







# Artificial intelligence fully automated analysis of handheld echocardiography in real-world patients with suspected heart failure

Ross T. Campbell<sup>1\*</sup>, Mark C. Petrie<sup>1</sup>, Kieran F. Docherty<sup>1</sup>,  
 Katriona J.M. Brooksbank<sup>1</sup>, Gemma McKinley<sup>2</sup>, Caroline Haig<sup>2</sup>,  
 Alex McConnachie<sup>2</sup>, Carolyn S.P. Lam<sup>3</sup>, Carly Adamson<sup>1</sup>, Elaine Butler<sup>1</sup>,  
 James P. Curtain<sup>1</sup>, Nick Hartshorne-Evans<sup>4</sup>, Fraser J. Graham<sup>5</sup>, Helen Hailey<sup>5</sup>,  
 John Jarvie<sup>1</sup>, Matthew M.Y. Lee<sup>1</sup>, Leeanne Macklin<sup>6</sup>, Kenneth Mangion<sup>5</sup>,  
 Aimee McCoubrey<sup>1</sup>, Kirsty McDowell<sup>1</sup>, Aileen McIntyre<sup>7</sup>, Sabrina Nordin<sup>5</sup>,  
 Joanna Osmanska<sup>1</sup>, Pierpaolo Pellicori<sup>1</sup>, Joanne Simpson<sup>5</sup>, Piotr Sonecki<sup>5</sup>,  
 Karen Taylor<sup>5</sup>, Daniel Taylor-Sweet<sup>1</sup>, Pamela Turnbull<sup>8</sup>, Paul Welsh<sup>1</sup>,  
 John J.V. McMurray<sup>1</sup>, Clare L. Murphy<sup>6</sup>, and David J. Lowe<sup>9,10</sup>

<sup>1</sup>School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, UK; <sup>2</sup>Robertson Centre for Biostatistics, School of Health and Wellbeing, University of Glasgow, Glasgow, UK; <sup>3</sup>National Heart Centre Singapore and Duke-National University of Singapore, Singapore, Singapore; <sup>4</sup>Pumping Marvellous, Preston, UK; <sup>5</sup>Cardiology Department, Queen Elizabeth University Hospital, Glasgow, UK; <sup>6</sup>Cardiology Department, Forth Valley Royal Hospital, Larbert, UK; <sup>7</sup>Research and Innovation department, NHS Greater Glasgow and Clyde, Glasgow, UK; <sup>8</sup>Cardiology Department, Glasgow Royal Infirmary, Glasgow, UK; <sup>9</sup>School of Health and Wellbeing, University of Glasgow, Glasgow, UK; and <sup>10</sup>Emergency Department, Queen Elizabeth University Hospital, Glasgow, UK

Received 8 April 2025; revised 8 June 2025; accepted 27 June 2025

## Aims

Echocardiography is a rate-limiting step in the timely diagnosis of heart failure (HF). Automated reporting of echocardiograms has the potential to streamline workflow. The aim of this study was to test the diagnostic accuracy of fully automated artificial intelligence (AI) analysis of images acquired using handheld echocardiography and its interchangeability with expert human-analysed cart-based echocardiograms in a real-world cohort with suspected HF.

