ESC HEART FAILURE
ESC Heart Failure (2023)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.14525

ORIGINAL ARTICLE

Webtool to enhance the accuracy of diagnostic algorithms for HFpEF: a prospective cross-over study

Jerremy Weerts[1]*, Hesam Amin[1,2], Arantxa Barandiarán Aizpurua[1], Andreas B. Gevaert[3,4], M. Louis Handoko[5,6], Jeroen Dauw[7], Han Naung Tun[8], Karl-Philipp Rommel[9], Frederik H. Verbrugge[10,11], Karl-Patrik Kresoja[9], Sandra Sanders-van Wijk[12], Hans-Peter Brunner-La Rocca[1], Antoni Bayés-Genís[13], Joost Lumens[14], Christian Knackstedt[1] and Vanessa P.M. van Empel[1]

[1]*Department of Cardiology, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University Medical Centre+ (MUMC+), PO Box 616, 6200 MD, Maastricht, The Netherlands;* [2]*Department of Cardiology, Thoraxcentrum Twente, Medisch Spectrum Twente, Enschede, The Netherlands;* [3]*Research Group Cardiovascular Diseases, GENCOR Department, University of Antwerp, Antwerp, Belgium;* [4]*Department of Cardiology, Antwerp University Hospital (UZA), Edegem, Belgium;* [5]*Department of Cardiology, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands;* [6]*Amsterdam Cardiovascular Sciences/Heart Failure and Arrhythmias, Amsterdam, The Netherlands;* [7]*Department of Cardiology, AZ Sint-Lucas, Ghent, Belgium;* [8]*Larner College of Medicine, University of Vermont, Burlington, VT, USA;* [9]*Department of Internal Medicine/Cardiology, Heart Center Leipzig, Leipzig University, Leipzig, Germany;* [10]*Centre for Cardiovascular Diseases, University Hospital Brussels, Jette, Belgium;* [11]*Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Belgium;* [12]*Department of Cardiology, Zuyderland Medical Center Heerlen, Heerlen, The Netherlands;* [13]*Heart Institute, Hospital Universitari Germans Trias I Pujol, CIBERCV, Badalona, Spain; and* [14]*Department of Biomedical Engineering, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University, Maastricht, The Netherlands*

Abstract

Aims Diagnosis of heart failure with preserved ejection fraction (HFpEF) can be challenging. This study aimed to evaluate the potential of a webtool to enhance the scoring accuracy when applying the complex HFA-PEFF and H_2FPEF algorithms, which are commonly used for diagnosing HFpEF.

Methods and results We developed an online tool, the HFpEF calculator, that enables the automatic calculation of current HFpEF algorithms. We assessed the accuracy of manual vs. automatic scoring, defined as the percentage of correct scores, in a cohort of cardiologists with varying clinical experience. Cardiologists scored eight online clinical cases using a triple cross-over design (i.e. two manual–two automatic–two manual–two automatic). Data were analysed in study completers (n = 55, 29% heart failure specialists, 42% general cardiologists, and 29% cardiology residents). Manually calculated scores were correct in 50% (HFA-PEFF: 50% [50–75]; H_2FPEF: 50% [38–50]). Correct scoring improved to 100% using the HFpEF calculator (HFA-PEFF: 100% [88–100], $P < 0.001$; H_2FPEF: 100% [75–100], $P < 0.001$). Time spent on clinical cases was similar between scoring methods (±4 min). When corrections for faulty algorithm scores were displayed, cardiologists changed their diagnostic decision in up to 67% of cases. At least 67% of cardiologists preferred using the online tool for future cases in clinical practice.

Conclusions Manual calculation of HFpEF diagnostic algorithms is often inaccurate. Using an automated webtool to calculate HFpEF algorithms significantly improved correct scoring. This new approach may impact the eventual diagnostic decision in up to two-thirds of cases, supporting its routine use in clinical practice.

Keywords Heart failure with preserved ejection fraction; HFpEF; Diastolic heart failure; Diagnosis; eHealth; Online system

Received: 21 July 2023; Accepted: 22 August 2023
**Correspondence to: Jerremy Weerts, Department of Cardiology, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University Medical Centre+ (MUMC+), PO Box 616, 6200 MD, Maastricht, The Netherlands. Tel: +31 43 3871592; Fax: +31 43 3875104. Email: jerremy.weerts@mumc.nl*

Introduction

Heart failure (HF) with preserved ejection fraction (HFpEF) is a challenging clinical syndrome to diagnose. Atrial fibrillation (AF) and obesity particularly confound the HFpEF diagnosis due to symptom overlap and their impact on echocardiographic parameters and circulating natriuretic peptides (NPs).[1] Expert consensus suggests using the HFA-PEFF and H_2FPEF algorithms to aid the HFpEF diagnosis.[1–3] These algorithms were developed to assess the probability of HFpEF and standardize its diagnosis. However, the resulting scores may be discrepant, and many patients are categorized as having

an intermediate likelihood for HFpEF.[4,5] Consequently, additional diagnostic investigations with exercise echocardiography or invasive right heart catheterization are often required to confirm an HFpEF diagnosis.

Moreover, manually applying the diagnostic HFpEF algorithms is time-consuming and prone to errors. The HFA-PEFF algorithm in particular is complex with 10 different parameters and different cut-offs depending on age, sex, and presence of AF.[6] This complexity may result in diagnostic errors. New ways to optimize the application of the HFpEF algorithms might be helpful.

This study aims to evaluate the accuracy of manually scoring the HFA-PEFF and H_2FPEF algorithms compared with using an online diagnostic HFpEF scoring tool. Additionally, we aim to examine whether using an online tool can reduce the time spent on clinical cases and how cardiologists' diagnostic decisions change when they are provided with accurate algorithm scores.

Methods

Study design and recruitment

We performed an investigator-initiated cross-over study based on online clinical cases that were assessed by cardiologists (Supporting Information, *Figure S1*). Cardiology residents (maximum of 33% of total inclusions), general cardiologists, and cardiologists with HF expertise were invited for online participation. Study team members from each independent affiliated hospital recruited participants through social media posts and direct emails. Participants knew the topic of the study but were unaware of the primary analyses to minimize bias (Supporting Information, *Figure S2*). All participants provided online consent to participate. The study was approved by the medical ethics review committee academisch ziekenhuis Maastricht (Number 2022-3422) and complied with the principles of the Declaration of Helsinki.

