ESC HEART FAILURE
ESC Heart Failure (2021)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13429

ORIGINAL RESEARCH ARTICLE

Risk factors and prediction models for incident heart failure with reduced and preserved ejection fraction

Liam Gaziano[1,2], Kelly Cho[1,3], Luc Djousse[1,3], Petra Schubert[1], Ashley Galloway[1], Yuk-Lam Ho[1], Katherine Kurgansky[1], David R. Gagnon[1,4], John P. Russo[1,5], Emanuele Di Angelantonio[2], Angela M. Wood[2], John Danesh[2], John Michael Gaziano[1,3], Adam S. Butterworth[2], Peter W.F. Wilson[6,7] and Jacob Joseph[1,8]*

[1]*Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), Cardiology Section, VA Boston Healthcare System, 1400 VFW Parkway, West Roxbury, Boston, MA 02132, USA;* [2]*BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK;* [3]*Department of Medicine, Division of Aging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA;* [4]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;* [5]*Landmark College, Putney, VT, USA;* [6]*Atlanta VA Medical Center, Decatur, GA, USA;* [7]*Department of Medicine, Division of Cardiovascular Disease, Emory University School of Medicine, Atlanta, GA, USA; and* [8]*Department of Medicine, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

Abstract

Aims This study aims to develop the first race-specific and sex-specific risk prediction models for heart failure with preserved (HFpEF) and reduced ejection fraction (HFrEF).

Methods and results We created a cohort of 1.8 million individuals who had an outpatient clinic visit between 2002 and 2007 within the Veterans Affairs (VA) Healthcare System and obtained information on HFpEF, HFrEF, and several risk factors from electronic health records (EHR). Variables were selected for the risk prediction models in a 'derivation cohort' that consisted of individuals with baseline date in 2002, 2003, or 2004 using a forward stepwise selection based on a change in C-index threshold. Discrimination and calibration were assessed in the remaining participants (internal 'validation cohort'). A total of 66 831 individuals developed HFpEF, and 92 233 developed HFrEF (52 679 and 71 463 in the derivation cohort) over a median of 11.1 years of follow-up. The HFpEF risk prediction model included age, diabetes, BMI, COPD, previous MI, antihypertensive treatment, SBP, smoking status, atrial fibrillation, and estimated glomerular filtration rate (eGFR), while the HFrEF model additionally included previous CAD. For the HFpEF model, C-indices were 0.74 (SE = 0.002) for white men, 0.76 (0.005) for black men, 0.79 (0.015) for white women, and 0.77 (0.026) for black women, compared with 0.72 (0.002), 0.72 (0.004), 0.77 (0.017), and 0.75 (0.028), respectively, for the HFrEF model. These risk prediction models were generally well calibrated in each race-specific and sex-specific stratum of the validation cohort.

Conclusions Our race-specific and sex-specific risk prediction models, which used easily obtainable clinical variables, can be a useful tool to implement preventive strategies or subtype-specific prevention trials in the nine million users of the VA healthcare system and the general population after external validation.

Keywords Risk prediction; Heart failure; Heart failure with reduced ejection fraction; Heart failure with preserved ejection fraction; Electronic health records

Received: 29 September 2020; Revised: 28 April 2021; Accepted: 2 May 2021
**Correspondence to: Jacob Joseph, Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), Cardiology Section, VA Boston Healthcare System, 1400 VFW Parkway, West Roxbury, MA 02132, USA. Phone: +1-857-203-6841. Email: jjoseph16@partners.org; jacob.joseph@va.gov*

Introduction

Heart failure (HF) is a common condition with an estimated prevalence of 6.2 million in the USA.[1] It carries a poor prognosis, including a 5 year mortality of 50%,[2] and adds substantially to healthcare costs.[3] The two major subtypes of HF, heart failure with preserved (HFpEF) and reduced ejection fraction (HFrEF), are increasingly regarded as distinct conditions.[4] In part, this distinction arises because patients with overt HFpEF and HFrEF present with distinct ventricular remodelling[5] and have responded differently to pharmacological agents in randomized trials on hospitalizations and

mortality outcomes.[6,7] Although guidelines have called for a greater focus on prevention of HF to curtail its growing burden,[8,9] the majority of prospective studies have not been able to differentially examine the risk of developing HFrEF or HFpEF. As a result, there is a lack of evidence on subtype-specific risk factors, preventing the development of risk prediction models for HFpEF and HFrEF. Previous studies[10,11] have developed such models, but these were not stratified by race or sex, which could be clinically important to predict disease. Hence, we used a large electronic health record (EHR) database from the Veterans Health Administration (VA) with more than 1.9 million individuals to create a prospective cohort to study incident HFpEF and HFrEF events. We compared associations between each HF subtype for a wide range of risk factors and developed the first race-specific and sex-specific risk prediction models for HFpEF and HFpEF.

Methods

Population

For this study, we utilized a previously described cohort[12] comprising all patients with at least one primary care visit at a VA facility who had at least one outpatient lipid result between 2002 and 2007 and a blood pressure measurement within 30 days of this index lipid testing. These restrictions reduce misclassification biases in that infrequent users of the VA are more likely to have systematically inaccurate measurements of baseline characteristics or outcomes.[13] Baseline entry into the cohort was the date of lipid testing. This cohort was created using data from the VA corporate data warehouse (CDW)[14] linked to the Centers for Medicaid and Medicare Services (CMS) and National Death Index (NDI)[15] databases. We excluded individuals with a history of severe co-morbidities at baseline including cancer, dialysis, and major mental health disorder (Supporting Information, *Table S1*), those of non-white or non-black race, those with HF before or up to 30 days after baseline, or those who died within 30 days of baseline date Those with baseline date in 2002, 2003, or 2004 were allocated to the 'derivation cohort' with the remaining individuals (baseline date in 2005 or 2006) serving as the internal 'validation cohort'.

Exposure definitions

This extensive EHR resource contains International Classification of Diseases (ICD) codes, lab results (sodium, potassium, creatinine, and major lipids), anthropometric measurements (blood pressure, height, and weight), medication information (blood pressure medications and statins, see Supporting Information, *Table S2* for more detail) and demographic variables (age, sex, ethnicity, and race) in a structured format. Diabetes mellitus was defined as the use of a diabetes medication prior to the baseline date plus either two ICD-9-CM diagnosis codes 250.xx or the use of at least one 250.xx code in combination with a VA primary care visit.[16] Smoking history was categorized as current, former, or never with an algorithm developed and validated within the VA medical records data.[17] We used ICD-9 codes to define previous MI (410–411.x), CAD (410–414.x), COPD (490–496.x, 510.x, and 781.5x), and atrial fibrillation (427.31). Anaemia was defined as a haemoglobin $<$13.5 g/dL for men, a haemoglobin $<$12 g/dL for women, or an ICD-9 code of anaemia (280–285.x). Estimated glomerular filtration rate (eGFR) was estimated from creatinine using the CKD-EPI equation.[18] For sodium, potassium, and creatinine, we selected the closest measurement in the EHR on or before baseline date up to a year prior. Medication status was defined as having an active prescription at the time of baseline, meaning their baseline date fell between the date the prescription was picked up/mailed and the end of their supply.