## Methods and results

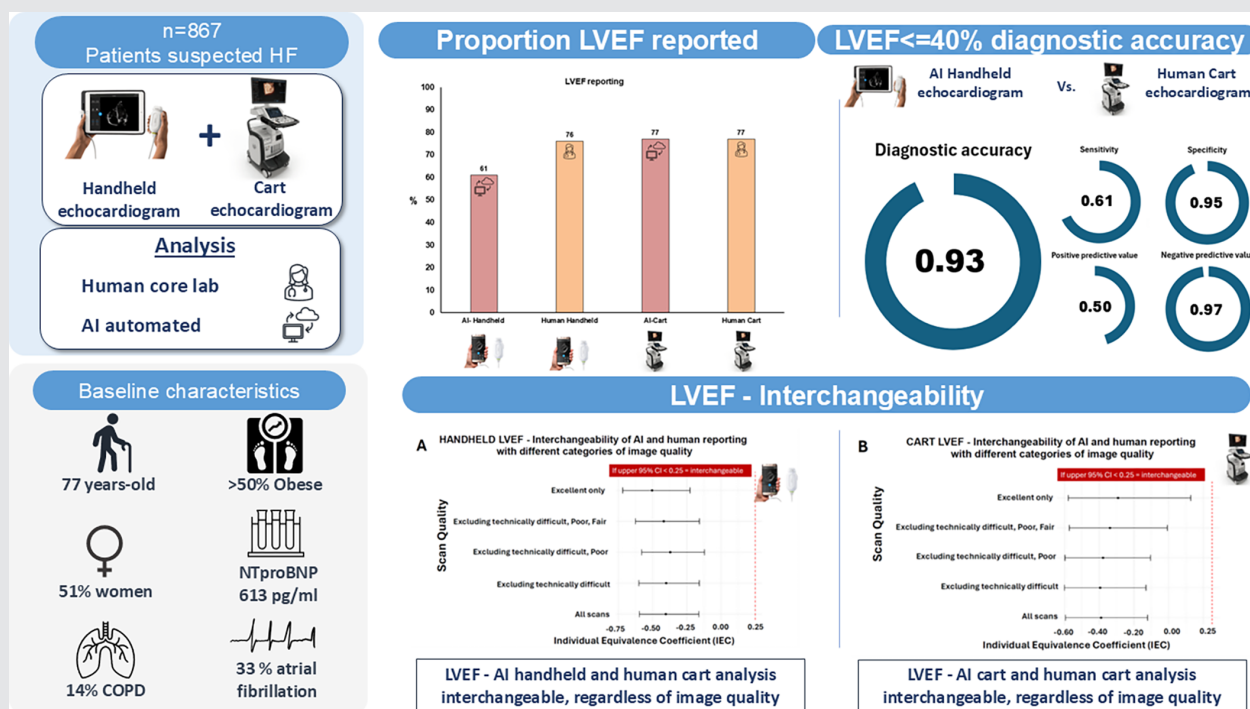
In this multicentre, prospective, observational study, patients with suspected HF had two echocardiograms: one handheld portable and one cart-based scan. Both echocardiograms were analysed using fully automated AI software and by human expert sonographers. The primary endpoint was the diagnostic accuracy of AI-automated analysis of handheld echocardiography to detect left ventricular ejection fraction (LVEF)  $\leq 40\%$ . Other endpoints included the interchangeability (assessed using individual equivalence coefficient [IEC]), between AI-automated and human analysis of cart-based LVEF. A total of 867 patients participated. The AI-automated analysis produced an LVEF in 61% of the handheld scans and 77% of the cart-based scans, compared to 76% and 77% of human analyses of the handheld and cart-based scans, respectively. The AI-automated analysis of handheld echocardiography had a diagnostic accuracy of 0.93 (95% confidence interval [CI] 0.90, 0.95) for identifying LVEF  $\leq 40\%$  (compared to the human analysis of cart-based transthoracic echocardiography scans). AI-automated analysis of LVEF on handheld devices was interchangeable with cart-based LVEF reported by two expert humans (IEC  $-0.40$ , 95% CI  $-0.60$ ,  $-0.16$ ).

## Conclusions

Artificial intelligence-automated analysis of handheld echocardiography had good diagnostic accuracy for detecting LVEF  $\leq 40\%$ . AI-automated analysis of LVEF of handheld scans was interchangeable with cart-based expert human analysis.

\*Corresponding author. School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow Cardiovascular Research Centre, 126 University Place, Glasgow G12 8TA, UK. Email: ross.campbell@glasgow.ac.uk

## Graphical Abstract



Artificial intelligence (AI) reporting of handheld echocardiography in a real-life cohort of patients with suspected heart failure (HF). COPD, chronic obstructive pulmonary disease; LVEF, left ventricular ejection fraction; NTproBNP, N-terminal pro-B-type natriuretic peptide.