Participants were first asked for demographic information and experience with HFpEF diagnosis and diagnostic algorithms. This included self-perceived knowledge of HFpEF diagnosis on an 11-point Likert scale, where 10 was marked as *completely familiar*. Then, eight clinical cases with suspicion of HFpEF (Supporting Information, *Table S1*) from a dedicated HFpEF outpatient clinic were structurally presented. The HFA-PEFF score ranged from 0 to 6 and the H_2FPEF score ranged from 1 to 9 in those eight clinical cases. None of the cases mentioned a previous HF hospitalization or results of exercise echocardiography or right heart catheterization. Participants were asked to diagnose a case based on the clinical scenario, then score the HFA-PEFF and H_2FPEF algorithms according to the initial publications,[6,7] and diagnose the case again (Supporting Information, *Methods* and *Figure S3*). The scoring method for the HFpEF algorithms was in a triple cross-over manner, alternating after every two cases between manual scoring and automatic scoring with the online tool (Supporting Information, *Figure S1*). After completing the eight cases, an optional survey was presented to assess the future diagnostic behaviour of the participants, including changes in diagnostic decisions if corrected scores from both algorithms were provided (Supporting Information, *Figure S4*). Participants who did not complete all eight clinical cases were reminded twice by email to complete the study and were excluded from analyses if fewer than four cases were completed.

Dedicated online research platform and diagnostic scoring tool

A dedicated online research platform was built using the Laravel v8 framework based on PHP 8.1 and JavaScript using MySQL as a relational database. This facilitated anonymized surveys for multiple-choice and open questions and provided a combined interface for clinical cases, figures, online tools, and questions (H. A. and J. W.). The website was accessible from all major browsers and devices. All answers were recorded, including time from opening to finishing a clinical case.

We developed an online scoring tool for the diagnostic HFpEF algorithms, the HFpEF calculator, as recommended by the European Society of Cardiology (ESC) and American Heart Association HF guidelines[1,2] (J. W. and H. A.). The HFpEF calculator (v2.2) collects all clinical variables to follow the specific criteria outlined by the ESC 2016 HF guidelines,[8] the HFA-PEFF algorithm,[6] and both categorical and numerical H_2FPEF algorithms.[7] The scoring tool was again assessed for its technical and medical output correctness for this study. Observers experienced in diagnosing HFpEF at a dedicated clinic confirmed the output of 18 cases with separate manual scoring after additional reviewing of the original publications and supplemental materials of both algorithms (J. W. and A. B. A.). The correctness was 100% (Bland–Altman bias of 0 with levels of agreement 0–0) (J. W. and A. B. A.). Using the same input for the tool through a batch analysis module (550 clinical cases submitted five times) provided identical output in all cases. The public version of the HFpEF calculator, which was not accessible during the study, can be found at cardiologytools.com/hfpef-calculator.

Study endpoint

The primary endpoint comprised accuracy evaluations of correct scoring of the HFA-PEFF and H_2FPEF algorithms between a manual and an automatic online scoring tool method.

Statistics

The endpoint was assessed per the intention-to-treat principle, that is, as if participants always used the scoring tool when provided. A paired Wilcoxon test was used to assess the accuracy differences (percentage of correct scores), as all differences were non-normally distributed (Shapiro's test P-value < 0.05). Correct score calculation differences were also evaluated between algorithms. The same test was used for categorical output from both diagnostic algorithms (low, intermediate, and high) to determine the correct calculations between the manual and online scoring tool methods. A stratified paired Wilcoxon test evaluated the primary outcome among the three participant groups (cardiology resident, general cardiologist, and HF specialist).

Confounders for lower score correctness of both algorithms were identified using univariable linear regression analyses for both scoring methods separately. Diagnostic agreement between participants was assessed based on their decision from the clinical scenario and after scoring both algorithms using the Fleiss' Kappa agreement test. Missing data were not imputed. Submitted scores that were clearly swapped between algorithms by participants were corrected before analyses (n = 7, i.e. scored HFA-PEFF 7 and H_2FPEF 4 instead of HFA-PEFF 4 and H_2FPEF 7) to minimize the effect of typographical errors not expected in clinical practice. Sensitivity analyses of the outcomes were performed in participants with ≥4 cases submitted. Data are presented as mean ± standard deviation, median [interquartile ranges], or number (percentage), as appropriate. A two-sided P-value < 0.05 was considered statistically significant. All data structuring and statistical analyses were performed using RStudio (v2023.03; R4.2.1), specifically using the packages *reshape2*[9] and *ggplot2*.[10]

Results

Characteristics of participating cardiologists

During the recruitment period between 19 December 2022 and 12 February 2023, 96 cardiologists enrolled in the study. The dropout rate was higher than anticipated, and the study recruitment was closed after we were notified that at least 35 participants completed all eight cases according to the sample size calculation (Supporting Information, *Methods*). Participants were given 1 month to finish their participation. The final dataset was collated on 15 March 2023 and included 59 (61%) cardiologists who had at least four cases completed and 55 (57%) all eight cases (Supporting Information, *Figure S1*). Cardiologists were equally distributed among HF expertise (cardiology resident, general cardiologist, and HF specialist) and sex and reported a median of 6 [3–10] years of clinical cardiology experience (*Table 1*). Most participants worked in Europe and at a university hospital. The self-perceived knowledge of HFpEF diagnosis at the start of the study was high, scoring 8 [7–8.5] on a 0–10 scale. A small number of participants (7%) indicated having never read the original publications of the HFA-PEFF or H_2FPEF algorithms, and 23% indicated not using either one in clinical care.

