Development and Validation of Machine Learning-Based Race-Specific Models to Predict 10-Year Risk of Heart Failure: A Multi-cohort Analysis

Matthew W. Segar, MD, MS[1,2], Byron C. Jaeger, PhD[3], Kershaw V. Patel, MD[1,4], Vijay Nambi, MD, PhD[5,6], Chiadi E. Ndumele, MD[7], Adolfo Correa, MD[8], Javed Butler, MD, MPH, MBA[8], Alvin Chandra, MD[1], Colby Ayers, MS[1], Shreya Rao, MD, MPH[1,2], Alana A. Lewis, MD[9], Laura M. Raffield, PhD[10], Carlos J. Rodriguez, MD, MPH[11], Erin D. Michos, MD, MHS[7], Christie M. Ballantyne, MD[5], Michael E. Hall, MD[8], Robert J. Mentz, MD[12], James A. de Lemos, MD[1], Ambarish Pandey, MD, MSc[1]

[1]Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA

[2]Parkland Health and Hospital System, Dallas, TX, USA

[3]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

[4]Department of Cardiology, Houston Methodist DeBakey Heart & Vascular Center, Houston, TX

[5]Michael E DeBakey Veterans Affairs Hospital and Baylor College of Medicine, Houston, TX, USA

[6]Section of Cardiology and Cardiovascular Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

[7]Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[8]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

[9]Division of Cardiology, Northwestern University, Chicago, IL, UA

Address for correspondence: Dr. Ambarish Pandey, MD, MSCS, Assistant Professor of Medicine, Department of Internal Medicine, Division of Cardiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390-9047, Tel: 214-645-2101; Pager: 214-344-0090, ambarish.pandey@utsouthwestern.edu.

[10]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

[11]Departments of Medicine, Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York

[12]Division of Cardiology, Duke Clinical Research Institute, Durham, North Carolina

Abstract

Background: Heart failure (HF) risk and the underlying risk factors vary by race. Traditional models for HF risk prediction treat race as a covariate in risk prediction and do not account for significant parameters such as cardiac biomarkers. Machine learning (ML) may offer advantages over traditional modeling techniques to develop race-specific HF risk prediction models and elucidate important contributors of HF development across races.

Methods: We performed a retrospective analysis of four large, community cohort studies (ARIC, DHS, JHS, and MESA) with adjudicated HF events. Participants were aged >40 years and free of HF at baseline. Race-specific ML models for HF risk prediction were developed in the JHS cohort (for Black race-specific model) and White adults from ARIC (for White rate-specific model). The models included 39 candidate variables across demographic, anthropometric, medical history, laboratory, and electrocardiographic domains. The ML models were externally validated and compared with prior established traditional and non-race specific ML models in race-specific subgroups of the pooled MESA/DHS cohort and Black participants of ARIC. Harrell's C-index and Greenwood-Nam-D'Agostino chi-square tests were used to assess discrimination and calibration, respectively.

Results: The ML models had excellent discrimination in the derivation cohorts for Black (N=4,141 in JHS, C-index=0.88) and White (N=7,858 in ARIC, C-index=0.89) participants. In the external validation cohorts, the race-specific ML model demonstrated adequate calibration and superior discrimination (C-indices=0.80–0.83 [for Black individuals] and 0.82 [for White individuals]) compared with established HF risk models or with non-race specific ML models derived using race as a covariate. Among the risk factors, natriuretic peptide levels were the most important predictor of HF risk across both races, followed by troponin levels in Black and EKG-based Cornell voltage in White individuals. Other key predictors of HF risk among Black individuals were glycemic parameters and socioeconomic factors. In contrast, prevalent cardiovascular (CV) disease and traditional CV risk factors were stronger predictors of HF risk in White adults.

Conclusion: Race-specific and ML-based HF risk models that integrate clinical, laboratory, and biomarker data demonstrated superior performance when compared with traditional HF risk and non-race specific ML models. This approach identifies distinct race-specific contributors of HF.

Keywords

heart failure; risk prediction; machine learning; race; epidemiology

INTRODUCTION

Heart failure (HF) affects over 64 million adults worldwide and accounts for nearly 6% of total healthcare expenditure in North America and Europe.[1, 2] Due to the growing burden of cardiovascular risk factors such as obesity and diabetes and more prolonged survival after myocardial infarction (MI) and incident HF, the prevalence is expected to increase.[3] The lifetime risk of developing HF is 20% to 46% and varies substantially by race.[4, 5] Race is a social construct and Black adults are more likely to develop HF compared with White individuals due in large part to structural racism, social inequality, poor healthcare access, and a greater downstream burden of cardiovascular risk factors such as hypertension and diabetes.[6–8] Additionally, Black adults have a higher prevalence of subclinical phenotypes such as left ventricular hypertrophy (LVH) and chronic myocardial injury assessed by high-sensitivity cardiac troponin (hs-cTn), which are associated with higher risk of HF.[9] In contrast, natriuretic peptide (NP) levels are lower among Black adults, further suggesting potential differences in the determinants of HF risk across racial groups.[10, 11] After the development of HF, prognosis is poor and Black adults appear to have significantly higher risk of adverse events, particularly HF hospitalization, compared with other racial groups.[12] Thus, there is a global need for effective approaches to the prevention of HF, but the optimal strategies may differ by race.[13]

Risk stratification tools to identify individuals at higher risk for developing HF are available and can help target those at high risk with appropriate preventive strategies.[14–19] However, most existing HF risk prediction models do not include key prognostic markers such as hs-cTn and NP-levels.[14, 15, 19] Furthermore, most existing HF risk prediction models, with the exception of the recently developed Pooled Cohort Equation to Prevent HF (PCP-HF) risk score[19], are not race-specific, have been developed in biracial or mixed ethnic cohorts, and only incorporate race as a covariate. Self-reported race as a covariate in such risk models may not completely capture the social and biological pathways that might predispose Black individuals to a higher risk of HF. Additionally, current HF risk prediction tools were developed using classical statistical modeling techniques constrained by assumptions such as distribution normality, noninformative or random censoring, and hazard risk linearity.[20] Machine learning (ML) methods utilize large, multidimensional data efficiently and may improve HF risk prediction.

In the present study, we developed HF risk prediction models in Black and White participants separately as a strategy to better elucidate the importance of risk factors that may be most relevant to the development of HF across races. Specifically, using ML techniques incorporating clinical parameters, including biomarker data, we developed and externally validated race-specific risk models to predict the 10-year risk of incident HF in Black and White adults. We compared the ML-based race-specific models' performance to the established HF risk prediction models, and non-race specific ML model derived using race as a covariate. We further identified the distinct associations of several risk factors and clinical characteristics with HF development in Black and White adults.

METHODS

Requests to access the Atherosclerosis Risk in Communities Study (ARIC), Dallas Heart Study (DHS), Jackson Heart Study (JHS), and Multi-Ethnic Study of Atherosclerosis (MESA) datasets can be obtained by the respective study coordinating institutions. Details on the software code used in this study are provided in the Supplemental Materials and available from the corresponding author upon reasonable request.

Study population

The present study included individual-level data from four cohort studies: ARIC, DHS, JHS, and MESA. The study design, enrollment criteria, participant characteristics, and event adjudication for each cohort have been described previously and are discussed briefly in the Supplemental Materials.[21–25] For the present analysis, the initial (visit 1) phases of JHS (2000 – 2004), DHS (2000 – 2002), MESA (2000 – 2004), and visit 4 for the ARIC cohort (1996 – 1998) are considered the baseline visits to allow similar follow-up time and consistent availability of biomarkers across cohorts. Adults less than 40 years of age or those with a history of HF at study baseline were excluded from each cohort (Supplemental Figure I).

Race-specific HF risk prediction models were developed in a derivation cohort and subsequently externally validated (Figure 1). Data from JHS were used to derive the HF risk prediction model for Black participants. The HF risk prediction model for Black adults was validated in two separate cohorts of 1) a subset of Black participants from ARIC, excluding participants who were also enrolled from Jackson, MS and were included in JHS, and 2) a pooled cohort of Black participants from MESA and DHS. The HF risk prediction model for White participants was developed using data from ARIC and externally validated among White participants from the pooled MESA/DHS dataset. Written informed consent was obtained from all study participants, and local institutional review boards approved the study protocols. The present analysis was approved by the coordinating centers for each cohort and was considered exempt from Institutional Review Board approval at the University of Texas Southwestern Medical Center, Dallas, Texas. The study followed the Transparent reporting of the multivariable prediction model for individual prognosis or diagnosis (TRIPOD) and Prediction model Risk Of Bias Assessment Tool (PROBAST) reporting guidelines.[26, 27]

Candidate clinical covariates

Participants from all four cohorts underwent comprehensive evaluations at the study baseline, including self-reported history, assessment of clinical and anthropometric characteristics by examination, and laboratory tests using standardized protocols as detailed in the Supplemental Materials. In the present analysis, 54 demographic, anthropometric, clinic, electrocardiographic, and laboratory variables assessed at the study baseline were considered candidate covariates. Categorical candidate variables were harmonized across study cohorts. Troponin was measured as hs-cTnT in MESA and DHS and hs-cTnI in JHS and ARIC cohorts using commercially available assays as reported previously.[28–31] NP levels were measured as B-type natriuretic peptide (BNP) in JHS and N-terminal pro-B-type

(NT-proBNP) in ARIC, MESA, and DHS studies as reported previously.[29, 30, 32, 33] NP and hs-cTn levels were each transformed into Z-scores as previously described to facilitate comparison across different assays.[34] In total, 39 predictor variables were included in the present analysis after excluding 12 covariates for >20% missingness and 3 for correlation coefficient > 0.70 (Supplemental Table I).

