardiogram (ECG) data: exclusions from prior heart failure, insufficient follow-up or lack of demographic record; training, testing, and validation splits
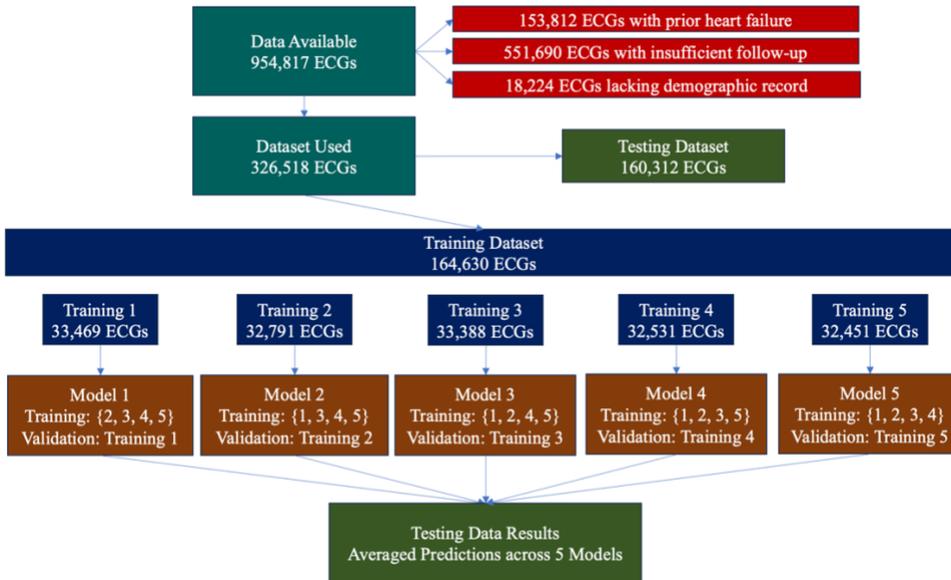
Figure S3: Approaches to reducing disparities in model performance: pre-processing, training modifications, and post-processing
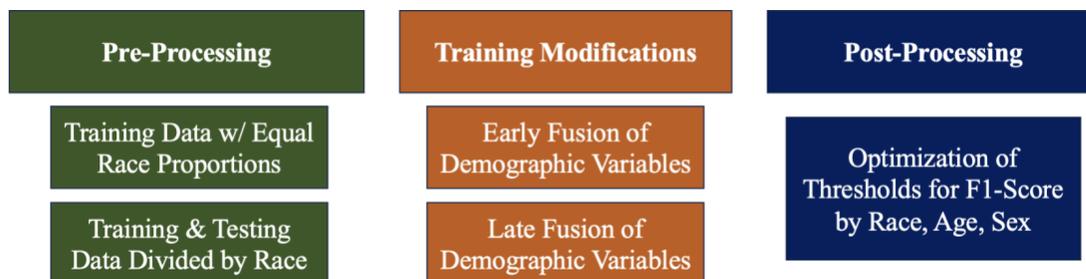

