ESC HEART FAILURE
ESC Heart Failure 2021; **8**: 2089–2102
Published online 11 March 2021 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13288

ORIGINAL RESEARCH ARTICLE

Diagnostic and prognostic implications of heart failure with preserved ejection fraction scoring systems

Vibhu Parcha[1], Gargya Malla[2], Rajat Kalra[3], Nirav Patel[4], Sandra Sanders-van Wijk[5], Ambarish Pandey[6], Sanjiv J. Shah[7], Garima Arora[1] and Pankaj Arora[1,8]*

[1]*Division of Cardiovascular Disease, University of Alabama at Birmingham, Birmingham, AL, USA;* [2]*Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA;* [3]*Cardiovascular Division, University of Minnesota, Minneapolis, MN, USA;* [4]*Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA;* [5]*Cardiovascular Research Institute Maastricht, Maastricht University Medical Centre (MUMC+), Maastricht, The Netherlands;* [6]*Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA;* [7]*Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; and* [8]*Section of Cardiology, Birmingham Veterans Affairs Medical Center, Birmingham, AL, USA*

Abstract

Aims We sought to compare the generalizability and prognostic implications of heart failure with preserved ejection fraction (HFpEF) scores (HFA-PEFF and H_2FPEF score) in Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) and Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Heart Failure with Preserved Ejection Fraction (RELAX) trial participants and matched controls from the Atherosclerosis Risk in Community (ARIC) study.

Methods and results Based on the respective scores, the study participants from the TOPCAT (*N* = 356), RELAX (*N* = 216), and ARIC (*N* = 379) studies were categorized as having a low, intermediate, or high likelihood of HFpEF. Age, sex, and race matched controls free of cardiovascular disease who had unexplained dyspnoea were used to evaluate the diagnostic performance. The prognostic value of scores was assessed using multivariable-adjusted Cox regression analyses. The median HFA-PEFF scores in the TOPCAT, RELAX, and ARIC studies were 5.0 [interquartile range (IQR): 5.0–6.0], 4.0 (IQR: 2.0–4.0), and 3.0 (IQR: 2.0–4.0), respectively. The median H_2FPEF scores in the three studies were 5.5 (IQR: 4.0–7.0), 6.0 (IQR: 4.0–7.0), and 3.0 (IQR: 2.0–5.0), respectively. A low HFA-PEFF and H_2FPEF score can rule out HFpEF with high sensitivity (99.5% and 99.6%, respectively) and negative predictive value (95.7% and 98.3%, respectively). A high HFA-PEFF and H_2FPEF score can rule-in HFpEF with good specificity (82.8% and 95.6%, respectively) and positive predictive value (79.9% and 90.4%, respectively). Among TOPCAT participants, the hazard for adverse cardiovascular events per point increase in HFA-PEFF and H_2FPEF score was 1.26 (95% confidence interval: 0.98–1.63) and 1.01 (95% confidence interval: 0.88–1.15), respectively. A higher H_2FPEF score was associated with lower peak oxygen intake in RELAX trial participants (adjusted *P* = 0.01).

Conclusions The HFA-PEFF and the H_2FPEF scores are reliable diagnostic tools for HFpEF. The prognostic utility of HFpEF scores requires further validation in larger rigorously phenotyped populations.

Keywords Cardiovascular outcomes; Diagnosis; Heart failure; Heart failure preserved ejection fraction; Prognosis

Received: 16 August 2020; Revised: 6 January 2021; Accepted: 18 February 2021
**Correspondence to: Pankaj Arora, Division of Cardiovascular Disease, 1670 University Boulevard, Volker Hall B140, University of Alabama at Birmingham, Birmingham, AL, 35294-0019, USA. Tel: 205-996-6630; Fax: 205-975-4720. Email: parora@uabmc.edu*

Introduction

Heart failure with preserved ejection fraction (HFpEF) accounts for over half of heart failure admissions.[1,2] With an aging population that has multiple coexisting comorbidities, the prevalence of HFpEF is likely to rise in the coming years.[1,2] Accurately diagnosing the heterogeneous disease state of HFpEF has been a significant limitation in providing effective therapeutic management for the patients.

Despite advances in the understanding of the HFpEF pathophysiology and the availability of imaging and cardiac biomarker assessment,[3–6] debate still exists regarding the roles of cardiac imaging parameters, biomarker levels, and comorbidity status for the diagnosis and prognostication of

HFpEF.[3,5–7] This has led to the development of two scoring systems for the evaluation of patients with dyspnoea to risk stratify their likelihood of having HFpEF: the H_2FPEF score[8] and the Heart Failure Association pre-test assessment, echocardiography and natriuretic peptide, functional testing, and final aetiology (HFA-PEFF)[9] diagnostic algorithm. The scoring systems differ in their usage of echocardiographic cut-offs, biomarker inclusion, invasive haemodynamic assessment, and the role of exercise testing.[8–10] Data suggest that these relatively new scoring systems have prognostic utility in addition to their diagnostic efficacy for HFpEF.[10,11] However, there are limited data comparing the diagnostic and prognostic efficacy of the two HFpEF scoring systems.

The participants of the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT)[12] and Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Heart Failure with Preserved Ejection Fraction (RELAX)[13] trials provide a unique opportunity to evaluate the application of the two scores in the HFpEF trial population and to examine their relationship with adverse cardiovascular outcomes. We sought to evaluate the generalizability of the HFA-PEFF and H_2FPEF scoring systems and their comparative prognostic utility among the TOPCAT and RELAX trial participants. We also compared the diagnostic utility of the scores in age-matched, sex-matched, and race-matched participants of the Atherosclerosis Risk in Communities (ARIC) study without any history of cardiovascular diseases.

Methods

The data for this study was obtained from the publicly available National Institute of Heart, Lung and, Blood Institute BioLINCC data repository. All participants of the TOPCAT and RELAX trials with available data were included for analysis. TOPCAT was a Phase 3, double-blinded multicentre randomized controlled trial that was conducted from August 2006 to January 2012 in the Americas, Russia, and the Republic of Georgia.[12] Due to previously reported concerns regarding enrolled patients in Russia/Georgia, we included only the TOPCAT patients from the Americas in our analysis. The RELAX trial was also a multicentre double-blinded randomized controlled trial that was conducted between October 2008 and February 2012 across the USA and Canada.[13,14] The ARIC study is a prospective cohort of US adults that were followed for cardiovascular events starting in 1987.[15] We included information from Visit 5 of this cohort (2011–2013), which is when the echocardiographic assessment was performed. The protocols were approved at the trial sites by the respective local institutional review boards.[12–14] All participants provided written, informed consent, and the studies complied with principles in the Declaration of Helsinki.

The details of the inclusion and exclusion criteria, trial methodology, and results have been published earlier and described in the supporting information methods.[12–14] Details of study population derivation are presented in supporting information *Figure S1*. The details of the ARIC study, including the identification controls in the ARIC, are described in supporting information methods. In brief, we performed 1:1 matching based on age (within 5 year range), race and sex, in ARIC participants who had unexplained dyspnoea without any history of cardiovascular disease (history of heart failure, coronary artery disease, or stroke).

HFA-PEFF score

The HFA-PEFF score, which is measured on a scale of 0–6 points, incorporates three domains—functional, morphological, and biomarker.[9,10] A high score (≥5 points) is considered diagnostic for HFpEF, and a low score (0–1 points) rules out the disease.[10] The score of 2–4 is intermediate, requiring further evaluation.[10] The additional steps, as described in the HFA-PEFF scoring algorithm of diastolic stress testing, invasive assessment, and etiological evaluation, were not assessed. The details of the HFA-PEFF score estimation have been described in supporting information methods and *Figure S2*. The comparison of the TOPCAT study population sub-group included in the current analysis with the rest of the trial population is summarized in *Table S1*.

H_2FPEF score

The details of the H_2FPEF score (0–9 points) estimation have been outlined in the supporting information methods and *Figure S3*. We modelled the score as both a discrete (low: 0–1 points, intermediate: 2–5 points, and high: ≥6 points) and continuous measure.[8]

Both HFA-PEFF and H_2FPEF scores were available in 264 participants of TOPCAT, 188 participants of RELAX, and 362 participants of the ARIC study. For prognostic evaluation of the scores, we used the primary endpoint of the TOPCAT trial, which was a composite of heart failure hospitalization, aborted cardiac arrest, or death due to cardiovascular causes. We also evaluated the relationship of the scores with the baseline peak VO_2 in the RELAX trial.