Outcome of interest

The primary outcome of interest for the present study was incident HF. Each cohort follow-up was landmarked at one year from baseline visit to account for reverse causation secondary to potential subclinical HF at study baseline. Each cohort had at least 10 years of follow-up for outcome assessment, and HF endpoints were censored at 10 years. Incident HF events were identified by the first hospitalization event with HF in each of the study cohorts and adjudicated according to standardized protocols as described in the Supplemental Materials.

Model development and evaluation

HF risk prediction models were developed separately for Black (JHS cohort) and White participants (ARIC cohort). The derivation cohort was randomly split into a training (50%) and testing dataset (50%). Imputation was performed on each dataset separately using random forest imputation to prevent data leakage across cohorts.[35] The performance of the best performing ML-based algorithms were further compared to models derived using a similar approach 1) with only routine clinical covariates (age, sex, systolic blood pressure [BP], diastolic BP, heart rate, body mass index [BMI], smoking history, history of cardiovascular disease [CVD], history of diabetes), 2) clinical plus routine laboratory covariates (hemoglobin A1c, total cholesterol, high-density lipoprotein cholesterol [HDL-c]), 3) with only NP levels and only hs-cTn levels as covariates, 4) model that was not race-specific and was developed using the overall cohort (i.e, not derived separately for Black and White adults) with race as a co-variate, 5) model that was developed among individuals without any missing data on all the parameters. The analysis overview is shown in Figure 1.

ML-based methods: In the training dataset, HF prediction models were developed incorporating the 39 covariates identified previously using the following six candidate ML-based and traditional Cox algorithms: 1) oblique random survival forests (RSF); 2) forward stepwise Cox regression; 3) lasso Cox regression; 4) ridge Cox regression; 5) boosted Cox regression; 6) gradient boosted trees. A brief description of each method is provided in the Supplemental Materials.

Variable selection: Variable selection was adjusted to optimize the number of covariates included in the HF risk prediction model. While oblique RSF can be used for risk prediction, random forest methods can also be used in variable selection.[36] Permutation-based selection was performed to quantify variable importance (VIMP). Briefly, a predictor variable's values were randomly permuted, and the importance of that variable was quantified by the resulting degradation in model prediction accuracy.[20] A high VIMP suggests that permutation of the variable worsens predictive accuracy and, thus, the variable is important in model prediction.

Given the diminishing returns of model performance with additional covariates, the top 20 variables in Black and White adults separately with the highest VIMP metrics were included in the ML-based algorithm with the highest discrimination.[37]

Statistical analysis

Internal validation in the development cohort was assessed using Monte Carlo cross-validation. The discriminatory performance of each model was evaluated in the validation dataset (50% of the derivation cohort) by calculating the Harrell concordance index (C-index).[38] The data transformation, imputation, training, and testing steps were repeated 1,000 times, with the final C-index reported as mean (95% confidence interval). The DeLong test was used to evaluate for differences between C-indices.[39] Calibration of the HF risk prediction model with the best discrimination was assessed using the Greenwood-Nam-D'Agostino statistic with a P-value > 0.05, indicating adequate calibration.[40]

External validation

The best performing HF risk prediction model in the internal validation dataset was evaluated in external datasets. The ML model for Black adults was externally validated in two separate cohorts of Black participants from 1) ARIC (excluding Black individuals enrolled in JHS) and 2) pooled MESA/DHS datasets. The ML model for White adults was externally validated in White participants from the pooled MESA/DHS cohort. The performance of ML-based models was further compared with the following well-established traditional Cox-regression based HF risk prediction models: the ARIC-HF biomarker, the Pooled Cohort Equations to Prevent Heart Failure (PCP-HF), and MESA-HF risk scores, each of which were developed in one the external validation cohorts.[17–19] Finally, decision curves, a measure of the clinical net benefit, were compared between the ML, ARIC-HF, PCP-HF, and MESA-HF models.[41] Decision curves measure the number of true positive cases identified without increasing the false positive rate at a specific risk threshold.

Sensitivity analysis

Several sensitivity and subgroup analyses were performed to assess the robustness of the ML-based race-specific HF risk prediction models. First, we compared the performance of the ML-based race-specific models in the other race (i.e., White race-specific model in Black participants and vice-versa) using discrimination, calibration metrics, and decision curve analysis. Second, to assess generalizability of the ML-based models in cohorts without biomarker data, we evaluated their performance in the validation cohorts with missing biomarker (NP and hs-cTn levels) data (i.e. setting the biomarker values to missing). Third, sensitivity analyses were performed to assess the performance of the ML-based model in a primary prevention population, excluding individuals with baseline atherosclerotic CVD. Fourth, subgroup analyses were performed to evaluate the performance of the ML-based race-specific model in sex-based (men and women) subgroups. Finally, the performance of the ML-based race-specific model was assessed to predict the risk of incident HF subtypes [HF with reduced ejection fraction (HFrEF) and preserved (HFpEF) ejection fraction]. For the HF subtype prediction analysis, the other HF subtype and mortality were considered as censoring events.

Contribution of specific risk factors to overall risk across races

To determine the proportion of HF incidence attributable to specific HF risk factor burden across races, we used Monte Carlo simulation of 1 million patients with risk factor burden similar to the ARIC cohort in both Black and White adults. A detailed description is provided in the Supplemental Materials. The predicted risk of each patient in the simulated cohorts was calculated using the best performing HF risk prediction model. The population attributable risk percentage (PARP) for each specific risk factor was estimated as the difference between the average risk in the unexposed group and the average risk in the overall population. For the PARP estimation, continuous variables were dichotomized into clinically meaningful categories using previously established cutoffs (BNP 30 pg/mL, NTproBNP 125 pg/mL, hs-cTn 14 ng/L). Analyses were performed using R version 3.6.3 with a P-value < 0.05 indicating significance.

RESULTS

Study cohort and baseline characteristics

The derivation cohort for Black and White adults consisted of 4141 participants from JHS and 7858 participants from ARIC, respectively (Supplemental Figure I). MESA and the DHS contributed 1848 and 973 participants to the external validation cohort for Black adults (n=2821), respectively, while ARIC contributed 1024 participants. The external validation cohort for White adults (n=3236) consisted of 2595 participants from MESA and 641 participants from DHS. During a maximum follow-up of 10 years, there were 288 [7.0%; male=99 (6.7%), female=189 (7.1%)] and 634 [8.1%; male=340 (9.4%), female=294 (7.0%)] HF events among Black and White adults in the derivation cohorts, respectively. In the external validation cohort, 284 (7.4%, ARIC cohort=169 [16.5%] and MESA/DHS=115 [4.1%]) Black and 100 (3.1%) White participants developed HF over a maximum of 10 years of follow-up (Supplemental Figure II).

Table 1 shows the baseline characteristics of Black and White participants across the derivation and validation cohorts. Black adults had a greater burden of cardiovascular risk factors, including a higher prevalence of hypertension and diabetes, higher body mass index, blood pressure, fasting plasma glucose (FPG), and greater chronic myocardial injury burden than White adults (Table 1). Conversely, White adults were more likely to be male and had higher rates of renal dysfunction and smoking than Black adults.

Derivation of prediction models

The 39 predictor covariates incorporated into the risk prediction models included demographics, anthropometrics, vital signs, social history, clinical history, medications, laboratory, and electrocardiographic parameters (Supplemental Table I). The C-indices for the ML-based race-specific HF risk prediction models are displayed in Table 2. Across both race groups, the oblique RSF (oRSF) model had the highest C-index of 0.88 (95% CI 0.86–0.91) among Black adults and 0.89 (95% CI 0.86–0.91) among White adults. Across both races the forward stepwise Cox regression model had the lowest C-index [Black adults: 0.76 (95% CI 0.71–0.80); White adults: 0.77 (95% CI 0.73–0.80)].

To improve clinical usability, we further constructed a model using the top 20 covariates identified by the VIMP metric in oRSF (Figure 2). Among Black adults, the C-index for the oRSF-top 20 model was 0.88 (95% CI 0.85–0.90) and was not significantly different from the oRSF model that included all 39 variables (DeLong p-value = 0.38). Similarly, among White participants, discrimination of the oRSF model incorporating all covariates was not different from that which included only the top 20 variables with the highest VIMP metrics [0.88 (95% CI 0.85–0.90)] (DeLong p-value = 0.22). Thus, the oRSF top-20 model was identified as the best performing model among Black and White adults (hereby referred to as the ML model). The oRSF models derived using only routine clinical variables or clinical and lab parameters had a lower overall C-index than the oRSF top-20 model (C-indices = 0.75–0.76) (Table 3). Additionally, oRSF models with biomarkers only (NP alone or hs-cTn alone) also had a lower overall C-index than the oRSF top-20 (Table 3). Performance of the oRSF top-20 model was also superior to the non-race specific model derived in the overall cohort including race as a covariate (Table 2). Finally, in a cohort of participants with no missing data (n=10,053, 83.8% of the derivation cohort), the complete data oRSF top-20 model performed similarly to the oRSF top-20 model with missing data (Table 3).

