The HFA-PEFF and H$_2$FPEF scores largely disagree in classifying patients with suspected heart failure with preserved ejection fraction

The H$_2$FPEF and HFA-PEFF scores have recently been proposed to solve the clinical dilemma of diagnosing heart failure with preserved ejection fraction (HFpEF).[1,2] The H$_2$FPEF score includes four clinical [age, body mass index (BMI), atrial fibrillation (AF) and hypertension] and two echocardiographic items (E/e′ and right ventricular pressure). The HFA-PEFF score contains minor and major criteria within three domains: functional (E/e′, e′, tricuspid regurgitation velocity, global longitudinal strain), morphological (left atrial volume index and parameters reflecting left ventricular hypertrophy) and natriuretic peptides. A H$_2$FPEF score ≥6 or a HFA-PEFF ≥5 points is considered diagnostic of HFpEF. A H$_2$FPEF score of 2−5 or a HFA-PEFF of 2−4 points classifies patients as having an intermediate likelihood of HFpEF wherein invasive haemodynamic evaluation — preferably with exercise — or exercise echocardiography is proposed by the authors.[1,2] Two studies validated each score separately in Western populations.[3,4] Recently in this Journal, the two scores showed rather comparable diagnostic performance in an Asian case-control cohort, although lower than in the cited Western populations.[5] Both scores were predictive of adverse outcome in a Western population, although the risk prediction was very discordant between the two.[6] We hypothesized that the two scores classify a significant proportion of suspected HFpEF patients differently in terms of likelihood categories.

We calculated the absolute H$_2$FPEF and HFA-PEFF scores[1,2] and their likelihood categories (*Figure 1*) in 363 consecutive patients with suspected HFpEF. In summary, all patients from our outpatient HFpEF clinic (2015−2019) with a left ventricular ejection fraction (LVEF) of ≥50% were included prospectively and underwent a comprehensive one-day diagnostic work-up including echocardiography, blood, exercise, and pulmonary function testing, sleep apnoea screening and Holter.[4] Exclusion criteria were: previously reduced LVEF <50%, familiar/genetic hypertrophic cardiomyopathy, significant valvular or congenital disease, constrictive pericarditis, or heart transplantation. The final HFpEF diagnosis was expert adjudicated (consented by ≥2 heart failure cardiologists: V.P.M.v.E., H.P.B.L.R., C.K., N.U.L.), considering all baseline investigations, previous heart failure hospitalization(s), congestion with positive response to diuretic therapy and previous natriuretic peptide levels (e.g. before diuretic therapy). In case of clinical uncertainty, invasive haemodynamic evaluation was performed (*n* = 79; 21.8%).

The area under the receiver operating characteristic curve (AUC) was calculated for each score and compared using DeLong test. Sensitivity, specificity, negative and positive predictive value were calculated for each score (H$_2$FPEF ≥6 and HFA-PEFF ≥5). Differential patient classification by the two scores in terms of likelihood categories was visualized by a Sankey diagram and tested by Wilcoxon signed-rank test. Imputation of missing values in items of the scores was described previously[4]; sensitivity analysis excluded patients with any imputed value (*n* = 68; 19%).

Patients diagnosed with HFpEF (*n* = 300, 83%) were older (76 ± 8 vs. 67 ± 11 years, *P* < 0.001) and more often suffered from AF (56% vs. 21%, *P* < 0.001), but did not differ in terms of sex (67% vs. 63% female, *P* = 0.66), arterial hypertension (86% vs. 76%, *P* = 0.08), or diabetes mellitus (33% vs. 24%, *P* = 0.18) vs. non-HFpEF patients (online supplementary *Table S1*). The final diagnosis in the non-HFpEF group was mostly pulmonary disease (*n* = 29; 46%), including obstructive sleep apnoea syndrome (*n* = 12) and obstructive, restrictive, interstitial and vascular pulmonary disease. Other diagnoses were obesity (*n* = 5), AF (*n* = 4), coronary artery disease (*n* = 3), deconditioning (*n* = 3), or a combination of factors (*n* = 6). Finally, five patients were asymptomatic hypertensive controls and five had a transient cause of dyspnoea (e.g. anaemia, infection).