Correct scoring of the HFA-PEFF and H_2FPEF algorithms

Overall, calculating the correct scores of the HFA-PEFF and H_2FPEF algorithms, irrespective of the scoring method, was considerably lower than expected based on clinical experience. The correct scores were 75% [62–88] and 75% [62–75], respectively, in all eight cases regardless of scoring method, with a trend towards more often correct scores in the HFA-PEFF algorithm (P = 0.068). However, the number of correct scores was comparable between the HFA-PEFF and H_2FPEF algorithms when only manual scoring of the algorithms was applied (50% [50–75] vs. 50% [38–50], P = 0.185). Similarly, adequate categorization of the scores regardless of the scoring method was suboptimal in all eight cases, showing a correct assessment with the HFA-PEFF algorithm in 88% [75–88] and a correct assessment with the H_2FPEF

Table 1 Demographics of the participating physicians

Variables, n (%)	Total (n = 55)
Age category	
25–34	23 (42)
35–44	21 (38)
45–54	10 (18)
Female/male	26 (47%)/28 (51%)
Region	
Central and South-East Asia	2 (4)
Europe	49 (89)
Middle East	2 (4)
North and South America	2 (4)
Professional status	
Cardiologist in training	16 (29)
General cardiologist	23 (42)
Heart failure cardiologist	16 (29)
Hospital type	
Non-referral hospital	6 (11)
Referral hospital	12 (22)
University hospital	37 (67)
Clinical work setting	n = 41
General cardiology clinic	18 (44)
Specialized non-heart failure clinic	4 (10)
Heart failure clinic	11 (27)
Specialized HFpEF clinic	8 (19)
Clinical use of HFA-PEFF/H_2FPEF algorithm	n = 40
No	9 (23)
Yes	13 (33)
Only when in doubt based on clinical context	18 (45)
Algorithm used	n = 31
Both HFA-PEFF and H_2FPEF	15 (48)
HFA-PEFF only	6 (19)
H_2FPEF only	10 (32)

HFpEF, heart failure with preserved ejection fraction.

algorithm in 88% [81–88]. The categories of the HFA-PEFF score were accurate for 54% of the low likelihood category, 90% of the intermediate likelihood category, and 87% of the high likelihood category. The H_2FPEF score categories were correct in 97% of the low, 63% of the intermediate, and 94% of the high likelihood.

Compared with manual scoring of the algorithms, providing participants access to the online tool improved the correct scoring of the HFA-PEFF algorithm (50% [50–75] vs. 100% [88–100], $P < 0.001$) and the H_2FPEF algorithm (50% [38–50] vs. 100% [75–100], $P < 0.001$) (*Figure 1A*). The correct resulting categories of scoring the diagnostic algorithms (low, intermediate, or high) also improved with the online tool for the HFA-PEFF algorithm (75% [50–75%] vs. 100% [100–100], $P < 0.001$) and the H_2FPEF algorithm (75% [75–75] vs. 100% [100–100], $P < 0.001$) (*Figure 1B*). Improvement of correct absolute scores and categories of both algorithms was consistent among cardiology residents, cardiologists, and HF specialists (all $P < 0.05$). Results from cardiologists with at least four clinical cases submitted are comparable with those with eight cases (Supporting Information, *Figure S5*).

Detailed reviews of participants with correct scores below 50% while the online tool was provided ($n = 7$) revealed that those participants either partially entered data in the online tool or, in one case, did not use the webtool at all.

Time spent on clinical cases

Participants spent a total of 32 [25–44] min processing eight clinical cases, with a median of 4.0 [3.1–5.5] min per case. The average time spent on clinical cases did not differ between the manual and online tool methods, respectively, 3.7 [2.7–5.9] vs. 4.0 [2.7–5.1] min ($P = 0.769$) (Supporting Information, *Figure S6*).

Confounding factors for incorrect scoring of diagnostic heart failure with preserved ejection fraction algorithms

Linear regression analyses identified that more incorrect manual scoring of the HFA-PEFF algorithm was associated with participants aged between 35 and 44 years (*Figure 2* and Supporting Information, *Table S2*). No apparent factors were associated with more incorrect manual scoring of the H_2FPEF algorithm (Supporting Information, *Table S3*).

As previously mentioned, some participants used the online tool incompletely or submitted incorrect values. Cardiologists working at university hospitals and those who used the scores in clinical practice tended to use the HFpEF calculator more completely and scored more accurately (*Figure 2*).

The amount of correct absolute scores varied widely per clinical case (Supporting Information, *Figure S7*). The least frequent correct score included a clinical case with successfully ablated paroxysmal AF and aged >75 years for the HFA-PEFF algorithm (Case 6). This illustrates that the lower cut-off values for morphological markers in patients aged >75 years are likely not often incorporated when scoring manually. It also suggests different interpretations for applying paroxysmal AF in the algorithm as AF or non-AF. Cases with the most inaccurate scoring for the H_2FPEF algorithm (Cases 2 and 5) suggested that cardiologists used lateral or average instead of septal E/e' values > 9.