External validation of prediction models among Black adults

Among Black adults, the overall C-index for the ML risk score was high in both external validation cohorts [ARIC: 0.80 (95% CI 0.75–0.84); MESA/DHS: 0.83 (95% CI 0.77–0.87)] (Table 4). Discrimination and calibration were acceptable in a cohort with missing biomarker data (Supplemental Table II). The ML risk score performance was also superior to other HF risk scores (ARIC-HF, MESA-HF, and PCP-HF) in external cohorts, including those which were used to derive these models (Table 4). Calibration was adequate for the ML model in both ARIC and MESA/DHS validation cohorts. In contrast, ARIC-HF and PCP-HF risk scores demonstrated good calibration among Black participants from the ARIC cohort, but not in the pooled cohort of MESA/DHS (Figure 2A-B). The MESA-HF risk score demonstrated adequate calibration in Black participants from both external validation cohorts (Figure 2A-B). In decision curve analysis, the ML model increased the number of heart failure events detected by 12, 17, and 19 per 1000-patients compared to the ARIC-HF, PCP-HF, and MESA-HF risk scores in the ARIC validation cohort (Supplemental Figure III). Similarly, in the pooled MESA/DHS validation cohort, the ML model detected an additional 6, 2, and 5 heart failure events per 1000-patients compared to the ARIC-HF, PCP-HF, and MESA-HF risk scores (Supplemental Figure III). In a primary prevention cohort of individuals without CVD, the ML model demonstrated improved discrimination and calibration compared with the other established risk scores (Supplemental Table III).

External validation of prediction models among White adults

In the external validation cohort of pooled White participants from DHS and MESA, the ML risk score had the greatest discrimination as assessed by the C-index of 0.82 (95% CI 0.78–0.86) followed by the MESA-HF, ARIC-HF, and PCP-HF risk scores, respectively (Table 4). Only the ML and MESA-HF risk scores had adequate calibration with similar predicted and observed risk of HF (p-value > 0.05) (Figure 2C). In decision curve analysis, the ML model detected an additional 3, 6, and 4 heart failure events per 1000-patients compared to the ARIC-HF, PCP-HF, and MESA-HF risk scores (Supplemental Figure III).

Discrimination and calibration were also acceptable in a cohort with missing biomarker data (Supplemental Table II). In a primary prevention cohort of individuals without CVD, the ML model demonstrated improved discrimination and calibration compared with the other established risk scores (Supplemental Table III).

Performance of Black- and White-specific models in the validation cohort of the other race

We evaluated the performance of the race-specific models in the other race (i.e., White model in Black participants and vice-versa) in the MESA/DHS external validation cohort. The discriminatory performance (C-index) of race-specific models was superior in race-concordant vs. race-discordant cohorts (C-index of the Black model in Black vs. White participants = 0.83 vs. 0.80; C-index of the White model in White vs. Black participants = 0.82 vs. 0.78). Similarly, validation in race-discordant cohorts resulted in significant miscalibration (GND p-value < 0.05 for all) (Supplemental Table IV). In decision curve analysis, Black race-specific models detected an additional 6 cases per-1000 participants in the Black vs. White cohort. Similarly, the White race-specific model detected an additional 4 cases per-1000 participants in the White vs. Black cohort (Supplemental Figure IV).

Performance of ML-based race-specific models in sex- and HF subtype-based groups

In sex-based subgroup analysis, the race-specific models demonstrated good and comparable performance in men and women across both races in the validation cohorts (Supplemental Table V). For the HF subtype outcomes, the Black race-specific model performed well in predicting risk of HFrEF [C-index = 0.84 (95% CI 0.79–0.90)] and HFpEF [C-index = 0.82 (95% CI 0.76–0.87)] in the MESA/DHS validation cohort (Supplemental Table V). Similarly, the White race-specific model demonstrated good performance in predicting the risk of HFrEF [C-index = 0.83 (95% CI 0.77–0.89)] and HFpEF [C-index = 0.79 (95% CI 0.74–0.85)] in the MESA/DHS cohort.

Differences in predictors of HF risk in Black vs. White adults

The variables identified as the top 20 most important parameters for HF risk prediction according to VIMP in Black and White adults are displayed in an alluvial plot (Figure 2). The key variables for predicting HF risk were mostly the same for both race groups except for body mass index, which was only included in the Black race model, and sex, which was only part of the White race model. NP level was the most important parameter in HF risk prediction for both race groups, followed by hs-cTn levels in Black individuals and ECG left ventricular mass measure (Cornell Voltage) in White individuals. Estimated GFR, cardiometabolic parameters (presence of diabetes, FPG), and socioeconomic characteristics (education and income levels) were more important predictors of HF risk in Black vs. White adults. The other important predictors of HF risk in White adults included prevalent CVD, traditional CVD risk factors (HTN, BP levels, HDL-c), and hs-cTn levels.

Among Black individuals, the highest PARP for incident HF was noted for elevated natriuretic peptide levels (NT-ProBNP 125 pg/mL, PARP: 21.1%) followed by cardiac troponin (hs-TnI 14 ng/L, PARP = 12.2%), annual income (upper-middle or above = 11.7%), and cardiometabolic parameters (FPG 140 mg/dL = 9.5%, presence of diabetes = 8.6%) (Figure 3). Among White individuals, the PARP for incident HF associated with

cardiac biomarkers was more modest (NT-ProBNP ≥ 125 = 17.9%, hs-cTn ≥ 14 ng/L = 6.8%), lower than that associated with other risk factors such as FPG ≥ 140 mg/dL (4.3%) and annual income (7.4%) and higher among current smoking (9.1%) and LVH (8.0%) (Figure 3).

DISCUSSION

In this individual-level analysis of four prospective cohort studies, we developed and validated race-specific ML-based models to predict the 10-year risk of incident HF. We evaluated 6 unique ML-based algorithms and identified the oblique RSF model as the best performing HF risk prediction tool with excellent discrimination and calibration in external validation cohorts superior to previously developed risk scores. We similarly observed that ML models derived using race as a covariate have inferior performance than race-specific models. Using ML to identify the most important predictors of HF risk across racial groups, we noted differences in the risk factor profiles and relative importance and contribution of key predictors to the overall risk among Black and White adults. The present study supports the use of the ML model to identify Black and White individuals at higher risk for developing HF who may benefit from effective preventive strategies. To improve the clinically utility of our study, we implemented the ML model on a publicly available website (www.cvriskscores.com) (Supplemental Figure V).

HF risk prediction and utilization of machine learning

A critical component of population-based strategies for preventing HF is identifying the highest risk individuals who may be targeted with effective preventive approaches. HF risk scores were previously developed in selected populations, including predominantly White adults in the Framingham Heart Study, older individuals in the Health ABC study, and more recently among adults with no prior history of CVD in the PCP-HF study.[14, 15, 19] Recent studies included more diverse populations and incorporated traditional risk factors into Cox models to estimate HF risk.[16, 18] Model performance has improved with the inclusion of biomarkers, specifically NT-proBNP in the ARIC-HF and MESA-HF models.[16–18] However, traditional statistical modeling with Cox proportional hazard regression is unable to include non-linear outcomes or missing or correlated variables.

ML utilizes less stringent techniques that allow for the inclusion of large amounts of multidimensional data and has been shown to predict adverse events in population-based cohort studies.[42] Furthermore, ML models are better equipped to handle missing data on covariates and non-linear interactions between covariates compared with traditional Cox models. In the current study, of 54 candidate covariates, 39 variables were incorporated in risk prediction models. While all 6 ML-based algorithms had excellent discrimination for HF risk prediction, the oRSF model had the best performance. The oRSF model incorporating the top 20 most important predictors of HF risk, which included clinical parameters and cardiac biomarkers, outperformed previously established risk prediction tools. The oRSF top-20 model outperformed traditional models, including those developed in the external validation cohorts, and included biomarkers as covariates (ARIC and MESA HF risk score).