Diagnostic accuracy in terms of AUC was good for both scores, albeit sensitivity and negative predictive value were limited. This was offset by a good specificity and positive predictive value (*Table 1*). In 41% of our cohort (*n* = 145; 42% when excluding imputed data), patients were classified into different likelihood categories, depending on the score used (*P* < 0.001, *Figure 1*). Patients with an assigned H$_2$FPEF category higher than their assigned HFA-PEFF category (H$_2$FPEF > HFA-PEFF) most often suffered from AF (82.0% vs. 14.0% and 61%,
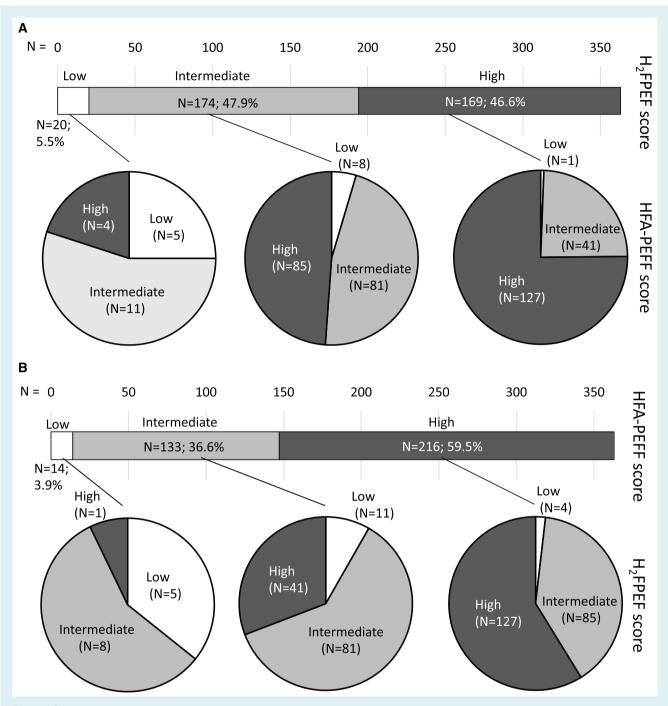
Figure 1 Reclassification of patients into likelihood categories (low, intermediate and high) from the H₂FPEF to the HFA-PEFF score (*A*) and vice versa (*B*). White = low likelihood; light-grey = intermediate likelihood; dark-grey = high likelihood of heart failure with preserved ejection fraction, estimated by each designated score.

$P < 0.001$) and had the highest BMI (33.5 ± 6.5 vs. 27.9 ± 4.9 and 31.0 ± 6.1, $P < 0.001$). Patients with an assigned H₂FPEF category lower than HFA-PEFF (H₂FPEF < HFA-PEFF) had lowest e′ [7.4 (5.6–8.8) vs. 8.7 (7.4–9.8) and 8.0 (6.7–9.4) cm/s, $P < 0.001$]. Patients assigned to the same category by both scores (H₂FPEF = HFA-PEFF) had the highest N-terminal pro B-type natriuretic peptide (NT-proBNP) [964 (220–1721) vs. 278 (175–545) and 448 (254–894) pg/L, $P < 0.001$], left atrial volume index [45 (32–60) vs. 36 (29–45) and 41 (35–50) mL/m², $P < 0.001$] and mitral E velocity [88 (67–109) vs. 78 (61–98) and 73 (55–88) cm/s, $P < 0.001$]. Age, sex, hypertension and left ventricular hypertrophy were not depending on differential/equivalent classification into likelihood categories by the two scores ($P > 0.20$).