Statistical analysis

Baseline characteristics between the scoring strata were compared separately for the three populations. Median and interquartile range (IQR) summarized the continuous data and were compared using the Wilcoxon rank-sum test or the Kruskal–Wallis test. Categorical data were compared using the χ^2 test or Fisher's exact test and reported as counts

and percentages. Sensitivity, specificity, positive predictive value (PPV), and negative predictive values (NPV) were estimated using the trial populations as HFpEF cases and the ARIC participants without any history of cardiovascular diseases as controls. We employed two approaches for assessing the diagnostic performance: (i) rule-in approach of taking only the high score category as 'positive' and (ii) rule-out approach taking intermediate and high-score categories as 'positive'. The diagnostic area under the curve (AUC) was also generated for the two scores. Sensitivity analyses of the diagnostic performance of the two scores were performed in the ambulatory trial population and using matched controls from ARIC. Among the TOPCAT trial participants, Cox proportional hazards models were used to assess the prognostic ability of both HFA-PEFF and H_2FPEF scores for adverse cardiovascular events. Analyses were adjusted for age, sex, race (Whites vs. non-White), body mass index, enrolment strata (hospitalization vs. natriuretic peptides), and treatment group.[16,17] The Kaplan–Meier method was used to plot the cumulative incidence of cardiovascular events by score categories. The log-rank test was used to compare the outcome across the categories of the two scoring systems. Outcome analyses were limited to TOPCAT participants due to few cardiovascular events among RELAX participants. Receiver operating characteristic curves were generated, and the AUC and Harrell's C-statistics were computed by the inclusion of the respective scores in the Cox models to compare the prognostic ability of the two scoring systems. Additionally, multivariable-adjusted logistic regression was performed to obtain the odds of the primary outcome and heart failure hospitalization outcome in the TOPCAT population. In a secondary analysis, we evaluated the relationship of the scores with the baseline peak VO_2 in the RELAX trial, adjusted for age, sex, race, and body mass index. We also performed logistic regression to assess the association of high HFpEF scores with peak $VO_2 < 15$ mL/kg/min. All statistical analyses were conducted using SAS 9.4 (Cary, N.C.). A two-sided Type I error of 0.05 was considered significant for all the reported analyses.

Results

HFA-PEFF score in the study populations

Among the TOPCAT trial participants (N = 356), the median HFA-PEFF score was 5.0 (IQR: 5.0–6.0). This was similar across sex [male: 5.0 (IQR: 5.0–6.0) vs. female: 5.0 (IQR: 5.0–6.0); P = 0.59], race [Whites: 5.0 (IQR: 5.0–6.0) vs. non-Whites: 5.0 (IQR: 5.0–6.0); P = 0.92], and treatment strategy [spironolactone: 5.0 (IQR: 5.0–6.0) vs. placebo: 5.0 (IQR: 5.0–6.0); P = 0.80] strata. The scores were higher in those enrolled by natriuretic peptide criteria [5.0 (IQR: 5.0–6.0)] compared with those enrolled by hospitalization criteria [5.0 (IQR: 5.0–5.0); P = 0.04]. The baseline characteristics of the TOPCAT participants stratified by HFA-PEFF score categories are illustrated in *Table 1*. In TOPCAT, 19.7% of participants had an intermediate score[2–4] and 80.3% had a high score (≥5).

Among the RELAX trial participants (N = 216), the median HFA-PEFF score was 4.0 (IQR: 2.0–4.0). The baseline characteristics of RELAX participants stratified by HFA-PEFF score categories are illustrated in *Table 1*. The median scores were similar across sex [male: 4.0 (IQR: 2.0–4.0) vs. female: 4.0 (IQR: 2.0–5.0); P = 0.41], race [Whites: 4.0 (IQR: 2.0–5.0) vs. non-Whites: 4.0 (IQR: 2.0–4.0); P = 0.80], and treatment strategy [sildenafil: 4.0 (IQR: 2.0–4.0) vs. placebo: 4.0 (IQR: 3.0–4.0); P = 0.14] strata. While only 2.8% of participants had a low score (<2), 76.9% of participants had an intermediate score,[2–4] and 20.4% of participants had a high score (≥5). The distribution of the score domains for both trial populations is presented in *Figure 1*.

In the ARIC population (N = 362), a population with any history of atherosclerotic cardiovascular diseases, the median HFA-PEFF score was 3.0 (IQR: 2.0–4.0). The scores were different across sex [male: 3.0 (IQR: 2.0–4.0) vs. female: 3.0 (IQR: 3.0–4.0); P = 0.005] strata but not race [Whites: 3.0 (IQR: 2.0–4.0) vs. non-Whites: 3.0 (IQR: 2.0–4.0); P = 0.28] strata. ARIC participants mostly had intermediate HFA-PEFF scores (68.5%). In the remaining participants, 21.8% had a high score and 9.7% had a low score (*Table 1*).

H_2FPEF score in the study populations

The median score of the TOPCAT population (N = 214) was 5.5 (IQR: 4.0–7.0) and was different by sex [male: 6.0 (IQR: 5.0–8.0) vs. female: 5.0 (IQR: 4.0–7.0); P = 0.02], and race [Whites: 5.0 (IQR: 4.0–6.0) vs. non-Whites: 6.0 (IQR: 4.0–8.0); P = 0.005]. The scores were similar in those enrolled by natriuretic peptide criteria [6.0 (IQR: 4.0–7.0)] compared with those enrolled by hospitalization criteria [5.0 (IQR: 4.0–7.0); P = 0.88]. In TOPCAT, 1.1% of participants had a low score, 48.9% of participants had an intermediate score, and 50% of participants had a high score. The baseline characteristics of the TOPCAT population stratified by H_2FPEF score categories are described in *Table 2*.

In the RELAX population (N = 188), the median score was 6.0 (IQR: 4.0–7.0). The scores were similar across sex [male: 6.0 (IQR: 4.0–7.0) vs. female: 6.0 (IQR: 4.0–7.0); P = 0.97], race [Whites: 6.0 (IQR: 4.0–7.0) vs. non-Whites: 6.0 (IQR: 4.0–7.0); P = 0.87], and treatment strategy [sildenafil: 6.0 (IQR: 5.0–7.0) vs. placebo: 6.0 (IQR: 4.0–7.0); P = 0.42] strata. Among the participants with available H_2FPEF score, 2.7% had a low score, 42.0% had an intermediate score, and 55.3% had a high score. The baseline characteristics of the RELAX trial participants by H_2FPEF score categories are described in *Table 2*.