Need for race-specific HF risk scores

Despite differences in HF epidemiology across races, most traditional HF risk prediction tools are not race-specific and incorporate self-reported race as a covariate adjustment in the risk prediction equation.[16] Self-reported Black race is a heterogeneous entity and a surrogate marker for several factors including but not limited to genetics, socioeconomic status, cultural elements, and often unmeasured societal experiences related to structured racism.[43, 44] The disproportionately higher risk of HF and other adverse outcomes in Black individuals is related to the downstream effects of all these components, which may not be adequately accounted for by adjusting for race in the risk models. Accordingly, there is a growing consensus to move away from race as a risk predictor approach.[45] In the present study, we developed and validated HF risk prediction models in Black and White participants separately as a strategy to better elucidate the importance of risk factors that may be most relevant to development of HF in Black individuals. The ML-based race-specific HF risk models developed in this study included individuals with as well as without baseline CVD and complement the existing race- and sex-specific Cox-model based PCP-HF risk score which is focused on the primary prevention population.[19] The ML-based race-specific models demonstrated superior performance in the race concordant validation cohorts as compared with the other race. Furthermore, the race-specific ML model for HF risk demonstrated superior performance than non-race specific ML models which included race as a covariate. Overall, these findings highlight the prognostic utility of race-specific risk assessment for HF.

HF risk factors and relative differences among Black and White adults

In addition to providing a race-specific HF risk stratification tool, our study also provides key insights into the differences in the risk factors for HF development among Black versus White adults. Across both races, we observed that NP levels were the most important predictor of HF risk. Biomarkers of chronic myocardial injury were a more important risk predictor in Black vs. White individuals. Furthermore, the PARP associated with cardiac biomarkers for HF risk was higher in Black vs. White individuals. This is particularly noteworthy because among Black individuals, chronic myocardial injury is more prevalent and has been associated with a disproportionately higher risk of HF.[28] Socioeconomic factors such as income and cardiometabolic parameters (FPG, presence of diabetes) were also of higher importance and associated with higher PARP for incident HF among Black vs. White individuals.

Besides differences in relative importance of predictor variables shared across both race-specific models, there were unique variables identified as key predictors of HF risk in Black vs. White adults. Sex was identified as a significant predictor variable only in White individuals and BMI was a significant for incident HF only in Black individuals. These differences may be related to differences in the epidemiology of HF development in Black vs. White adults. For example, as noted in our study and by others,[6] sex differences in risk of HF in Black adults are more modest as compared with those noted among White adults, likely due to high baseline risk of HF among Black women.[9] This may explain why sex is one of the top predictors of HF risk only in the White race model.[6] It is also plausible that the contribution of certain factors to HF risk may be accounted for by other covariates and

the interactions between covariates. For example, the contribution of female sex to HF risk in Black individuals may be accounted for by BMI and other cardiometabolic parameters, which are already in the model.[46, 47]

Role of biomarkers in risk prediction using ML-based models

Our ML-based models also highlight the role of biomarker assessment, in addition to routinely captured clinical risk factors, for improved risk stratification. Across both race groups, NP levels were identified as the single most important predictor of HF risk. Additionally, hs-cTn was a significant predictor of HF risk but more so in Black versus White adults. We observed improved discrimination with the oRSF top-20 model as compared with the ML models with clinical parameters only as well as those with biomarkers only. This is particularly important as high-sensitivity assays for troponin are currently available and can be routinely assessed in outpatient setting. Additionally, the 2017 ACC/AHA HF Focused Update recommends NP screening as part of a HF prevention strategy.[48] Complementing our study findings, previous studies have also demonstrated the utility of cardiac biomarkers in identifying individuals with risk factors such as hypertension, diabetes, pre-diabetes who may be at a higher risk of developing HF and may benefit from targeted preventive strategies.[49, 50] However, cardiac biomarkers may not be routinely available among all at-risk participants, particularly those without existing CVD. The ML models demonstrated robust performance in subgroups with missing biomarker data and among individuals without CVD at baseline, highlighting their generalizability across different clinical settings. A pragmatic application of the ML-based models for HF risk prediction in the real-world may involve a multi-step approach tailored to the population in consideration. For example, in the general primary prevention population without biomarker data, clinical risk factors can be used to predict risk as the first step followed by selective biomarker testing to refine risk prediction by capturing residual risk of HF.

Clinical implications and future directions

The ML model for HF risk prediction developed in the present study can identify high-risk individuals who are most likely to benefit from effective preventive therapies. These strategies may include lifestyle interventions to improve cardiorespiratory fitness and weight loss,[51, 52] intensive blood pressure control,[53] and initiation of pharmacotherapies such as sodium-glucose cotransporter-2 inhibitors among individuals with type 2 diabetes.[54]

A key next step for the practical implementation of the ML-based race-specific model in the clinical setting is its integration with electronic health records (EHR). This will not only improve accessibility and real-world application of these ML-based models but also allow for better and automated capture of data on the risk prediction variables. To this end, continued improvements in EHR software will allow ML models to be developed and stored on a secure cloud server and implemented in the EHR through the use of an application programming interface (API) (Supplemental Figure VI). The next steps to facilitate integration of the ML-based race-specific models with the EHR include storing the models on a secure cloud-computing server, developing an API to access our race-specific models, and scaling the platform to provide real-time estimates of a patient's HF risk.

Limitations

Our study has several limitations. First, we only included risk factors that were consistently available across all cohorts at the baseline visit in the derivation cohorts. Data on important predictors of HF risk such as echocardiographic characteristics were not included in the derivation of the risk prediction model. However, the ML model included a wide range of clinical variables including traditional risk factors as well as biomarkers with established prognostic importance that have not been included in other HF risk prediction tools.[19] While echocardiographic parameters are able to accurately risk stratify HF prognosis, echocardiograms are not routinely obtained in patients without suspected HF. It is thus possible the ML models can improve resource allocation of echocardiograms for individuals at highest risk for HF. Second, we could not account for several social risk factors including structural racism, access to care, and racial disparities in management of CV risk factors which may influence the risk of HF among Black participants or for individuals who identify as neither White nor Black. Future cohort studies that better capture the effects of long-standing societal racism and disparities may improve our HF risk model performance by incorporating objective measures of these microaggressions and allopathic load. Similarly, data on relevant lifestyle factors such as physical activity, cardiorespiratory fitness, and adiposity measures, which may differ across races and are associated with risk of HF, was not available.[51, 55, 56] However, the ML-based models demonstrate robust performance even in absence of data on key parameters and across different subgroups and sensitivity analysis. Furthermore, through the use of an API, the ML-based models can be continuously updated as these data become available in additional cohorts. Finally, HF event adjudication differed slightly across all cohorts, and HF subtypes were not systematically captured. However, each cohort study has its own well-established, pre-specified criteria for HF event ascertainment, and the model performance was tested separately in each specific cohort and demonstrated good discrimination and calibration.

Conclusion

In conclusion, we used ML to develop race-specific models to predict the 10-year risk of incident HF and externally validated these findings in large population-based cohorts. The ML models displayed excellent discrimination and calibration among Black and White adults. Finally, using ML to identify the most important predictors of HF risk across racial groups, we found a similar profile of risk factors with substantial differences in the relative importance and contribution to the overall risk of specific parameters among Black and White adults.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Nonstandard Abbreviations and Acronyms

ARIC	Atherosclerosis Risk in Communities
DHS	Dallas Heart Study
GND	Greenwood-Nam-D'Agostino
hs-cTn	high-sensitivity cardiac troponin
JHS	Jackson Heart Study
MESA	Multi-Ethnic Study of Atherosclerosis
ML	machine learning
NP	natriuretic peptide
oRSF	oblique random survival forest
PARP	population attributable risk percentage
VIMP	variable importance

REFERENCES

1. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, Nodari S, Lam CSP, Sato N, Shah AN, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol 2014;63:1123–1133. [PubMed: 24491689]

2. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation 2020;141:e139–e596. [PubMed: 31992061]

3. Khera R, Kondamudi N, Zhong L, Vaduganathan M, Parker J, Das SR, Grodin JL, Halm EA, Berry JD and Pandey A. Temporal Trends in Heart Failure Incidence Among Medicare

Beneficiaries Across Risk Factor Strata, 2011 to 2016. JAMA Netw Open 2020;3:e2022190. [PubMed: 33095250]

4. Huffman MD, Berry JD, Ning H, Dyer AR, Garside DB, Cai X, Daviglus ML and Lloyd-Jones DM. Lifetime risk for heart failure among white and black Americans: cardiovascular lifetime risk pooling project. J Am Coll Cardiol 2013;61:1510–7. [PubMed: 23500287]

5. Pandey A, Omar W, Ayers C, LaMonte M, Klein L, Allen NB, Kuller LH, Greenland P, Eaton CB, Gottdiener JS, et al. Sex and Race Differences in Lifetime Risk of Heart Failure With Preserved Ejection Fraction and Heart Failure With Reduced Ejection Fraction. Circulation 2018;137:1814–1823. [PubMed: 29352072]

6. Chang PP, Wruck LM, Shahar E, Rossi JS, Loehr LR, Russell SD, Agarwal SK, Konety SH, Rodriguez CJ and Rosamond WD. Trends in Hospitalizations and Survival of Acute Decompensated Heart Failure in Four US Communities (2005–2014): ARIC Study Community Surveillance. Circulation 2018;138:12–24. [PubMed: 29519849]

7. Kalogeropoulos AP, Georgiopoulou VV, deFilippi CR, Gottdiener JS, Butler J and Cardiovascular Health S. Echocardiography, natriuretic peptides, and risk for incident heart failure in older adults: the Cardiovascular Health Study. JACC Cardiovasc Imaging 2012;5:131–40. [PubMed: 22340818]