Outcome definition

Ejection fraction (EF) values were extracted from echocardiogram reports, radiology reports, and clinical notes (to ensure capture of EF values measured outside the VA) using a validated natural language processing tool in a process described elsewhere[19] and used in previous studies.[20,21] HF patients were identified as those with an ICD-9 code of 428.x or ICD-10 code of I50.x. We defined incident HFpEF as the first instance of an HF code within 6 months of an echocardiogram EF $\geq$ 50% and incident HFrEF as the first instance of an echocardiogram EF $\leq$ 40% in patients with an HF code. If patients met criteria for both subtypes, the subtype that presented first was considered so that outcomes were mutually exclusive. Those with HF but no EF value (unclassified HF; *N* = 51 561; 23% of all HF cases) and those with EF between 40% and 50% (mid-range HF; *N* = 15 083, 7% of all HF cases) were included in the at-risk population but censored at first HF event. We chose not to consider mid-range HF as either HFpEF or HFrEF because they are likely a mix of the two aetiologies.[22]

Statistical analysis

In the derivation cohort, we used Cox regression adjusted for age, sex, race, and ethnicity to estimate the hazard ratios (HR) and 95% confidence intervals (95% CI) between risk factors and each HF subtype, censoring at the other subtype of HF, unclassified HF, mid-range HF, death, or end of follow-up (31 December 2016), whichever came first. A cause-specific Cox model was chosen over a competing risk

model used in previous studies[10,11] because results between the approaches were similar, but cause-specific Cox models were much less computationally intensive, especially for a cohort of this size.

We selected the variables to be included in the HFrEF and HFpEF models independently, using the entire derivation cohort (i.e. not stratified by race and sex). Continuous variables that displayed non-linear associations were modelled using fractional polynomials, initially in a model that included age, sex, and race, then estimated again after the final model was selected. To select variables, we started with an initial reference model that included age, sex, and race and calculated the change in Harrell's C-index[23–25] of adding each risk factor listed in *Table 1* individually (except for individual antihypertensive medications). In a forward stepwise manner, we added the variable that most increased the C-index to the reference model until no variable had an improvement greater than 0.002. Inspection of log–log plots indicated that the proportional hazards assumption was met for all variables in both final models. We took an available-case approach, which restricted to those with complete data on the reference model and tested variable.

After variable selection, we used Cox regression to estimate regression coefficients and baseline survival (Supporting Information, *Tables S3*–S5) within each race-stratum and sex-stratum of the derivation cohort, which were used to calculate predicted 10 year risk for each individual in the validation cohort. In each race-specific and sex-specific stratum, we evaluated discrimination of the models in the validation cohort using Harrell's C-index and evaluated calibration by plotting mean quantiles (deciles for men, quintiles for women) of predicted 10 year risk against observed 10 year risk derived from Kaplan–Meier curves. We compared our models with an existing one for HFpEF,[10] the multi-cohort International Collaboration on Heart Failure Subtypes (ICHFS) HFpEF model.

Table 1 Baseline characteristics in the derivation cohort

	White men	Black men	White women	Black women
N total	1 145 867	173 246	33 419	10 081
Age (SD)[a], years	63.6 (9.7)	57.3 (10.3)	56.5 (11.4)	49.3 (7.4)
Hispanic or Latino ethnicity	53 253 (4.7)	3420 (2.0)	1064 (3.2)	109 (1.1)
SBP (SD)[a], mmHg	137.5 (18.7)	139.4 (19.8)	132.9 (19.5)	133.1 (19.3)
DBP (SD), mmHg	77.2 (10.8)	80.6 (11.8)	75.1 (10.7)	78.0 (11.4)
BMI (SD), kg/m^2	29.3 (5.3)	29.2 (5.7)	29.6 (6.7)	30.5 (6.2)
N missing	3213	292	98	6
HDL-c (SD)[a], mg/dL	44.1 (12.9)	48.7 (15.4)	55.4 (16.1)	57.3 (16.8)
LDL-c (SD), mg/dL	117.1 (34.6)	121.0 (37.2)	126.3 (36.4)	123.5 (37.0)
N missing	55 463	6711	1244	340
Total cholesterol (SD), mg/dL	193.1 (40.8)	196.2 (42.1)	211.8 (41.8)	202.3 (41.1)
N missing	541	135	13	8
Ln-triglycerides (SD), mg/dL	5.0 (0.6)	4.8 (0.6)	4.9 (0.6)	4.6 (0.5)
N missing	50 308	6932	1565	417
eGFR (SD), mL/min/1.73 m^2	75.2 (17.3)	85.4 (21.0)	79.0 (18.4)	93.1 (20.7)
N missing	168 397	22 397	5436	1401
Sodium (SD), mmol/L	139.5 (2.9)	139.6 (2.8)	139.4 (2.7)	139.3 (3.1)
N missing	195 678	22 986	5623	1410
Potassium (SD), mmol/L	4.4 (0.4)	4.2 (0.4)	4.2 (0.4)	4.0 (0.4)
N missing	190 636	22 611	5576	1402
Current smoker (%)[a]	185 842 (16.2)	41 973 (24.2)	6728 (20.1)	2101 (20.8)
Statin medication (%)[a]	386 851 (33.8)	34 214 (19.9)	6657 (19.9)	993 (9.9)
Antihypertensive medication (%)[a]	603 223 (52.6)	92 404 (53.3)	12 524 (37.5)	4078 (40.5)
ACE inhibitors (%)[a]	333 151 (29.1)	47 912 (27.7)	5196 (15.6)	1423 (14.1)
ARB (%)[a]	31 994 (2.8)	4126 (2.4)	731 (2.2)	188 (2.7)
Beta-blockers (%)[a]	248 749 (21.7)	26 634 (15.4)	4591 (13.7)	1166 (11.6)
Calcium channel blockers (%)[a]	180 163 (15.7)	38 432 (22.2)	3536 (10.6)	1431 (14.2)
Diuretics (%)[a]	209 979 (18.3)	44 609 (25.8)	5866 (17.6)	2399 (23.8)
Previous MI (%)[a]	71 787 (6.3)	6225 (3.6)	749 (2.2)	116 (1.2)
Previous CAD (%)[a]	297 358 (26.0)	21 250 (12.3)	2876 (8.6)	442 (4.4)
Diabetes (%)[a]	265 236 (23.2)	45 385 (26.2)	4098 (12.3)	1280 (12.7)
COPD (%)[a]	151 750 (13.2)	17 725 (10.2)	5453 (16.3)	1364 (13.5)
Anaemia (%)[a]	118 516 (10.3)	33 797 (19.5)	2863 (8.6)	2096 (20.8)
Atrial fibrillation (%)[a]	51 401 (4.5)	2074 (1.2)	584 (1.8)	35 (0.4)

[a]No missing.

ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; BMI, body mass index; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HDL-c, high density lipoprotein cholesterol; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LDL-c, low density lipoprotein cholesterol; MI, myocardial infarction; SBP, systolic blood pressure.

DOI: 10.1002/ehf2.13429

Analyses were carried out using Stata 15 (StataCorp LLC, College Station, Texas).

Research ethics statement

This study was approved by the Institutional Review Boards at VA Boston Healthcare System and Emory University. This study was restricted to secondary data analysis, and thus, the requirement for informed consent from study participants was waived.

Results

Among Veterans 40–80 years of age at baseline, 66 831 developed HFpEF, and 92 233 developed HFrEF (52 679 and 71 463, respectively, in the derivation cohort). *Table 1* shows baseline characteristics within each race-specific and sex-specific stratum of the derivation cohort (see Supporting Information, *Tables S6* for baseline characteristics of the validation cohort).

Risk factors for heart failure with preserved ejection fraction and heart failure with reduced ejection fraction

Associations between risk factors and each HF subtype adjusted for age, sex, race, and ethnicity are shown in *Table 2*. Many associations were in similar direction for HFpEF and HFrEF, but there were notable quantitative differences for some important cardiometabolic risk factors. Specifically, systolic blood pressure (SBP), body mass index (BMI), antihypertensive medication use, and COPD were stronger risk factors for HFpEF, while male sex, previous MI, and previous CAD were stronger risk factors for HFrEF.

Risk prediction models for heart failure with preserved ejection fraction and heart failure with reduced ejection fraction

The HFpEF risk prediction model included age, diabetes, BMI, COPD, previous MI, antihypertensive treatment, SBP,

Table 2 Risk factors for HFpEF and HFrEF in the derivation cohort

	HFpEF	HFrEF
	HR (95% CI)	HR (95% CI)
Age (per 10 years)	1.38 (1.37–1.40)	1.28 (1.27–1.29)
Female sex	0.89 (0.85–0.94)	0.44 (0.41–0.47)
Black race	1.37 (1.34–1.40)	1.36 (1.34–1.39)
Hispanic or Latino ethnicity	0.99 (0.95–1.04)	1.07 (1.03–1.07)
SBP (per 20 mmHg)	1.26 (1.25–1.27)	1.12 (1.11–1.13)
DBP (per 10 mmHg)	1.00 (1.00–1.01)	0.99 (0.98–1.00)
BMI (per 5 kg/m^2)	1.48 (1.47–1.49)	1.20 (1.20–1.21)
HDL-c (per 15 mg/dL)	0.83 (0.82–0.84)	0.81 (0.80–0.81)
LDL-c (per 35 mg/dL)	0.88 0.87–0.88)	0.91 (0.90–0.92)
Total cholesterol (per 40 mg/dL)	0.92 (0.91–0.93)	0.95 (0.94–0.96)
Ln-triglycerides (per 0.5 Ln-mg/dL)	1.16 (1.15–1.17)	1.15 (1.14–1.16)
eGFR (per 15 mL/min/1.73 m^2)	0.88 (0.87–0.88)	0.90 (0.89–0.90)
Sodium (per 3 mmol/L)	0.88 (0.88–0.89)	0.90 (0.89–0.90)
Potassium (per 0.5 mmol/L)	0.95 (0.94–0.96)	1.02 (1.01–1.02)
Current smoker (versus never)	1.52 (1.48–1.57)	1.80 (1.75–1.85)
Statin medication	1.20 (1.18–1.23)	1.31 (1.29–1.33)
Antihypertensive medication	2.28 (2.25–2.33)	1.98 (1.95–2.02)
ACE inhibitors	1.86 (1.82–1.89)	1.72 (1.69–1.74)
ARB	1.56 (1.50–1.63)	1.35 (1.30–1.40)
Beta-blockers	1.79 (1.76–1.83)	1.82 (1.79–1.85)
Calcium channel blockers	1.74 (1.71–1.78)	1.45 (1.43–1.48)
Diuretics	2.14 (2.11–2.19)	1.69 (1.66–1.72)
Previous MI	2.85 (2.78–2.93)	4.18 (4.10–4.26)
Previous CAD	1.73 (1.69–1.76)	2.69 (2.65–2.74)
Diabetes	2.31 (2.27–2.35)	2.07 (2.04–2.11)
COPD	2.14 (2.09–2.19)	1.58 (1.55–1.62)
Atrial fibrillation	2.24 (2.17–2.32)	2.21 (2.15–2.27)
Anaemia	1.61 (1.57–1.65)	1.45 (1.42–1.48)

Adjusted for age, sex, race, and ethnicity

ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; BMI, body mass index; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HDL-c, high density lipoprotein cholesterol; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LDL-c, low density lipoprotein cholesterol; MI, myocardial infarction; SBP, systolic blood pressure.

DOI: 10.1002/ehf2.13429

smoking status, atrial fibrillation, and eGFR, while the HFrEF model additionally included previous CAD. Age, BMI, SBP, and eGFR were modelled non-linearly for both HFpEF and HFrEF, each using second degree polynomials. In both models, we tested for inclusion of interactions terms for each risk factor with age, but none increased C-index by more than 0.002. The order in which variables were added during variable selection and the corresponding C-index for each model are shown in the Supporting Information, *Figure S1*. Race-specific and sex-specific regression coefficients for each model are shown in the Supporting Information, *Table S3*.

In the validation cohort, the HFpEF model C-indices were 0.74 (SE = 0.002) for white men, 0.76 (0.005) for black men, 0.79 (0.015) for white women, and 0.77 (0.026) for black women, and 0.72 (0.002), 0.72 (0.004), 0.77 (0.017), and 0.75 (0.028), respectively, for the HFrEF model (*Table 3*). In this population, our model discriminated better than the ICHFS HFpEF model, which had C-indices of 0.66 (0.003), 0.71 (0.006), 0.78 (0.015), and 0.73 (0.025) for white men, black men, white women, and black women, respectively. We could not assess the performance of the ICHFS HFrEF model because it included left ventricular hypertrophy (LVH) and left bundle branch block (LBBB), variables that were not available in our dataset. Our risk prediction models were generally well calibrated in each race-specific and sex-specific stratum of the internal validation cohort (*Figure 1*). We developed an online tool that calculates one's predicted 10 year risk of HFrEF and HFpEF (https://bos-mav.github.io/HFRiskCalc/).