Figure S4: (a) Modifications to model architecture for early fusion of demographic variables (b) Modifications to model architecture for late fusion of demographic variables
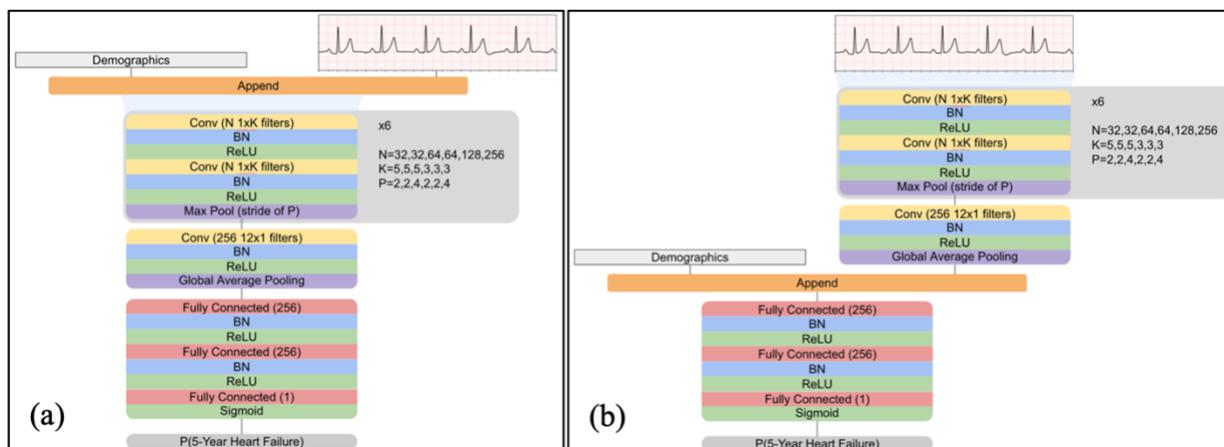
Figure S5. Age-related AUC (area under the curve) trends stratified by race and gender
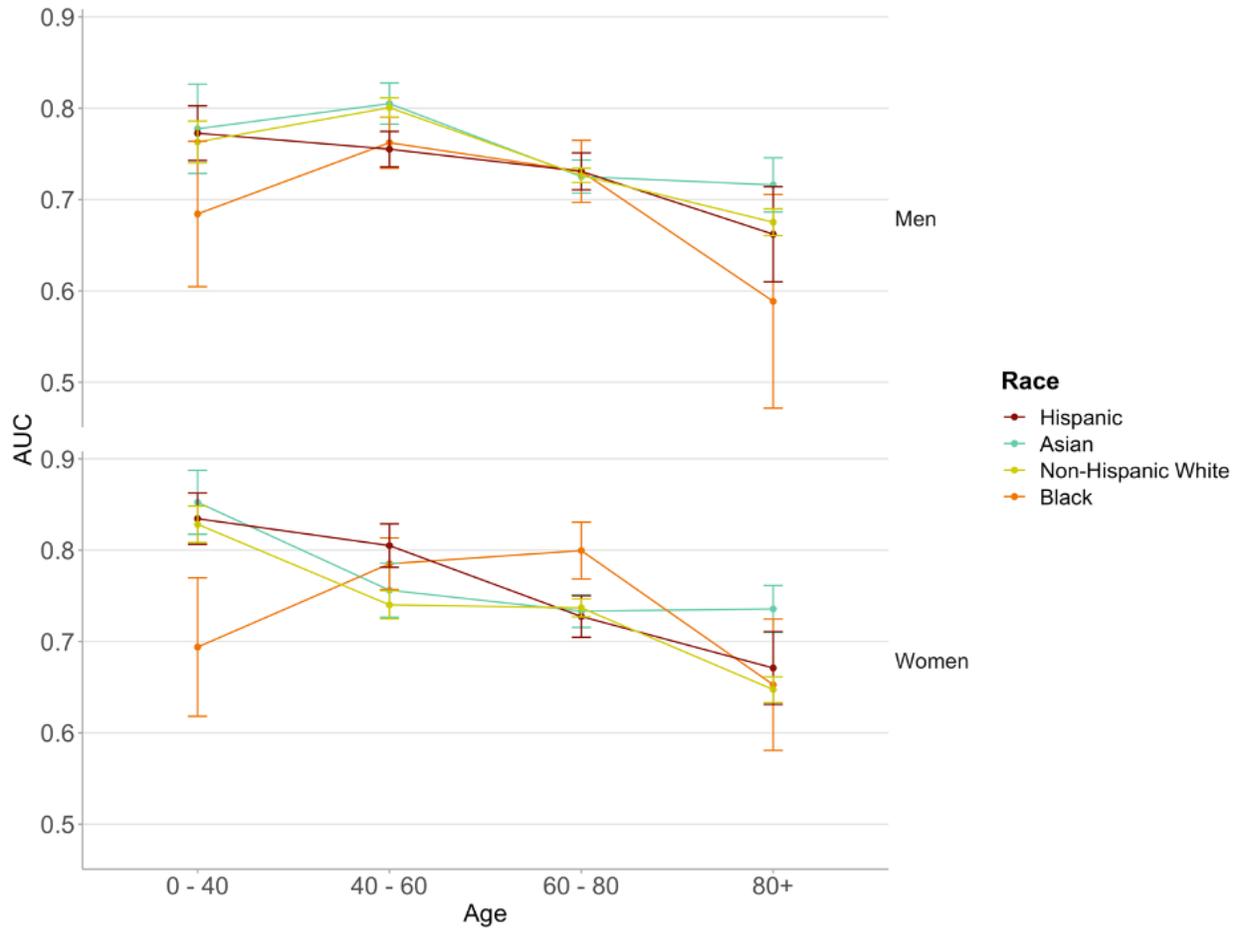
Figure S6. Receiver operating characteristic curves stratified by gender across age groups
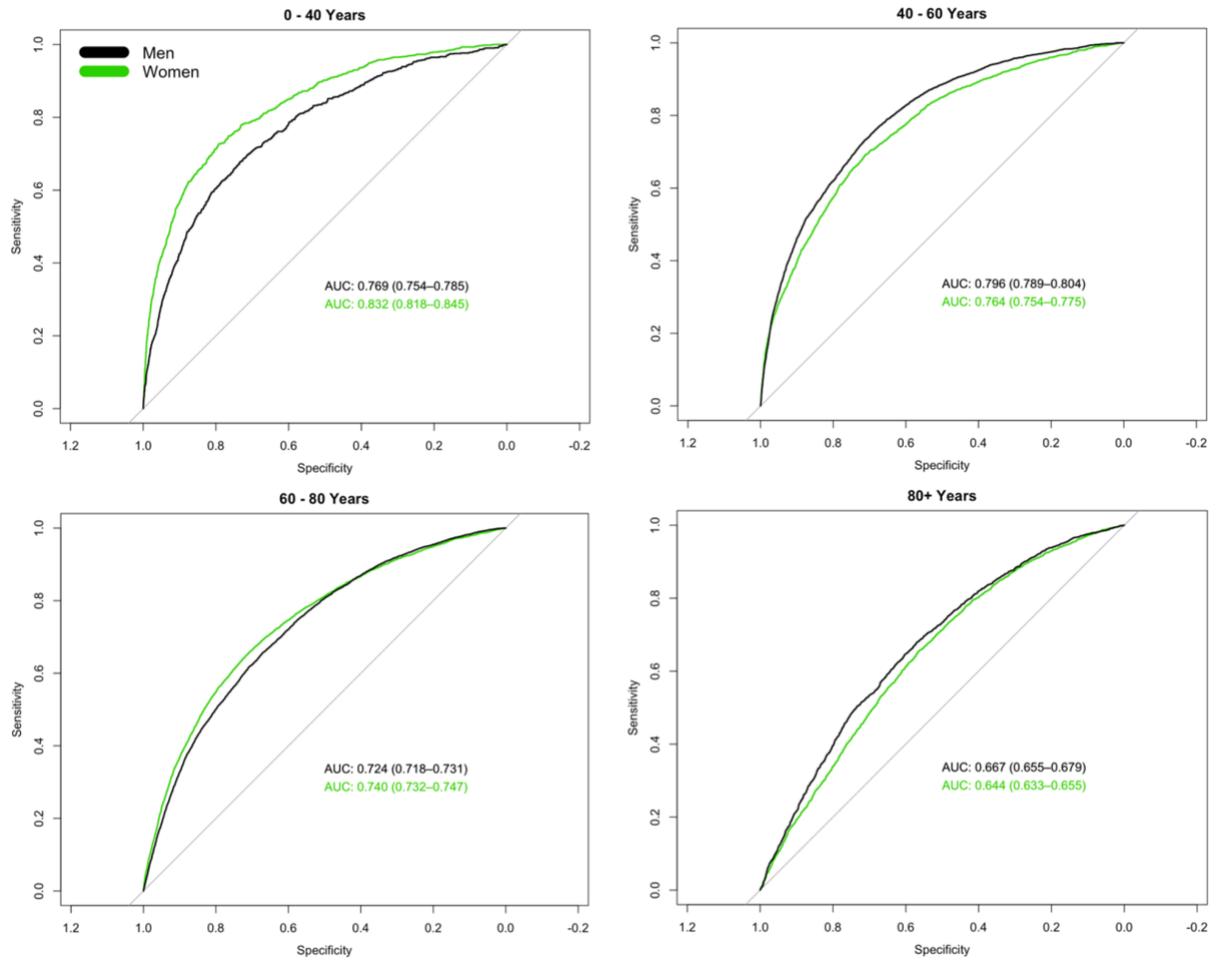
Figure S7. Probability distributions of cases with and without incident heart failure (HF) by race and gender
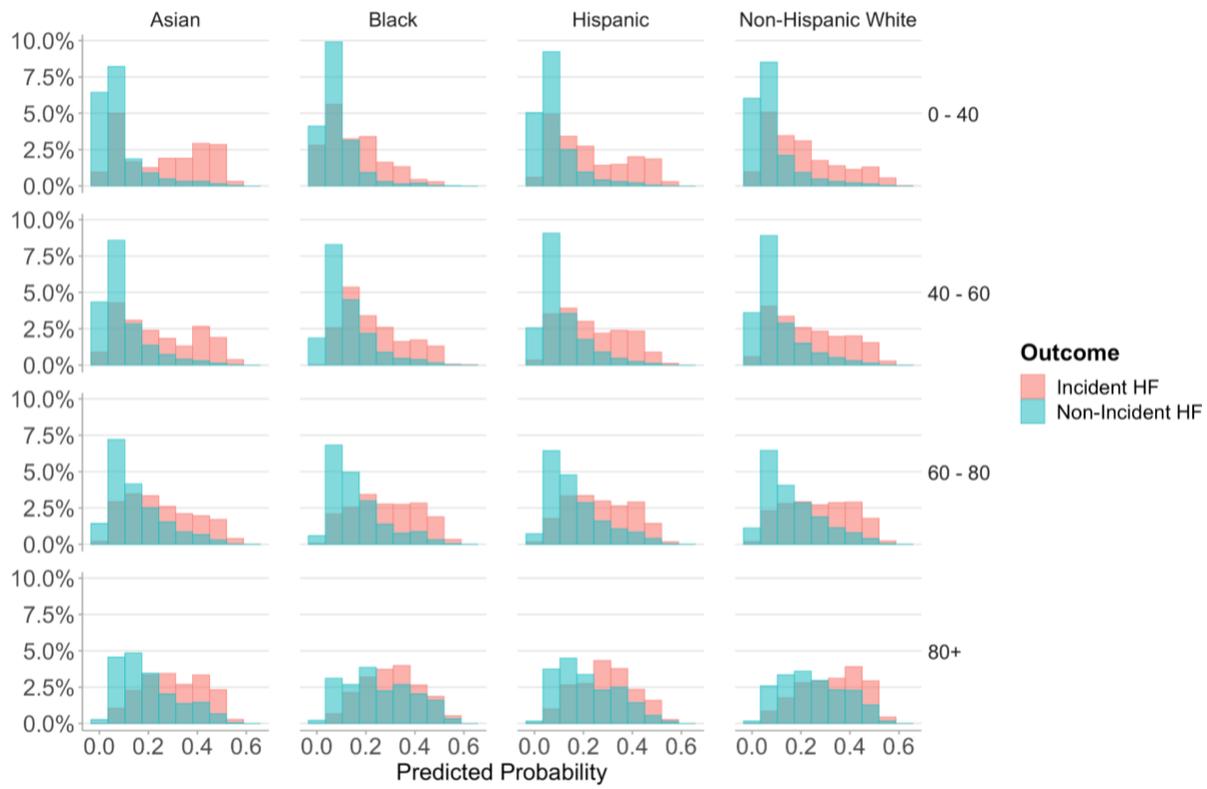
Figure S8. Receiver operating characteristic curves stratified by race and age from a model

