

As a library, NLM provides access to scientific literature. Inclusion in an NLM database does not imply endorsement of, or agreement with, the contents by NLM or the National Institutes of Health.

Learn more: [PMC Disclaimer](#) | [PMC Copyright Notice](#)



JAMA Cardiol. 2022 Jul 6;7(8):844–854. doi: [10.1001/jamacardio.2022.1900](https://doi.org/10.1001/jamacardio.2022.1900)

Machine Learning–Based Models Incorporating Social Determinants of Health vs Traditional Models for Predicting In-Hospital Mortality in Patients With Heart Failure

[Matthew W Segar](#)¹, [Jennifer L Hall](#)², [Pardeep S Jhund](#)³, [Tiffany M Powell-Wiley](#)^{4,5}, [Alanna A Morris](#)⁶, [David Kao](#)⁷, [Gregg C Fonarow](#)^{8,9}, [Rosalba Hernandez](#)¹⁰, [Nasrien E Ibrahim](#)¹¹, [Christine Rutan](#)¹², [Ann Marie Navar](#)^{13,14}, [Laura M Stevens](#)^{2,7}, [Ambarish Pandey](#)¹³,

[Author information](#) [Article notes](#) [Copyright and License information](#)

PMCID: PMC9260645 PMID: [35793094](https://pubmed.ncbi.nlm.nih.gov/35793094/)

See commentary "[Replacing Race With Social Determinants of Health in Risk Prediction—Getting It Right](#)" on page 856.

See commentary "[Machine Learning and Social Determinants of Health-An Opportunity to Move Beyond Race for Inpatient Risk Prediction in Patients With Heart Failure.](#)" with doi:

[10.1001/jamacardio.2022.1924](https://doi.org/10.1001/jamacardio.2022.1924).

Key Points

Question

Do machine learning (ML)–based models that incorporate social determinants of health (SDOH) improve the prediction of in-hospital mortality among patients with heart failure (HF)?

Findings

In this cohort study, ML models developed in the Get With The Guidelines–Heart Failure (GWTG-HF) registry using race-specific and race-agnostic approaches were associated with an improvement in the prediction of in-hospital mortality after hospitalization for HF compared with the existing and rederived logistic regression models. The addition of SDOH was associated with an improvement in the performance and prognostic utility of the ML models in Black patients but not in non-Black patients.

Meaning

The findings indicate that ML models incorporating SDOH may improve risk prediction of in-hospital mortality after hospitalization for HF, particularly in Black adults.

This cohort study develops and validates a machine learning–based model incorporating social determinants of health (SDOH) for predicting heart failure mortality.

Abstract

Importance

Traditional models for predicting in-hospital mortality for patients with heart failure (HF) have used logistic regression and do not account for social determinants of health (SDOH).

Objective

To develop and validate novel machine learning (ML) models for HF mortality that incorporate SDOH.

Design, Setting, and Participants

This retrospective study used the data from the Get With The Guidelines–Heart Failure (GWTG-HF) registry to identify HF hospitalizations between January 1, 2010, and December 31, 2020. The study included patients with acute decompensated HF who were hospitalized at the GWTG-HF participating centers during the study period. Data analysis was performed January 6, 2021, to April 26, 2022. External validation was performed in the hospitalization cohort from the Atherosclerosis Risk in Communities (ARIC) study between 2005 and 2014.

Main Outcomes and Measures

Random forest-based ML approaches were used to develop race-specific and race-agnostic models for predicting in-hospital mortality. Performance was assessed using C index (discrimination), regression slopes for observed vs predicted mortality rates (calibration), and decision curves for prognostic utility.

Results

The training data set included 123 634 hospitalized patients with HF who were enrolled in the GWTG-HF registry (mean [SD] age, 71 [13] years; 58 356 [47.2%] female individuals; 65 278 [52.8%] male individuals. Patients were analyzed in 2 categories: Black (23 453 [19.0%]) and non-Black (2121 [2.1%] Asian; 91 154 [91.0%] White, and 6906 [6.9%] other race and ethnicity). The ML models demonstrated excellent performance in the internal testing subset ($n = 82\,420$) (C statistic, 0.81 for Black patients and 0.82 for non-Black patients) and in the real-world-like cohort with less than 50% missingness on covariates ($n = 553\,506$; C statistic, 0.74 for Black patients and 0.75 for non-Black patients). In the external validation cohort (ARIC registry; $n = 1205$ Black patients and 2264 non-Black patients), ML models demonstrated high discrimination and adequate calibration (C statistic, 0.79 and 0.80, respectively). Furthermore, the performance of the ML models was superior to the traditional GWTG-HF risk score model (C index, 0.69 for both race groups) and other rederived logistic regression models using race as a covariate. The performance of the ML models was identical using the race-specific and race-agnostic approaches in the GWTG-HF and external validation cohorts. In the GWTG-HF cohort, the addition of zip code-level SDOH parameters to the ML model with clinical covariates only was associated with better discrimination, prognostic utility (assessed using decision curves), and model reclassification metrics in Black patients (net reclassification improvement, 0.22 [95% CI, 0.14-0.30]; $P < .001$) but not in non-Black patients.

Conclusions and Relevance

ML models for HF mortality demonstrated superior performance to the traditional and rederived logistic regressions models using race as a covariate. The addition of SDOH parameters improved the prognostic utility of prediction models in Black patients but not non-Black patients in the GWTG-HF registry.

Introduction

Heart failure (HF) hospitalization confers a high mortality risk, with in-hospital mortality rates approaching 5%.¹ In-hospital mortality rates vary substantially by race and ethnicity, and there is a growing need to develop risk-prediction tools to better identify high-risk individuals across races and ethnicities.^{2,3}

Multiple clinical risk prediction tools are available to estimate in-hospital mortality risk among individuals hospitalized with HF, including Get With The Guidelines–Heart Failure (GWTG-HF), Acute Decompensated Heart Failure National Registry (ADHERE), and Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF) risk scores.^{4,5,6} However, most commonly implemented tools for predicting mortality risk incorporate race as a covariate, assigning a lower risk to Black individuals compared with individuals of other races. Concerns have been raised about this race-based approach that assigns lower risk to Black patients and thus potentially raises the threshold required for risk-based allocation of clinical therapies and adds to the existing disparities in HF care.⁷ Moreover, including race solely as a covariate in such risk models may not completely capture the societal factors contributing to racial disparities in outcomes among patients with HF. Thus, novel approaches to risk prediction are needed that do not use race as a biological risk factor and better account for social determinants of health (SDOH) in the risk assessment.

Race-specific risk prediction is one such approach used previously to predict the risk of atherosclerotic cardiovascular disease and incident HF in community-dwelling individuals.^{8,9,10} Race-specific risk-prediction models acknowledge that outcomes are different between races and look at risk gradients within each race strata, thus allowing for better capture of unique race-specific risk predictors. The race-agnostic approach to predicting risk is another strategy that does not consider race as a covariate while developing the risk model in the overall cohort. Race-agnostic approaches have been recently evaluated to assess biological parameters, such as kidney function.^{11,12} In this study, we developed and evaluated race-specific and race-agnostic models incorporating clinical and SDOH parameters to predict in-hospital mortality risk among patients hospitalized with HF. We hypothesized that improved and more equitable risk prediction can be achieved when risk is assessed without race as a biological covariate and accounts for SDOH. Consistent with our prior approaches, we used the machine learning (ML)–based random forest technique to develop the race-specific risk-prediction models.^{10,13}

Methods

Study Population for Model Development

The present study used data from the American Heart Association (AHA) GWTG-HF registry. Details of the GWTG-HF program have been reported previously^{14,15} and are summarized in the eMethods and eFigure 1 in the [Supplement](#). For the present analysis, 677 140 patients from 634 hospitals between January 1, 2010, and December 31, 2020, were considered for model development and validation. A total of 206 054 participants had less than 15% missing data on relevant covariates, 123 634 (60%) of whom were included in the model training and 82 420 (40%) for internal testing subsets. An additional cohort of 471 086 participants with less than 50% missingness was added to the internal validation cohort to test the performance of the derived model in a real-world–like data set where patients often have multiple missing model covariates (n = 553 506). Participating GWTG-HF centers obtained institutional review board approval and are granted a waiver for informed consent under the common rule. IQVIA (Parsippany, New Jersey) served as the data collection and coordination center. The American Heart Association Precision Medicine Platform was used for data analysis.