## Keywords

Heart failure • Diagnosis • Artificial intelligence • Echocardiography • Point of care • Handheld

## Introduction

Transthoracic echocardiography (TTE) is the guideline-recommended diagnostic test for heart failure (HF) and guides the initiation of disease-modifying and prognostic therapies based on left ventricular ejection fraction (LVEF).<sup>1</sup> In patients presenting in the community with suspected HF, a delayed diagnosis is associated with worse outcomes including hospitalization due to HF and increased mortality.<sup>2</sup> The main rate-limiting step in the timely diagnosis of HF is access to TTE. Delays in TTE are often driven by a supply–demand mismatch, with the number of referrals overwhelming the ability of many healthcare systems to provide this important diagnostic test in a timely fashion. The demand for TTE is likely to worsen with ageing populations and the increasing prevalence of HF.<sup>3</sup>

Transthoracic echocardiography is a complex diagnostic examination which takes an expert approximately 45 minutes to acquire and interpret images, usually in a secondary care setting.<sup>4</sup> Artificial intelligence (AI) algorithms which perform fully automated analysis of TTE images are as accurate as the measurements made

by human experts and could accelerate workflow and increase capacity when incorporated into a HF diagnostic pathway.<sup>5,6</sup> The combination of automated AI analysis and portable handheld TTE devices has the potential to increase the availability of diagnostic testing for HF in the community. However, the performance of these algorithms and their use along with handheld TTE devices requires the assessment of their performance in unselected ‘real-world’ populations of patients with suspected (as opposed to confirmed) HF, many of whom may have comorbidities such as chronic obstructive pulmonary disease and obesity that may reduce image quality. Handheld echocardiography has many potential benefits including ease of transport to environments where patients with suspected HF have clinical interactions (e.g. primary care and pharmacies). Handheld devices are less expensive than cart machines.

In this study, we tested the diagnostic accuracy and interchangeability of automated AI reporting of images acquired using handheld TTE with human analysis (both routine clinical care and core laboratory-based analyses) of cart-based TTE in a real-world cohort of patients with suspected HF.

## Methods

### Study design, setting and participants

This was a multicentre, prospective, observational study conducted at five outpatient sites in Glasgow, United Kingdom. Consecutive patients who were attending a clinical HF diagnostic pathway (online supplementary Figure S1) were prospectively screened for inclusion. The clinical HF diagnostic pathway is between primary care and secondary care and patients are referred by their general practitioner with a clinical suspicion of HF. The exclusion criteria were a prior diagnosis of HF or if the patient was unable or unwilling to provide informed consent. No patients were excluded based on their echocardiogram image quality. The current study was designed before the COVID-19 pandemic and carried out during the pandemic. This study was performed according to the UK Policy Framework for Health and Social Care Research and the Declaration of Helsinki and was approved by the West of Scotland Research Ethics Committee (20/SW/0182) and the Health Research Authority. All patients provided written informed consent. This study is registered at ClinicalTrials.gov identifier: NCT04724200. N-terminal pro-B-type natriuretic peptide (NT-proBNP) testing was repeated at the time of the study assessment (Roche CARDIAC proBNP+ assay run on the Roche cobas h232 instrument).

### Echocardiography

Patients who consented had two echocardiograms performed by one sonographer on the same day at the same clinic visit in the following order:

1. A handheld echocardiogram, using a Phillips® Lumify scanner. The scanning protocol for the handheld included two-dimensional (including color flow Doppler) parasternal long-axis, apical two-, three-, and four-chamber views.
2. A conventional TTE scan ('cart-based' scan), performed according to current guidelines<sup>4,7</sup> using a GE® Vivid E95 scanner. All TTEs were performed by accredited sonographers or cardiologists (either European Association of Cardiovascular Imaging or British Society of Echocardiography accredited) (HH, KT, PT, PP, FG).