8. Bahrami H, Kronmal R, Bluemke DA, Olson J, Shea S, Liu K, Burke GL and Lima JA. Differences in the incidence of congestive heart failure by ethnicity: the multi-ethnic study of atherosclerosis. Arch Intern Med 2008;168:2138–45. [PubMed: 18955644]

9. Lewis AA, Ayers CR, Selvin E, Neeland I, Ballantyne CM, Nambi V, Pandey A, Powell-Wiley TM, Drazner MH, Carnethon MR, et al. Racial Differences in Malignant Left Ventricular Hypertrophy and Incidence of Heart Failure: A Multicohort Study. Circulation 2020;141:957–967. [PubMed: 31931608]

10. Wang TJ, Wollert KC, Larson MG, Coglianese E, McCabe EL, Cheng S, Ho JE, Fradley MG, Ghorbani A, Xanthakis V, et al. Prognostic utility of novel biomarkers of cardiovascular stress: the Framingham Heart Study. Circulation 2012;126:1596–604. [PubMed: 22907935]

11. Gupta DK, Claggett B, Wells Q, Cheng S, Li M, Maruthur N, Selvin E, Coresh J, Konety S, Butler KR, et al. Racial differences in circulating natriuretic peptide levels: the atherosclerosis risk in communities study. J Am Heart Assoc 2015;4:e001831. [PubMed: 25999400]

12. Lewis EF, Claggett B, Shah AM, Liu J, Shah SJ, Anand I, O'Meara E, Sweitzer NK, Rouleau JL, Fang JC, et al. Racial Differences in Characteristics and Outcomes of Patients With Heart Failure and Preserved Ejection Fraction in the Treatment of Preserved Cardiac Function Heart Failure Trial. Circ Heart Fail 2018;11:e004457. [PubMed: 29664406]

13. Schocken DD, Benjamin EJ, Fonarow GC, Krumholz HM, Levy D, Mensah GA, Narula J, Shor ES, Young JB, Hong Y, et al. Prevention of heart failure: a scientific statement from the American Heart Association Councils on Epidemiology and Prevention, Clinical Cardiology, Cardiovascular Nursing, and High Blood Pressure Research; Quality of Care and Outcomes Research Interdisciplinary Working Group; and Functional Genomics and Translational Biology Interdisciplinary Working Group. Circulation 2008;117:2544–65. [PubMed: 18391114]

14. Kannel WB, D'Agostino RB, Silbershatz H, Belanger AJ, Wilson PW and Levy D. Profile for estimating risk of heart failure. Arch Intern Med 1999;159:1197–204. [PubMed: 10371227]

15. Butler J, Kalogeropoulos A, Georgiopoulou V, Belue R, Rodondi N, Garcia M, Bauer DC, Satterfield S, Smith AL, Vaccarino V, et al. Incident heart failure prediction in the elderly: the health ABC heart failure score. Circ Heart Fail 2008;1:125–33. [PubMed: 19777072]

16. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, Folsom AR, He M, Hoogeveen RC, Ni H, et al. Prediction of incident heart failure in general practice: the Atherosclerosis Risk in Communities (ARIC) Study. Circ Heart Fail 2012;5:422–9. [PubMed: 22589298]

17. Nambi V, Liu X, Chambless LE, de Lemos JA, Virani SS, Agarwal S, Boerwinkle E, Hoogeveen RC, Aguilar D, Astor BC, et al. Troponin T and N-terminal pro-B-type natriuretic peptide: a biomarker approach to predict heart failure risk--the atherosclerosis risk in communities study. Clin Chem 2013;59:1802–10. [PubMed: 24036936]

18. Chahal H, Bluemke DA, Wu CO, McClelland R, Liu K, Shea SJ, Burke G, Balfour P, Herrington D, Shi P, et al. Heart failure risk prediction in the Multi-Ethnic Study of Atherosclerosis. Heart 2015;101:58–64. [PubMed: 25381326]

19. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, Mentz RJ, O'Brien E, Correa A, Suthahar N, et al. 10-Year Risk Equations for Incident Heart Failure in the General Population. J Am Coll Cardiol 2019;73:2388–2397. [PubMed: 31097157]

20. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, Basit M, Kannan V, Grodin JL, Everett B, et al. Machine Learning to Predict the Risk of Incident Heart Failure Hospitalization Among Patients With Diabetes: The WATCH-DM Risk Score. Diabetes Care 2019;42:2298–2306. [PubMed: 31519694]

21. ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 1989;129:687–702. [PubMed: 2646917]

22. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871–81. [PubMed: 12397006]

23. Keku E, Rosamond W, Taylor HA Jr, Garrison R, Wyatt SB, Richard M, Jenkins B, Reeves L and Sarpong D. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. Ethn Dis 2005;15:S6-62–70.

24. Taylor HA Jr The Jackson Heart Study: an overview. Ethn Dis 2005;15:S6-1–3.

25. Victor RG, Haley RW, Willett DL, Peshock RM, Vaeth PC, Leonard D, Basit M, Cooper RS, Iannacchione VG, Visscher WA, et al. The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. Am J Cardiol 2004;93:1473–80. [PubMed: 15194016]

26. Collins GS, Reitsma JB, Altman DG and Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63. [PubMed: 25560714]

27. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S and Groupdagger P. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019;170:51–58. [PubMed: 30596875]

28. Pandey A, Keshvani N, Ayers C, Correa A, Drazner MH, Lewis A, Rodriguez CJ, Hall ME, Fox ER, Mentz RJ, et al. Association of Cardiac Injury and Malignant Left Ventricular Hypertrophy With Risk of Heart Failure in African Americans: The Jackson Heart Study. JAMA Cardiol 2019;4:51–58. [PubMed: 30566191]

29. de Lemos JA, Ayers CR, Levine BD, deFilippi CR, Wang TJ, Hundley WG, Berry JD, Seliger SL, McGuire DK, Ouyang P, et al. Multimodality Strategy for Cardiovascular Risk Assessment: Performance in 2 Population-Based Cohorts. Circulation 2017;135:2119–2132. [PubMed: 28360032]

30. Peters MN, Seliger SL, Christenson RH, Hong-Zohlman SN, Daniels LB, Lima JAC, de Lemos JA, Neeland IJ and deFilippi CR. "Malignant" Left Ventricular Hypertrophy Identifies Subjects at High Risk for Progression to Asymptomatic Left Ventricular Dysfunction, Heart Failure, and Death: MESA (Multi-Ethnic Study of Atherosclerosis). J Am Heart Assoc 2018;7:e006619. [PubMed: 29437599]

31. McEvoy JW, Chen Y, Ndumele CE, Solomon SD, Nambi V, Ballantyne CM, Blumenthal RS, Coresh J and Selvin E. Six-Year Change in High-Sensitivity Cardiac Troponin T and Risk of Subsequent Coronary Heart Disease, Heart Failure, and Death. JAMA Cardiol 2016;1:519–28. [PubMed: 27439107]

32. Hussain A, Sun W, Deswal A, de Lemos JA, McEvoy JW, Hoogeveen RC, Matsushita K, Aguilar D, Bozkurt B, Virani SS, et al. Association of NT-ProBNP, Blood Pressure, and Cardiovascular Events: The ARIC Study. J Am Coll Cardiol 2021;77:559–571. [PubMed: 33538254]

33. Fox ER, Musani SK, Bidulescu A, Nagarajarao HS, Samdarshi TE, Gebreab SY, Sung JH, Steffes MW, Wang TJ, Taylor HA, et al. Relation of obesity to circulating B-type natriuretic peptide concentrations in blacks: the Jackson Heart Study. Circulation 2011;124:1021–7. [PubMed: 21824924]

34. Anand IS, Claggett B, Liu J, Shah AM, Rector TS, Shah SJ, Desai AS, O'Meara E, Fleg JL, Pfeffer MA, et al. Interaction Between Spironolactone and Natriuretic Peptides in Patients With Heart Failure and Preserved Ejection Fraction: From the TOPCAT Trial. JACC Heart Fail 2017;5:241–252. [PubMed: 28359411]

35. Stekhoven DJ and Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28:112–8. [PubMed: 22039212]

36. Jaeger BC, Long DL, Long DM, Sims M, Szychowski JM, Min Y-I, McClure LA, Howard G and Simon N. Oblique random survival forests. Ann Appl Stat 2019;13:1847–1883. [PubMed: 36704751]

37. Anderson JE, Lassiter R, Bickler SW, Talamini MA and Chang DC. Brief tool to measure risk-adjusted surgical outcomes in resource-limited hospitals. Arch Surg 2012;147:798–803. [PubMed: 22987164]

38. Harrell FE Jr., Califf RM, Pryor DB, Lee KL and Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543–6. [PubMed: 7069920]

39. DeLong ER, DeLong DM and Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45. [PubMed: 3203132]

40. Rahman MS, Ambler G, Choodari-Oskooei B and Omar RZ. Review and evaluation of performance measures for survival prediction models in external validation settings. BMC Med Res Methodol 2017;17:60. [PubMed: 28420338]