Table 1 Baseline characteristics: HFA-PEFF score categories

Characteristics	TOPCAT			RELAX			
	Intermediate (2–4 points) (N = 70)	High (≥5 points) (N = 286)	P value	Low (0–1 points) (N = 6)	Intermediate (2–4 points) (N = 166)	High (≥5 points) (N = 44)	P value
Age (years)	68 (59–76)	73 (64–80)	0.001	59.5 (51.0–67.0)	68.5 (62.0–77.0)	69.0 (64.0–77.0)	0.25
Female sex	39 (55.7)	128 (44.8)	0.10	1 (16.7)	87 (52.4)	16 (36.4)	0.057
White race	51 (72.9)	2020 (70.6)	0.71	6 (100.0)	153 (92.2)	38 (86.4)	0.41
BMI (kg/m^2)	35.7 (28.6–42.7)	32.6 (27.7–37.8)	0.03	36.9 (33.3–39.8)	33.1 (28.3–39.0)	33.5 (29.9–37.0)	0.65
Systolic blood pressure (mmHg)	120.5 (110–133)	129 (118–139)	0.002	126 (116–132)	125 (112–137)	129 (117–141)	0.46
eGFR (mL/min/1.73 m^2)	62.1 (47.1–86.1)	59.9 (47.0–72.7)	0.34	80.6 (57.5–81.2)	63.9 (48.5–82.2)	61.9 (42.1–82.3)	0.47
NT-proBNP (ng/L)	1,068 (484–1,462)	882 (571–2081)	0.66	56.6 (25.9–465.7)	660.2 (219.7–1424.5)	922.3 (490.3–1919.0)	0.003
BNP (ng/L)	152 (64–380)	289.5 (172–490)	<0.001	—	249.0 (106.0–443.0)	328.5 (241.5–478.0)	0.09
Medical history							
Atrial fibrillation	27 (38.6)	71 (24.9)	0.04	3 (50.0)	86 (51.8)	22 (50.0)	0.97
COPD	9 (12.9)	56 (19.7)	0.19	2 (33.3)	32 (19.3)	8 (18.2)	0.67
Diabetes	30 (42.9)	144 (50.5)	0.25	3 (50.0)	70 (42.2)	20 (45.5)	0.80
Dyslipidaemia	42 (60.0)	203 (71.2)	0.07	5 (83.3)	125 (75.3)	30 (68.2)	0.56
Hypertension	63 (90.0)	261 (91.6)	0.68	6 (100.0)	137 (82.5)	40 (90.9)	0.31
Peripheral arterial disease	7 (10.0)	34 (11.9)	0.65	0 (0)	26 (15.7)	5 (11.4)	0.64
Medication use							
ACE inhibitors/ARBs	53 (75.7)	222 (7.6)	0.73	5 (83.3)	108 (65.1)	32 (72.7)	0.50
Beta-blockers	54 (77.1)	244 (85.3)	0.097	5 (83.3)	122 (73.5)	37 (84.1)	0.36
Calcium channel blockers	29 (41.4)	114 (39.9)	0.81	0 (0)	50 (30.1)	16 (36.4)	0.21
Diuretics	60 (85.7)	261 (91.3)	0.16	4 (66.7)	139 (83.7)	43 (97.7)	0.009
Statin	44 (62.9)	202 (70.6)	0.21	4 (66.7)	105 (63.3)	29 (65.9)	0.95
NYHA Class III/IV	35 (50.0)	121 (42.6)	0.26	3 (50.0)	91 (54.8)	21 (47.7)	0.71
Echocardiographic findings							
E/e' ratio	11.6 (8.6–14.9)	13.9 (10.6–18.1)	0.006	9.4 (8.8–10.0)	15.7 (11.7–22.0)	20.0 (13.3–26.7)	<0.0001
LA volume index (mL/m^2)	23.1 (19.7–26.8)	30.1 (23.2–39.5)	<0.0001	25.1 (22.3–28.6)	43.0 (34.8–55.8)	50.9 (36.8–60.9)	0.008
LV ejection fraction (%)	61.5 (57.6–66.1)	61.0 (57.6–13.2)	0.91	67.8 (58.8–79.0)	60.8 (51.6–72.8)	73.7 (59.3–93.2)	0.52
LV end-diastolic volume index (mL/m^2)	38.1 (32.7–48.4)	43.7 (35.8–51.9)	0.01	—	49.7 (43.7–57.8)	51.7 (44.1–68.0)	0.56
LV mass index (g/m^2)	84.6 (71.7–95.0)	106.7 (88.2–124.7)	<0.0001	67.8 (58.8–79.0)	60.8 (51.6–72.8)	73.7 (59.3–93.2)	0.02
Pulmonary artery systolic pressure (mmHg)	33.9 (28.6–41.9)	36.5 (31.0–45.1)	0.08	28.1 (26.2–30.0)	41.0 (32.0–51.3)	47.2 (39.2–50.1)	0.058

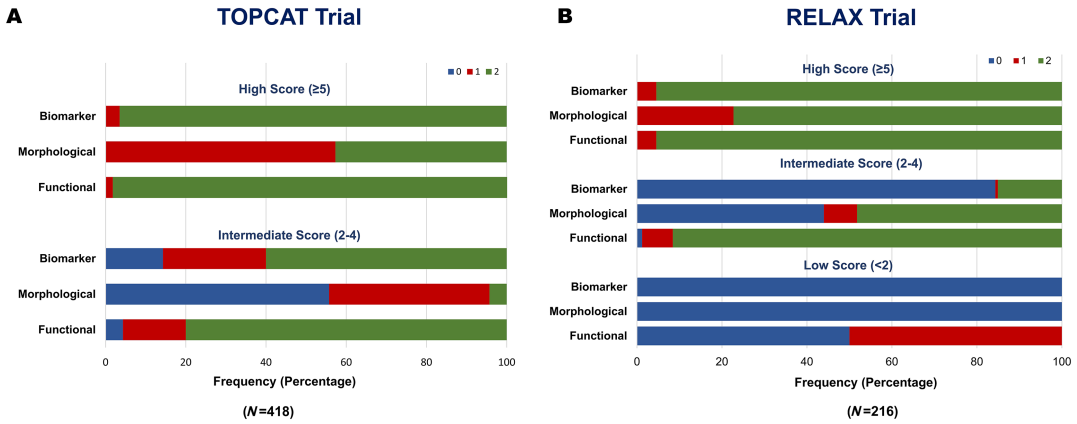
DOI: 10.1002/ehf2.13288

Table 1 (continued)

Characteristics	ARIC			
	Low (0–1 points) (*N* = 35)	Intermediate (2–4 points) (*N* = 248)	High (≥5 points) (*N* = 79)	*P* value
Age (years)	72 (70–75)	74 (71–79.5)	79 (73–84)	<0.0001
Female sex	15 (42.9)	152 (61.3)	52 (65.8)	0.06
White race	30 (85.7)	193 (77.8)	68 (86.1)	0.19
BMI (kg/m^2)	30.3 (27.5–34.2)	29.6 (26.5–33.3)	28.9 (25.1–32.3)	0.13
Systolic blood pressure (mmHg)	120 (117.5–135)	128.8 (117.5–142.5)	134.5 (117.5–149)	0.03
eGFR (mL/min/1.73 m^2)	75.3 (59.0–84.5)	72.7 (60.4–84.6)	63.4 (50.1–80.3)	0.001
NT-proBNP (ng/L)	62.2 (35.2–76.6)	99.2 (66.5–165.7)	378.6 (256.3–608.5)	<0.001
BNP (ng/L)	—	—	—	—
Medical history				
Atrial fibrillation	0 (0.0)	13 (5.2)	11 (13.9)	0.01
COPD	—	—	—	—
Diabetes	13 (37.1)	79 (31.9)	12 (15.4)	0.007
Dyslipidaemia	23 (65.7)	191 (77.0)	59 (74.7)	0.33
Hypertension	24 (70.6)	183 (74.1)	60 (76.9)	0.73
Peripheral arterial disease	0 (0.0)	13 (6.81)	5 (9.3)	0.25
Medication use				
ACE inhibitors/ARBs	12 (34.3)	84 (33.9)	23 (29.1)	0.74
Beta-blockers	2 (5.7)	60 (24.2)	32 (40.5)	<0.001
Calcium channel blockers	6 (17.1)	54 (21.8)	19 (24.1)	0.73
Diuretics	14 (40.0)	78 (31.5)	24 (30.4)	0.56
Statin	15 (42.9)	112 (45.2)	32 (40.5)	0.76
NYHA Class III/IV	—	—	—	—
Echocardiographic findings				
E/e′ ratio	8.0 (7.2–8.7)	11.0 (9.1–13.0)	11.1 (9.4–13.5)	<0.001
LA volume index (mL/m^2)	20.6 (17.6–23.9)	23.4 (19.0–27.8)	31.3 (26.6–40.5)	<0.001
LV ejection fraction (%)	66.0 (63.5–70.1)	66.2 (62.7–69.6)	66.3 (62.2–70.6)	0.98
LV end-diastolic volume index (mL/m^2)	37.1 (33.9–40.1)	39.8 (35.1–46.8)	41.0 (34.7–50.3)	0.01
LV mass index (g/m^2)	69.1 (60.3–76.5)	74.8 (64.1–85.2)	80.6 (69.3–93.0)	<0.001
Pulmonary artery systolic pressure (mmHg)	28.2 (26.5–31.0)	32.1 (28.3–35.0)	35.9 (32.9–40.8)	<0.001

Abbreviations: ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; ARIC, Atherosclerosis Risk in Community; BMI, body mass index; BNP, B-type natriuretic peptide; COPD, chronic obstructive pulmonary disease; eGFR, estimated glomerular filtration rate; HF, heart failure; LA, left atrial; LV, left ventricular; NP, natriuretic peptide; NT-proBNP: N terminal pro-BNP; NYHA, New York Heart Association; RELAX, Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Heart Failure with Preserved Ejection Fraction; TOPCAT, Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist.
Note: No individual with low (0–1 points) HFAPEFF score in TOPCAT. Values presented as count (percentage) or median (interquartile range).