trained on a dataset with equal racial representation.
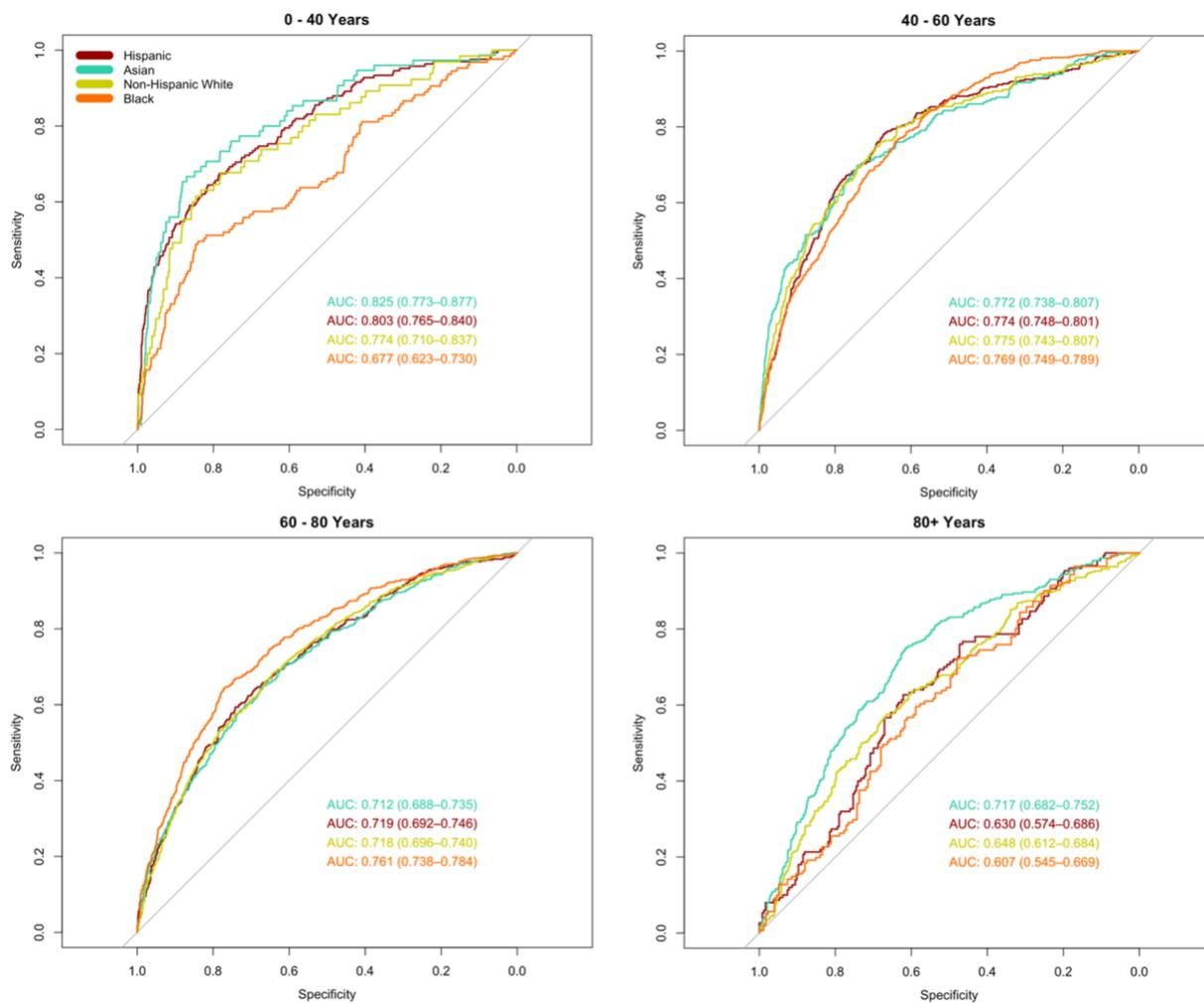
Figure S9. Receiver operating characteristic curves stratified by race and age from models trained separately for each racial group.
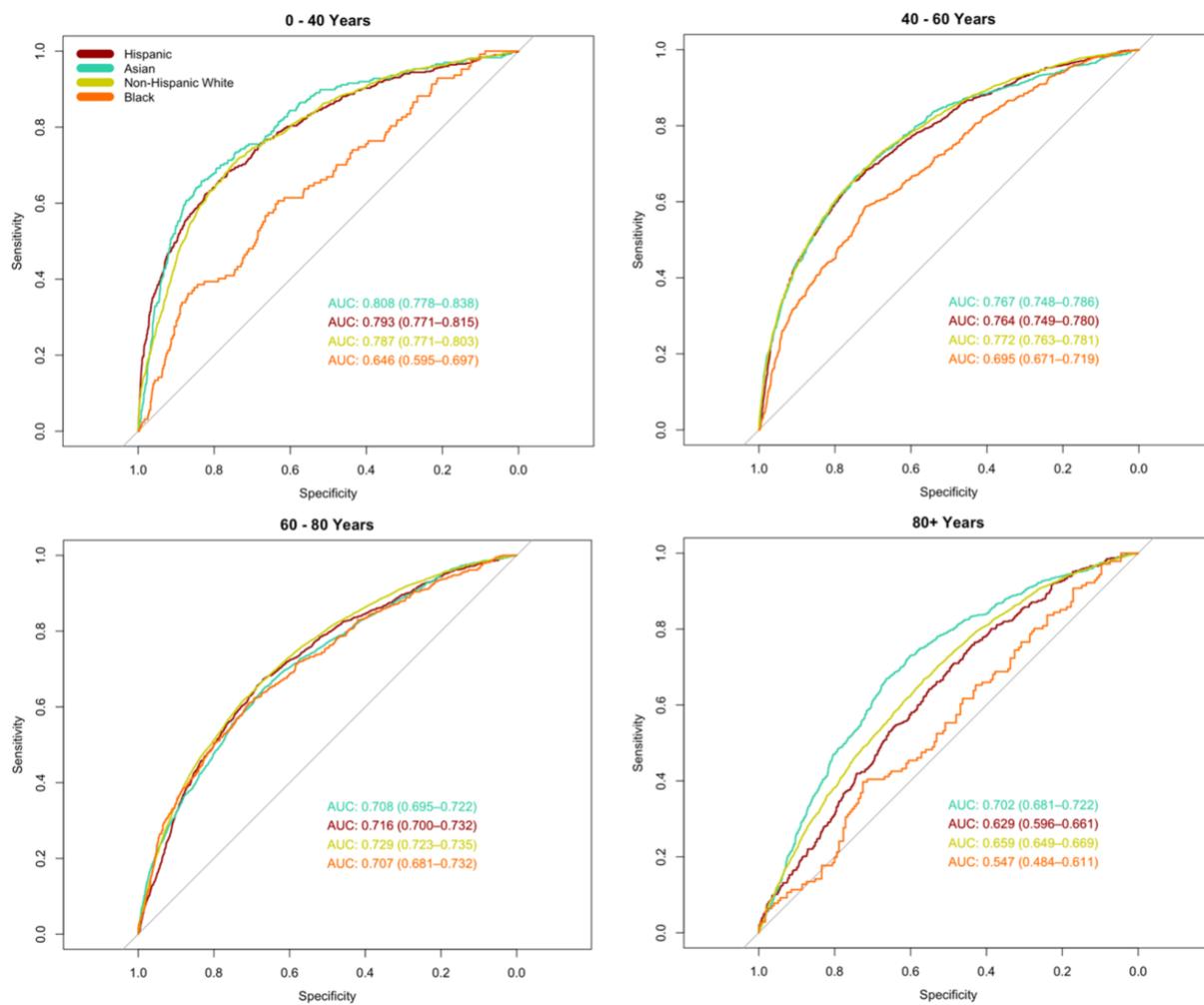
Figure S10. Receiver operating characteristic curves from models trained and tested separately by age group
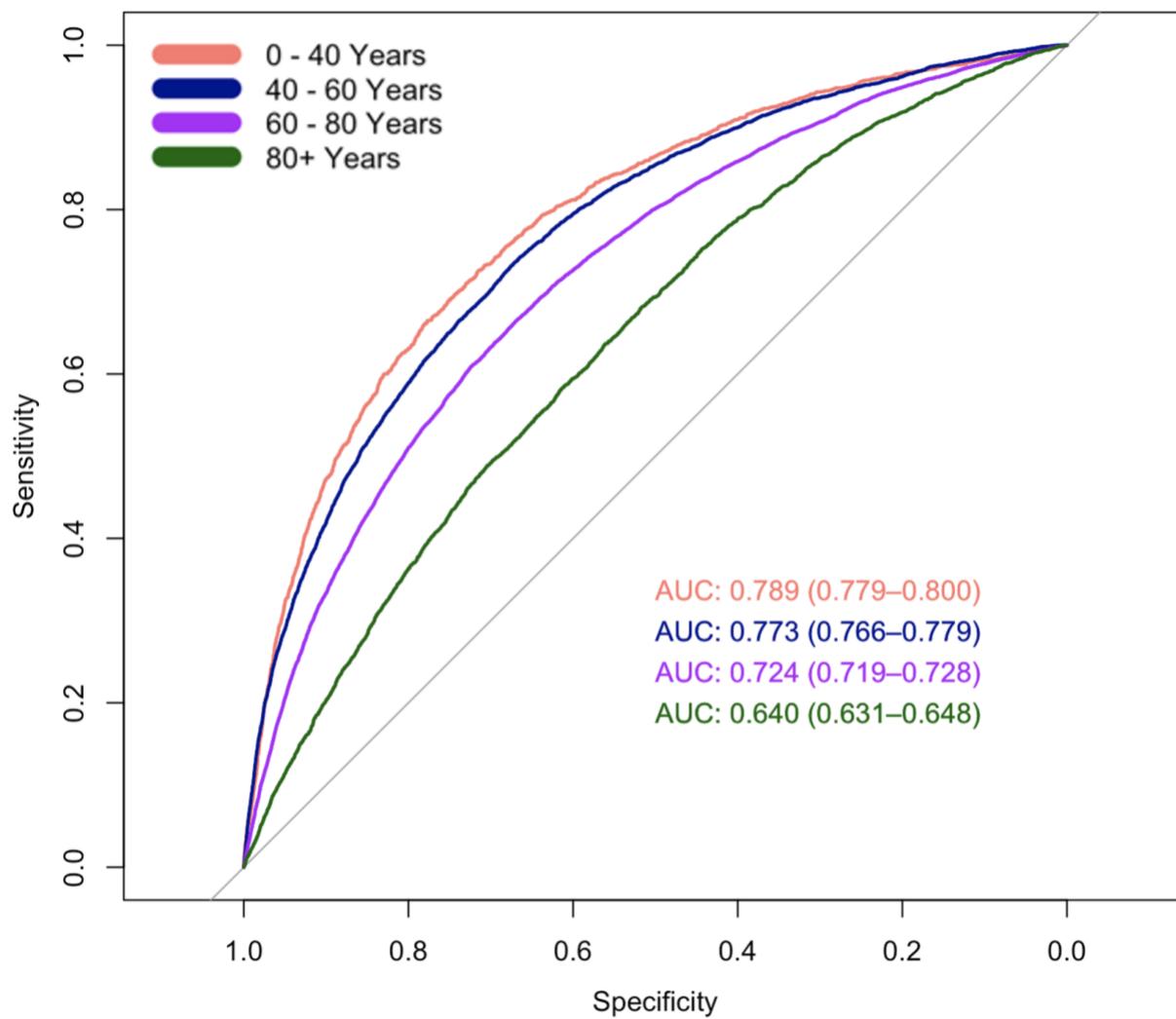
Figure S11. Optimizing threshold for F1-score by subgroups of race, age, and gender
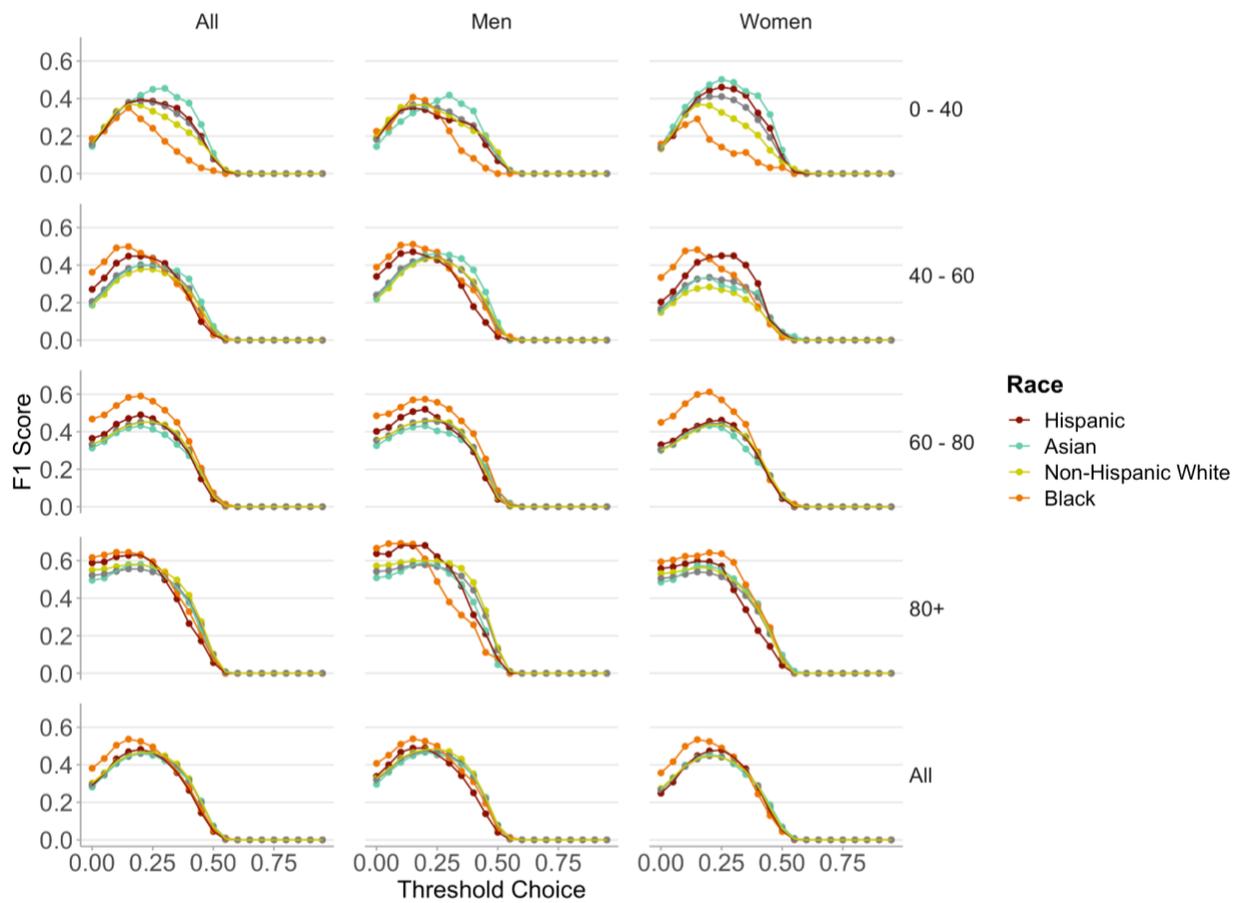