41. Vickers AJ and Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74. [PubMed: 17099194]

42. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, Gomes AS, Folsom AR, Shea S, Guallar E, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. Circ Res 2017;121:1092–1101. [PubMed: 28794054]

43. Kaufman JS and Cooper RS. Race in epidemiology: new tools, old problems. Ann Epidemiol 2008;18:119–23. [PubMed: 18083541]

44. Youmans QR, Hastings-Spaine L, Princewill O, Shobayo T and Okwuosa IS. Disparities in cardiovascular care: Past, present, and solutions. Cleve Clin J Med 2019;86:621–632. [PubMed: 31498768]

45. Vyas DA, Eisenstein LG and Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. The New England journal of medicine 2020;383:874–882. [PubMed: 32853499]

46. Ebong IA, Goff DC Jr., Rodriguez CJ, Chen H, Bluemke DA, Szklo M and Bertoni AG. The relationship between measures of obesity and incident heart failure: the multi-ethnic study of atherosclerosis. Obesity (Silver Spring) 2013;21:1915–22. [PubMed: 23441088]

47. Loehr LR, Rosamond WD, Poole C, McNeill AM, Chang PP, Folsom AR, Chambless LE and Heiss G. Association of multiple anthropometrics of overweight and obesity with incident heart failure: the Atherosclerosis Risk in Communities study. Circ Heart Fail 2009;2:18–24. [PubMed: 19808311]

48. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr., Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2017;136:e137–e161. [PubMed: 28455343]

49. Pandey A, Vaduganathan M, Patel KV, Ayers C, Ballantyne CM, Kosiborod MN, Carnethon M, DeFilippi C, McGuire DK, Khan SS, et al. Biomarker-Based Risk Prediction of Incident Heart Failure in Pre-Diabetes and Diabetes. JACC Heart Fail 2021;9:215–223. [PubMed: 33422434]

50. Pandey A, Patel KV, Vongpatanasin W, Ayers C, Berry JD, Mentz RJ, Blaha MJ, McEvoy JW, Muntner P, Vaduganathan M, et al. Incorporation of Biomarkers Into Risk Assessment for Allocation of Antihypertensive Medication According to the 2017 ACC/AHA High Blood Pressure Guideline: A Pooled Cohort Analysis. Circulation 2019;140:2076–2088. [PubMed: 31707797]

51. Pandey A, Patel KV, Bahnson JL, Gaussoin SA, Martin CK, Balasubramanyam A, Johnson KC, McGuire DK, Bertoni AG, Kitzman D, et al. Association of Intensive Lifestyle Intervention, Fitness, and Body Mass Index With Risk of Heart Failure in Overweight or Obese Adults With Type 2 Diabetes Mellitus: An Analysis From the Look AHEAD Trial. Circulation 2020;141:1295–1306. [PubMed: 32134326]

52. Patel KV, Bahnson JL, Gaussoin SA, Johnson KC, Pi-Sunyer X, White U, Olson KL, Bertoni AG, Kitzman DW, Berry JD, et al. Association of Baseline and Longitudinal Changes in Body Composition Measures With Risk of Heart Failure and Myocardial Infarction in Type 2 Diabetes: Findings From the Look AHEAD Trial. Circulation 2020;142:2420–2430. [PubMed: 33164570]

53. Sprint Research Group Wright JT Jr., Williamson JD Whelton PK, Snyder JK Sink KM, Rocco MV Reboussin DM, Rahman M, Oparil S, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. The New England journal of medicine 2015;373:2103–16. [PubMed: 26551272]

54. Zelniker TA, Wiviott SD, Raz I, Im K, Goodrich EL, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Furtado RHM, et al. SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. Lancet 2019;393:31–39. [PubMed: 30424892]

55. Pandey A, Park BD, Ayers C, Das SR, Lakoski S, Matulevicius S, de Lemos JA and Berry JD. Determinants of Racial/Ethnic Differences in Cardiorespiratory Fitness (from the Dallas Heart Study). Am J Cardiol 2016;118:499–503. [PubMed: 27349903]

56. Pandey A, Kondamudi N, Patel KV, Ayers C, Simek S, Hall ME, Musani SK, Blackshear C, Mentz RJ, Khan H, et al. Association Between Regional Adipose Tissue Distribution and Risk of Heart Failure Among Blacks. Circ Heart Fail 2018;11:e005629. [PubMed: 30571193]

57. Rosamond WD, Chang PP, Baggett C, Johnson A, Bertoni AG, Shahar E, Deswal A, Heiss G and Chambless LE. Classification of heart failure in the atherosclerosis risk in communities (ARIC) study: a comparison of diagnostic criteria. Circ Heart Fail 2012;5:152–9. [PubMed: 22271752]

58. de Lemos JA, Ayers CR, Levine BD, deFilippi CR, Wang TJ, Hundley WG, Berry JD, Seliger SL, McGuire DK, Ouyang P, et al. Multimodality Strategy for Cardiovascular Risk Assessment: Performance in 2 Population-Based Cohorts. Circulation 2017;135:2119–2132. [PubMed: 28360032]

59. Ishwaran H, Kogalur U, Blackstone E and MS L. Random survival forests. Ann Appl Stat 2008;2:841–860.

60. Tibshirani R Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological) 1996;58:267–288.

61. Simon N, Friedman J, Hastie T and Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw 2011;39:1–13.

62. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Statist 2001;29:1189–1232.

63. Binder H CoxBoost: Cox models by likelihood based boosting for a single survival endpoint or competing risks 2020; R package version 1.4. https://CRAN.R-project.org/package=CoxBoost

64. Friedman J, Hastie T and Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1–22. [PubMed: 20808728]

65. Jaeger B obliqueRSF: Oblique Random Forests for Right-Censored Time-to-Event Data 2018; R package version 0.1.1. https://CRAN.R-project.org/package=obliqueRSF

66. Chen T and Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery 2016;786–794. DOI: 10.1145/2939672.2939785.

CLINICAL PERSPECTIVE

What is new?

- Race-specific, ML-based HF risk models that integrate clinical, laboratory, and biomarker data demonstrate superior performance compared with traditional HF risk and non-race specific ML models.

- Among the risk factors, natriuretic peptide levels were the most important predictor of HF risk across both races, followed by troponin levels, glycemic parameters, and socioeconomic factors in Black adults.

- The EKG-based measure of Cornell voltage and left ventricular hypertrophy, prevalent cardiovascular disease, and traditional cardiovascular risk factors were stronger predictors of HF risk in White adults.

What are the clinical implications?

- The race-specific machine learning models for heart failure risk prediction can identify high-risk individuals who are most likely to benefit from effective approaches to heart failure prevention.