Figure 1 Frequency distribution of HFA-PEFF score domains in Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) and Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Heart Failure with Preserved Ejection Fraction (RELAX) trials. Panel (A) represents the HFA-PEFF score categories in the TOPCAT trial, while Panel (B) represents the HFA-PEFF score categories in the RELAX trial.



In the ARIC population (N = 362), the median H_2FPEF score was 3.0 (IQR: 2.0–4.0). The score was similar across sex [male: 3.0 (IQR: 2.0–5.0) vs. female: 3.0 (IQR: 2.0–5.0); P = 0.83], and across race [Whites: 3.0 (IQR: 2.0–5.0) vs. non-Whites: 4.0 (IQR: 2.0–5.0); P = 0.12]. The ARIC participants generally had low (8.0%) or intermediate H_2FPEF scores (84.5%), and only 7.5% had high H_2FPEF score (*Table 2*).

HFA-PEFF vs. H_2FPEF score

In participants with both scores calculated (N in TOPCAT: 264; RELAX: 188), the application of the HFA-PEFF score categories to patients categorized using H_2FPEF score led to a reclassification of 52.7% of TOPCAT and 50.5% of RELAX trial participants (*Figures 2* and S4).

In the TOPCAT trial, 81.4% of patients with an intermediate likelihood of HFpEF based on H_2FPEF score (score 2–5) were reclassified as having a high likelihood based on the HFA-PEFF scale (score ≥5). Among those with a high H_2FPEF score (score ≥6), 23.5% were reclassified into the intermediate category (*Figure 2*). In the RELAX trial, 13.9% of patients with an intermediate likelihood of HFpEF on the H_2FPEF scale (score 2–5) were reclassified as having a high likelihood of HFpEF on the HFA-PEFF scale. The majority of participants with high H_2FPEF scores was reclassified as having intermediate (70.2%) or low HFA-PEFF scores (2.9%) (*Figure 2*).

Diagnostic value of the HFA-PEFF and H_2FPEF scores

We presumed that all TOPCAT and RELAX participants had HFpEF and none of the ARIC participants had heart failure. Both the scores generally performed well for diagnostic purposes when assessed in the study populations (*Table 3*). Using a rule-in approach, the specificity and PPV of the HFA-PEFF score were 78.2% and 76.2%, respectively. The specificity and PPV of the H_2FPEF score rule-in approach were 92.5% and 85.6%, respectively. The sensitivity and NPV of the HFA-PEFF score using a rule-out approach were 99.7% and 97.2%, respectively. The H_2FPEF score had a sensitivity of 99.6% and an NPV of 96.7% using the rule-out approach (*Table 3*). The diagnostic performance of the rule-in and rule-out approach of the scores in respective trials is shown in Tables S2–S3. The diagnostic performance of the AUC curves for the two scores are depicted in *Figure S5*.

In sensitivity analyses, the scores were computed in the ambulatory cohort of TOPCAT and RELAX trials. In 215 TOPCAT patients with either score available, 82.8% and 17.2% had a high and intermediate HFA-PEFF score, respectively. While 2.0%, 47.1%, and 51.0% had low, intermediate, and high H_2FPEF scores in TOPCAT ambulatory cohort. In 137 ambulatory RELAX patients, 3.4%, 78.8%, and 17.5% had low, intermediate, and high HFA-PEFF score, respectively. Similar to TOPCAT, 4.1%, 43.9% and, 52.0% had low, intermediate, and high H_2FPEF scores. Using the rule-in approach in these patients, the HFA-PEFF score specificity and PPV were 76.8% and 74.9%, respectively. The H_2FPEF score specificity and PPV were 89.7% and 80.8%, respectively. The sensitivity and NPV of the HFA-PEFF score using a rule-out approach in these patients were 99.6% and 95.2%, respectively. The H_2FPEF score had a sensitivity of 99.4% and an NPV of 94.7% using the rule-in approach (Table S4).

Prognostic value of the HFA-PEFF and H_2FPEF scores in TOPCAT trial

For every one-point increase in the HFA-PEFF score, the hazard for adverse cardiovascular events increased by 26%

Table 2 Baseline characteristics: H_2FPEF score categories

Characteristics	TOPCAT				RELAX			
	Low (0–1 points) $N = 3$	Intermediate (2–5 points) $N = 129$	High (≥6 points) $N = 132$	P value	Low (0–1 points) $N = 5$	Intermediate (2–5 points) $N = 79$	High (≥6 points) $N = 104$	P value
Age (years)	59 (57–71)	69 (10–79)	73 (66–80)	0.004	60.0 (59.0–61.0)	67.0 (59.0–73.0)	71.0 (65.0–80.0)	0.001
Female sex	2 (66.7)	70 (54.3)	56 (42.4)	0.10	2 (40.0)	38 (48.1)	51 (49.0)	0.99
White race	3 (100)	82 (63.6)	104 (78.8)	0.01	5 (100.0)	71 (89.9)	95 (91.4)	0.88
BMI (kg/m^2)	22.5 (18.5–28.8)	31.9 (26.8–37.7)	33.3 (29.2–38.8)	0.03	28.2 (28.1–28.4)	33.3 (28.0–39.1)	33.1 (28.9–38.1)	0.08
Systolic blood pressure (mmHg)	118 (110–128)	131.5 (121–142)	122 (111–135)	0.002	116 (112–124)	128 (114–140)	128 (116–140)	0.20
eGFR (mL/min/1.73 m^2)	91.1 (81.6–109.5)	62.1 (47.5–75.6)	58.2 (45.6–71.6)	0.03	73.6 (57.7–92.1)	68.2 (47.1–87.9)	61.1 (46.0–75.9)	0.25
NT-proBNP (ng/L)	1734.5 (528–2,941)	791 (475–1820)	1,017 (682–1,661)	0.70	63.3 (48.7–81.5)	366.0 (94.3–778.3)	1075.0 (507.8–1776.0)	<0.0001
BNP (ng/L)	108 (108–108)	239.5 (135–429)	323 (173–563)	0.08	235.0 (235.0–235.0)	254.0 (93.0–476.0)	364.0 (253.0–494.5)	0.22
Medical history								
Atrial fibrillation	0 (0.0)	3 (2.3)	71 (53.8)	<0.0001	0 (0.0)	5 (6.3)	88 (84.6)	<0.0001
COPD	0 (0.0)	31 (24.2)	19 (14.4)	0.11	0 (0.0)	14 (17.7)	23 (22.1)	0.55
Diabetes	1 (33.3)	70 (54.7)	60 (45.5)	0.36	1 (20.0)	34 (43.0)	45 (43.3)	0.69
Dyslipidaemia	1 (33.3)	87 (68.0)	96 (72.7)	0.21	5 (100.0)	58 (73.4)	81 (77.9)	0.48
Hypertension	2 (66.7)	121 (94.5)	117 (88.6)	0.07	4 (80.0)	67 (84.8)	91 (87.5)	0.59
Peripheral arterial disease	0 (0.0)	17 (13.3)	12 (9.1)	0.53	1 (20.0)	12 (15.2)	12 (11.5)	0.46
Medication use								
ACE inhibitors/ARBs	2 (66.7)	101 (78.3)	103 (78.0)	0.79	3 (60.0)	52 (65.8)	70 (67.3)	0.93
Beta-blockers	3 (100)	110 (85.3)	107 (8.1)	0.66	1 (20.0)	55 (69.6)	87 (83.7)	0.002
Calcium channel blockers	1 (33.3)	51 (39.5)	63 (47,7)	0.40	0 (0.0)	21 (26.6)	38 (36.5)	0.12
Diuretics	1 (33.3)	112 (86.8)	124 (93.9)	0.004	2 (40.0)	58 (73.4)	99 (95.2)	<0.0001
Statin	1 (33.3)	92 (71,3)	87 (65.9)	0.23	5 (100.0)	49 (62.0)	71 (68.3)	0.22
NYHA Class III/IV	1 (33.3)	58 (45.3)	66 (50.4)	0.62	2 (40.0)	40 (50.6)	56 (53.9)	0.78
Echocardiographic findings								
E/e' ratio	7.6 (7.2–9.0)	13.6 (9.3–17.6)	13.6 (10.4–18.3)	0.04	8.9 (7.5–10.0)	15.0 (10.0–20.0)	16.8 (13.3–25.0)	0.002
LA volume index (mL/m^2)	25.6 (24.–26.4)	26.8 (20.3–35.2)	31.5 (26.2–40.7)	<0.0001	24.9 (24.9–24.9)	38.9 (31.8–44.8)	52.7 (40.0–62.4)	<0.0001
LV ejection fraction (%)	60.0 (51.7–71.3)	62.2 (58.2–66.6)	60.9 (57.6–65.7)	0.47	57.0 (54.0–68.0)	58.0 (56.0–67.0)	58.0 (54.0–63.0)	0.29
LV end-diastolic volume index (mL/m^2)	41.7 (36.6–93.9)	45.8 (36.9–51.8)	39.9 (33.8–50.8)	0.047	53.0 (48.2–57.7)	50.2 (43.8–63.8)	51.6 (48.0–63.7	0.99
LV mass index (g/m^2)	146.9 (67.7–169.1)	98.9 (82.3–123.8)	99.6 (82.2–118.9)	0.57	51.4 (48.2–54.2)	63.1 (53.8–77.2)	60.1 (55.3–75.0)	0.08
Pulmonary artery systolic pressure (mmHg)	—	32.5 (28.6–39.5)	37.6 (31.2–46.9)	0.004	32.0 (32.0–32.0)	34.2 (29.4–41.0)	46.0 (36.4–53.6)	<0.001