Candidate Variables

Recorded data in the GWTG-HF registry encompass a range of domains, including patient demographics, vital signs, socioeconomic status, medical history, laboratory values, cardiac biomarkers, and electrocardiography and ejection fraction. Details about the candidate variables used for the risk-prediction model developed are provided in the eMethods and eTable 1 in the [Supplement](#).

Zip Code– and Hospital-Level SDOH

Among 123 634 participants in the derivation cohort, 64 573 (52.2%) had an admission year of 2015 or later and recorded residential zip codes available to link with publicly available zip code–level measures of SDOH detailed in eTable 2 and the eMethods in the [Supplement](#). All zip code–level data on SDOH were for the patients' residence. Additionally, hospital-level measures of geography, sole community hospital, essential hospital membership, and disproportionate share hospital metrics were included as described previously^{16,17} and in the eMethods in the [Supplement](#). Race was self-reported in questionnaires with standardized answer choices: Asian, Black, White, and other.

Study Outcome

Our primary outcome of interest was in-hospital mortality. Mortality events were captured as documented on the case report form for participants in GWTG-HF.

External Validation Cohort

The performance of derived risk models was assessed in an external validation cohort of participants from the Atherosclerosis Risk in Communities (ARIC) study obtained from the National Heart, Lung, and Blood Institute BioLINCC data repository. Details of the ARIC study have been previously reported^{18,19} and are described in the eMethods in the [Supplement](#). Among 3612 candidate hospitalizations, 3469 patients with HF (1205 among Black patients and 2264 among non-Black patients) were included in the final cohort after excluding participants who were discharged to hospice (n = 69), left against medical advice (n = 4), or were discharged on comfort care (n = 70).

Statistical Analysis

Model Development and Metrics for Performance Assessment

Race-specific and race-agnostic models were developed using random forest machine learning (ML) techniques described previously and detailed in the eMethods in the [Supplement](#). The race-specific models were developed separately for Black participants and non-Black participants (subsequently referred to as the race-specific ML model). Variable selection was performed independently for Black participants and non-Black participants in the training data set of the GWTG-HF registry as described in the eMethods in the [Supplement](#). The race-agnostic model was developed in the entire training cohort, excluding race as a candidate covariate from the variable selection. Multiple metrics were used to assess model performance in the testing data sets of the GWTG-HF registry and the external validation cohort. Discrimination was evaluated using the C index with 95% CIs determined using bootstrapping with 2000 replicates.²⁰ Differences in C indices were compared across different models using the DeLong test.²¹ Consistent with the recent literature on risk prediction,^{22,23} calibration was assessed using the Brier score, representing the mean squared error between the observed and predicted risk and calibration slopes.^{24,25} Additionally, a regression slope of the observed mortality rates was calculated across deciles of predicted mortality rates. A lower Brier score indicates calibration intercept closer to 0, and calibration slope closer to 1 indicates better performance. Observed and predicted risks across deciles of predicted risk were also reported in the validation cohorts as additional measures of model calibration. Reclassification was reported using categorical net reclassification improvement (at race-specific event rate risk threshold) and integrated discrimination index.^{26,27} Decision curve analysis, a measure of the true-positive cases identified without an increase in the false-positive rate, was used to assess the clinical net benefit with the model across thresholds of risk.²⁸

Model Performance in the Internal Validation Cohort (GWTG-HF)

The generalizability of the ML models was assessed in cohorts of participants with less than 15% missing data and less than 50% missing data in model covariates. Subgroup analyses were performed to evaluate the performance of the ML models in age-based (≤ 70 years or > 70 years), sex-based (male and female), race-based (Asian, White, and other), ethnicity-based (Hispanic and non-Hispanic), ejection fraction–based (HF with reduced and preserved ejection fraction, using the 50% ejection fraction cutoff), and socioeconomic status–based (median income of $\geq \$54\,471$ vs $< \$54\,471$) subgroups.

Model Performance in the External Validation Cohort (ARIC Cohort)

We compared the performance of the race-specific and race-agnostic ML models with the models that used race as a covariate. This included the original GWTG-HF risk score and a logistic regression model with race as a covariate that was rederived in the GWTG-HF registry data used in the present study. We also compared the performance of the ML models vs a race-specific logistic regression model. Details of the logistic regression model are described in the eMethods in the [Supplement](#). Sensitivity analyses were also performed to evaluate the performance of an additional ML model with race as a covariate. The importance of race in risk prediction using the ML model with race as a covariate was assessed using the minimum depth metric.

Model Performance With the Incorporation of SDOH

To evaluate whether incorporating SDOH might improve risk prediction of race-specific or race-agnostic ML models, the random forest ML model was rederived using an expanded pool of covariates that included patient-level clinical data, patient-level insurance status, and zip code–based SDOH parameters (65 covariates: 38 clinical and 27 SDOH) for patients admitted in 2015 and later and with available zip code–level data. Because participant zip codes were not available in the ARIC external validation cohort, the clinical and socioeconomic models were validated only in the GWTG-HF internal validation cohort with less than 50% missingness. Subgroup analyses were performed by disproportionate share hospital status. Finally, to determine the proportion of in-hospital mortality associated with specific clinical and socioeconomic risk factors across races, we used the Greenland-Drescher method for calculating population-attributable risk percentage as detailed in the eMethods in the [Supplement](#).²⁹ Analyses were performed using R version 4.0.2 (R Foundation) with a 2-tailed *P* value $< .05$ indicating significance.

Results

The training cohort consisted of 123 634 participants (mean [SD] age, 71 [13] years; 58 356 [47.2%] female individuals and 65 278 [52.8%] male individuals), of whom 2121 (2.1%) were Asian; 23 453 (19.0%), Black; 91 154 (91.0%), White; and 6906 (6.9%), other race and ethnicity.

[Table 1](#) shows the baseline characteristics of Black participants and non-Black participants in the GWTG-HF training data set. More Black patients were female; Black patients were also generally younger; had a higher prevalence of hypertension, obesity, and kidney dysfunction; and had higher levels of cardiac biomarkers, including natriuretic peptide levels and troponin ([Table 1](#)). Non-Black patients were more likely to have a history of coronary artery disease and diabetes at presentation. Black patients were more likely to lack health insurance coverage and had lower median zip code household income.

Table 1. Baseline Characteristics of Black Patients and Non-Black Patients in the Get With The Guidelines–Heart Failure Derivation Cohort.