Figure S12. Sample electrocardiograms showing a) Non-Hispanic White man with high predicted probability who developed heart failure (HF), b) Non-Hispanic White woman with low predicted probability who did not develop heart failure, c) Asian man with high predicted probability who did not develop heart failure, d) Non-Hispanic White woman aged >80 years with high predicted probability who did not develop HF, e) Black woman aged 0-40 with low predicted probability who developed heart failure, and f) Black woman aged 0-40 with high predicted probability who did not develop heart failure
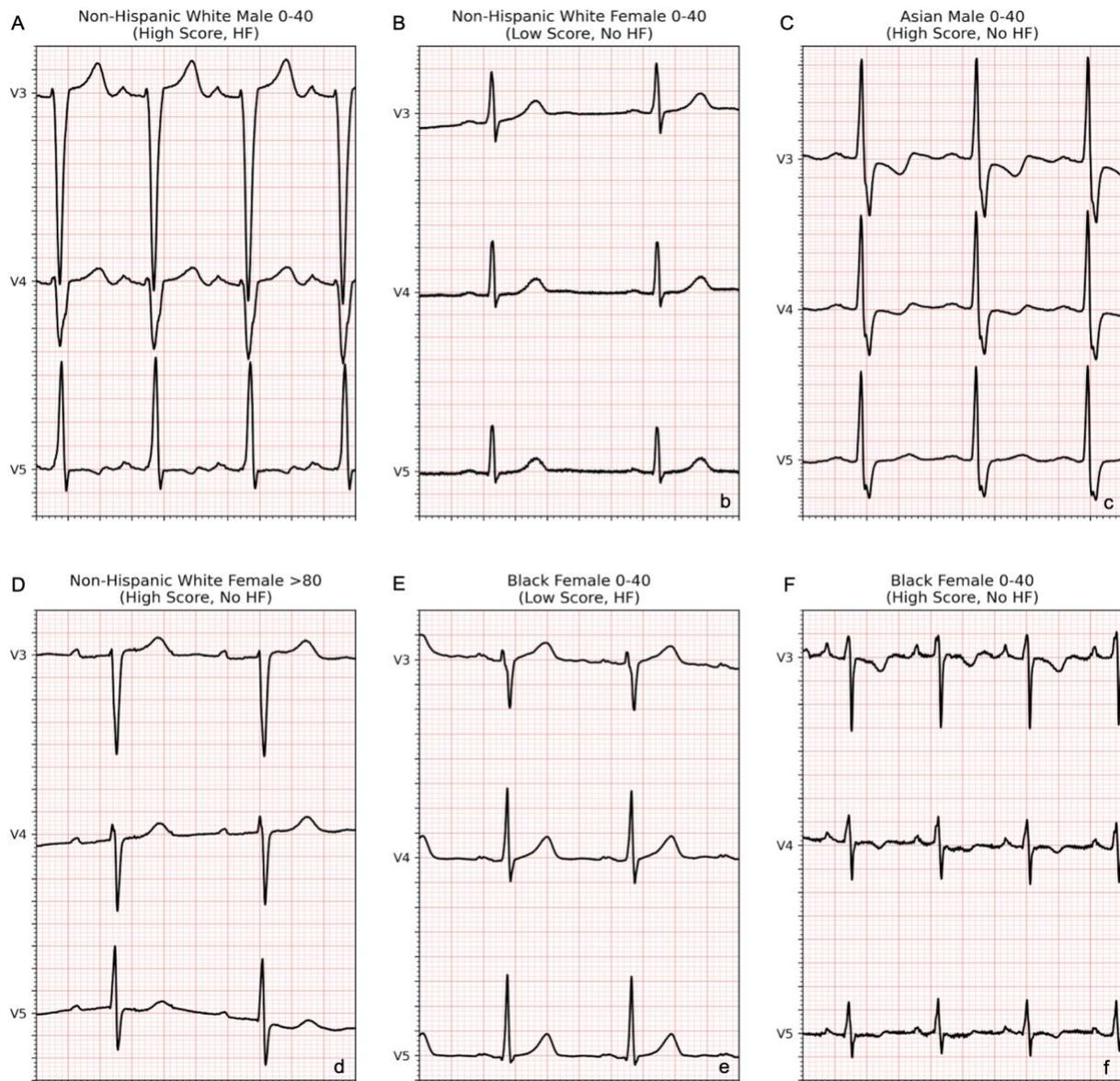
Figure S13: Sensitivity analyses using BNP (B-type natriuretic peptide) as part of heart failure diagnosis for age-based disparities in model performance
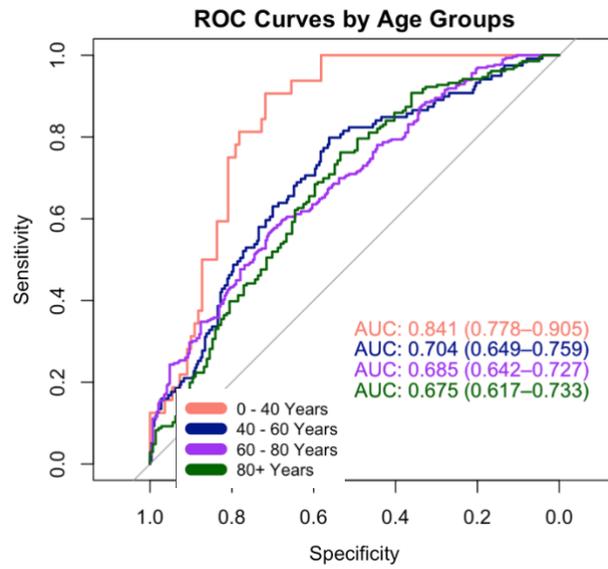
Figure S14: AUC (area under the curve) values and receiver operating characteristic curves stratified by race and age using a blanking period of three days from electrocardiogram collection
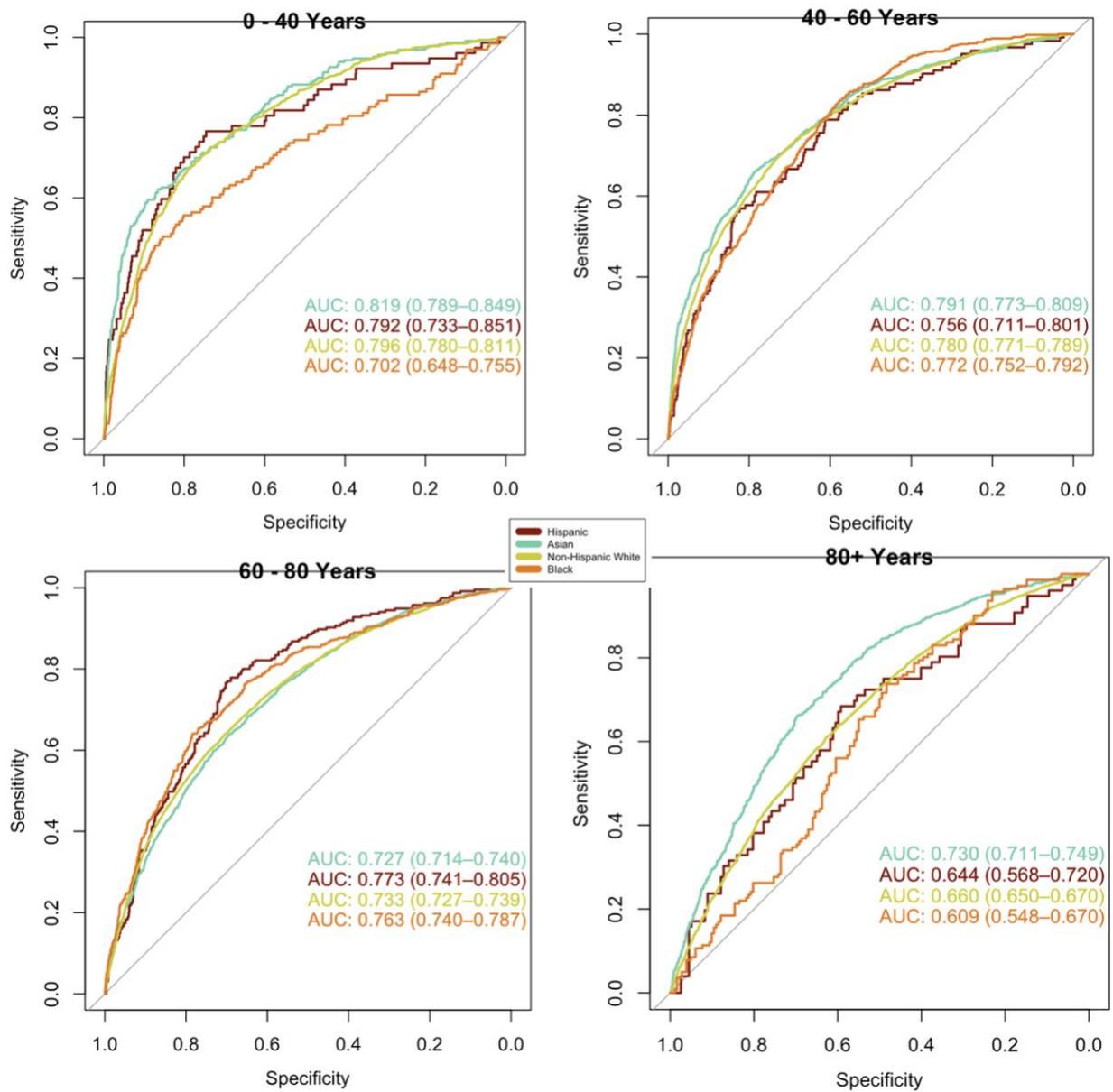
Table S1. Demographics of heart failure (HF) cases by race, gender, and age