Current and future diagnostic behaviour

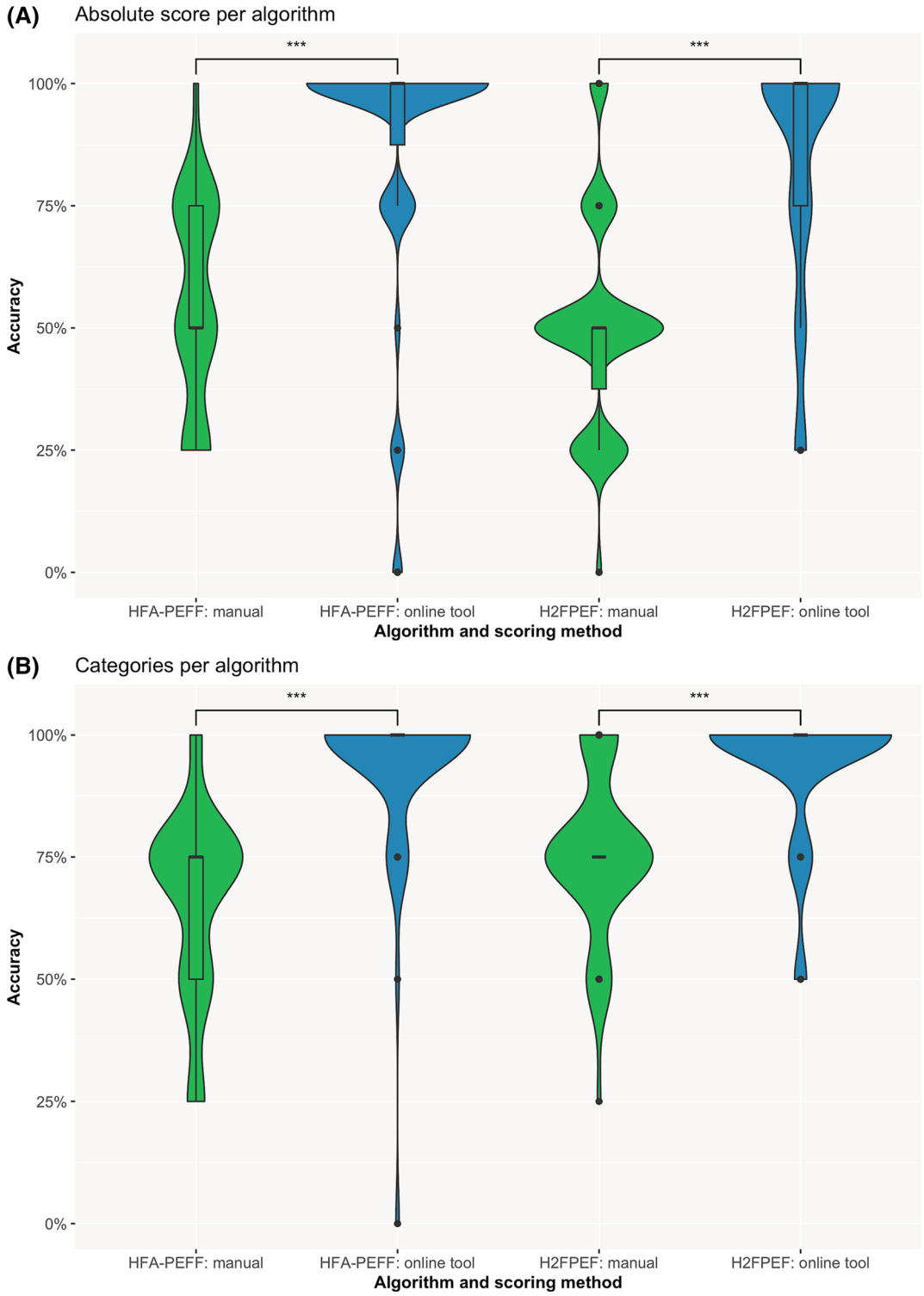
The diagnostic consensus of each case varied widely when cardiologists were asked for a diagnostic decision solely based on the clinical context before the algorithms (Fleiss' Kappa 0.374, $P < 0.001$) (*Figure 3*). After employing both HFpEF algorithms, cardiologists changed the diagnostic decision by 12% [12–31], and the diagnostic consensus improved (Fleiss' Kappa 0.483, $P < 0.001$). The highest agreements remained for the 'Yes HFpEF' diagnosis (Kappa 0.694) and the 'No HFpEF' diagnosis (Kappa 0.477). Diagnostic decisions were best aligned when a clinical case scored either high in both algorithms (Cases 1 and 4) or low (Case 5) in both algorithms.

After finishing all eight clinical cases, participants were shown the correct score in cases where a different score was initially submitted (4 [3–5] times) and were asked again for their diagnostic decision. The changes in diagnostic decisions mainly comprised more HFpEF diagnoses and fewer no HFpEF diagnoses (*Figure 4*). When correct scores were provided, participants changed their diagnostic decision in 0–67% of cases. In addition, 37 (67%) of all participants indicated to would want to use the online tool for future clinical cases, 5 (9%) would use both algorithms manually, 5 (9%) only the H_2FPEF algorithm manually, and 3 (5%) only the HFA-PEFF algorithm manually. The remaining participants would use both algorithms when in clinical doubt but did not specify their preferred method, 2 (4%), or did not answer this question, 3 (5%).

Discussion

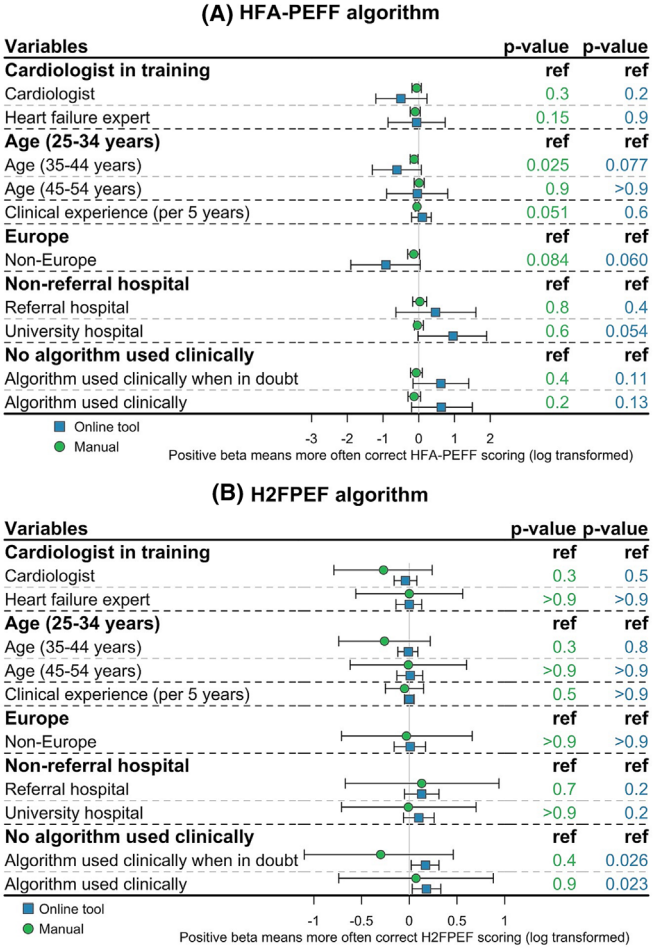
The present study investigated the potential clinical value of an online tool to enhance the accuracy of implementing the HFA-PEFF and H_2FPEF algorithms for diagnosing HFpEF. Correct scoring was substantially lower than expected when assessing the two algorithms manually but increased significantly when an online scoring tool was provided. The accurate assessment was similar among HF specialists,

Figure 1 Accuracy of scoring the HFA-PEFF or H_2FPEF algorithm by cardiologists based on the percentage of correct scores. The amount of correct absolute scores from both the HFA-PEFF and H_2FPEF algorithms significantly improved using an online tool instead of manual scoring (all $P < 0.001$) (A). Clinical cases were also better categorized in both algorithms using an online tool (B).