The ICHFS HFpEF model appeared to underpredict risk, particularly in black men and women (*Figure 1A*). For black men, the ICHFS HFpEF model had a mean predicted 10 year risk of 5.2% in the top decile, while mean observed 10 year risk was 9.3%. For black women the mean predicted 10 year risk was 2.5% in the top quintile compared with a mean observed 10 year risk of 3.8%.

Predicted 10 year risk for heart failure with preserved ejection fraction versus heart failure with reduced ejection fraction

Distributions of 10 year risk of HFpEF and HFrEF are shown in *Figure 2*. While predicted 10 year risk estimates were strongly correlated between the HF subtypes (Pearson's $r = 0.86$ in the whole validation cohort), there were differences in absolute risk. Forty-five per cent of white men and 52% of black men with predicted 10 year risk of HFrEF over 5% had HFpEF risk below 5%. For white and black women, 66% and 77%, respectively, of those with HFpEF risk above 5% had HFrEF risk below 5% (Supporting Information, *Table S7*).

Discussion

Using an EHR-based cohort of over 1.9 million individuals, we quantified the associations of several risk factors with incident HFpEF and HFrEF events, identifying quantitative differences in the magnitude of estimates. Furthermore, we developed the first race-specific and sex-specific risk prediction models for HFpEF and HFrEF, using a number of clinical risk factors routinely recorded in the EHR that will enable easy applicability.

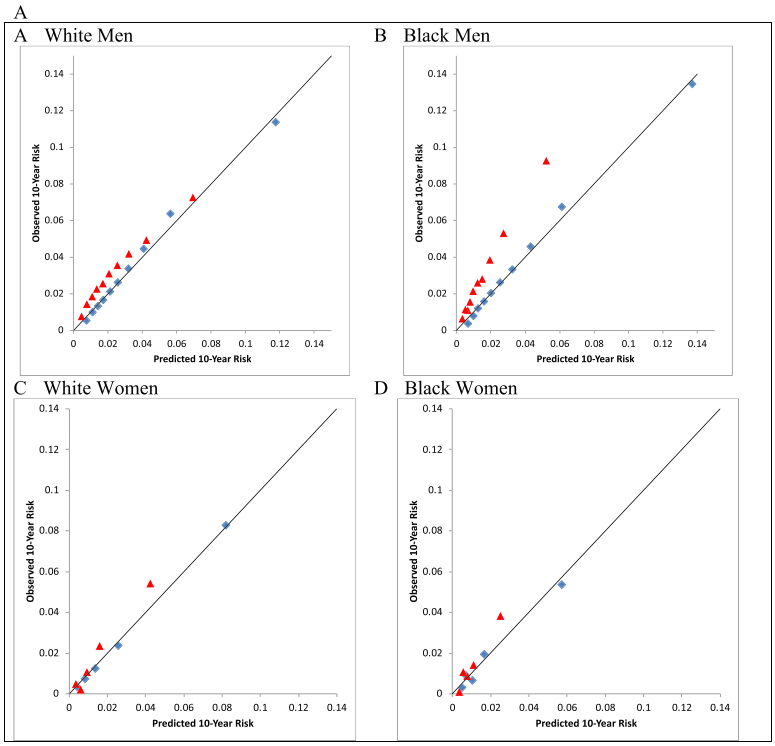
Previous prospective studies examining risk factors for incident HF subtypes[10,11,26–30] have been moderately powered and mostly confined to the four prospective studies that make up the International Collaboration on Heart Failure Subtypes (ICHFS).[10] In analysis of >100 000 incident HF events, we found results that were generally concordant with those previous purpose-designed cohorts in relation to direction and magnitude for traditional cardiovascular risk factors with HF subtypes (Supporting Information, *Table S8*), although our study was able to document these associations with greater precision. The similarity of findings between our

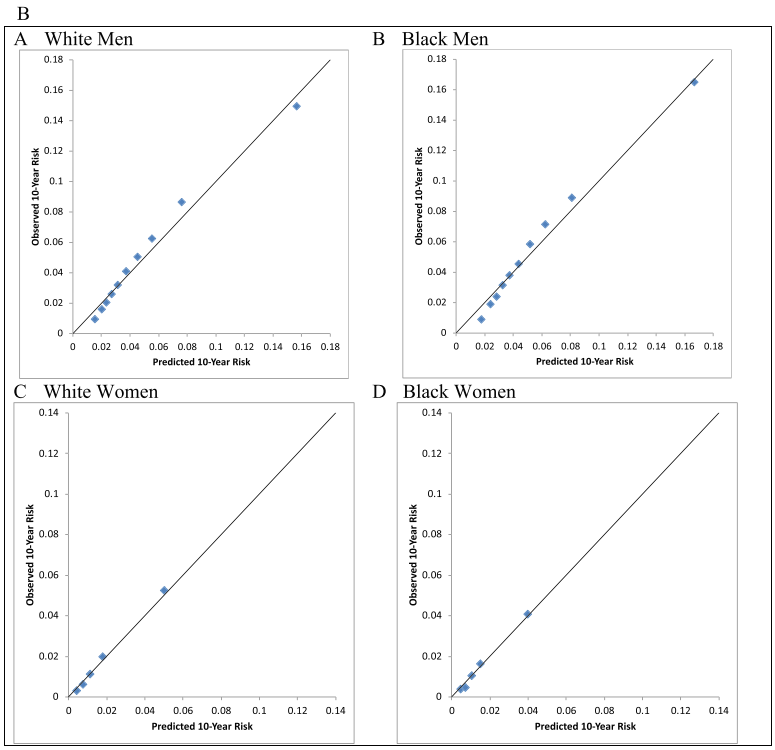
Table 3 C-indices (SE) for HFpEF and HFrEF risk prediction models

	Derivation cohort			Validation cohort		
	N total	Events	C-index (SE)	N total	Events	C-index (SE)
HFpEF						
White men	974 622	37 349	0.730 (0.001)	393 269	10 541	0.739 (0.002)
Black men	150 578	6978	0.753 (0.003)	73 978	1968	0.764 (0.005)
White women	27 899	963	0.790 (0.007)	13 911	242	0.793 (0.015)
Black women	8674	254	0.775 (0.014)	6074	81	0.772 (0.026)
HFrEF						
White men	974 622	51 153	0.715 (0.001)	393 269	15 327	0.724 (0.002)
Black men	150 578	9772	0.706 (0.003)	73 978	3235	0.719 (0.004)
White women	27 899	635	0.761 (0.009)	13 911	212	0.769 (0.017)
Black women	8674	202	0.728 (0.019)	6074	80	0.745 (0.028)