Race	Age	Gender	Total	HF Cases	Percent w/ HF
Asian	0 - 40	Men	1335	104	7.8
		Women	1683	133	7.9
	40 - 60	Men	3503	437	12.5
		Women	3541	303	8.6
	60 - 80	Men	4575	888	19.4
		Women	5045	894	17.7
	>80	Men	1225	418	34.1
		Women	1555	498	32.0
Black	0 - 40	Men	519	66	12.7
		Women	713	61	8.6
	40 - 60	Men	1329	321	24.2
		Women	1312	263	20.0
	60 - 80	Men	975	312	32.0
		Women	973	282	29.0
	>80	Men	96	48	50.0
		Women	220	93	42.3
Hispanic	0 - 40	Men	2400	248	10.3
		Women	3539	256	7.2
	40 - 60	Men	3327	680	20.4
		Women	3636	412	11.3

	60 - 80	Men	2716	682	25.1
		Women	3124	619	19.8
	>80	Men	421	197	46.8
		Women	711	275	38.7
Non-Hispanic White	0 - 40	Men	4346	472	10.9
		Women	5749	405	7.0
	40 - 60	Men	15315	1875	12.2
		Women	13414	1069	8.0
	60 - 80	Men	22822	4940	21.6
		Women	17656	3113	17.6
	>80	Men	5233	2098	40.1
		Women	6117	2214	36.2