DOI: 10.1002/ehf2.14525

Figure 2 (A, B) Forest plot for confounders of accurate heart failure with preserved ejection fraction algorithms scoring.



general cardiologists, and cardiology residents. In addition, diagnostic decisions for HFpEF diagnosis varied widely between cardiologists based solely on clinical information and improved when diagnostic algorithms were used. Subsequently, the diagnostic behaviour of cardiologists changed in up to 67% of cases when correct scores were provided. Although time per case remained similar for both scoring methods, most participants preferred using the online tool over manually calculating the HFpEF scores in future clinical practice.

This study aligns with previous studies and ongoing developments to use technical solutions to facilitate healthcare processes, mainly to be more accurate on an individual patient level and possibly to reduce the time of simple tasks.[11,12] The variability of physician performance impacts decisions, such as internal cardioverter defibrillator indication for patients with HF with reduced ejection fraction based on left ventricular ejection fraction or for hypertrophic cardiomyopathy patients using maximal wall thickness,[13,14] and technical solutions may overcome parts of these hurdles. Along the same lines, enhancing the accurate scoring of HFpEF diagnostic algorithms by implementing online tools can decrease variability and is a first step to facilitating more accurate medicine.

The results of our study point to the challenges in diagnosing HFpEF. Both the HFA-PEFF and H_2FPEF algorithms are widely used in clinical research and are adopted in the most recent clinical HF guidelines.[1,2] However, we found suboptimal categorization of each diagnostic algorithm due to incorrect scoring, which may partially have contributed to the varying diagnostic decisions by cardiologists in identical cases. In addition, we report that the current diagnostic algorithms are not consistently applied in hospitals, even though most of our participants worked at university hospitals. Hence, clinical study populations and real-world outpatient populations may differ substantially. This observation enhances the urge to validate positive trials in patients with HFpEF in real-world clinical settings when such diagnostic algorithms are used. Moreover, it suggests that more simplified approaches for diagnosing HFpEF are needed. Promising tools

Figure 3 Distribution of diagnostic decisions per clinical case before and after heart failure with preserved ejection fraction (HFpEF) algorithms. Diagnostic decisions were distributed for all participants in each clinical case before the HFpEF algorithms were scored (A). A more uniform distribution of diagnostic decisions is seen after using both the HFA-PEFF and H_2FPEF algorithms (B).

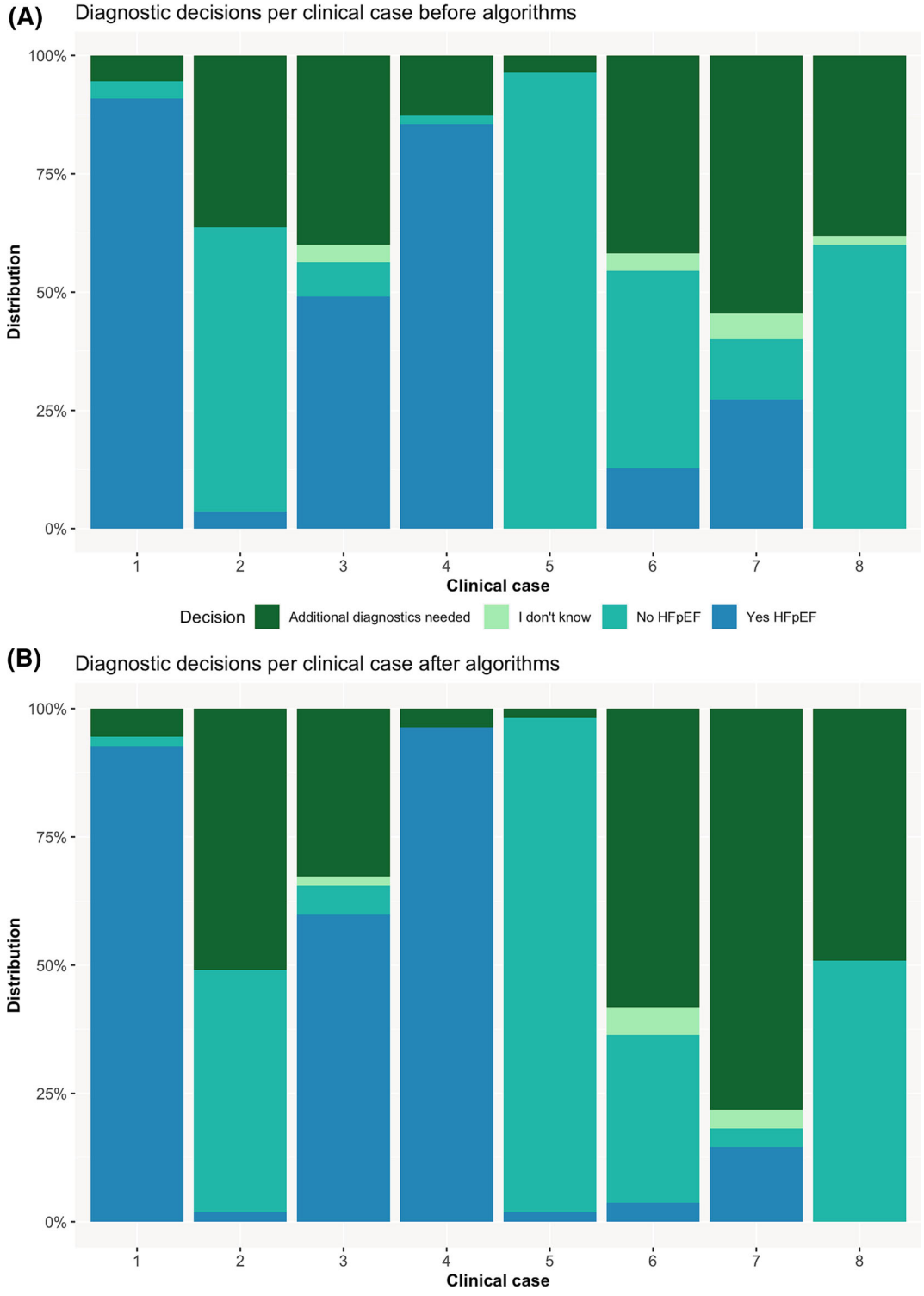
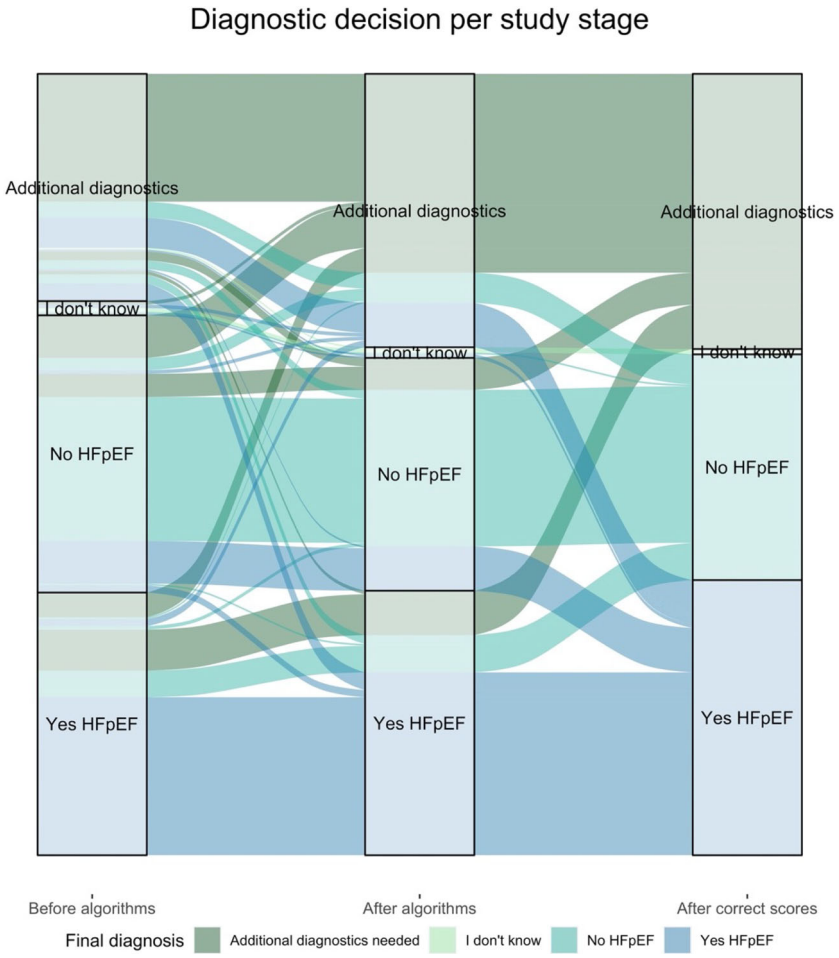