In HFpEF validation cohort, model includes age, diabetes, BMI, COPD, previous MI, antihypertensive treatment, SBP, smoking status, atrial fibrillation, and eGFR. In HFrEF validation cohort, model includes age, diabetes, BMI, COPD, previous MI, previous CAD, antihypertensive treatment, SBP, smoking status, atrial fibrillation and eGFR

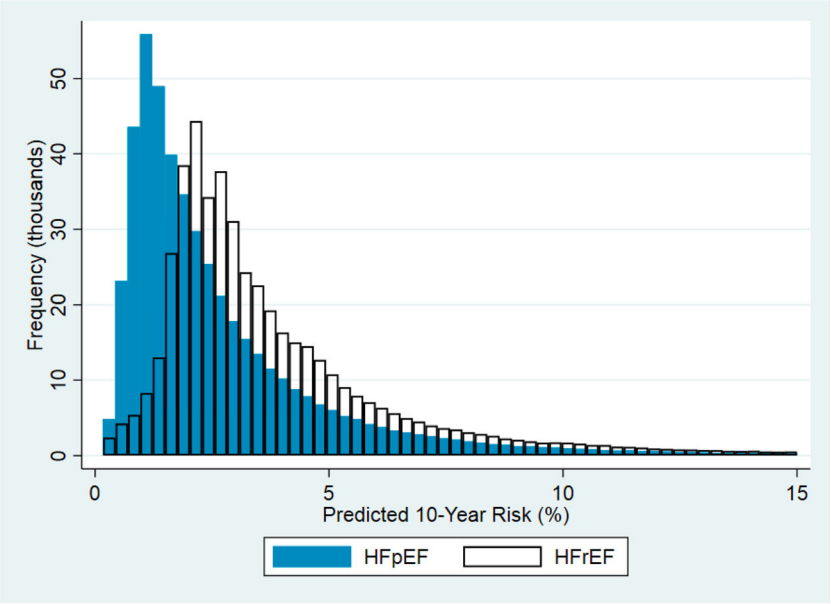
Figure 1 (A) Calibration plots in the validation cohort for HFpEF for (a) white men, (b) black men, (c) white women, and (d) black women. Mean predicted 10 year risk of HFpEF within each quantile (deciles for men and quintiles for women) against observed 10 year risk derived from Kaplan–Meier curves. Blue diamonds represent the current analysis, red triangles represent results using the International Collaboration on Heart Failure Subtypes model variables, weights (beta-coefficients), and baseline survival. (B) Calibration in the validation cohort for HFrEF for (a) white men, (b) black men, (c) white women, and (d) black women. Mean predicted 10 year risk of HFrEF within each quantile (deciles for men and quintiles for women) against observed 10 year risk derived from Kaplan–Meier curves.





DOI: 10.1002/ehf2.13429

Figure 2 Distributions of predicted 10 year risk of HFpEF and HFrEF in the validation cohort. Ten year risk of HFpEF and HFrEF was estimated using race-specific and sex-specific weights and baseline survival. Blue and white bars represent HFpEF and HFrEF, respectively. Values of predicted 10 year risk were truncated at 15%.



EHR-based cohort and purpose-designed cohorts may indicate that the risk factors for HF subtypes are reliably captured by EHRs and potentially generalizable across populations.

For many risk factors, the present study is the first or largest prospective study to assess relationships with incident HFrEF and HFpEF. While triglycerides levels were positively associated with both HFpEF and HFrEF, LDL-c levels were inversely associated with both outcomes. The meaning of this LDL-c association, which was also reported for total incident HF by a Danish cohort of 113 554 individuals,[31] is not clear, but it could be due to confounding by statin usage. A previous study has reported that lower serum sodium at time of diagnosis can predict mortality and hospitalization in those with existing HFpEF,[21] and our study observed an inverse association between sodium and incident events as well. Our study has also noted associations of less well studied risk factors with incident HFpEF and HFrEF, including COPD, anaemia, eGFR, and specific blood pressure medications.

A key advantage of the present study's statistical power was the ability to detect quantitative differences in risk factor profiles between HFrEF and HFpEF. While there were similarities in the direction of associations, there are notable differences in the magnitude of associations for a number of important cardiometabolic risk factors, like sex, SBP, antihypertensive medications, BMI, previous CAD/MI, and COPD. Our results add to the existing evidence that HFrEF and HFpEF may result from distinct pathophysiologic processes.[4] It is likely that CAD through myocardial injury or hibernation is a causal precedent for HFrEF. In contrast, some have hypothesized a mechanism for HFpEF in which obesity, diabetes, and COPD induce a proinflammatory reaction in the myocardium, which may lead to HFpEF.[32]

While multiple risk prediction models exist for total HF,[33–35] only two previous papers have developed models for HF subtypes. One[11] relied exclusively on CMS data and had a high proportion (62%) of unspecified HF and the other[10] combined data from the ICHFS cohorts. Our HFpEF model contained all of the ICHFS HFpEF model variables, as well as diabetes, COPD, smoking status, atrial fibrillation, and eGFR. The additional variables and regression coefficients more tailored to our population lead to better discrimination than the ICHFS HFpEF model in each race-specific and sex-specific stratum. The ICHFS HFrEF model and our HFrEF model were similar in that both contained age, SBP, BMI, antihypertensive treatment, diabetes, smoking status, and previous MI. However, the ICHFS HFrEF model additionally included LVH and LBBB, which were not available in our dataset, and our HFrEF model additionally included COPD, previous CAD, atrial fibrillation, and eGFR, none of which were assessed in their analysis. Our study was designed to use clinical variables that are available in any healthcare system and are not indicative of already present heart disease (such as LBBB and LVH). Therefore, we did not include biomarkers such as high sensitivity troponin and natriuretic peptides that have been previously evaluated in purpose-designed prospective cohort studies and shown to improve discrimination.[29]

Stratifying by race and sex appears to be important for accurate prediction of HFpEF and HFrEF. Because the ICHFS

models were derived from cohorts comprising almost 95% white European ancestry, this earlier work could not evaluate the impact of stratification by race and sex. Consequently, the ICHFS HFpEF model underpredicted risk in black men and women, which in the clinical setting, may lead to undertreatment and poorer outcomes. Because the burden of HF is generally greater in black individuals than white individuals in both our cohort and previous studies,[36,37] it is important to have accurate prediction of HF to inform clinical decisions in these understudied populations.