- Implementation and integration of race-specific ML-based models for HF risk prediction with electronic health record systems can improve the accessibility and real-world application of these tools.
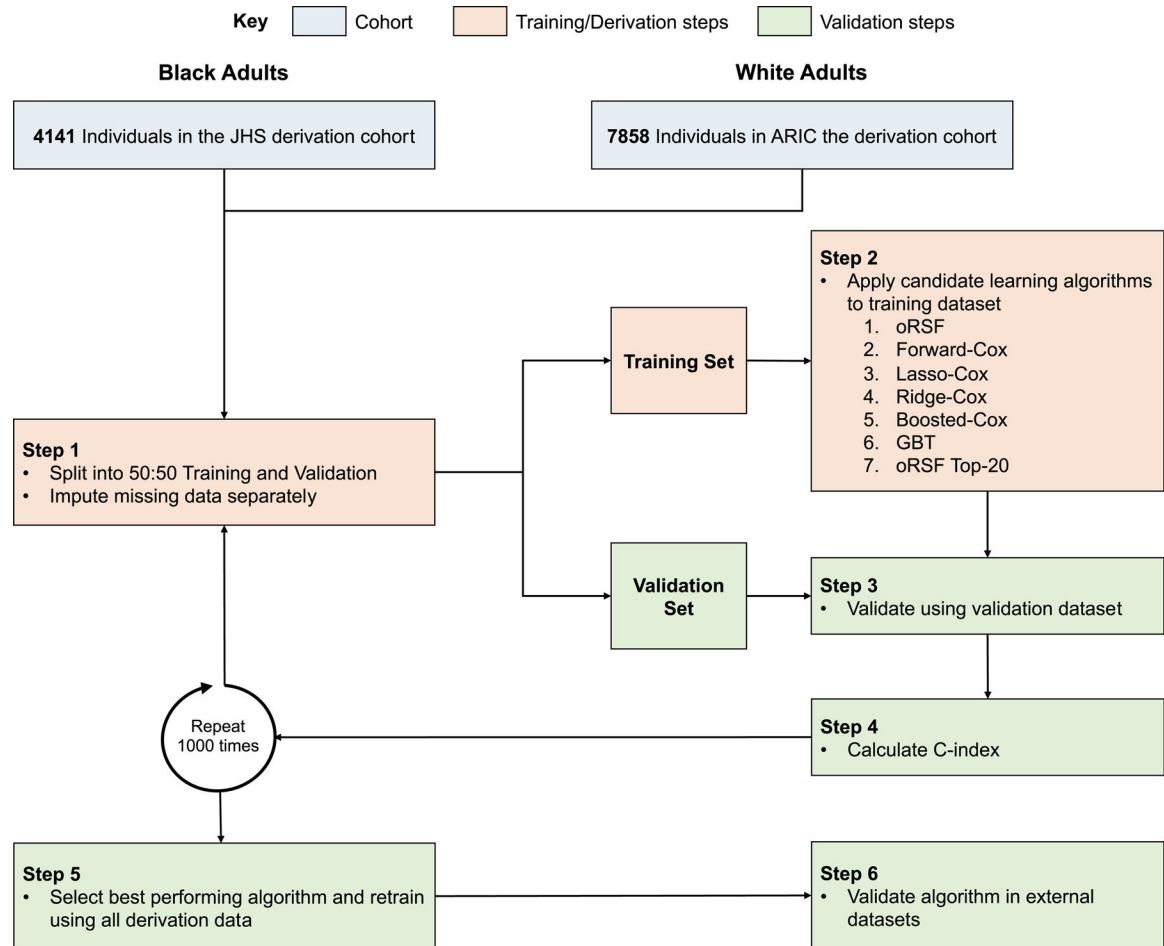
Figure 1. Analysis overview for identifying best performing risk prediction model
Abbreviations: ARIC, Atherosclerosis Risk in Communities Study; GBT, gradient boosted trees; HF, heart failure; JHS, Jackson Heart Study; oRSF, oblique random survival forest
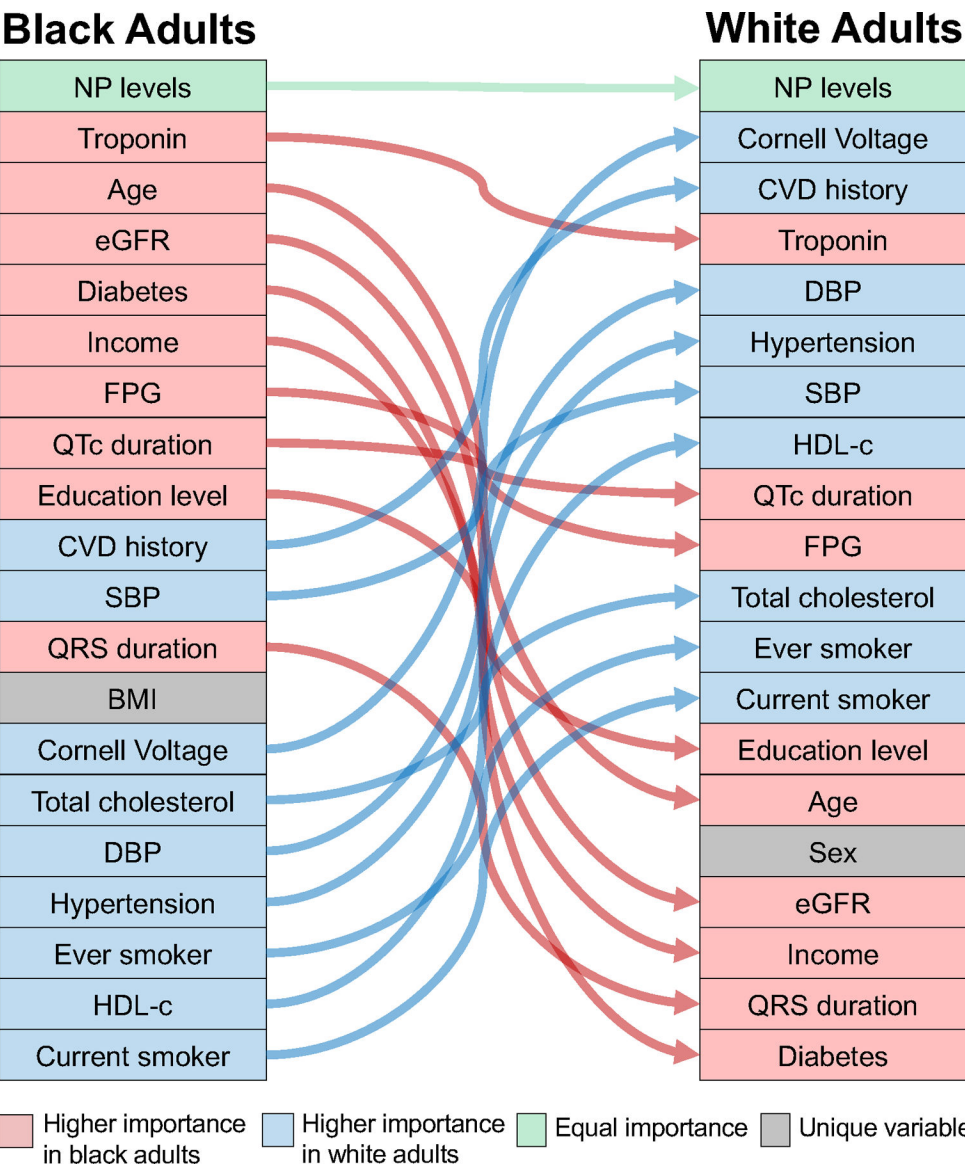
Figure 2.
Alluvial plot of the 20 most important parameters for HF risk prediction identified by the variable importance metric in the oblique random survival forest model in the derivation cohort.
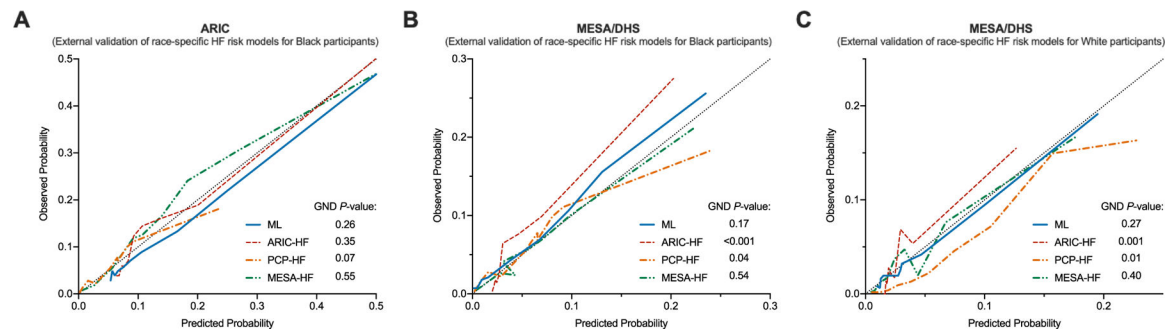
Figure 3.
Calibration of the models for predicting 10-year risk of developing HF in (Panel A) Black participants in the ARIC, (Panel B) Black participants in the pooled MESA/DHS, and (Panel C) White participants in the pooled MESA/DHS validation cohorts.
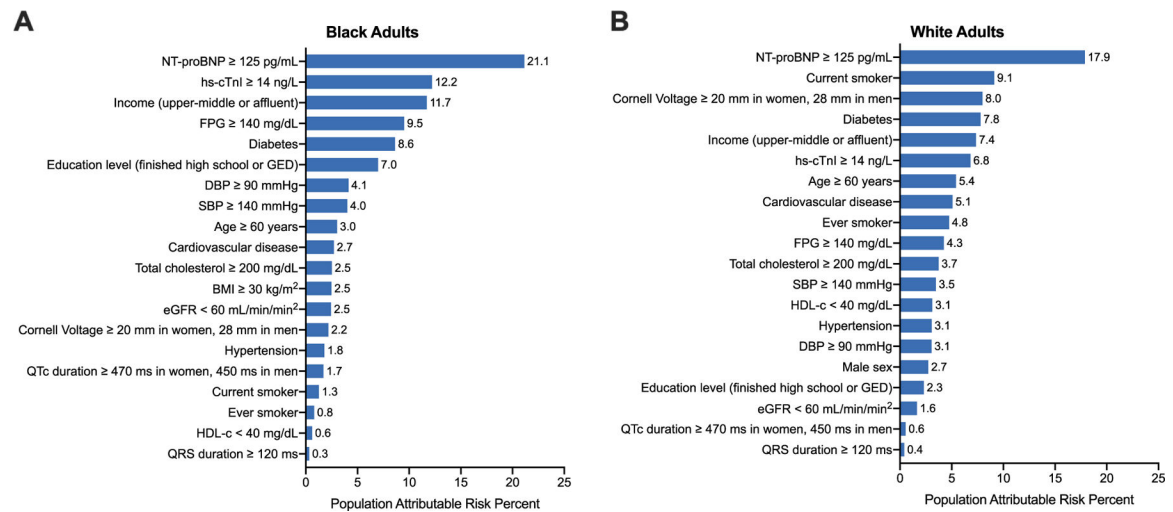
Figure 4. Results of the Monte Carlo simulation to calculate the population attributable risk percent of each covariate among (Panel A) Black and (Panel B) White adults.
Simulation values were obtained based on patient characteristics from Black and White adults from the ARIC cohort. The referent group was the group opposite the listed cutoff. Abbreviations: BMI, body mass index; CVD, cardiovascular disease; FPG, fasting plasma glucose; HDL-c, high density lipoprotein cholesterol; NP, natriuretic peptide

Table 1.