Table S2. Changes in positive predictive value (PPV), negative predictive value (NPV), and recall using thresholds for binary cutoff optimized via individual F1-score vs overall optimal threshold (0.20)

Race	Gender	Age	Threshold	F1-Score Change	NPV Change	PPV Change	Recall Change
Asian	Men	0 - 40	0.30	0.06	0.00	0.13	-0.07
		40 - 60	0.25	0.01	-0.01	0.07	-0.08
		60 - 80	0.20	0.00	0.00	0.00	0.00
		>80	0.20	0.00	0.00	0.00	0.00
	Women	0 - 40	0.25	0.03	0.00	0.08	-0.07
		40 - 60	0.20	0.00	0.00	0.00	0.00
		60 - 80	0.20	0.00	0.00	0.00	0.00
		>80	0.15	0.00	0.06	-0.04	0.15
Black	Men	0 - 40	0.15	0.02	0.01	-0.08	0.14
		40 - 60	0.15	0.03	0.04	-0.05	0.18
		60 - 80	0.20	0.00	0.00	0.00	0.00
		>80	0.10	0.08	0.20	-0.01	0.27
	Women	0 - 40	0.15	0.11	0.02	0.03	0.21
		40 - 60	0.15	0.05	0.03	-0.06	0.20
		60 - 80	0.20	0.00	0.00	0.00	0.00
		>80	0.20	0.00	0.00	0.00	0.00
Hispanic	Men	0 - 40	0.15	0.01	0.01	-0.04	0.13
		40 - 60	0.15	0.02	0.03	-0.04	0.16