Characteristic	No. (%)		P value
	Black (n = 23 453)	Non-Black (n = 100 181)	
Age, mean (SD), y	63 (15)	74 (14)	<.001
Female	11 158 (47.6)	47 198 (47.1)	.203
Male	12 295 (52.4)	52 983 (52.9)	
Race and ethnicity ^a			
Asian	NA	2121 (2.1)	NA
Hispanic	211 (0.9)	10 668 (10.6)	<.001
White	NA	91 154 (91.0)	NA
Other	NA	6906 (6.9)	NA
Clinical characteristics			
Systolic BP, mm Hg	148 (31)	141 (29)	<.001
BMI	32.4 (10.3)	30.1 (8.9)	<.001
Medicaid insurance	6854 (29.8)	11 624 (11.8)	<.001
Current smoking	6694 (28.5)	15 063 (15.0)	<.001
Hypertension	20 701 (88.3)	82 710 (82.6)	<.001
Coronary artery disease	8018 (34.2)	50 313 (50.2)	<.001
Diabetes	11 462 (48.9)	45 346 (45.3)	<.001
Sodium, mEq/L	138.9 (4.0)	137.8 (4.5)	<.001
Creatinine, mg/dL	2.0 (2.0)	1.6 (1.3)	<.001
Hemoglobin, g/dL	11.7 (2.2)	11.9 (2.2)	<.001
Plasma glucose, mg/dL	133 (46)	133 (49)	.68
Total cholesterol, mg/dL	143 (31)	138 (27)	<.001
HDLC, mg/dL	42 (10)	41 (9)	<.001
BNP, median (IQR), pg/mL	921 (399-1919)	793 (390-1556)	<.001
N-terminal pro-B-type natriuretic peptide, median (IQR), pg/mL	4670 (1946-11 456)	4818 (2148-10 965)	.119

Characteristic	No. (%)		P value
	Black (n = 23 453)	Non-Black (n = 100 181)	
Abnormal troponin	8350 (35.6)	29 545 (29.5)	<.001
QRS duration, ms	107 (28)	115 (33)	<.001
Social determinants of health characteristics (residential zip code–based) ^a			
Disproportionate share hospital	8764 (67.0)	25 801 (50.1)	<.001
Median household income, \$	47 526 (18 012)	61 026 (20 569)	<.001
Adults without high school degree	15.6 (7.6)	11.7 (8.1)	<.001
Poverty rate	22.1 (10.3)	13.8 (8.1)	<.001
Adults not employed	4.8 (2.2)	3.2 (1.6)	<.001
Vacancy rate	14.0 (7.6)	11.8 (9.4)	<.001
Born outside the US	10.3 (10.6)	10.1 (11.9)	.09
Distress score	69.7 (25.7)	48.2 (27.8)	<.001

[Open in a new tab](#)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); BNP, B-type natriuretic peptide; BP, blood pressure; HDLC, high-density lipoprotein cholesterol; NT-proBNP, N-terminal pro-B-type natriuretic peptide.

SI conversion factors: To convert BNP to nanograms per liter, multiply by 1; cholesterol to millimoles per liter, multiply by 0.0259; creatinine to micromoles per liter, multiply by 88.4; glucose to millimoles per liter, multiply by 0.0555; hemoglobin to grams per liter, multiply by 10; sodium to millimoles per liter, multiply by 1.

^a Patients reported race and ethnicity by choosing one of these options.

^b Social determinants of health characteristics are for 64 573 participants with available residential zip code and admission year 2015 or later.

Development and Performance of Race-Specific and Race-Agnostic ML Models in the GWTG-HF Cohort

The ranked variables for the race-specific and race-agnostic models according to variable importance are displayed in eFigure 2 in the [Supplement](#). No improvement in C index was observed with more than 20 covariates in an ML model (eFigure 3 in the [Supplement](#)). The testing subset included 15 634 Black patients and 66 786 non-Black patients (eTable 3 in the [Supplement](#)) with in-hospital event rates of 1.7% (n = 269) and 3.1% (n = 2082), respectively. Across participants in the Black and non-Black groups, the race-specific ML model demonstrated excellent discrimination performance (C index among Black patients, 0.81 [95% CI, 0.79-0.83] and non-Black patients, 0.82 [95% CI, 0.81-0.83]) and adequate calibration (eTable 4 and eFigure 4 in the [Supplement](#)). The performance of race-agnostic ML models among Black patients and non-Black patients was comparable with that observed for race-specific models (eTable 4 and eFigure 4 in the [Supplement](#)).

In the validation data set with up to 50% missingness in covariates (107 508 Black patients [19.4%] and 445 998 non-Black patients [80.6%]), the Black and non-Black race-specific ML models demonstrated high discrimination (C index, 0.74 [95% CI, 0.71-0.77] and 0.75 [95% CI, 0.73-0.78], respectively) and adequate calibration (Brier score, 16 and 31 $\times 10^{-3}$, respectively), comparable with that noted in the testing subset (eTable 4 and eFigure 5 in the [Supplement](#)). Similar results were also observed with the race-agnostic ML model (eTable 4 in the [Supplement](#)). In subgroup analysis, the race-specific and race-agnostic ML models also demonstrated good and comparable discrimination and calibration performance across age-based (≤ 70 years or > 70 years), sex-based, ethnicity-based, ejection fraction, and socioeconomic status–based subgroups (eTables 5 and 6 in the [Supplement](#)).

External Validation of Race-Specific and Race-Agnostic ML Models and Their Performance vs Models With Race as a Covariate

We externally validated the ML models in a cohort of participants with hospitalization for HF from the ARIC study (n = 3469; 1205 Black patients [34.7%] and 2264 White patients [65.3%]) (eTable 3 in the [Supplement](#)). All non-Black patients self-identified as White. In-hospital mortality rates were 2.0% (n = 24) and 3.1% (n = 70) for Black patients and White patients, respectively. Compared with the GWTG-HF cohort, more participants in the ARIC cohort were female; they were also generally older and had higher rates of cardiovascular disease risk factors and higher levels of abnormal cardiac biomarkers. Among Black patients, the race-specific ML models demonstrated superior discrimination (C index = 0.79 [95% CI, 0.77-0.81]) and calibration (Brier, 19 $\times 10^{-3}$) compared with the GWTG-HF risk (C index, 0.69 [95% CI, 0.67-0.71]; difference, 0.10 [95% CI, 0.07-0.13]), rederived logistic regression with race as a covariate (C index = 0.71 [95% CI, 0.69-0.72]; difference, 0.09 [95% CI, 0.06-0.11]), and race-specific logistic regression models (C index = 0.74 [95% CI, 0.72-0.76]; difference, 0.05 [95% CI, 0.02-0.08]) ([Table 2](#); [Figure 1](#); eFigure 6 in the [Supplement](#)). A similar pattern of results was

observed among non-Black patients, with consistently high and superior performance of the race-specific ML models (C index = 0.80 [95% CI, 0.79-0.81]) compared with other models using race as a covariate and the race-specific logistic regression model ([Table 2](#); [Figure 1](#)).

Table 2. Discrimination and Calibration Performance of Risk Prediction Models for Predicting In-Hospital Mortality Among Patients With Heart Failure^a.

Factor	Discrimination, C index (95% CI)	Calibration		
		Brier score (95% CI), ×10 ⁻⁵	Intercept	Slope
Black patients (n = 1205)				
Race-specific ML model	0.79 (0.77-0.81)	19 (11-28)	−0.09	0.95
Race-agnostic ML model	0.79 (0.77-0.81)	20 (11-29)	−0.13	0.94
ML model (race as a covariate)	0.79 (0.77-0.81)	19 (11-29)	−0.09	0.94
GWTG risk score ^{b}	0.69 (0.67-0.71)	30 (23-38)	−0.50	0.78
LR model (race as a covariate) ^{b}	0.71 (0.69-0.72)	29 (23-40)	−0.25	0.79
Race-specific LR model ^{b}	0.74 (0.72-0.76)	24 (18-33)	−0.14	0.88
Non-Black patients (n = 2264)				
Race-specific ML model	0.80 (0.79-0.81)	16 (12-19)	−0.04	0.90
Race-agnostic ML model	0.80 (0.79-0.81)	16 (12-18)	−0.05	0.92
ML model (race as a covariate)	0.80 (0.79-0.81)	16 (12-19)	−0.04	0.90
GWTG risk score ^{b}	0.69 (0.68-0.72)	23 (20-27)	−0.19	0.83
LR model (race as a covariate) ^{b}	0.70 (0.67-0.73)	28 (25-31)	−0.16	0.82
Race-specific LR model ^{b}	0.74 (0.73-0.76)	24 (20-27)	−0.10	0.91