Figure 4 Sankey plot depicting the changes in diagnostic decisions of all participants for all clinical cases combined. The decisions are displayed for each study stage separately (*n* = 440 per stage). If no mistake was made, the decision showed at 'After correct score' was copied from the prior stage. HFpEF, heart failure with preserved ejection fraction.

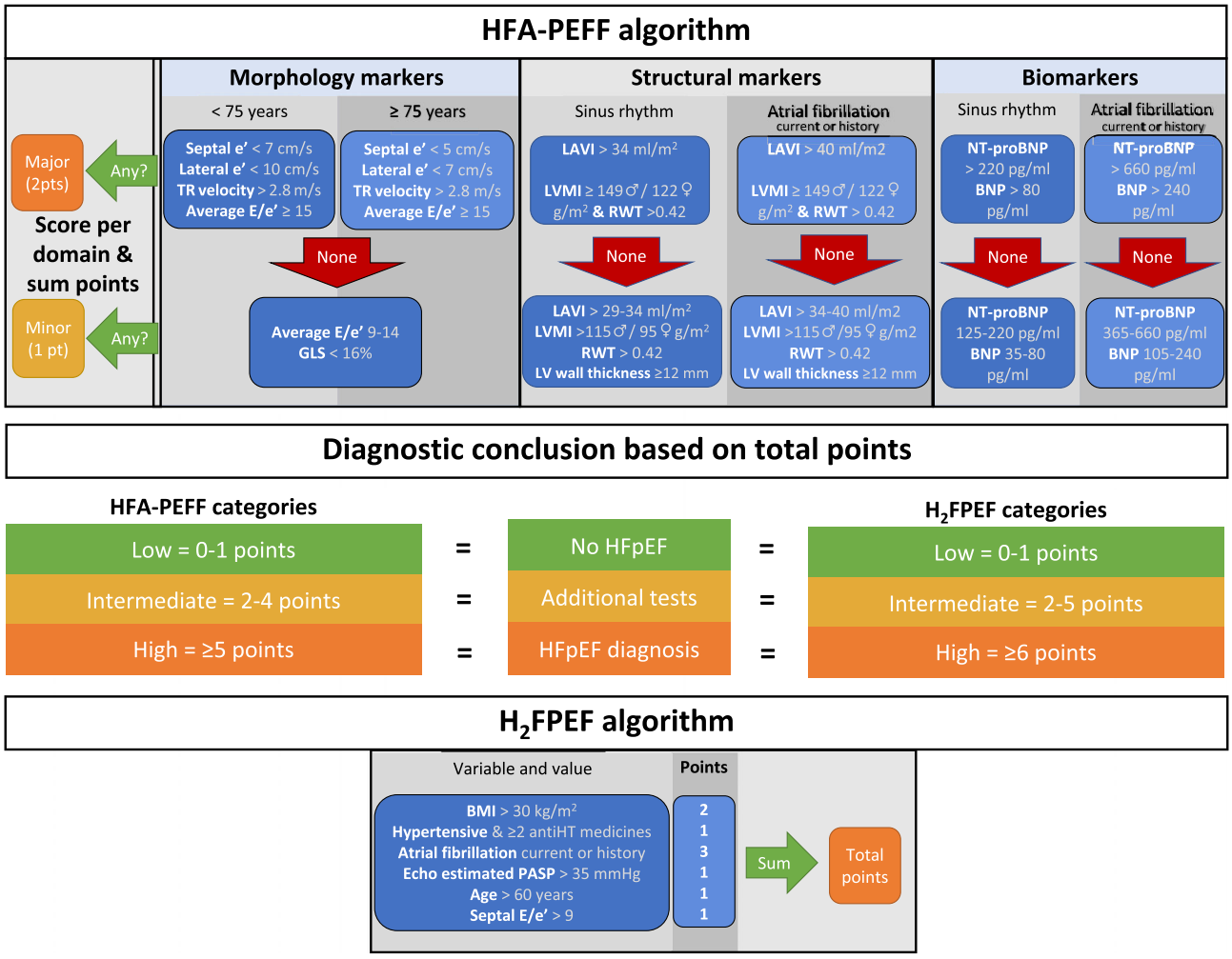


for simplifying HFpEF diagnosis may include markers of atrial dysfunction, mainly left atrial strain,[15,16] alongside dedicated HFpEF clinics with incorporated exercise testing.[17] However, until new simplified approaches have been developed and validated, tools that facilitate the proper use of the current algorithms are desired. As such, we propose using online tools, such as our HFpEF calculator, or at least a comprehensive graphic overview (*Figure 5*).