The HFA-PEFF and H₂FPEF score opened a new era in the diagnosis of HFpEF by substituting the classical binary diagnostic approach with a likelihood estimation of HFpEF. As shown by this report and others, the high-likelihood cut-off of either score is quite accurate to diagnose HFpEF, while

Table 1 Diagnostic performance of the H$_2$FPEF and HFA-PEFF scores

	AUC (95% CI)	Cut-off	Sensitivity	Specificity	NPV	PPV
H$_2$FPEF	0.77 (0.71−0.83)	≥6	52.7%	82.5%	26.8%	93.5%
HFA-PEFF	0.88 (0.82−0.93)*	≥5	70.0%	90.5%	38.8%	97.2%

Results were similar when excluding imputed data. AUC, area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.
*$P < 0.009$ vs. H$_2$FPEF.

sensitivity is limited.[3,5] Both scores assign a substantial proportion of suspected HFpEF patients as intermediate likelihood, wherein additional diagnostics are proposed. Arbitrarily, 41% of suspected HFpEF patients are classified differently by one vs. the other score. Thus, depending on which score is used, completely different patients will be referred for additional testing or allocated as having HFpEF. This limits the clinical applicability of the scores and demonstrates the ongoing diagnostic uncertainty in HFpEF. As expected, AF and BMI were main drivers of the discrepancy between the scores — being key items of the H$_2$FPEF score whilst AF raises thresholds for HFpEF in the HFA-PEFF score. Despite age affecting the scores in different directions, it was not related to discrepant classification. NT-proBNP was highest when the two scores agreed, although it is only included in the HFA-PEFF score.

It strengthens our study that it is performed in a prospective, consecutive cohort. It included a rather selected population, yet both scores were actually designed to be used after a pre-test assessment to raise pre-test probability. We recognize that the considered gold standard of invasive haemodynamic testing was not applied to all patients. This cohort is however a reflection of clinical reality and our protocol included all fundamental elements of HFpEF and its differential diagnosis.

In conclusion, until a more uniform and accurate classification is available, the H$_2$FPEF and HFA-PEFF scores can be used for estimating HFpEF likelihood whilst keeping in mind the large discrepancy between them.

Combining the two can be insightful in daily practice and is currently applied in our specialized centre.

Supplementary Information

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

Funding

Sandra Sanders-van Wijk[1]*[†],
Arantxa Barandiarán Aizpurua[1][†],
Hans-Peter Brunner-La Rocca[1],
Michiel T.H.M. Henkens[1],
Jerremy Weerts[1],
Christian Knackstedt[1],
Nicole Uszko-Lencer[1],
Stephane Heymans[1,2,3], and
Vanessa van Empel[1]

[1]Department of Cardiology, Maastricht University Medical Centre (MUMC+), Cardiovascular Research Institute Maastricht (CARIM), Maastricht, The Netherlands; [2]Department of Cardiovascular Research, University of Leuven, Leuven, Belgium; and [3]Netherlands Heart Institute (ICIN), Utrecht, The Netherlands
*Email: sandra.van.wijk@mumc.nl
[†]These authors contributed equally.

References

1. 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 J Heart Fail* 2020;**22**:391−412.
2. Reddy YN, 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.
3. Sepehrvand N, Alemayehu W, Dyck GJ, Dyck JR, Anderson T, Howlett J, Paterson I, McAlister FA, Ezekowitz JA. External validation of the H$_2$F-PEF model in diagnosing patients with heart failure and preserved ejection fraction. *Circulation* 2019;**139**:2377−2379.
4. Barandiaran Aizpurua A, Sanders-van Wijk S, Brunner-La Rocca HP, Henkens M, Heymans S, Beussink-Nelson L, Shah SJ, van Empel VP. Validation of the HFA-PEFF score for the diagnosis of heart failure with preserved ejection fraction. *Eur J Heart Fail* 2020;**22**:413−421.
5. Ouwerkerk W, Tromp J, Jin X, Jaufeerally F, Yeo PS, Leong KT, Ong HY, Ling LH, Loh SY, Sim D, Lee S, Soon D, Chin C, Richards AM, Lam CS. Heart failure with preserved ejection fraction diagnostic scores in an Asian population. *Eur J Heart Fail* 2020;**22**:1737−1739.
6. Selvaraj S, Myhre PL, Vaduganathan M, Claggett BL, Matsushita K, Kitzman DW, Borlaug BA, Shah AM, Solomon SD. Application of diagnostic algorithms for heart failure with preserved ejection fraction to the community. *JACC Heart Fail* 2020;**8**:640−653.