Table 2 (continued)

Characteristics	ARIC			
	Low (0–1 points) *N* = 29	Intermediate (2–5 points) *N* = 306	High (≥6 points) *N* = 27	*P* value
Age (years)	72 (69–77)	75 (71–80)	77 (74–81)	0.07
Female sex	15 (51.7)	188 (61.4)	16 (59.3)	0.59
White race	24 (82.8)	244 (79.7)	23 (85.2)	0.75
BMI (kg/m^2)	26.3 (24.5–28.7)	29.6 (26.5–33.4)	32.8 (30.9–35.4)	<0.001
Systolic blood pressure (mmHg)	122.5 (115–132.5)	130.5 (118.5–143.0)	126 (108.5–144)	0.15
eGFR (mL/min/1.73 m^2)	73.5 (63.1–81.3)	71.0 (58.5–84.5)	65.8 (52.7–76.3)	0.13
NT-proBNP (ng/L)	82.8 (47.7–179.9)	113.6 (70.6–235.4)	357.9 (140.2–1,109)	<0.001
BNP (ng/L)	—	—	—	—
Medical history				
Atrial fibrillation	0 (0.0)	7 (2.3)	17 (63.0)	<0.001
COPD	—	—	—	—
Diabetes	5 (17.2)	88 (28.9)	11 (40.7)	0.15
Dyslipidaemia	20 (69.0)	232 (75.8)	21 (77.8)	0.68
Hypertension	13 (46.4)	231 (76.0)	23 (85.2)	<0.001
Peripheral arterial disease	1 (4.4)	16 (6.9)	1 (4.8)	0.84
Medication use				
ACE inhibitors/ARBs	6 (20.7)	100 (32.7)	13 (48.2)	0.09
Beta-blockers	1 (3.5)	75 (24.5)	18 (66.7)	<0.001
Calcium channel blockers	1 (3.5)	67 (21.9)	11 (40.7)	0.002
Diuretics	1 (3.5)	105 (34.3)	10 (37.0)	<0.001
Statin	13 (44.8)	135 (44.1)	11 (40.7)	0.31
NYHA Class III/IV	—	—	—	—
Echocardiographic findings				
E/e′ ratio	7.9 (7.4–8.7)	11.0 (9.1–13.0)	11.6 (9.3–13.4)	<0.001
LA volume index (mL/m^2)	22.2 (20.4–25.7)	24.2 (19.4–29.6)	30.9 (21.1–42.3)	0.002
LV ejection fraction (%)	66.0 (62.9–69.7)	66.3 (62.7–70.0)	66.1 (61.6–69.4)	0.63
LV end-diastolic volume index (mL/m^2)	39.5 (34.8–46.0)	39.8 (35.1–46.2)	38.2 (31.5–50.3)	0.63
LV mass index (g/m^2)	69.1 (57.6–80.2)	75.0 (65.4–85.9)	81.5 (71.6–93.8)	0.008
Pulmonary artery systolic pressure (mmHg)	27.1 (26.5–30.6)	32.5 (28.5–35.8)	36.3 (35.4–41.7)	<0.001

Abbreviations: ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; ARIC, Atherosclerosis Risk in Community; BMI, body mass index; BNP, B-type natriuretic peptide; COPD, chronic obstructive pulmonary disease; eGFR, estimated glomerular filtration rate; HF, heart failure; LA, left atrial; LV, left ventricular; NP, natriuretic peptide; NT-proBNP: N terminal pro-BNP; NYHA, New York Heart Association; RELAX, Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Heart Failure with Preserved Ejection Fraction; TOPCAT, Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist.
Note: Values presented as count (percentage) or median (interquartile range).

Figure 2 Reclassification of HFA-PEFF score categories from H_2FPEF score categories in (A) Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) and (B) Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (RELAX) trials. The arrows in red represent upward reclassification, and the curves in green represent a downward reclassification. The arrows in yellow show the reclassification in the same score category. The arrows in green indicate a downward reclassification.

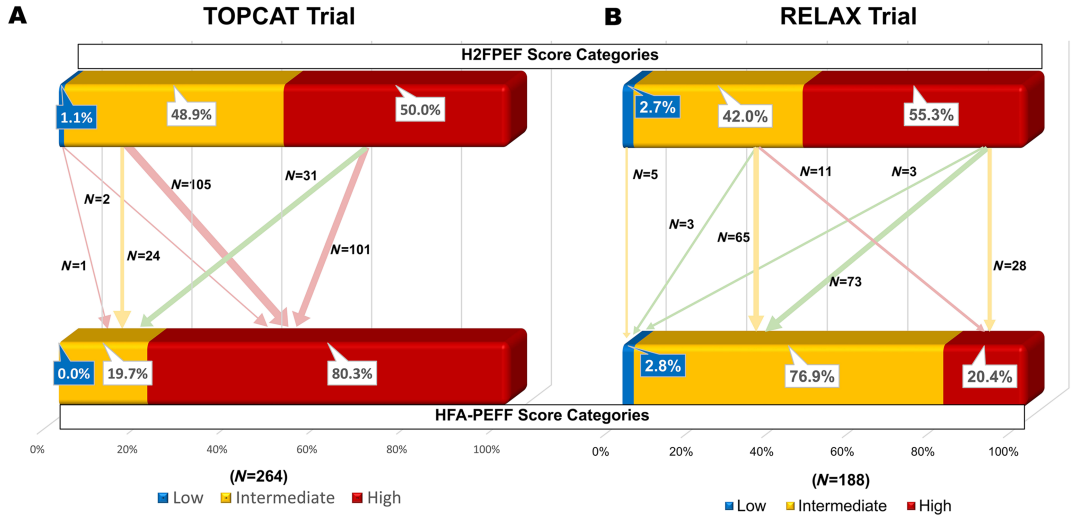


Table 3 Diagnostic performance of HFA-PEFF and H_2FPEF scores

	HFA-PEFF score[a]	
	Rule-in approach[b] (%)	Rule-out approach[c] (%)
Sensitivity	69.9	99.7
Specificity	78.2	9.7
Positive predictive value	76.2	52.5
Negative predictive value	72.2	97.2
	H_2FPEF score[d]	
	Rule-in approach[b] (%)	Rule-out approach[c] (%)
Sensitivity	58.1	99.6
Specificity	92.5	8.0
Positive predictive value	85.6	45.3
Negative predictive value	74.3	96.7

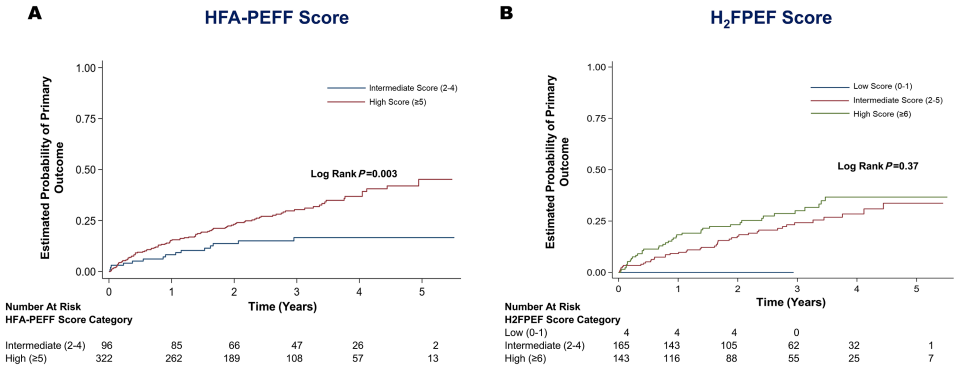
[a]Includes 379 participants from Atherosclerotic Risk in Communities (ARIC) study and 379 participants from Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) and Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Heart Failure with Preserved Ejection Fraction (RELAX) trials.
[b]For the rule-in approach, only the high-risk category is considered as a 'positive' test result.
[c]For the rule-out approach, both the intermediate-likelihood and high-likelihood categories are considered as a 'positive' test result.
[d]Includes 379 participants from ARIC study and 291 participants from TOPCAT and RELAX trials.