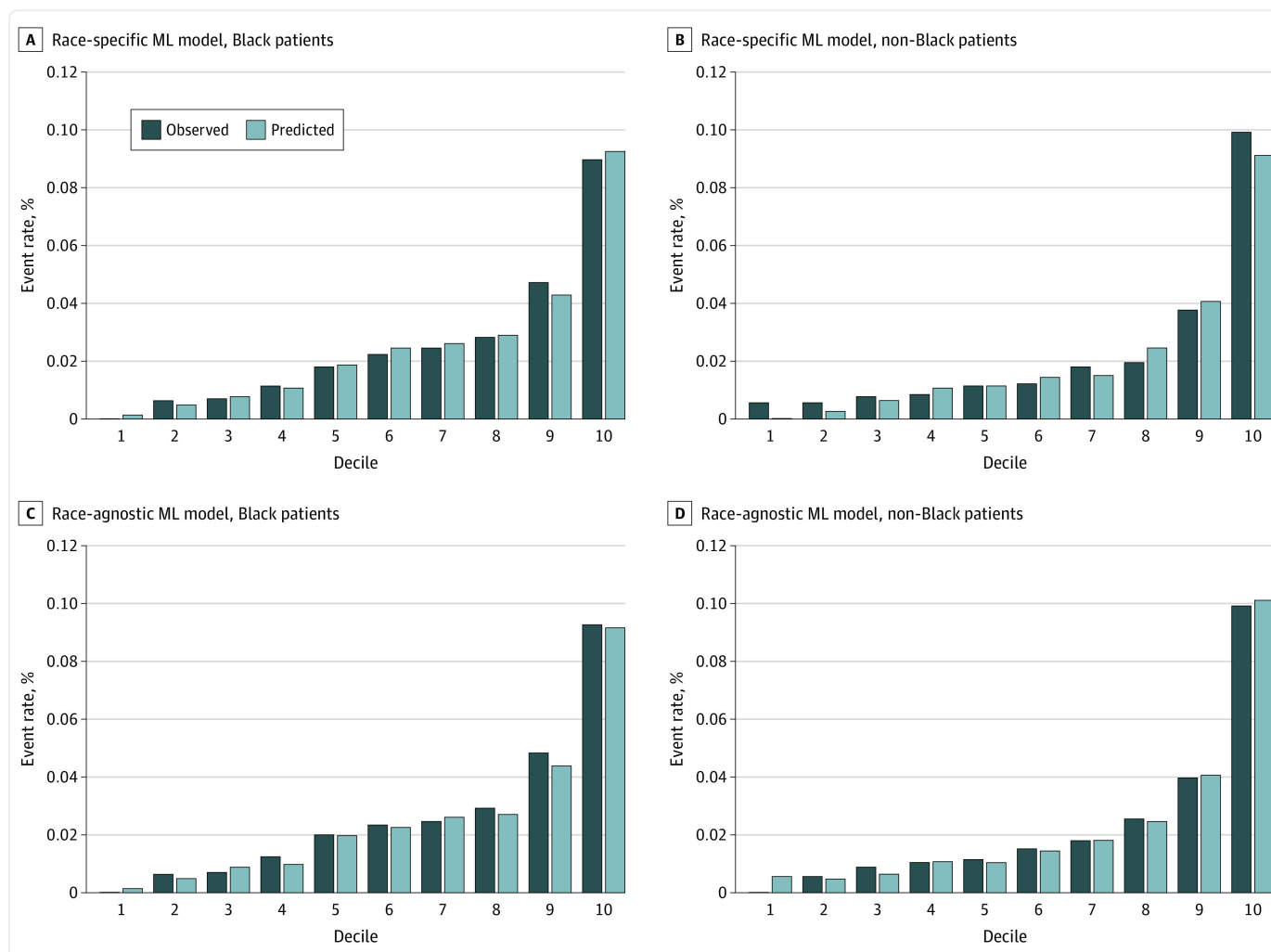
[Open in a new tab](#)

Abbreviations: ARIC, Atherosclerosis Risk in Communities; GWTG-HF, Get With The Guidelines–Heart Failure; LR, logistic regression; ML, machine learning.

^a A higher C index and lower Brier score indicate better performance. Among calibration slope measures, an intercept closer to 0 and slope closer to 1 indicates better calibration.

^b Indicates significant difference in C indices (DeLong test *P* value <.005) compared with the race-specific ML model.

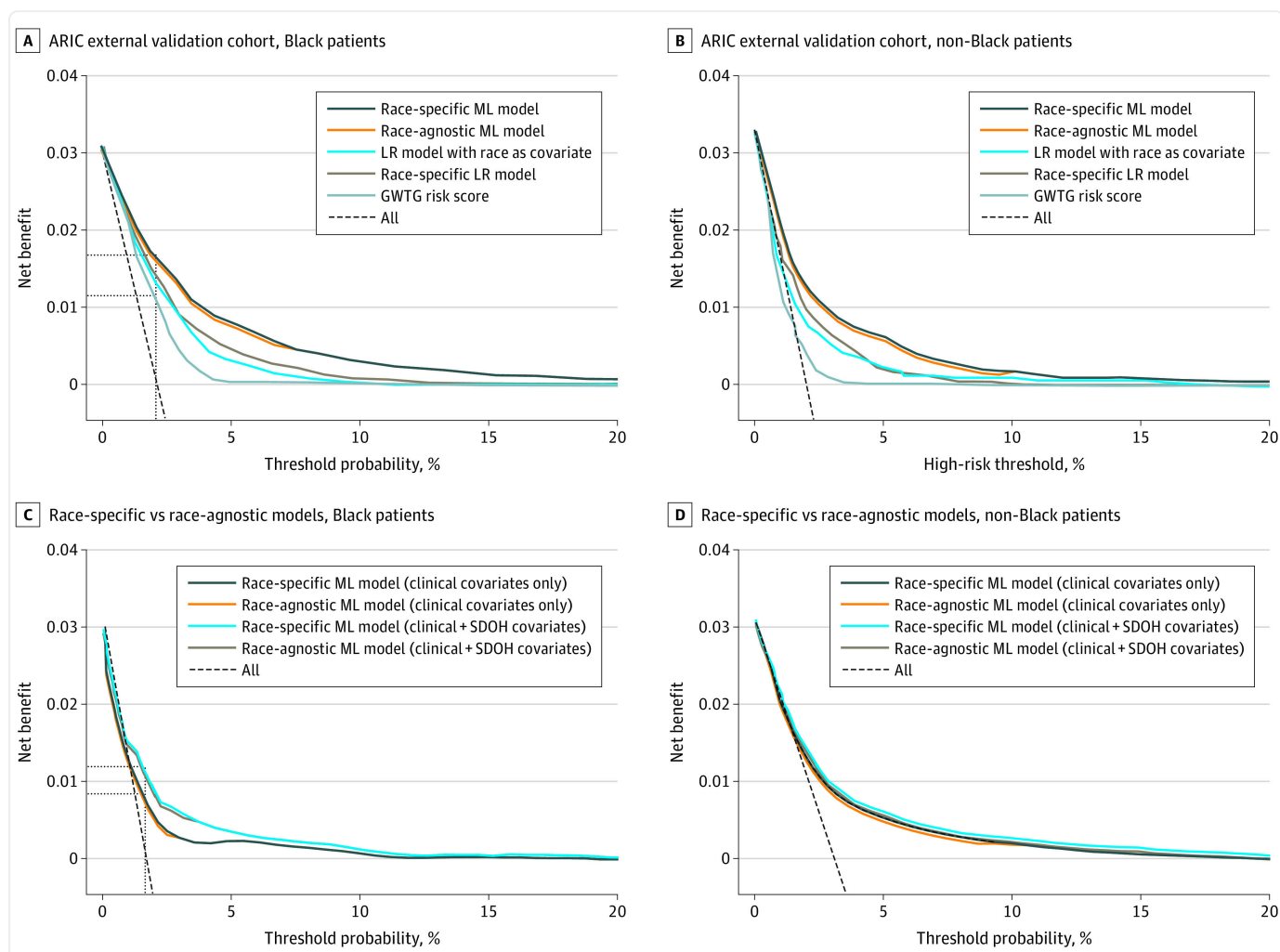
Figure 1. Observed vs Predicted Probability of In-Hospital Mortality for the Race-Specific vs Race-Agnostic Machine Learning (ML) Models.



