Three rules to find them all: Clinical risk scores in heart failure with preserved ejection fraction

A.Akerman[1], W. Hawkes[1], N. Al-Roub[2], C. Scott[3], C. Angell-James[2], H. Piotrowska[1], P. Leeson[4], G. Woodward[1], J. Strom[2], P. Pellikka[3], R. Upton[1]

[1]Ultromics Ltd, Oxford, UK; [2]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, US; [3]Mayo Clinic, Rochester, US; [4] University of Oxford, Oxford, UK

Purpose: Multiple scoring rules and clinical decision aids exist to support the identification of heart failure with preserved ejection fraction (HFpEF) but it remains frequently unrecognised or misdiagnosed. The purpose of the current study was to validate multi-site diagnostic performance of clinical scores, independently and combined, in identifying HFpEF .

Methods: Independent patients undergoing clinically indicated echocardiograms at Mayo Clinic (1) and Beth Israel Deaconess Medical Center (2) were retrospectively identified. Risk of HFpEF was assessed according to three validated algorithms; the H2FPEF and HFA-PEFF scores, and EchoGo Heart Failure (Ultromics). The H2FPEF score and HFA-PEFF score are multiparametric clinical models, and EchoGo Heart Failure is an AI computer vision model using a single echocardiographic video input. The continuous outputs from the H2FPEF score and EchoGo Heart Failure were combined with the HFA-PEFF categorical score (logistic regression) to provide a unique prediction ("Three Scores"). Discrimination, calibration, classification, and clinical utility were assessed.

Results: Compared with patients without HFpEF (n=886), patients with HFpEF (n=894) were slightly older (73 vs. 68 y), had more comorbidities, and more pronounced cardiac dysfunction. The AI model and H2FPEF continuous score demonstrated high discrimination (AUROC; **Figure 1**), and similar calibration (**Figures 2-4**), both of which were improved when combining all three scores. EchoGo Heart Failure categorised 50.2% of patients as high likelihood of HFpEF, 40.2% as low likelihood of HFpEF, and 9.6% of patients as intermediate (**Table 2**). The H2FPEF score and HFA-PEFF score categorised 27.8% and 22.4% patients as high likelihood of HFpEF, 11.5% and 23.1% as low likelihood of HFpEF, and 60.8% and 54.5% of patients as intermediate (respectively; **Table 2**). The Three Scores combined demonstrated high sensitivity (91%) and specificity (82%). At a decision threshold probability of 30%, managing patients based on EchoGo Heart Failure output resulted in 27% more correct decisions than H2FPEF score, but the combined Three Scores increased correct decisions by a further 12% (**Central Figure**).

Conclusion: The integration of existing clinical scores and AI models may be the most valuable approach to diagnosing heart failure with preserved ejection fraction.

Utilizing all available information from clinical risk scores and decision support aids could increase utility in the management of HFpEF

AI can support traditional methods in the diagnosis of heterogenous and often missed/misdiagnosed cardiomyopathies

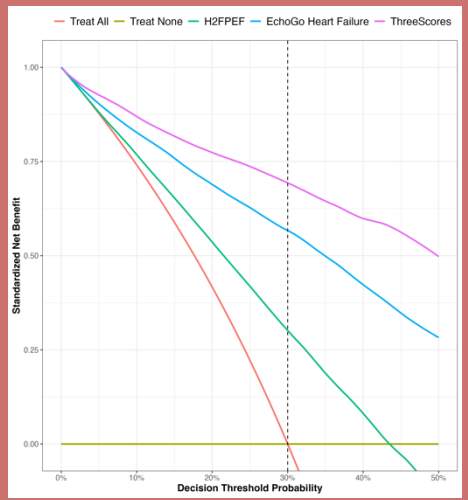

Table 1. Descriptive Statistics

Class	Age (y)	Female	AF	DM	HTN	BMI (kg/m2)	LVMi (g/m2)	LAVi (mL/m2)	EF (%)	E/e'	TRv (s)	H2FPEF (%)	EchoGo HF (%)
Case	73±13	53%	36%	48%	87%	30.3±7.0	119±35.9	45.6±15.3	62±6	19±7	3.0±0.5	30±31	80±22
Control	68±17	52%	14%	18%	54%	28.2±6.5	85.5±23.0	30.0±26.9	63±7	10±4	2.5±0.4	45±25	82±23

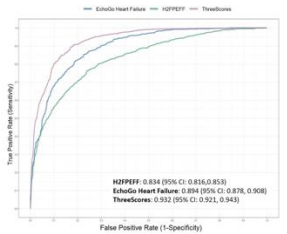

Figure 1. Area under the receiver operator characteristic curve (AUROC)

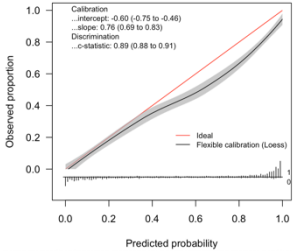

Figure 2. Flexible calibration curve for EchoGo Heart Failure

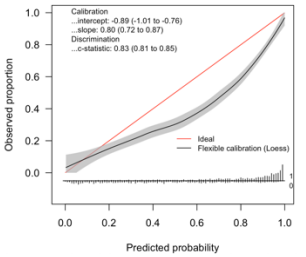

Figure 3. Flexible calibration curve for H2FPEF Score

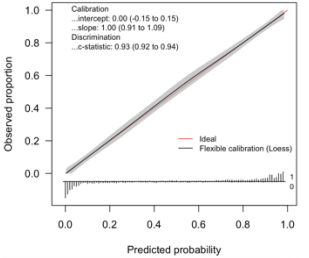

Figure 4. Flexible calibration curve for Three Scores

Table 2. Classification Statistics

Classification Statistic (%)	EchoGo Heart Failure	H2FPEF Score	HFA-PEFF Score	Three Scores
Intermediate	9.6	60.8	54.5	0
Sensitivity (+ve/-ve)	89.0 (86.2, 91.3)	98.0 (96.4, 99.4)	96.6 (94.5, 98.3)	90.7 (88.6, 92.8)
Sensitivity (all)	83.4 (81.8, 84.7)	49.8 (47.8, 52.1)	47.0 (44.9, 49.2)	
Specificity (+ve/-ve)	78.8 (74.7, 81.9)	66.0 (58.7, 72.3)	92.6 (89.9, 95.1)	81.7 (78.7, 84.8)
Specificity (all)	71.5 (70.5, 72.8)	25.3 (24.5, 26.4)	49.8 (48.2, 51.6)	
Negative Predictive Value	87.4 (84.9, 90.2)	96.1 (93.2, 98.8)	96.8 (95.0, 98.5)	89.9 (87.7, 92.1)
Positive Predictive Value	81.2 (78.2, 84.4)	79.6 (75.1, 83.4)	92.0 (89.1, 94.9)	83.1 (80.4, 86.0)

References: *1.https://www.jacc.org/doi/10.1016/j.jacadv.2023.100452. 2.https://www.nature.com/articles/s41467-025-58283-7*

Heart Failure & World Congress on Acute Heart Failure 2025