[hazard ratio (HR): 1.26, 95% confidence interval (CI) 0.98–1.63]. The hazard ratio for one-point increase in the H_2FPEF score was 1.01 (95% CI: 0.88–1.15). A similar association was seen when the scores were assessed as a categorical measure. The Kaplan–Meier curves for the risk of adverse cardiovascular outcomes across various strata of HFA-PEFF and H_2FPEF scores are shown in *Figure 3*. In multivariable-adjusted logistic regression, the odds of the primary outcome and for heart failure hospitalizations in those with high HFA-PEFF scores were 2.06 (95% CI: 1.06–4.03) and 2.07 (95% CI: 0.99–4.33), respectively. In those with high H_2FPEF score, the odds of the primary outcome and of heart failure hospitalizations were 1.12 (95% CI: 0.62–2.03) and 1.08 (95% CI: 0.57–2.06), respectively.

Among TOPCAT participants who had both scores available (*N* = 264), every one-point increase in the HFA-PEFF score led to an increase in the hazard for adverse cardiovascular events by 42% (HR: 1.42, 95% CI: 1.04–1.93). The predictive ability (estimated by C-statistics) for the incidence of adverse cardiovascular events from the multivariate-adjusted model was 0.69 (95% CI: 0.63–0.74). This improved with the inclusion of either the HFA-PEFF score (C-statistic: 0.70, 95% CI: 0.65–0.75) or the H_2FPEF score (C-statistic: 0.70, 95% CI: 0.63–0.77) to the multivariate model (*Figure S6*). The C-statistics for HFA-PEFF and H_2FPEF score alone were 0.54 (95% CI: 0.49–0.58) and 0.53 (95% CI: 0.48–0.59), respectively.

In multivariable-adjusted logistic regression among TOPCAT participants with both scores available (*N* = 264), the odds of the primary outcome and for heart failure hospitalizations in those with high HFA-PEFF scores were 1.97 (95% CI: 1.00–3.86) and 2.00 (95% CI: 0.94–4.17), respectively. In those with high H_2FPEF score, the odds of the primary outcome and for heart failure hospitalizations in those with high HFA-PEFF scores were 0.96 (95% CI: 0.53–1.73) and 0.86 (95% CI: 0.46–1.63), respectively.

Figure 3 Kaplan–Meier curves for the risk of primary outcome in Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) trial: stratified by HFA-PEFF and H_2FPEF score categories. The figure depicts Kaplan–Meier curves for the primary outcome of the TOPCAT trial, stratified by the HFA-PEFF score categories (Panel A) and H_2FPEF score categories (Panel B).



Relationship of the HFA-PEFF and H_2FPEF scores with exercise capacity in RELAX trial

A higher H_2FPEF score was associated with decreased VO_2 maximum in unadjusted (β: −0.51, $P < 0.001$) and adjusted models (β: −0.26, $P = 0.01$). The β-estimate for the association of HFAPEFF score with VO_2 maximum was −0.10 ($P = 0.50$) and −0.02 ($P = 0.90$) in unadjusted and adjusted models, respectively (*Figures S7* and *S8*). The relationship of the scores with baseline VO_2 maximum stratified by prior heart hospitalization is described in *Figures* S9 and *S10*. A high H_2FPEF score was associated with a peak VO_2 of <15 mL/kg/min (odds ratio: 2.0; 95% CI: 0.93–4.15%). The odds of peak VO_2 being <15 mL/kg/min in those with high HFA-PEFF scores were 0.76 (95% CI: 0.32–1.79).

Discussion

This study evaluated the generalizability of the HFA-PEFF and the H_2FPEF score in the HFpEF participants of the TOPCAT and RELAX trials. Both scores predict a high likelihood of HFpEF in the trial cohorts. Both the scores are effective in ruling out HFpEF and may also be used to rule-in the diagnosis of HFpEF. In addition to providing diagnostic probability, the HFA-PEFF and H_2FPEF scores may also predict the risk for the development of adverse cardiovascular events. The H_2FPEF score was robustly associated with decreased exercise capacity. A large proportion of individuals classified in a particular H_2FPEF score category were reclassified after the application of the HFA-PEFF score. A large proportion of participants from TOPCAT, RELAX, and ARIC were categorized as having an intermediate likelihood for HFpEF diagnosis by both the scores, indicating the need for additional testing in these participants.

Heart failure with preserved ejection fraction is increasingly being recognized as a multi-systemic disorder with a heterogeneous population substrate.[18–21] Cardiac dysfunction exists along with multiple comorbidities contributing to the complex etiopathogenesis and diverse phenotypes of HFpEF.[18–21] The diagnosis of HFpEF patients has been challenging due to the distinct clinical phenotypes and absence of invasive haemodynamic assessment in the routine care of patients.[18,20] The H_2FPEF score, which was derived from a single-centre cohort of dyspnoea patients undergoing invasive haemodynamic exercise assessment, has been validated in other HFpEF populations.[8,10] The recently developed and validated HFA-PEFF algorithm provides a comprehensive stepwise approach for the diagnosis of HFpEF in suspected patients.[9,10]

It is of great interest to us that both scoring systems exhibit diagnostic utility despite their contrasting focus on cardiac and extracardiac features. While both scoring systems emphasize the importance of filling pressures by assigning points for elevated E/e' ratios and tricuspid regurgitant velocity, the H_2FPEF score seemingly has a greater focus on clinical features such as obesity, hypertension, and age. Due to the simplified scoring and high validity,[8,11,16,17] the H_2FPEF score may be readily employed in routine clinical practice. However, the H_2FPEF score does not account for gender, natriuretic peptide levels, and cardiac morphological features, such as left atrial volume index and left ventricular mass index. This may limit the diagnostic and prognostic applicability of the scoring system.[8] We also validate the previously reported association of H_2FPEF score with exercise capacity.[22] This indicates that the H_2FPEF score may be a reliable surrogate for underlying functional capacity in patients with HFpEF, which is not reliably estimated in clinical settings.