[Open in a new tab](#)

In reclassification analysis, the race-specific ML model demonstrated improved net reclassification improvement and integrated discrimination index to the original GWTG-HF score in Black individuals and non-Black individuals (eTable 7 in the [Supplement](#)). In decision curve analyses, the race-specific ML model detected an additional 3 to 6 mortality events per 1000 Black patients ([Figure 2A](#)) and 2 to 9 events per 1000 non-Black patients compared with other models using race as a covariate ([Figure 2B](#)). Performance of the race-agnostic ML model in the external validation cohort was comparable with that of the race-specific models across both race groups ([Table 2](#); [Figure 2](#)).

Figure 2. Decision Curve Analysis in the Atherosclerosis Risk in Communities (ARIC) External Validation Cohort Demonstrating Prognostic Utility of the Different In-Hospital Mortality Models.



[Open in a new tab](#)

The vertical lines indicate a risk threshold at the cohort event rate. The horizontal lines indicate the net benefit with each model at a respective risk threshold. For example, the machine learning (ML) model detected an additional 6 and 9 mortality events per 1000 patients in Black and non-Black patient cohorts, respectively, compared with the Get With The Guidelines–Heart Failure (GWTG) risk score. LR indicates logistic regression.

Notably, race-specific logistic regression had superior discrimination and reclassification compared with logistic regression with race as a covariate model in Black patients and non-Black patients (C index difference, 0.04 [95% CI, 0.01-0.06] and 0.04 [95% CI, 0.01-0.08]; net reclassification improvement, 0.18 [95% CI, 0.02-0.29] and 0.22 [95% CI, -0.04 to 0.42],

respectively) ([Table 2](#); eTable 7 in the [Supplement](#)). Sensitivity analysis comparing the performance of ML models using a race-specific approach vs race as a covariate demonstrated comparable C indices among Black patients and non-Black patients. In the ML models with race as a covariate, race featured in the top 5 predictor variables based on the minimum depth.

Among Black patients in the ARIC validation cohort, the GWTG risk score predicted 5.7% of patients with an estimated risk above the 5% threshold (eFigure 7 in the [Supplement](#)). Conversely, the race-specific and race-agnostic ML models identified a significantly higher proportion of patients above the different risk thresholds (5% risk threshold: GWTG score = 5.7%; race-specific ML model = 12.4%; race-agnostic ML model = 15.3%; χ^2 P value <.001).

Incorporation of SDOH Into the ML Models

In the GWTG-HF validation cohort, among 13 088 Black patients, the race-specific ML model that included SDOH demonstrated improved discrimination and calibration (C indices, 0.77 [95% CI, 0.75-0.79]; intercept, -0.07; slope, 0.93) than the models with clinical covariates only (C indices, 0.73 [95% CI, 0.71-0.75]; intercept, -0.18; slope, 0.85; difference, 0.04 [95% CI, 0.01-0.07]) (eFigure 8 in the [Supplement](#)). Among reclassification metrics, the addition of SDOH parameters to the race-specific ML model with clinical covariates was associated with a significant improvement in upwards reclassification (net reclassification improvement, 0.22 [95% CI, 0.14-0.30]; integrated discrimination index, 0.007 [95% CI, 0.005-0.01]). In the decision curve analysis, the race-specific ML model with clinical and SDOH covariates (vs clinical covariates only) detected an additional 3 events per 1000 patients ([Figure 2C](#)). Similar results were observed with the race-agnostic ML model (C index of 0.76 [95% CI, 0.75-0.78]; difference, 0.01 [95% CI, -0.02 to 0.03]). In subgroup analysis, the race-specific and race-agnostic ML models demonstrated good and comparable discrimination and calibration performance across disproportionate share hospital–based subgroups (eTable 8 in the [Supplement](#)).

Conversely, among 51 485 non-Black patients, inclusion of SDOH to the race-specific ML models was not associated with a significant improvement in risk prediction performance with comparable discrimination (C indices, 0.75 [95% CI, 0.73-0.77]; difference, 0.01 [95% CI, -0.03 to 0.04]), calibration (intercept, -0.39; slope, 0.93), prognostic utility (no additional mortality events detected by decision curve analysis), and reclassification (net reclassification improvement, -0.01 [95% CI, -0.05 to 0.03]; integrated discrimination index, 0.003 [95% CI, -0.002 to 0.006]) ([Figure 2D](#); eFigure 8 in the [Supplement](#)). Similar results were observed with the race-agnostic ML model (C index, 0.75 [95% CI, 0.73-0.76]; difference, 0.005 [95% CI, -0.02 to 0.03]).

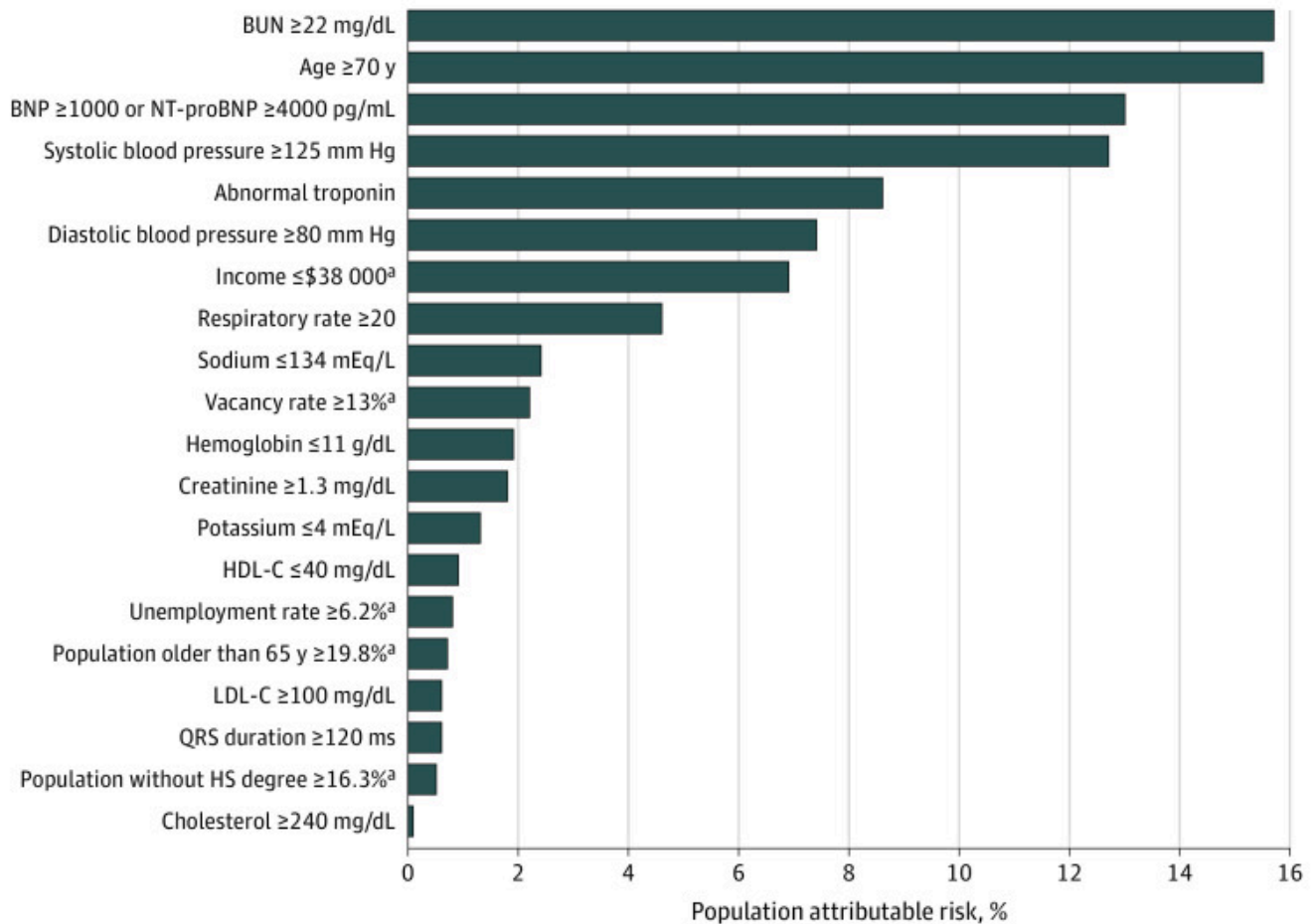
Race-Specific Determinants of In-Hospital Mortality