We found that over 45% of men with high predicted 10 year risk (above 5%) of developing HFrEF had low predicted 10 year risk (below 5%) of HFpEF, and over 65% of women with high predicted 10 year risk of HFpEF had low predicted 10 year risk of HFrEF, underscoring the need for subtype-specific risk prediction models. This is not surprising given that, although the HFpEF and HFrEF models contained almost the same variables, the relative importance (i.e. the order in which the variables were added to models) and weighting for some differed considerably. ACE inhibitors (ACEI),[38] angiotensin-receptor blockers (ARB),[39] angiotensin receptor neprilysin inhibitors (ARNI),[40,41] beta-blockers,[42] mineralocorticoid receptor antagonists (MRA),[43,44] and SGLT2 inhibitors[45] reduce morbidity and mortality in those with overt HFrEF, but no treatments have been definitely proven to benefit HFpEF. Hence, prevention of HFpEF is crucial to lessen its burden. Previous trials have shown that HF prevention is possible,[46–49] but given the differential effects of therapies in overt HFpEF and HFrEF, it is likely that primary prevention requires subtype-specific approaches as well. Our risk models would allow for efficient screening using the EHR to identify high-risk individuals who may be good candidates for primary prevention trials of each subtype at relatively low cost because they use variables already routinely measured in clinical practice.

Strengths and limitations

There are a number of strengths of this study. It is the largest investigation of incident events of HF subtypes, and by using the rich EHR data in the VA, we were able to examine a wide range of potential HF risk factors. We were able to determine who had active prescriptions at baseline, including information on specific antihypertensive medications. By analysing the EHR data, we did not rely on any self-reported definitions, which can inadequately characterize predictors,[50,51] and HF diagnosis.[52] Because our baseline dates ranged from 2002 to 2007, we have a more modern cohort than previous studies, more relevant to the current clinical setting. Lastly, based on our established expertise in curating and subtyping HF in the VA databases,[53] we were able to accurately capture incident cases of HFpEF and HFrEF, including those that happened outside the VA though linkage with CMS databases.

The major limitation of our study is that we could not externally validate our risk prediction models in fully independent populations, primarily because no other large-scale dataset includes phenotyped HF subtypes and all the variables in our risk prediction models. The relative proportion of women and minorities was low, but this resource included a larger number of black men, white women, and black women than any previous study looking at incident HFrEF and HFpEF. Excluding individuals with missing variables in the models could introduce selection bias, but those individuals made up a small proportion of the dataset (13%) and had similar traditional cardiovascular risk factors to those included in the risk prediction analysis (Supporting Information, *Table S9*).

Conclusions

Using a large EHR database, we identified quantitatively different associations of several cardiometabolic exposures with HFpEF and HFrEF. For HF subtypes, this EHR-based cohort behaved similarly to traditional prospective cohorts, possibly indicating that its generalizability could extend to healthcare systems outside the VA. We developed the first race-specific and sex-specific models for HFrpEF and HFrEF, which performed well in our diverse population and used variables easily obtainable in clinical practice. A large proportion of men with high risk of HFrEF had low risk of HFpEF and vice versa for women, indicating a need for subtype specific models. These risk prediction models could help to identify individuals at high-risk of future HF events for clinical intervention or prevention trials in the nine million regular users of the VA or in the general population once externally validated in independent datasets. An online tool was developed to allow for implementation of this risk calculator in the clinical setting.

Acknowledgements

This research was supported by VA Merit Awards I01 CX001025 and I01 CX001922-01. This publication does not represent the views of the Department of Veterans Affairs or the U.S. Government. Support for VA/CMS data provided by the Department of Veterans Affairs, VA Health Services Research and Development Service, VA Information Resource Center (Project Numbers SDR 02-237 and 98-004).

Conflict of interest

The authors have no conflicts of interest to declare.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Exclusion Criteria for Comorbidities at Baseline.
Table S2. Definitions of Medications.
Figure S1a. Order in which variables were added to the HFrEF model during variable selection, and C-index after each addition.
Figure S1b. Order in which variables were added to the HFrEF model during variable selection, and C-index after each addition.
Table S3a. Race- and Sex-Specific Regression Coefficients for HFpEF Risk Prediction Model.
Table S3b. Race- and Sex-Specific Regression Coefficients for HFrEF Risk Prediction Model.
Table S4. Race- and Sex-Specific Means of Transformed Continuous Variables in HFpEF and HFrEF Models. These were used to mean-center continuous variables.
Table S5. Race- and Sex-Specific Baseline 10-year Survival.
Table S6. Baseline Characteristics in the Validation Cohort.
Table S7a. 10-Year Risk Categories of HFpEF versus HFrEF for each race- and sex-specific stratum. Each cell represents percentage of the race- and sex-specific stratum of the validation cohort.
Table S7b. 10-Year Risk Categories of HFpEF versus HFrEF in those aged 40 to 60 for each race- and sex-specific stratum. Each cell represents percentage of the race- and sex-specific stratum of the validation cohort.
Table S7c. 10-Year Risk Categories of HFpEF versus HFrEF in those aged 60 to 80 for each race- and sex-specific stratum. Each cell represents percentage of the race- and sex-specific stratum of the validation cohort.
Table S8a. Risk Factors for HFpEF in the ICHFS Derivation Cohorts and the VA Derivation Cohort.
Table S8b. Risk Factors for HFrEF in the ICHFS Derivation Cohorts and the VA Derivation Cohort.
Table S9. Baseline characteristics for individuals included in the risk prediction analysis versus those excluded due to missingness.