In comparison, the HFA-PEFF scoring system assigns a greater focus on cardiac morphological features, biomarkers, and functional testing. The HFA-PEFF algorithm suggests

diastolic stress testing or invasive haemodynamic assessment for those in the intermediate score category.[9] We found in our investigation that this may amount to a large proportion of patients—approximately 25% in TOPCAT and approximately 75% in the RELAX trial. Additionally, the HFA-PEFF score does not account for body mass index. This may be important given the recognition of the obese-HFpEF phenotype as a distinct pathophysiological phenotype.[14,23,24] The diagnostic weightage and importance of the numerous individual components in the multi-step algorithm have been subject to deliberations.[10,25]

Our findings substantiate the diagnostic validity of the HFA-PEFF and H_2FPEF scoring systems for the identification of HFpEF patients.[10,11,16,17] We also advance these findings by comparing the two study cohorts with a group of individuals that had a low pre-test probability of cardiovascular disease to identify test characteristics. The differences in scores within the TOPCAT and RELAX cohorts highlight the significant heterogeneity in the populations identified and classified by the two scores. The scores may capture non-overlapping populations when applied to the same population, thereby highlighting the need for a unified, comprehensive approach that incorporates the features of both the scores. The reclassification and differences in the two trial populations also underscore the heterogeneous population recruitment in HFpEF trials. This heterogeneity has been proposed as being a contributor to the neutral trial results reported over the last few years.[26] Additionally, our findings regarding the significant geographical heterogeneity in the probability of HFpEF in TOPCAT participants add to the existing evidence of regional variations in the recruited population.[16,17,27] The inclusion of patients based on N terminal pro brain natriuretic peptide (NT-proBNP) as one of the inclusion criteria for TOPCAT preferentially biases the population to a higher HFA-PEFF score because NT-proBNP is accounted for in the HFA-PEFF score, but not in the H_2FPEF score. Given the variable association of NT-proBNP with both the diagnosis and symptomatic burden in HFpEF,[3,28] the patients diagnosed with HFpEF without elevated NT-proBNP (such as those in RELAX), the HFA-PEFF score may have a lower discriminatory value. However, the presented analyses indicate that NT-proBNP may have prognostic utility in HFpEF. Both the HFA-PEFF and H_2FPEF scores have been reported to have important prognostic implications in predicting adverse clinical events in patients with HFpEF.[10,11,16,17] Our results differ from the prior works in the TOPCAT trial due to differences in the population selection and analytical approach. The findings from our study add to the growing evidence for the prognostic value of both the scoring systems and their use in daily practice for risk stratification of HFpEF patients.[10,11,16,17] The reliance on obesity in the H_2FPEF score and the resulting statistical obesity paradox may have also accounted for lack of association between H_2FPEF score and adverse outcomes.

Our findings carry importance on a public health level. Both scoring systems have clear clinical utility with the potential to allow for the differentiation of HFpEF from many overlapping cardiorespiratory syndromes. In an era where cost-conscious care is of paramount importance, an accurate diagnosis could yield more efficient resource utilization for expensive and invasive investigations such as right heart catheterization and diastolic stress testing. Our efforts to compare the diagnostic utility of these scoring systems also allow clinicians to decide upon an optimal method for the HFpEF diagnosis. Both the scores may be useful to rule-out HFpEF and streamline the further investigation of aetiology in the exclusion of HFpEF. Standardized and accurate diagnosis of HFpEF would allow for uniform determination of patient eligibility, risk enrichment, and evaluation of therapeutic efficacy in future global HFpEF clinical trials. This may limit the heterogeneity of the trial participants, which has been frequently cited as a limiting factor for showing adequate effect size with clinical interventions.[27] Finally, the prognostic value of both scoring systems may also have significant benefits for clinical practice. This may potentially shape shared patient–physician decision making to determine the optimal treatment approach that is suited to individual patients.

By describing the test characteristics of each of the scoring systems, our data also inform directions for future research. There has been debate regarding the weighting of the major and minor criteria. The inclusion of global longitudinal strain and the arbitrary cut-off values for the left atrial volume index and left ventricular mass index have been suggested to require further prospective examination before inclusion into a globally valid diagnostic score.[10,25] Moreover, there may be limited applicability of the HFA-PEFF algorithm to individuals who cannot exercise and/or have contraindications to invasive assessment.[25] Resting echocardiographic evaluation along with biomarker assessment (galectin-3) may also provide a reliable method for diagnostic and prognostic assessment of suspected HFpEF patients unable to exercise.[29] Lastly, our findings should provide the impetus for future research that examines whether treatment changes lead to changes in individual patients' risk scores and, ultimately, their hazard of adverse events.

Study limitations

Our study has several limitations. All participants enrolled in the trials were presumed to be HFpEF patients by their enrolment definitions in the respective trials. The definitions of HFpEF and the inclusion criteria differed in the two trials, which may have contributed to the differences in the scores and the diagnostic performance in the respective trial populations. It is notable that the two systems score the patients on a different scale, and the same points cannot be directly

DOI: 10.1002/ehf2.13288

compared between both the systems. For example, 5 points on the HFA-PEFF score (range: 0–6 points) are not equivalent to 5 points on the H_2FPEF score (range: 0–9 points). There are key differences in the components of the two scores such as the requirement of natriuretic peptide levels as a key component for computation of the HFA-PEFF score but not the H_2FPEF score. The use of control population from ARIC without any history of cardiovascular disease and with unexplained dyspnoea helped in improving the clinical representativeness of the study population. The present study is subject to selection bias as only a subset of the trial population with available data was included in the analyses. The echocardiographic cohort of the TOPCAT trial was a subset of the total trial population, and even fewer had the global longitudinal strain data available. Due to the study design, we could not estimate the results of the Step 3 of the HFA-PEFF algorithm, that is, invasive and/or exercise testing. Also, the small sample size might have had an impact on the study's ability to assess the diagnostic and prognostic value of the scores adequately. Hence, a larger cohort of suspected HFpEF patients with gold standard invasive haemodynamic metrics would aid in verifying the diagnostic validity and prognostic ability of the scoring systems. Findings from such studies could have significant implications in routine patient care and could help determine eligibility for clinical trials.

Conclusion

Despite recent advances in the understanding of the pathophysiology of HFpEF, the diagnosis of HFpEF remains challenging. Both the HFA-PEFF and H_2FPEF scoring systems are reliable diagnostic tools for HFpEF patients. Further research in large, diverse, and well-phenotyped population-based cohorts is needed to substantiate the validity of the scoring systems and to simplify the diagnosis in those with intermediate-likelihood of HFpEF based on these scores.

Acknowledgements

We would like to thank the TOPCAT, RELAX, and ARIC investigators for making the study data available publically through NHLBI BioLINCC.

Conflict of interest

None of the other authors had any conflicts of interest or financial disclosures to declare.

Funding

This work is supported by the National Institutes of Health Mentored Patient-Oriented Research Award (5K23 HL146887-02) to Dr Pankaj Arora.

Supporting information

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

Table S1. Baseline Characteristics of Study Population and Remaining TOPCAT Trial Participants.
Table S2. Diagnostic Performance of HFA-PEFF and H2FPEF Scores in TOPCAT Trial.
Table S3. Diagnostic Performance of HFA-PEFF and H2FPEF Scores in RELAX Trial.
Figure S1. Study Population Derivation.
Figure S2. Calculation of HFA-PEFF Score.
Figure S3. Calculation of H_2FPEF Score.
Figure S4. Distribution of HFA-PEFF and H_2FPEF Score in TOPCAT and RELAX Trial.
Figure S5. Diagnostic Performance of HFA-PEFF and H_2FPEF Score.
Figure S6. Receiver Operating Characteristic Curve for the Primary Outcome of TOPCAT Trial.
Figure S7. Relationship of H_2FPEF Score with Baseline VO_2 Maximum in RELAX Trial.
Figure S8. Relationship of HFA-PEFF Score with Baseline VO_2 Maximum in RELAX Trial.
Figure S9. Relationship of H_2FPEF Score with Baseline VO_2 Maximum in RELAX Trial Patients with No Prior Heart Failure Hospitalizations.
Figure S10. Relationship of HFA-PEFF Score with Baseline VO_2 Maximum in RELAX Trial Patients with $\geq$1 Prior Heart Failure Hospitalizations.