		60 - 80	0.20	0.00	0.00	0.00	0.00
		>80	0.10	0.00	0.11	-0.06	0.16
	Women	0 - 40	0.25	0.02	0.00	0.09	-0.08
		40 - 60	0.25	0.01	-0.01	0.10	-0.11
		60 - 80	0.25	0.01	-0.01	0.06	-0.10
		>80	0.15	0.00	0.04	-0.05	0.13
Non-Hispanic White	Men	0 - 40	0.15	0.00	0.01	-0.04	0.10
		40 - 60	0.25	0.01	-0.01	0.06	-0.11
		60 - 80	0.25	0.00	-0.02	0.04	-0.11
		>80	0.20	0.00	0.00	0.00	0.00
	Women	0 - 40	0.15	0.01	0.01	-0.04	0.16
		40 - 60	0.20	0.00	0.00	0.00	0.00
		60 - 80	0.25	0.00	-0.01	0.05	-0.12
		>80	0.15	0.00	0.04	-0.03	0.11

Table S3: Average age across racial groups

Race	Mean Age [95% CI]
Asian	60.7 [60.6, 60.9]
Black	53.5 [53.2, 53.8]
Hispanic	51.2 [51.0, 51.4]
Non-Hispanic White	61.5 [61.4, 61.5]

Table S4: Fractions excluded due to prior heart failure (HF) and loss to follow-up by race, age, and gender

Age	Gender	Race	Fraction w/ Prior HF	Fraction Lost to Follow-Up
0 - 40	Men	Asian	0.14	0.56
		Black	0.16	0.54
		Hispanic White	0.15	0.52
		Non-Hispanic White	0.11	0.55
	Women	Asian	0.07	0.50
		Black	0.14	0.51
		Hispanic White	0.09	0.48
		Non-Hispanic White	0.09	0.51
40-60	Men	Asian	0.18	0.48
		Black	0.29	0.54
		Hispanic White	0.19	0.48
		Non-Hispanic White	0.17	0.46
	Women	Asian	0.11	0.49
		Black	0.21	0.44
		Hispanic White	0.12	0.49
		Non-Hispanic White	0.12	0.44
60 - 80	Men	Asian	0.23	0.52
		Black	0.36	0.57
		Hispanic White	0.20	0.48
		Non-Hispanic White	0.24	0.50

	Women	Asian	0.17	0.47
		Black	0.29	0.56
		Hispanic White	0.17	0.50
		Non-Hispanic White	0.17	0.49
80+	Men	Asian	0.31	0.62
		Black	0.33	0.62
		Hispanic White	0.31	0.63
		Non-Hispanic White	0.34	0.61
	Women	Asian	0.27	0.60
		Black	0.35	0.63
		Hispanic White	0.24	0.55
		Non-Hispanic White	0.28	0.58

Table S5: Fraction of heart failure diagnoses with elevated BNP (B-type natriuretic peptide)

		Fraction with Elevated BNP	n
Age	0 – 40	0.71 [0.67 – 0.76]	381
	40 – 60	0.81 [0.79 – 0.83]	1211
	60 – 80	0.89 [0.88 – 0.90]	1923
	80+	0.97 [0.96 – 0.98]	966
Race	Asian	0.84 [0.82 – 0.86]	864
	Black	0.82 [0.77 – 0.85]	355
	Hispanic	0.83 [0.80 – 0.86]	651

	Non-Hispanic White	0.90 [0.89 – 0.91]	2611
Gender	Men	0.86 [0.85 – 0.88]	2137
	Women	0.88 [0.87 – 0.89]	2344

Table S6: Average number of interaction days with the healthcare system per patient by race and age group

Race	Age	Average # Visit Days per Patient
Asian	0 - 40	72 [68, 76]
Black		70 [63, 78]
Hispanic White		67 [60, 74]
Non-Hispanic White		68 [65, 70]
Asian	40 - 60	87 [84, 91]
Black		78 [73, 84]
Hispanic White		77 [70, 83]
Non-Hispanic White		75 [74, 77]
Asian	60 - 80	92 [89, 95]
Black		90 [84, 96]
Hispanic White		68 [62, 75]
Non-Hispanic White		83 [82, 84]
Asian	80+	75 [70, 80]
Black		68 [58, 78]
Hispanic White		59 [47, 72]
Non-Hispanic White		66 [64, 68]

Table S7: Ratio of predicted risk of heart failure between primary model trained on electrocardiogram data compared to linear regression model trained on clinical risk factors (prior CAD (coronary artery disease) or MI (myocardial infarction), prior stroke, diabetes, blood pressure, BMI (body mass index), and smoking history) and demographic variables (race, gender, age)

		Ratio of Model Risk Predictions
Age	0 - 40	10.5 [10.3 – 10.7]
	40 - 60	8.5 [8.4 – 8.6]
	60 - 80	5.6 [5.6 – 5.7]
	80+	3.3 [3.2 – 3.3]
Race	Asian	7.2 [7.1 – 7.4]
	Black	5.9 [5.7 – 6.1]
	Hispanic White	7.1 [6.8 – 7.3]
	Non-Hispanic White	6.6 [6.6 – 6.7]
Sex	Men	6.2 [6.2 – 6.3]
	Women	7.3 [7.2 – 7.4]