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, Djousse L, Elkind MSV, Ferguson JF, Fornage M, Khan SS, Kissela BM, Knutson KL, Kwan TW, Lackland DT, Lewis TT, Lichtman JH, Longenecker CT, Loop MS, Lutsey PL, Martin SS, Matsushita K, Moran AE, Mussolino ME, Perak AM, Rosamond WD, Roth GA, Sampson UKA, Satou GM, Schroeder EB, Shah SH, Shay CM, Spartano NL, Stokes A, Tirschwell DL, VanWagner L, Tsao CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation* 2020; **141**: e139–e596.
2. Roger VL, Weston SA, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ. Trends in heart failure incidence and survival in a community-based population. *JAMA* 2004; **292**: 344–350.
3. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, Ikonomidis JS, Khavjou O, Konstam MA, Maddox TM, Nichol G, Pham M, Piña IL, Trogdon JG. American Heart Association Advocacy Coordinating Committee, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Radiology and Intervention, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circ Heart Fail* NIH Public Access 2013; **6**: 606–619.
4. Borlaug BA, Redfield MM. Diastolic and systolic heart failure are distinct phenotypes within the heart failure spectrum. *Circulation* 2011; **123**: 2006–2014.
5. Fukuta H, Little WC. Contribution of systolic and diastolic abnormalities to heart failure with a normal and a reduced ejection fraction. *Prog Cardiovasc Dis* 2007; **49**: 229–240.
6. Emdin CA, Callender T, Cao J, McMurray JJV, Rahimi K. Meta-analysis of large-scale randomized trials to determine the effectiveness of inhibition of the renin-angiotensin aldosterone system in heart failure. *Am J Cardiol* 2015; **116**: 155–161.
7. Campbell RT, Jhund PS, Castagno D, Hawkins NM, Petrie MC, McMurray JJV. What have we learned about patients with heart failure and preserved ejection fraction from DIG-PEF, CHARM-preserved, and I-preserve? *J Am Coll Cardiol* 2012; **60**: 2349–2356.
8. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, Hollenberg SM, Lindenfeld J, Masoudi FA, McBride PE, Peterson PN, Stevenson LW, Westlake C. 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.
9. Shah SJ, Borlaug BA, Kitzman DW, McCulloch AD, Blaxall BC, Agarwal R, Chirinos JA, Collins S, Deo RC, Gladwin MT, Granzier H, Hummel SL, Kass DA, Redfield MM, Sam F, Wang TJ, Desvigne-Nickens P, Adhikari BB. Research priorities for heart failure with preserved ejection fraction. *Circulation* 2020; **141**: 1001–1026.
10. Ho JE, Enserro D, Brouwers FP, Kizer JR, Shah SJ, Psaty BM, Bartz TM, Santhanakrishnan R, Lee DS, Chan C, Liu K, Blaha MJ, Hillege HL. Predicting heart failure with preserved and reduced ejection fraction: the international collaboration on heart failure subtypes. *Circ Heart Fail* 2016; **9**: e003116.
11. Lee MP, Glynn RJ, Schneeweiss S, Lin KJ, Patorno E, Barberio J, Levin R, Evers T, Wang SV, Desai RJ. Risk factors for heart failure with preserved or reduced ejection fraction among Medicare beneficiaries: application of competing risks analysis and gradient boosted model. *Clin Epidemiol* 2020; **12**: 607–616.
12. Vassy JL, Lu B, Ho YL, Galloway A, Raghavan S, Honerlaw J, Tarko L, Russo J, Qazi S, Orkaby AR, Tanukonda V,

Djousse L, Gaziano JM, Gagnon DR, Cho K, Wilson PWF. Estimation of atherosclerotic cardiovascular disease risk among patients in the Veterans Affairs Health Care System. *JAMA Netw open NLM (Medline)* 2020; **3**: e208236.
13. Vassy JL, Ho Y-L, Honerlaw J, Cho K, Gaziano JM, Wilson PWF, Gagnon DR. Yield and bias in defining a cohort study baseline from electronic health record data. *J Biomed Inform* 2018; **78**: 54–59.
14. Price LE, Shea K, Gephart S. The Veterans Affairs's corporate data warehouse. *Nurs Adm Q* 2015; **39**: 311–318.
15. Center of Excellence for Suicide Prevention. Joint Department of Veterans Affairs (VA) and Department of Defense (DoD) Suicide Data Repository – National Death Index (NDI). 2019. http://vaww.virec.research.va.gov/Mortality/Overview.htm. (Accessed: 24 July 2019).
16. Raghavan S, Vassy JL, Ho Y, Song RJ, Gagnon DR, Cho K, Wilson PWF, Phillips LS. Diabetes mellitus–related all-cause and cardiovascular mortality in a national cohort of adults. *J Am Heart Assoc* 2019; **8**: e011295.
17. Song RJ, Ho YL, Nguyen XMT, Honerlaw J, Quaden R, Gaziano JM, Concato J, Cho K, Gagnon DR. Development of an electronic health record-based algorithm for smoking status using the million veteran program (MVP) cohort survey response. *Circulation* 2016; **134**: A18809.
18. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, Kusek JW, Manzi J, van Lente F, Zhang YL, Coresh J, Levey AS, Investigators CKD-EPI. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* 2012; **367**: 20–29.
19. Patterson OV, Freiberg MS, Skanderson M, Fodeh S, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. *BMC Cardiovasc Disord BioMed Central* 2017; **17**: 151.
20. Freiberg MS, Chang CCH, Skanderson M, Patterson OV, DuVall SL, Brandt CA, So-Armah KA, Vasan RS, Oursler KA, Gottdiener J, Gottlieb S, Leaf D, Rodriguez-Barradas M, Tracy RP, Gibert CL, Rimland D, Bedimo RJ, Brown ST, Goetz MB, Warner A, Crothers K, Tindle HA, Alcorn C, Bachmann JM, Justice AC, Butt AA. Association between HIV infection and the risk of heart failure with reduced ejection fraction and preserved ejection fraction in the antiretroviral therapy era: results from the veterans aging cohort study. *JAMA Cardiol* American Medical Association 2017; **2**: 536–546.
21. Patel YR, Kurgansky KE, Imran TF, Orkaby AR, McLean RR, Ho Y, Cho K, Gaziano JM, Djousse L, Gagnon DR, Joseph J. Prognostic significance of baseline serum sodium in heart failure with preserved ejection fraction. *J Am Heart Assoc* 2018; **7**.
22. Nauta JF, Hummel YM, van Melle JP, van der Meer P, Lam CSP, Ponikowski P, Voors AA. What have we learned about heart failure with mid-range ejection fraction one year after its introduction? *Eur J Heart Fail* 2017; **19**: 1569–1573.
23. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982; **247**: 2543–2546.
24. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**: 361–387.
25. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004; **23**: 2109–2123.
26. Brouwers FP, de Boer RA, van der Harst P, Voors AA, Gansevoort RT, Bakker SJ, Hillege HL, van Veldhuisen DJ, van Gilst WH. Incidence and epidemiology of new onset heart failure with preserved vs. reduced ejection fraction in a community-based cohort: 11-year follow-up of PREVEND. *Eur Heart J* 2013; **34**: 1424–1431.
27. Ho JE, Lyass A, Lee DS, Vasan RS, Kannel WB, Larson MG, Levy D. Predictors of new-onset heart failure. *Circ Heart Fail* 2013; **6**: 279–286.
28. Seliger SL, de Lemos J, Neeland IJ, Christenson R, Gottdiener J, Drazner MH, Berry J, Sorkin J, deFilippi C. Older adults, 'malignant' left ventricular hypertrophy, and associated cardiac-specific biomarker phenotypes to identify the differential risk of new-onset reduced versus preserved ejection fraction heart failure. *JACC Heart Fail* 2015; **3**: 445–455.
29. de Boer RA, Nayor M, de Filippi CR, Enserro D, Bhambhani V, Kizer JR, Blaha MJ, Brouwers FP, Cushman M, Lima JAC, Bahrami H, van der Harst P, Wang TJ, Gansevoort RT, Fox CS, Gaggin HK, Kop WJ, Liu K, Vasan RS, Psaty BM, Lee DS, Hillege HL, Bartz TM, Benjamin EJ, Chan C, Allison M, Gardin JM, Januzzi JL, Shah SJ, Levy D. Association of cardiovascular biomarkers with incident heart failure with preserved and reduced ejection fraction. *JAMA Cardiol* 2018; **3**: 215.
30. Myhre PL, Claggett B, Ballantyne CM, Selvin E, Røsjø H, Omland T, Solomon SD, Skali H, Shah AM. Association between circulating troponin concentrations, left ventricular systolic and diastolic functions, and incident heart failure in older adults. *JAMA Cardiol* 2019.
31. Varbo A, Nordestgaard BG. Nonfasting triglycerides, low-density lipoprotein cholesterol, and heart failure risk. *Arterioscler Thromb Vasc Biol* 2018; **38**: 464–472.
32. Paulus WJ, Tschöpe C. A novel paradigm for heart failure with preserved ejection fraction. *J Am Coll Cardiol* 2013; **62**: 263–271.
33. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, Mentz RJ, O'Brien E, Correa A, Suthahar N, de Boer RA, Wilkins JT, Lloyd-Jones DM. 10-year risk equations for incident heart failure in the general population. *J Am Coll Cardiol* 2019; **73**: 2388–2397.
34. Butler J, Kalogeropoulos A, Georgiopoulou V, Belue R, Rodondi N, Garcia M, Bauer DC, Satterfield S, Smith AL, Vaccarino V, Newman AB, Harris TB, Wilson PWF, Kritchevsky SB, Health ABC Study. Incident heart failure prediction in the elderly. *Circ Heart Fail* 2008; **1**: 125–133.
35. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, Folsom AR, He M, Hoogeveen RC, Ni H, Quibrera PM, Rosamond WD, Russell SD, Shahar E, Heiss G. Prediction of incident heart failure in general practice. *Circ Heart Fail* 2012; **5**: 422–429.
36. Bibbins-Domingo K, Pletcher MJ, Lin F, Vittinghoff E, Gardin JM, Arynchyn A, Lewis CE, Williams OD, Hulley SB. Racial differences in incident heart failure among young adults. *N Engl J Med* 2009; **360**: 1179–1190.
37. Bahrami H, Kronmal R, Bluemke DA, Olson J, Shea S, Liu K, Burke GL, Lima JAC. Differences in the incidence of congestive heart failure by ethnicity. *Arch Intern Med* 2008; **168**: 2138–2145.
38. Flather MD, Yusuf S, Køber L, Pfeffer M, Hall A, Murray G, Torp-Pedersen C, Ball S, Pogue J, Moyé L, Braunwald E. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. ACE-inhibitor myocardial infarction collaborative group. *Lancet (London, England)* 2000; **355**: 1575–1581.
39. Cohn JN, Tognoni G, Valsartan Heart Failure Trial Investigators. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. *N Engl J Med* 2001; **345**: 1667–1675.
40. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K, Zile MR, PARADIGM-HF Investigators and Committees. Angiotensin–neprilysin inhibition versus enalapril in heart failure. *N Engl J Med* 2014; **371**: 993–1004.
41. Velazquez EJ, Morrow DA, DeVore AD, Duffy CI, Ambrosy AP, McCague K, Rocha R, Braunwald E. Angiotensin–neprilysin inhibition in acute decompensated heart failure. *N Engl J Med* 2019; **380**: 539–548.
42. Packer M, Coats AJS, Fowler MB, Katus HA, Krum H, Mohacsi P, Rouleau JL, Tendera M, Castaigne A, Roecker EB, Schultz MK, Staiger C, Curtin EL, DeMets DL. Effect of carvedilol on