References

1. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med* 2006; **355**: 251–259.
2. Steinberg BA, Zhao X, Heidenreich PA, Peterson ED, Bhatt DL, Cannon CP, Hernandez AF, Fonarow GC, Get With the Guidelines Scientific Advisory Committee and Investigators. Trends in patients hospitalized with heart failure

and preserved left ventricular ejection fraction: prevalence, therapies, and outcomes. *Circulation* 2012; **126**: 65–75.
3. Anjan VY, Loftus TM, Burke MA, Akhter N, Fonarow GC, Gheorghiade M, Shah SJ. Prevalence, clinical phenotype, and outcomes associated with normal B-type natriuretic peptide levels in heart failure with preserved ejection fraction. *Am J Cardiol* 2012; **110**: 870–876.
4. Lancellotti P, Galderisi M, Edvardsen T, Donal E, Goliasch G, Cardim N, Magne J, Laginha S, Hagendorff A, Haland TF, Aaberge L, Martinez C, Rapacciuolo A, Santoro C, Ilardi F, Postolache A, Dulgheru R, Mateescu AD, Beladan CC, Deleanu D, Marchetta S, Auffret V, Schwammenthal E, Habib G, Popescu BA. Echo-Doppler estimation of left ventricular filling pressure: results of the multicentre EACVI Euro-Filling study. *Eur Heart J Cardiovasc Imaging* 2017; **18**: 961–968.
5. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, Dokainish H, Edvardsen T, Flachskampf FA, Gillebert TC, Klein AL, Lancellotti P, Marino P, Oh JK, Alexandru Popescu B, Waggoner AD, Houston T, Oslo N, Phoenix A, Nashville T, Hamilton OC, Uppsala S, Ghent LB, Cleveland O, Novara I, Rochester M, Bucharest R, St. Louis M. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2016; **17**: 1321–1360.
6. Roberts E, Ludman AJ, Dworzynski K, Al-Mohammad A, Cowie MR, McMurray JJ, Mant J, Failure NGDGfAH. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. *BMJ* 2015; **350**: h910.
7. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* 2015; **131**: 269–279.
8. Reddy YNV, Carter RE, Obokata M, Redfield MM, Borlaug BA. A simple, evidence-based approach to help guide diagnosis of heart failure with preserved ejection fraction. *Circulation* 2018; **138**: 861–870.
9. Pieske B, Tschope C, de Boer RA, Fraser AG, Anker SD, Donal E, Edelmann F, Fu M, Guazzi M, Lam CSP, Lancellotti P, Melenovsky V, Morris DA, Nagel E, Pieske-Kraigher E, Ponikowski P, Solomon SD, Vasan RS, Rutten FH, Voors AA, Ruschitzka F, Paulus WJ, Seferovic P, Filippatos G. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). *Eur Heart J* 2019; **40**: 3297–3317.
10. Barandiaran Aizpurua A, Sanders-van Wijk S, Brunner-La Rocca HP, Henkens M, Heymans S, Beussink-Nelson L, Shah SJ, van Empel VPM. Validation of the HFA-PEFF score for the diagnosis of heart failure with preserved ejection fraction. *Eur J Heart Fail* 2020; **22**: 413–421.
11. Sueta D, Yamamoto E, Nishihara T, Tokitsu T, Fujisue K, Oike F, Takae M, Usuku H, Takashio S, Arima Y, Suzuki S, Nakamura T, Ito M, Kanazawa H, Sakamoto K, Kaikita K, Tsujita K. H_2FPEF score as a prognostic value in HFpEF patients. *Am J Hypertens* 2019; **32**: 1082–1090.
12. Pitt B, Pfeffer MA, Assmann SF, Boineau R, Anand IS, Claggett B, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Harty B, Heitner JF, Kenwood CT, Lewis EF, O'Meara E, Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, Yang S, McKinlay SM, Investigators T. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med* 2014; **370**: 1383–1392.
13. Redfield MM, Chen HH, Borlaug BA, Semigran MJ, Lee KL, Lewis G, LeWinter MM, Rouleau JL, Bull DA, Mann DL, Deswal A, Stevenson LW, Givertz MM, Ofili EO, O'Connor CM, Felker GM, Goldsmith SR, Bart BA, McNulty SE, Ibarra JC, Lin G, Oh JK, Patel MR, Kim RJ, Tracy RP, Velazquez EJ, Anstrom KJ, Hernandez AF, Mascette AM, Braunwald E, Trial R. Effect of phosphodiesterase-5 inhibition on exercise capacity and clinical status in heart failure with preserved ejection fraction: a randomized clinical trial. *JAMA* 2013; **309**: 1268–1277.
14. Redfield MM, Borlaug BA, Lewis GD, Mohammed SF, Semigran MJ, Lewinter MM, Deswal A, Hernandez AF, Lee KL, Braunwald E, Heart Failure Clinical Research N. PhosphdiesteRasE-5 Inhibition to Improve CLinical Status and EXercise Capacity in Diastolic Heart Failure (RELAX) trial: rationale and design. *Circ Heart Fail* 2012; **5**: 653–659.
15. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* 1989; **129**: 687–702.
16. Myhre PL, Vaduganathan M, Claggett BL, Lam CSP, Desai AS, Anand IS, Sweitzer NK, Fang JC, O'Meara E, Shah SJ, Shah AM, Lewis EF, Rouleau J, Pitt B, Solomon SD. Application of the H_2FPEF score to a global clinical trial of patients with heart failure with preserved ejection fraction: the TOPCAT trial. *Eur J Heart Fail* 2019; **21**: 1288–1291.
17. Segar MW, Patel KV, Berry JD, Grodin JL, Generalizability PA. Implications of the H_2FPEF score in a cohort of patients with heart failure with preserved ejection fraction. *Circulation* 2019; **139**: 1851–1853.
18. Shah SJ, Kitzman DW, Borlaug BA, van Heerebeek L, Zile MR, Kass DA, Paulus WJ. Phenotype-specific treatment of heart failure with preserved ejection fraction: a multiorgan roadmap. *Circulation* 2016; **134**: 73–90.
19. Ather S, Chan W, Bozkurt B, Aguilar D, Ramasubbu K, Zachariah AA, Wehrens XH, Deswal A. Impact of noncardiac comorbidities on morbidity and mortality in a predominantly male population with heart failure and preserved versus reduced ejection fraction. *J Am Coll Cardiol* 2012; **59**: 998–1005.
20. Shah SJ. Precision medicine for heart failure with preserved ejection fraction: an overview. *J Cardiovasc Transl Res* 2017; **10**: 233–244.
21. Parcha V, Patel N, Kalra R, Bhargava A, Prabhu SD, Arora G, Arora P. Clinical, demographic, and imaging correlates of anemia in heart failure with preserved ejection fraction (from the RELAX trial). *Am J Cardiol* 2020; **125**: 1870–1878.
22. Takahari K, Hidaka T, Ueda Y, Izumi K, Harada Y, Susawa H, Kinoshita M, Itakura K, Utsunomiya H, Kihara Y. H_2FPEF score for the prediction of exercise intolerance and abnormal hemodynamics in japanese—evaluation by exercise stress echocardiography combined with cardiopulmonary exercise testing. *Circ J* 2019; **83**: 2487–2493.
23. Packer M, Kitzman DW. Obesity-related heart failure with a preserved ejection fraction: the mechanistic rationale for combining inhibitors of aldosterone, neprilysin, and sodium-glucose cotransporter-2. *JACC Heart Fail* 2018; **6**: 633–639.
24. Reddy YNV, Lewis GD, Shah SJ, Obokata M, Abou-Ezzedine OF, Fudim M, Sun JL, Chakraborty H, McNulty S, LeWinter MM, Mann DL, Stevenson LW, Redfield MM, Borlaug BA. Characterization of the obese phenotype of heart failure with preserved ejection fraction: a RELAX trial ancillary study. *Mayo Clin Proc* 2019; **94**: 1199–1209.
25. Tadic M, Cuspidi C. Hypertension, diastolic stress test, and HFpEF: does new scoring system change something? *J Clin Hypertens (Greenwich)* 2019; **21**: 1905–1907.
26. Parikh KS, Sharma K, Fiuzat M, Surks HK, George JT, Honarpour N, Depre C, Desvigne-Nickens P, Nkulikiyinka R, Lewis GD, Gomberg-Maitland M, O'Connor CM, Stockbridge N, Califf RM, Konstam MA, Januzzi JL Jr, Solomon SD, Borlaug BA, Shah SJ, Redfield MM, Felker GM. Heart failure with preserved ejection fraction expert panel report: current controversies and implications for clinical trials. *JACC Heart Fail* 2018; **6**: 619–632.
27. Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Heitner JF, Lewis EF, O'Meara E, Rouleau JL,

Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, McKinlay SM, Pitt B. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. *Circulation* 2015; **131**: 34–42.

28. Reddy YNV, Rikhi A, Obokata M, Shah SJ, Lewis GD, AbouEzzedine OF, Dunlay S, McNulty S, Chakraborty H, Stevenson LW, Redfield MM, Borlaug BA. Quality of life in heart failure with preserved ejection fraction: importance of obesity, functional capacity, and physical inactivity. *Eur J Heart Fail* 2020.
29. Kosmala W, Przewlocka-Kosmala M, Rojek A, Marwick TH. Comparison of the diastolic stress test with a combined resting echocardiography and biomarker approach to patients with exertional dyspnea: diagnostic and prognostic implications. *JACC Cardiovasc Imaging* 2019; **12**: 771–780.