Baseline characteristics of Black and White participants stratified by derivation and validation cohorts.

	Black adults			White adults	
	Derivation cohort	*Validation cohorts*		*Derivation cohort*	*Validation cohort*
	JHS (n = 4141)	**ARIC (n = 1024)**	**MESA/DHS (n = 2821)**	**ARIC (n = 7858)**	**MESA/DHS (n = 3236)**
Age, years	58.1 (10.5)	62.4 (5.9)	58.3 (10.4)	63.0 (5.6)	60.2 (10.8)
Male	1468 (35.5)	412 (40.2)	1219 (43.2)	3633 (46.2)	1540 (47.6)
Heart rate, bpm	64.2 (10.5)	65.3 (11.5)	67.5 (12.2)	61.8 (9.7)	64.9 (10.5)
SBP, mmHg	128.6 (16.7)	135.3 (20.4)	132.8 (21.2)	125.6 (18.1)	123.5 (19.3)
BMI, kg/m^2	31.6 (6.9)	30.2 (6.6)	30.4 (6.3)	28.1 (5.1)	27.9 (5.4)
Hypertension	2498 (60.3)	604 (59.4)	1277 (45.9)	2416 (30.9)	1001 (31.2)
Diabetes	929 (22.7)	301 (30.7)	492 (17.5)	996 (12.7)	198 (6.1)
Cardiovascular disease	397 (9.6)	65 (6.6)	132 (4.7)	542 (6.9)	34 (1.1)
Current smoker	506 (12.3)	212 (21.4)	652 (23.2)	1093 (13.9)	448 (13.9)
Ever smoker	1374 (33.2)	555 (55.9)	1502 (53.5)	4655 (59.3)	1761 (54.6)
Education level					
Less than HS	880 (21.3)	366 (35.8)	399 (14.2)	1080 (13.8)	167 (5.2)
HS graduate/GED	750 (18.2)	246 (24.1)	871 (31.0)	2873 (36.6)	742 (23.0)
College/Vocational	2492 (60.5)	409 (40.1)	1539 (54.8)	3897 (49.6)	2318 (71.8)
Income					
< $25,000	1308 (31.6)	568 (55.5)	872 (35.4)	1751 (22.3)	485 (15.6)
$25,000 - $49,999	962 (23.2)	216 (21.1)	802 (32.6)	2948 (37.5)	852 (27.4)
$50,000	1871 (45.2)	240 (23.4)	789 (32.0)	3159 (40.2)	1773 (57.0)
eGFR, mL/min/1.73m^2	92.1 (20.2)	91.6 (21.1)	90.7 (21.5)	84.9 (14.0)	78.3 (7.5)
FPG, mg/dL	100.6 (32.3)	114.5 (46.2)	103.5 (39.1)	106.5 (29.3)	92.6 (22.3)
HDL-c, mg/dL	52.4 (14.8)	53.7 (18.2)	52.4 (15.2)	49.3 (16.3)	51.8 (15.8)
Total cholesterol, mg/dL	201.7 (39.8)	199.4 (37.7)	187.4 (38.2)	201.6 (36.2)	194.2 (35.8)
BNP, pg/mL [*]	16.7 (2.7–19.9)	-	-	-	-
NT-proBNP, pg/mL [*]	-	49.1 (20.7–107.5)	43.5 (19.2–94.1)	69.5 (34.9–131.8)	99.1 (25.3–110.2)
Troponin [*,†]	6.0 (2.2–4.7)	6.1 (2.2–7.8)	6.9 (3.0–7.6)	6.5 (2.7–8.6)	5.7 (3–7)
QRS duration, ms	93.9 (14.2)	93.2 (15.5)	92.5 (12.2)	93.5 (14.1)	92.5 (12.2)
QTc duration, ms	426.8 (26.8)	423.2 (25.7)	427.1 (22.7)	419.9 (20.1)	427.1 (22.7)
Cornell Voltage, mV	1502.4 (647.8)	1537.8 (656.2)	1193.4 (501.0)	1216.0 (506.1)	1193.4 (501.0)

Values are prior to imputation and displayed as mean (standard deviation) or n (%).

*Displayed as mean (quartile 1 – quartile 3)

[†]Troponin was recorded as high-sensitivity cardiac troponin-I in JHS and ARIC and high-sensitivity troponin-T MESA and DHS (units are ng/L).

Abbreviations: ARIC, Atherosclerosis Risk in Communities Study; BMI, body mass index; BNP, B-type natriuretic peptide; DHS, Dallas Heart Study; FPG, fasting plasma glucose; HDL-c, high density lipoprotein cholesterol; HS, high school; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; NT-proBNP, N-terminal pro-B type natriuretic peptide; SBP, systolic blood pressure.

Table 2.

Discrimination of the models for predicting 10-year risk of developing HF in the derivation cohort among Black and White adults.

| | Race Specific | | | | | | | Not Race Specific |
	oRSF	oRSF (Top-20)	Forward Cox	Lasso Cox	Ridge Cox	Boosted Cox	GBT	oRSF
Black adults	0.88 (0.86–0.91)	0.88 (0.85–0.90)	0.76 (0.71–0.80)	0.80 (0.75–0.84)	0.81 (0.50–0.85)	0.81 (0.74–0.87)	0.86 (0.84–0.89)	0.81 (0.78–0.83)
White adults	0.89 (0.86–0.91)	0.88 (0.85–0.90)	0.77 (0.73–0.80)	0.78 (0.55–0.84)	0.78 (0.50–0.82)	0.82 (0.71–0.87)	0.87 (0.85–0.90)	0.80 (0.76–0.85)

Data presented are C-index (95% confidence interval).

The following models were analyzed: race specific oblique random survival forests specific (oRSF), oRSF with the top 20 most important variables (oRSF top-20), forward stepwise Cox regression (forward Cox), Lasso Cox regression (Lasso Cox), Ridge cox regression (ridge Cox), boosted Cox regression (boosted Cox), gradient boosted trees (GBT), and oRSF with race as a covariate (oRSF not race specific). Confidence intervals are the 95% ranges among 1000 bootstrapped replicates.

Table 3.

Discrimination of the reference ML model, clinical variables, NP only, troponin only, and combined biomarkers model for predicting 10-year risk of developing HF in the derivation cohort among Black and White adults.

	Top-20 (reference)	Clinical variables	Clinical + laboratory variables	NP only	Troponin only	Complete data
Black adults	0.88 (0.85–0.90)	0.75 (0.73–0.78)	0.75 (0.73–0.78)	0.70 (0.66–0.74)	0.70 (0.66–0.72)	0.88 (0.85–0.90)
White adults	0.88 (0.85–0.90)	0.76 (0.71–0.80)	0.76 (0.71–0.81)	0.64 (0.61–0.66)	0.69 (0.66–0.71)	0.88 (0.85–0.90)

Data presented are C-index (95% confidence interval).

The following models were analyzed: race specific oblique random survival forests (oRSF) with the top 20 most important variables (oRSF top-20) [reference], oRSF with only clinical variables (age, sex, systolic blood pressure, diastolic blood pressure, heart rate, body mass index, smoking history, history cardiovascular disease, history of diabetes); oRSF with clinical and laboratory variables (clinical variables + hemoglobin A1c, total cholesterol, high-density lipoprotein-cholesterol), oRSF with natriuretic peptide (NP) levels only, oRSF with troponin levels only, and oRSF with complete data (i.e, no missing data). Of the 11,999 participants in the derivation cohort, 10,053 (83.8%) had no missing data. Confidence intervals are the 95% ranges among 1000 bootstrapped replicates.

Table 4.

Discrimination and calibration of the models for predicting 10-year risk of developing HF in the validation cohorts among Black and White adults.

	ARIC			MESA/DHS		
	C-index (95% CI)	GND chi-square (p-value)	DeLong's test (p-value)[*]	C-index (95% CI)	GND chi-square (p-value)	DeLong's test (p-value)[*]
Black Adults						
ML risk score	0.80 (0.75–0.84)	10.1 (0.26)	Ref.	0.83 (0.77–0.87)	11.7 (0.17)	Ref.
ARIC-HF risk score	0.77 (0.73–0.80)	8.9 (0.35)	<0.001	0.80 (0.76–0.84)	29.8 (<0.001)	0.01
PCP-HF risk score	0.73 (0.69–0.77)	14.4 (0.07)	<0.001	0.75 (0.71–0.79)	16.1 (0.04)	<0.001
MESA-HF risk score	0.72 (0.67–0.75)	6.9 (0.55)	<0.001	0.78 (0.74–0.82)	6.0 (0.54)	0.006
White Adults						
ML risk score				0.82 (0.78–0.86)	9.9 (0.27)	Ref.
ARIC-HF risk score		-		0.79 (0.76–0.81)	25.5 (0.001)	0.008
PCP-HF risk score				0.75 (0.71–0.79)	19.1 (0.01)	<0.001
MESA-HF risk score				0.80 (0.76–0.83)	8.33 (0.40)	0.044

[*] DeLong's test of C-index compared to ML risk score model