survival in severe chronic heart failure. *N Engl J Med* 2001; **344**: 1651–1658.
43. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N Engl J Med* 1999; **341**: 709–717.
44. Zannad F, McMurray JJV, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pocock SJ, Pitt B, EMPHASIS-HF Study Group. Eplerenone in patients with systolic heart failure and mild symptoms. *N Engl J Med* 2011; **364**: 11–21.
45. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Ponikowski P, Sabatine MS, Anand IS, Bělohlávek J, Böhm M, Chiang C-E, Chopra VK, de Boer RA, Desai AS, Diez M, Drozdz J, Dukát A, Ge J, Howlett JG, Katova T, Kitakaze M, Ljungman CEA, Merkely B, Nicolau JC, O'Meara E, Petrie MC, Vinh PN, Schou M, Tereshchenko S. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med* 2019: NEJMoa1911303.
46. Ledwidge M, Gallagher J, Conlon C, Tallon E, O'Connell E, Dawkins I, Watson C, O'Hanlon R, Bermingham M, Patle A, Badabhagni MR, Murtagh G, Voon V, Tilson L, Barry M, McDonald L, Maurer B, McDonald K. Natriuretic peptide–based screening and collaborative care for heart failure. *JAMA* 2013; **310**: 66–74.
47. Moser M, Hebert PR. Prevention of disease progression, left ventricular hypertrophy and congestive heart failure in hypertension treatment trials. *J Am Coll Cardiol* 1996; **27**: 1214–1218.
48. Kostis JB, Davis BR, Cutler J, Grimm RH, Berge KG, Cohen JD, Lacy CR, Perry HM, Blaufox MD, Wassertheil-Smoller S, Black HR, Schron E, Berkson DM, Curb JD, Smith WM, McDonald R, Applegate WB. Prevention of heart failure by antihypertensive drug treatment in older persons with isolated systolic hypertension. SHEP Cooperative Research Group. *JAMA* 1997; **278**: 212–216.
49. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, Zelniker TA, Kuder JF, Murphy SA, Bhatt DL, Leiter LA, McGuire DK, Wilding JPH, Ruff CT, Gause-Nilsson IAM, Fredriksson M, Johansson PA, Langkilde A-M, Sabatine MS. DECLARE–TIMI 58 Investigators. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2019; **380**: 347–357.
50. Gonçalves VSS, Andrade KRC, Carvalho KMB, Silva MT, Pereira MG, Galvao TF. Accuracy of self-reported hypertension. *J Hypertens* 2018; **36**: 970–978.
51. Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. *Am J Prev Med* 1999; **17**: 211–229.
52. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004; **57**: 1096–1103.
53. Patel YR, Robbins JM, Kurgansky KE, Imran T, Orkaby AR, McLean RR, Ho Y-L, Cho K, Michael Gaziano J, Djousse L, Gagnon DR, Joseph J. Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. *BMC Cardiovasc Disord* 2018; **18**: 128.