When physicians manually score either of the HFpEF algorithms, attention should be paid to some potential errors for accurate scoring. The HFA-PEFF scoring uses different cut-off values for patients aged >75 years (morphology markers), AF (left atrial volume index and B-type NP), and sex (left ventricular mass index) (*Figure 5*). Moreover, although the H_2FPEF algorithm is considered easy to score, absolute scores were correct in 50% and score categories in 75% in our study. These mistakes were likely largely attributed to the use of lateral or average E/e' instead of septal E/e' and are easily overcome, when measured.[18] The septal or average E/e' may also contribute to estimating elevated cardiac filling pressures but likely require higher cut-off values for diagnosing HFpEF than septal E/e'.[19]

Exploration of confounding factors for adequate algorithm scoring did not show remarkable results from the participant level. Some cardiologists incompletely used the online tool, which led to lower accuracies. The participating cardiologists mostly commented on the interpretation of paroxysmal AF for HFpEF diagnosis. It is well known that AF affects many of the clinical variables used in both algorithms.[6,20] Although rhythm therapies can facilitate left atrial reverse remodelling over the course of months, recurrence and progressive electrical and mechanical remodelling also take place.[21] Moreover, AF-related symptoms frequently can be attributed to HF, and AF episodes may be unnoticed. The longitudinal interaction between HFpEF and AF clearly remains a knowledge gap that should be addressed. Given the progressive nature

Figure 5 Graphical summary of the HFA-PEFF and H_2FPEF algorithms to diagnose heart failure with preserved ejection fraction (HFpEF). BMI, body mass index; GLS, global longitudinal strain; HT, hypertensive; LAVI, left atrial volume index; LV, left ventricular; LVMI, left ventricular mass index; NT-proBNP, N-terminal pro-brain natriuretic peptide; PASP, pulmonary artery systolic pressure; RWT, relative wall thickness; TR, tricuspid regurgitation.



of AF and its episodes that may be unnoticed, we propose to score patients with a paroxysmal AF history as AF in the HFA-PEFF and H_2FPEF algorithms.

Strengths and limitations

Our study has several strengths and limitations. The study was able to quickly enrol participants interested in the study using the study team's social media and personal networks, showing the high feasibility of this kind of implementation study. Although two personal reminders were sent to participants, only 56% completed the entire study. However, results were comparable between those who finished at least four or all eight cases. Even more importantly, the inclusion of more participants would have hardly influenced the main conclusions of this study, that is, the inaccuracy of manually scoring the two algorithms. Likely, participating cardiologists were more interested in HFpEF than non-participants. Given this selection bias, we expect the use of the diagnostic algorithms and their correct scoring to be lower outside this study. As such, an implementation tool to support HFpEF diagnosis could be even more beneficial outside a study setting. Even though this study tried to approximate real-life settings through online clinical cases, the actual difference in correct scoring or time between using the diagnostic HFpEF algorithms manually or with an online tool may differ in clinical practice. It could not be tracked if participants did simultaneous actions outside the study that could impact the time spent on a clinical case. Moreover, all relevant clinical information was provided in a single window, whereas electronic patient records often require different windows to be checked. Also, the online tool still relies on the manual transfer of clinical data, with the potential for typographical

errors. As such, using an online tool rather than manually scoring a diagnostic algorithm may demand less multitasking for physicians. Optimally, such a tool is directly integrated into hospital information systems as a clinical decision support system.[22]

Conclusions

Applying either the HFA-PEFF or H_2FPEF algorithm in clinical cases suspected of HFpEF yielded suboptimal accuracy due to miscalculations, which was significantly improved by an online scoring tool. The diagnostic decisions of cardiologists changed when provided with correct algorithm scores in up to two-thirds of cases. As such, improving the diagnosis and management of patients with HFpEF in clinical practice could be supported by using tools to score the HFpEF algorithms automatically.

Acknowledgements

We thank Michiel T. H. M. Henkens, MD, PhD, for supporting forest plot visualizations. We acknowledge all the cardiologists who attended this online study.

Conflict of interest

J.W. runs a web-development business, outside the submitted work. H.A. runs a web-development business, outside the submitted work. A.B.G. reports lecture/advisory board fees paid to his institution by Abbott, AstraZeneca, Boehringer Ingelheim, and Novartis, outside the submitted work. A.B.G. has received personal fees from Abbott, AstraZeneca, Boehringer Ingelheim, and Novartis, outside the submitted work. M.L.H. reports grants and personal fees from Novartis, Boehringer Ingelheim, and Vifor and personal fees from Daiichi Sankyo, AstraZeneca, Bayer, MSD, and Quin, outside the submitted work. J.D. has received speaker fees from Bayer, Boehringer Ingelheim, and AstraZeneca, outside the submitted work. J.D. reports personal fees from Bayer, Boehringer Ingelheim, and AstraZeneca, outside the submitted work. K.-P.K. reports travel grants by Edwards Lifesciences, outside the submitted work. H.-P.B.-L.R. reports grants and personal fees from Novartis, Roche Diagnostics, and Vifor and personal fees from Boehringer Ingelheim and AstraZeneca, outside the submitted work. A.B.-G. reports personal fees from Abbott, AstraZeneca, Boehringer Ingelheim, Novartis, Vifor, Critical Diagnostics, and Roche Diagnostics, outside the submitted work. C.K. reports advisory board fees paid to his institution by Pfizer, Alnylam, Philips/TOMTEC Imaging Systems, and Bristol Myers Squibb and research grants by AstraZeneca, Philips/TOMTEC Imaging Systems, and Boehringer Ingelheim, outside the submitted work. The other authors report no relationships that could be construed as a conflict of interest with the submitted work.

Funding

None.