Using the race-specific ML model among Black patients, multiple SDOH parameters were identified as strong predictors of in-hospital mortality, with 5 such parameters among the top 20 predictors ([Figure 3](#)). Overall, the population-attributable risk percentage for in-hospital mortality associated with all SDOH parameters was 11.6% among Black patients. In contrast, among non-Black patients, only 1 SDOH parameter featured in the top 20 risk predictors with a total population-attributable risk percentage of 0.5% for in-hospital mortality using the race-specific ML model. Among clinical risk factors, measures of kidney function, blood pressure, natriuretic peptide, troponin, and age were among the top predictors of in-hospital mortality across both race groups ([Figure 3](#)).

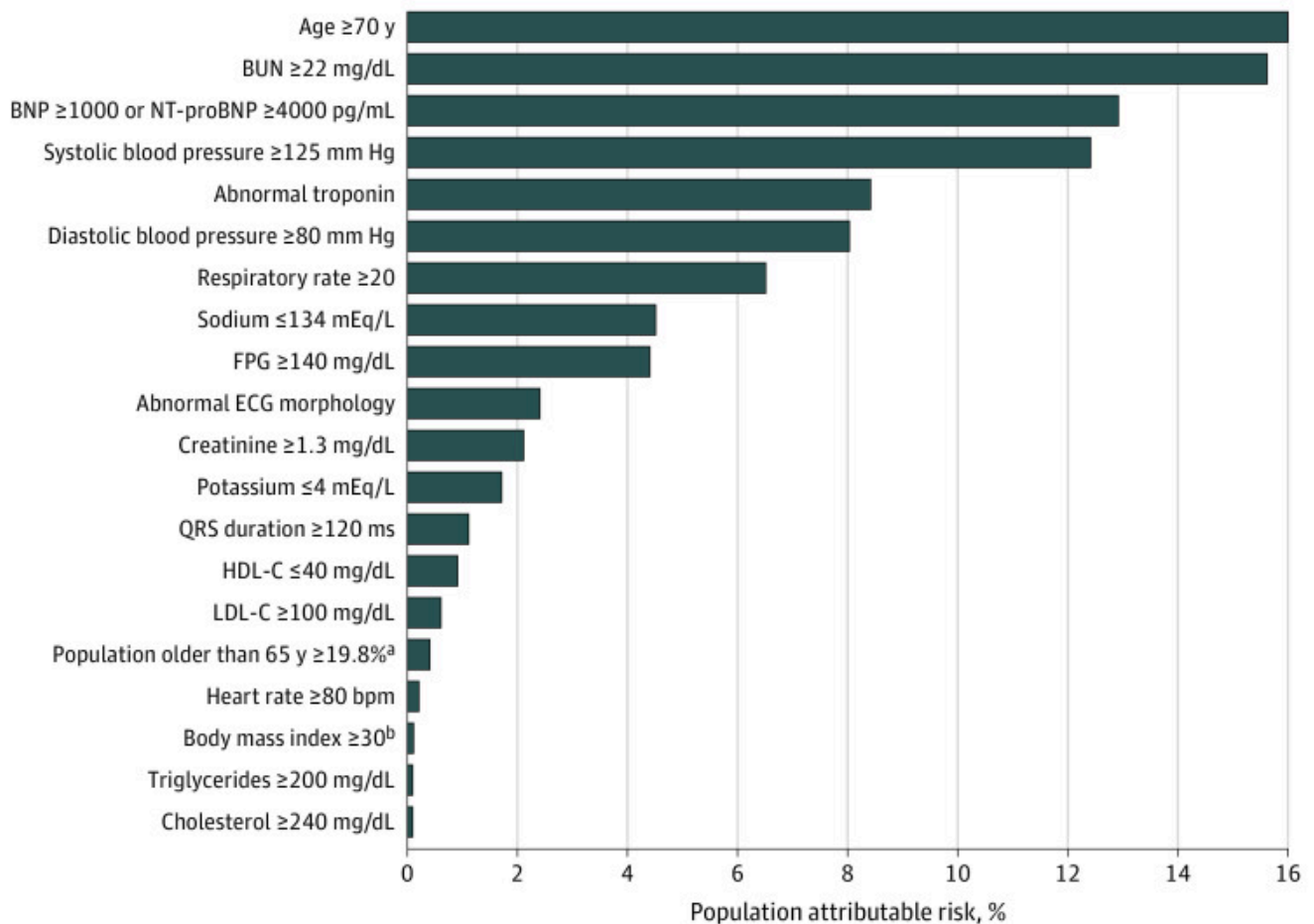
Figure 3. Population-Attributable Risk of Each Covariate Among Black and Non-Black Patients.



A Black patients



B Non-Black patients



[Open in a new tab](#)

BNP indicates B-type natriuretic peptide; BUN, blood urea nitrogen; COPD, chronic obstructive pulmonary disease; ECG, electrocardiogram; FPG, fasting plasma glucose; HS, high school; NT-proBNP, N-terminal pro-B-type natriuretic peptide; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

SI conversion factors: To convert BNP to nanograms per liter, multiply by 1; cholesterol to millimoles per liter, multiply by 0.0259; creatinine to micromoles per liter, multiply by 88.4; glucose to millimoles per liter, multiply by 0.0555; hemoglobin to grams per liter, multiply by 10; potassium to millimoles per liter, multiply by 1; sodium to millimoles per liter, multiply by 1.

^aIndicates social determinants of health parameters.

^bBody mass index was calculated as weight in kilograms divided by height in meters squared.

Discussion

In this cohort study, we developed and validated ML-based race-specific and race-agnostic risk models to predict in-hospital mortality among individuals with hospitalization for HF. We observed that the race-specific and race-agnostic ML-based models demonstrated excellent performance in the testing data sets, including those with substantial missingness in model covariates. Furthermore, the ML-based models had superior discrimination, calibration, and clinical utility in the external validation cohort than the original GWTG-HF risk scores and other rederived logistic regression models using race as a covariate. The addition of zip code-level SDOH to the ML model was associated with an improvement in risk reclassification and prognostic utility of the model in Black patients. We also observed significant race-specific differences in the population-attributable risk of in-hospital mortality associated with the SDOH with a significantly greater contribution of these parameters to the overall in-hospital mortality risk in Black patients vs non-Black patients. Overall, the present study demonstrates the potential utility of ML models for better and more equitable prediction of in-hospital mortality risk among Black patients and non-Black patients hospitalized for HF.

Novel Approach to Risk Prediction: Role of Machine Learning and Race-Specific Approach

The most significant advancement with our risk models is the use of ML-based approach to risk prediction. Several models exist for predicting the risk of adverse outcomes among patients with HF hospitalization. Established models, such as the GWTG-HF, OPTIMIZE-HF, ADHERE, and AHFI (Acute Heart Failure Index) risk scores, use traditional statistical modeling techniques, provide acceptable risk stratification, and have been well validated in external cohorts.^{4,5,30} A summary of prior risk-prediction models is provided in eTable 9 in the [Supplement](#). Besides traditional risk-prediction models, some previous studies have also developed ML-based models to predict in-hospital mortality. However, these studies have been mainly developed in non-US-based, ethnically homogenous cohorts.³¹ The ML models developed in the present study offer several advantages. First, they incorporate well-established prognostic biomarkers for HF in risk prediction not included in previous risk models and are thus better predicting individual-level risk. Second, the ML-based approach allows for greater generalizability, better tolerance to missing data, and more accurate risk prediction in external cohorts. This is evidenced by the ML model demonstrating adequate performance in a cohort with up to 50% missingness. With the addition of an application programming interface to improve implementation,¹⁰ the ML models offer an opportunity for real-world, electronic health record–based risk prediction.

In addition to using the ML-based approach to risk prediction, other aspects of our risk models are noteworthy. In the present study, we developed race-specific models for in-hospital mortality. We observed superior performance of the race-specific logistic regression model compared with models using race as a covariate, highlighting the potential utility of race-specific risk-prediction models. However, when using an ML-based approach, the discrimination and calibration metrics for the race-specific ML model were comparable with the ML model using race as a covariate. This is related to the modeling approach used by random forest learning methods, which assigned race a lower minimal depth (higher importance). Thus, the random forest model takes a race-specific approach, even when race is included as a covariate, creating a decision tree after the first few nodes, comparable with a race-specific model. Furthermore, even with the use of a race-agnostic approach (without race as a covariate), the performance of the ML model was comparable with the race-specific model. Taken together, the ML-based approach represents the most novel aspect of our risk model that is associated with improved and more equitable prediction of individual-level risk across race groups. Thus, even though the risk of mortality among Black patients was lower than non-Black patients in our study cohorts, the proportion of Black patients identified to be above specific risk thresholds was higher with the ML models than with the traditional GWTG-HF model. Future studies are needed to determine if similar ML models may facilitate more equitable risk-based allocation of care. This is particularly relevant considering the concerns raised about the potential unintended effects of assigning lower risk to Black patients using the existing risk models that use race as a covariate, which may add to existing disparities in HF care.⁷

Incorporation of SDOH to Improve Risk Prediction

SDOH are stronger predictors of in-hospital mortality in Black patients vs non-Black patients with HF.^{[32,33,34](#)} Previous studies have observed improved model performance incorporating SDOH data in the risk prediction equations.^{[35,36,37](#)} In the present study, we observed that zip code–level SDOH contributed more than 11% of the total in-hospital mortality risk in Black patients compared with 0.5% in non-Black patients. Consistent with the greater relative importance of SDOH in predicting the risk of in-hospital mortality in Black patients, we observed a significant improvement in risk reclassification and calibration with the addition of these parameters in Black patients but not non-Black patients. Furthermore, the model performance for the clinical and SDOH model was comparable among patients hospitalized at disproportionate share vs non–disproportionate share hospitals, highlighting the generalizability of the risk models among patients hospitalized in low- and high-resource hospitals.

While the improvement in risk prediction with the incorporation of zip code–level SDOH parameters is encouraging, future studies are needed to understand better how SDOH factors can be better incorporated for risk prediction in HF patients. First, it would be more informative to include individual-level data on SDOH rather than zip code–level data alone. Second, the ML-based approach used in the present study allowed us to evaluate the clinical and SDOH model in a data set with up to 50% missingness in clinical parameters. It is plausible that with better capture of clinical parameters in the model, the relative improvement in model performance with the incorporation of zip code–level SDOH may be attenuated.

Limitations

Our study has some notable limitations. First, we only included variables that were regularly captured in the GWTG registry. Data on certain laboratory measures, such as hemoglobin A1c and lipid profiles, which are associated with increased mortality risk,^{[38](#)} had significant missingness and were excluded from the candidate covariates. Second, the race-specific models were developed using self-reported race. Individuals who may not identify with a specific race on the GWTG-HF data form may not be accurately represented. However, sensitivity analysis across available races and ethnicities showed similar performance. Third, only zip code–level SDOH data were available for the present analysis. Incorporating participant-level SDOH parameters in risk models may further improve their predictive performance. Fourth, race was self-reported data on genetic ancestry were not available. Fifth, we could not externally validate the clinical and SDOH models given the lack of zip code data in the ARIC cohort.

Conclusions

The race-specific and race-agnostic ML models to predict in-hospital mortality among patients with HF demonstrated superior discrimination and calibration in Black patients and non-Black patients and outperformed traditional logistic regression models with race as a covariate. Furthermore, incorporating zip code–level SDOH parameters into the risk prediction ML models improved their performance among Black patients but not non-Black patients. Future studies are needed to determine whether race-specific and race-agnostic ML models may improve risk prediction, resource allocation, and care outcomes among Black patients with HF.

Supplement.

eMethods

eReferences

eTable 1. Candidate covariates and their respective domain

eTable 2. ZIP-code level social determinants of health parameters that were considered for predicting in-hospital mortality following HF hospitalization

eTable 3. Baseline characteristics of participants in the internal and external validation cohorts by race

eTable 4. Discrimination and calibration performance of the race-specific models for predicting in-hospital mortality among patients with heart failure in the internal GWTG testing cohort with complete data available and with up to 50% missingness in the covariate data

eTable 5. Discrimination and calibration performance of the models for predicting in-hospital mortality among patients with heart failure in the internal GWTG validation cohort across age, sex, and socioeconomic status-based subgroups

eTable 6. Discrimination and calibration performance of the non-Black race-specific and race-agnostic models for predicting in-hospital mortality among patients with heart failure with different self-identified race/ethnicities

eTable 7. Reclassification metrics in the ARIC external validation between the ML model and the original GWTG risk score

eTable 8. Discrimination and calibration performance of the models for predicting in-hospital mortality among patients with heart failure in the internal GWTG validation cohort across disproportional share hospital-based subgroups

eTable 9. Comparison of models to predict risk of in-hospital mortality among patients with hospitalization for heart failure

eFigure 1. CONSORT diagram

eFigure 2. Variable importance of Black and non-Black patients determined by the VIMP metric of a race-specific random forest model with 20 bootstrap replicates

eFigure 3. Area under the receiver operating characteristics and precision-recall curve for increasing number of variables in a random forest model to predict in-hospital mortality in the overall cohort

eFigure 4. Observed vs. predicted probability of in-hospital mortality for the race-specific ML models

eFigure 5. Observed vs. predicted probability of in-hospital mortality for the race-agnostic ML models

eFigure 6. Observed vs. predicted probability of in-hospital mortality for the GWTG-HF risk score

eFigure 7. Among Black participants in the ARIC external validation cohort, percentage of participants with a predicted risk above the specific risk thresholds between the original GWTG risk score and the race-specific ML model

eFigure 8. Observed vs. predicted probability of in-hospital mortality for the race-specific ML + social determinants of health models

[Click here for additional data file.](#) (1MB, pdf)

References

1. Abraham WT, Adams KF, Fonarow GC, et al. ; ADHERE Scientific Advisory Committee and Investigators; ADHERE Study Group . In-hospital mortality in patients with acute decompensated heart failure requiring intravenous vasoactive medications: an analysis from the Acute Decompensated Heart Failure National Registry (ADHERE). J Am Coll Cardiol. 2005;46(1):57-64. doi: 10.1016/j.jacc.2005.03.051 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

2. Chang PP, Chambless LE, Shahar E, et al. Incidence and survival of hospitalized acute decompensated heart failure in four US communities (from the Atherosclerosis Risk in Communities Study). *Am J Cardiol*. 2014;113(3):504-510. doi: 10.1016/j.amjcard.2013.10.032 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
3. Kamath SA, Drazner MH, Wynne J, Fonarow GC, Yancy CW. Characteristics and outcomes in African American patients with decompensated heart failure. *Arch Intern Med*. 2008;168(11):1152-1158. doi: 10.1001/archinte.168.11.1152 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
4. Peterson PN, Rumsfeld JS, Liang L, et al. ; American Heart Association Get With The Guidelines-Heart Failure Program . A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association Get With The Guidelines program. *Circ Cardiovasc Qual Outcomes*. 2010;3(1):25-32. doi: 10.1161/CIRCOUTCOMES.109.854877 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
5. Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ; ADHERE Scientific Advisory Committee, Study Group, and Investigators . Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA*. 2005;293(5):572-580. doi: 10.1001/jama.293.5.572 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
6. Abraham WT, Fonarow GC, Albert NM, et al. ; OPTIMIZE-HF Investigators and Coordinators . Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). *J Am Coll Cardiol*. 2008;52(5):347-356. doi: 10.1016/j.jacc.2008.04.028 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
7. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874-882. doi: 10.1056/NEJMms2004740 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
8. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(25 Pt B):2935-2959. doi: 10.1016/j.jacc.2013.11.005 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
9. 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. doi: 10.1016/j.jacc.2019.02.057 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