Data availability statement

The full study protocol and data underlying this article will be shared on reasonable request to the corresponding author.

Supporting information

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

Figure S1. Study flow diagram.
Figure S2. Screenshot of informed consent page.
Figure S3. Screenshot of a clinical case in the study platform.
Figure S4. Screenshot of the survey held after finishing the 8 clinical cases.
Figure S5. Accuracy of scoring the HFA-PEFF and H_2FPEF algorithm in participants with $\geq$4 cases processed.
Figure S6. Time spent on assessment of the online clinical cases by participants.
Figure S7. The total amount of correct HFpEF algorithm scoring per clinical case of all participants.
Table S1. Clinical characteristics of presented patients in the cases.
Table S2. Confounders for accurately scoring the HFA-PEFF algorithm.
Table S3. Confounders for accurately scoring the H_2FPEF algorithm.

References

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, *et al*. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2021;**2021**: 3599-3726. doi:10.1093/eurheartj/ehab368
2. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, *et al*. 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2022;

145:e895-e1032. doi:10.1161/CIR.0000000000001063
3. Borlaug BA, Sharma K, Shah SJ, Ho JE. Heart failure with preserved ejection fraction. *J Am Coll Cardiol* 2023;**81**:1810-1834. doi:10.1016/j.jacc.2023.01.049
4. Sanders-van Wijk S, Barandiarán Aizpurua A, Brunner-La Rocca H-P, Henkens MTHM, Weerts J, Knackstedt C, *et al*. The HFA-PEFF and H_2FPEF scores largely disagree in classifying patients with suspected heart failure with preserved ejection fraction. *Eur J Heart Fail* 2021;**23**:838-840. doi:10.1002/ejhf.2019
5. Reddy YNV, Kaye DM, Handoko ML, van de Bovenkamp AA, Tedford RJ, Keck C, *et al*. Diagnosis of heart failure with preserved ejection fraction among patients with unexplained dyspnea. *JAMA Cardiol* 2022;**7**:891-899. doi:10.1001/jamacardio.2022.1916
6. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, *et al*. 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. doi:10.1093/eurheartj/ehz641
7. 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. doi:10.1161/CIRCULATIONAHA.118.034646
8. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, *et al*. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;**2016**:2129-2200. doi:10.1093/eurheartj/ehw128
9. Wickham H. Reshaping data with the reshape package. *J Stat Softw* 2007;**21**:1-20. doi:10.18637/jss.v021.i12
10. Wickham H. *ggplot2: Elegant Graphics for Data Analysis | SpringerLink*. New York, NY: Springer New York; 2009.
11. Amin H, Weerts J, Brunner-La Rocca H-P, Knackstedt C, Sanders-van Wijk S. Future perspective of heart failure care: Benefits and bottlenecks of artificial intelligence and eHealth. *Future Cardiol* 2021;**17**:917-921. doi:10.2217/fca-2021-0008
12. Lee DS, Straus SE, Farkouh ME, Austin PC, Taljaard M, Chong A, *et al*. Trial of an intervention to improve acute heart failure outcomes. *N Engl J Med* 2023;**388**:22-32. doi:10.1056/NEJMoa2211680
13. Kaufmann BA, Min SY, Goetschalckx K, Bernheim AM, Buser PT, Pfisterer ME, *et al*. How reliable are left ventricular ejection fraction cut offs assessed by echocardiography for clinical decision making in patients with heart failure? *Int J Cardiovasc Imaging* 2013;**29**:581-588. doi:10.1007/s10554-012-0122-5
14. Captur G, Manisty CH, Raman B, Marchi A, Wong TC, Ariga R, *et al*. Maximal wall thickness measurement in hypertrophic cardiomyopathy. *JACC Cardiovasc Imaging* 2021;**14**:2123-2134. doi:10.1016/j.jcmg.2021.03.032
15. Venkateshvaran A, Tureli HO, Faxén UL, Lund LH, Tossavainen E, Lindqvist P. Left atrial reservoir strain improves diagnostic accuracy of the 2016 ASE/EACVI diastolic algorithm in patients with preserved left ventricular ejection fraction: Insights from the KARUM haemodynamic database. *Eur Heart J Cardiovasc Imaging* 2022;**23**:1157-1168. doi:10.1093/ehjci/jeac036
16. Frydas A, Morris DA, Belyavskiy E, Radhakrishnan A-K, Kropf M, Tadic M, *et al*. Left atrial strain as sensitive marker of left ventricular diastolic dysfunction in heart failure. *ESC Heart Fail* 2020;**7**:1956-1965. doi:10.1002/ehf2.12820
17. Verwerft J, Soens L, Wynants J, Meysman M, Jogani S, Plein D, *et al*. Heart failure with preserved ejection fraction: Relevance of a dedicated dyspnoea clinic. *Eur Heart J* 2023;**44**:ehad141. doi:10.1093/eurheartj/ehad141
18. Sitges M, Ajmone Marsan N, Cameli M, D'Andrea A, Carvalho RF, Holte E, *et al*. EACVI survey on the evaluation of left ventricular diastolic function. *Eur Heart J Cardiovasc Imaging* 2021;**22**:1098-1105. doi:10.1093/ehjci/jeab087
19. Mitter SS, Shah SJ, Thomas JD. A test in context: E/A and E/e' to assess diastolic dysfunction and LV filling pressure. *J Am Coll Cardiol* 2017;**69**:1451-1464. doi:10.1016/j.jacc.2016.12.037
20. Weerts J, Barandiarán Aizpurua A, Henkens MT, Lyon A, van Mourik MJ, van Gemert MR, *et al*. The prognostic impact of mechanical atrial dysfunction and atrial fibrillation in heart failure with preserved ejection fraction. *Eur Heart J-Cardiovasc Imaging* 2022;**23**:74-84. doi:10.1093/ehjci/jeab222
21. Lo L-W, Chen S-A. Cardiac remodeling after atrial fibrillation ablation. *J Atr Fibrillation* 2013;**6**:877. doi:10.4022/jafib.877
22. Peters LJ, Torres-Castaño A, van Etten-Jamaludin FS, Perestelo Perez L, Ubbink DT. What helps the successful implementation of digital decision aids supporting shared decision-making in cardiovascular diseases? A systematic review. *Eur Heart J - Digit Health* 2023;**4**:53-62. doi:10.1093/ehjdh/ztac070