10. Segar MW, Jaeger BC, Patel KV, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation*. 2021;143(24):2370-2383. doi: 10.1161/CIRCULATIONAHA.120.053134 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
11. Inker LA, Eneanya ND, Coresh J, et al. ; Chronic Kidney Disease Epidemiology Collaboration . New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med*. 2021;385(19):1737-1749. doi: 10.1056/NEJMoa2102953 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
12. Diao JA, Wu GJ, Taylor HA, et al. Clinical implications of removing race from estimates of kidney function. *JAMA*. 2021;325(2):184-186. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
13. Segar MW, Vaduganathan M, Patel KV, 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(12):2298-2306. doi: 10.2337/dc19-0587 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
14. Hong Y, LaBresh KA. Overview of the American Heart Association “Get With The Guidelines” programs: coronary heart disease, stroke, and heart failure. *Crit Pathw Cardiol*. 2006;5(4):179-186. doi: 10.1097/01.hpc.0000243588.00012.79 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
15. Smaha LA; American Heart Association . The American Heart Association Get With The Guidelines program. *Am Heart J*. 2004;148(5)(suppl):S46-S48. doi: 10.1016/j.ahj.2004.09.015 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
16. Americas Essential Hospitals . Accessed November 2, 2021. <http://essentialhospitals.org>
17. Disproportionate Share Hospital . Accessed November 2, 2021. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/dsh> . <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/dsh>
18. ARIC Investigators . The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. the ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702. doi: 10.1093/oxfordjournals.aje.a115184 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
19. Caughey MC, Sueta CA, Stearns SC, Shah AM, Rosamond WD, Chang PP. Recurrent acute decompensated heart failure admissions for patients with reduced versus

preserved ejection fraction (from the Atherosclerosis Risk in Communities study). *Am J Cardiol*. 2018;122(1):108-114. doi: 10.1016/j.amjcard.2018.03.011 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

20. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36. doi: 10.1148/radiology.143.1.7063747 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845. doi: 10.2307/2531595 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

22. Elliott J, Bodinier B, Bond TA, et al. Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. *JAMA*. 2020;323(7):636-645. doi: 10.1001/jama.2019.22241 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

23. Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiol*. 2021;6(6):633-641. doi: 10.1001/jamacardio.2021.0122 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

24. Brier G. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78(78):1-3. doi: [[DOI](#)] [[Google Scholar](#)]

25. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol*. 2010;63(8):938-939. doi: 10.1016/j.jclinepi.2009.11.009 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

26. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med*. 2014;160(2):122-131. doi: 10.7326/M13-1522 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

27. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11-21. doi: 10.1002/sim.4085 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

28. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574. doi: 10.1177/0272989X06295361 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

29. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics*. 1993;49(3):865-872. doi: 10.2307/2532206 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
30. Auble TE, Hsieh M, Gardner W, et al. A prediction rule to identify low-risk patients with heart failure. *Acad Emerg Med*. 2005;12(6):514-521. doi: 10.1197/j.aem.2004.11.026 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
31. Kwon JM, Kim KH, Jeon KH, et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PLoS One*. 2019;14(7):e0219302. doi: 10.1371/journal.pone.0219302 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
32. Basu J, Hanchate A, Bierman A. Racial/Ethnic disparities in readmissions in US hospitals: the role of insurance coverage. *Inquiry*. 2018;55:46958018774180. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
33. Rodriguez-Gutierrez R, Herrin J, Lipska KJ, Montori VM, Shah ND, McCoy RG. Racial and ethnic differences in 30-day hospital readmissions among US adults with diabetes. *JAMA Netw Open*. 2019;2(10):e1913249. doi: 10.1001/jamanetworkopen.2019.13249 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
34. Patel SA, Krasnow M, Long K, Shirey T, Dickert N, Morris AA. Excess 30-day heart failure readmissions and mortality in Black patients increases with neighborhood deprivation. *Circ Heart Fail*. 2020;13(12):e007947. doi: 10.1161/CIRCHEARTFAILURE.120.007947 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
35. Hammond G, Johnston K, Huang K, Joynt Maddox KE. Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death. *Circ Cardiovasc Qual Outcomes*. 2020;13(6):e006752. doi: 10.1161/CIRCOUTCOMES.120.006752 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
36. Dalton JE, Perzynski AT, Zidar DA, et al. Accuracy of cardiovascular risk prediction varies by neighborhood socioeconomic position: a retrospective cohort study. *Ann Intern Med*. 2017;167(7):456-464. doi: 10.7326/M16-2543 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
37. Bhavsar NA, Gao A, Phelan M, Pagidipati NJ, Goldstein BA. Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data. *JAMA Netw Open*. 2018;1(5):e182716. doi:

10.1001/jamanetworkopen.2018.2716 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

38. Aguilar D, Bozkurt B, Ramasubbu K, Deswal A. Relationship of hemoglobin A1C and mortality in heart failure patients with diabetes. J Am Coll Cardiol. 2009;54(5):422-428. doi: 10.1016/j.jacc.2009.04.049 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

Associated Data

This section collects any data citations, data availability statements, or supplementary materials included in this article.

Supplementary Materials

Supplement.

eMethods

eReferences

eTable 1. Candidate covariates and their respective domain

eTable 2. ZIP-code level social determinants of health parameters that were considered for predicting in-hospital mortality following HF hospitalization

eTable 3. Baseline characteristics of participants in the internal and external validation cohorts by race

eTable 4. Discrimination and calibration performance of the race-specific models for predicting in-hospital mortality among patients with heart failure in the internal GWTG testing cohort with complete data available and with up to 50% missingness in the covariate data

eTable 5. Discrimination and calibration performance of the models for predicting in-hospital mortality among patients with heart failure in the internal GWTG validation cohort across age, sex, and socioeconomic status-based subgroups

eTable 6. Discrimination and calibration performance of the non-Black race-specific and race-agnostic models for predicting in-hospital mortality among patients with heart failure with different self-identified race/ethnicities

eTable 7. Reclassification metrics in the ARIC external validation between the ML model and the original GWTG risk score

eTable 8. Discrimination and calibration performance of the models for predicting in-hospital mortality among patients with heart failure in the internal GWTG validation cohort across disproportional share hospital-based subgroups

eTable 9. Comparison of models to predict risk of in-hospital mortality among patients with hospitalization for heart failure

eFigure 1. CONSORT diagram

eFigure 2. Variable importance of Black and non-Black patients determined by the VIMP metric of a race-specific random forest model with 20 bootstrap replicates

eFigure 3. Area under the receiver operating characteristics and precision-recall curve for increasing number of variables in a random forest model to predict in-hospital mortality in the overall cohort

eFigure 4. Observed vs. predicted probability of in-hospital mortality for the race-specific ML models

eFigure 5. Observed vs. predicted probability of in-hospital mortality for the race-agnostic ML models

eFigure 6. Observed vs. predicted probability of in-hospital mortality for the GWTG-HF risk score

eFigure 7. Among Black participants in the ARIC external validation cohort, percentage of participants with a predicted risk above the specific risk thresholds between the original GWTG risk score and the race-specific ML model

eFigure 8. Observed vs. predicted probability of in-hospital mortality for the race-specific ML + social determinants of health models

[Click here for additional data file.](#) (1MB, pdf)

Articles from JAMA Cardiology are provided here courtesy of **American Medical Association**