### Echocardiogram analysis

The following echocardiogram analyses were performed:

1. Handheld echocardiography reported by Us2.ai® (commercially available, CE marked and Food and Drug Administration-approved software for automated AI analysis, version 1.4.1)
2. Handheld echocardiography reported by two blinded core laboratory sonographers.
3. Cart-based echocardiogram reported by the clinical sonographer as part of usual care at the time of clinic appointment.
4. Cart-based echocardiogram reported by a blinded core laboratory sonographer.
5. Cart-based echocardiogram reported by Us2.ai automated AI analysis

Core laboratory assessors were blinded to all measurement performed by the usual care sonographer. The Us2.ai software allows for fully automated and unsupervised analysis, where the entire echocardiogram (in Digital Imaging and Communications in Medicine format) is analysed, with no need for manual classification of image sequence. The

development, training, and validation of the deep-learning algorithm and workflow are described in detail elsewhere.<sup>5,6</sup> An example of the automated report is shown in online supplementary Figure S2.

The clinical TTE report performed during usual care was obtained using GE® EchoPac (version 2.04) and manually reported in keeping with contemporary guidelines.<sup>7</sup> Simpson's biplane measurement of LVEF was performed manually (i.e. not using the semi-automated analysis available in EchoPac). Blinded core laboratory sonographers (RTC/KM/SN/MMYL/PP/JC/PS/JJ) used StudyCast® software. Echocardiograms analysed by humans in the core laboratory were anonymized.

### Outcomes

#### Primary endpoint

The primary endpoint was the diagnostic accuracy of AI-automated analysis of handheld TTE to identify an LVEF  $\leq 40\%$ . The 'gold standard' used for the purposes of comparison with the AI-automated analysis of handheld TTE was the assessment of LVEF during the human analysis of the cart-based TTE images as part of usual clinical care. Diagnostic accuracy was defined as: (true positive + true negative)/(true positive + true negative + false positive + false negative). Sensitivity, specificity, negative and positive predictive values and area under the receiver operating characteristic (ROC) curve were also calculated.

#### Secondary endpoints

Secondary endpoints included: (i) the interchangeability (assessed by individual equivalence coefficient [IEC]) of LVEF as measured by AI-automated handheld TTE and human reporting of cart-based TTE (both usual care clinical and core laboratory reports); and (ii) the interchangeability of echocardiographic measures of HF with preserved ejection fraction (HFpEF) between AI-automated and human reporting of cart-based TTE (both usual care clinical and core lab reports).

#### Exploratory endpoints

Exploratory endpoints included the interchangeability of AI-automated cart-based LVEF compared to human reporting of cart-based LVEF (both usual care clinical and core laboratory reports).

The proportion of all scans where an LVEF was calculated by each echocardiography method was also reported. The agreement between AI-automated analyses of LVEF (for both handheld and cart-based images) and human core laboratory reports, and the agreement between AI-automated assessment of echocardiographic measures of HFpEF and human core laboratory reports (for cart-based images) were reported.

### Statistical analyses

The primary endpoint of the diagnostic accuracy of AI-automated analysis of handheld TTE to identify an LVEF  $\leq 40\%$  compared to human clinical assessment of LVEF  $\leq 40\%$  was assessed by diagnostic accuracy, negative and positive predictive value, false positive, false negative, sensitivity, specificity, and area under the ROC curve.

The secondary endpoints comparing the interchangeability of LVEF as measured by AI-automated analysis and human analyses of cart-based TTE were assessed by the IEC. The IEC is a measure of the level of disagreement between AI-automated measurements and human measurements relative to the disagreement among the human measurements. For the purposes of analysis of the IEC in

this study, human measurements refer to the results provided by the human-based usual care clinical and human-based core laboratory TTE reports (i.e. two human reports). The IEC can be calculated using the following formula:  $IEC = [Q_{TR} - Q_{RR}] / (Q_{RR} / 2)$ .  $Q_{RR}$  refers to the mean of the squared differences between two pairs of within-patient human (reference) measurements, and  $Q_{TR}$  refers to the mean of the squared differences between within-patient measurements from the automated analysis and each of two human measurements. An IEC of less than 0 (i.e. a negative value) indicates less variability between AI-automated analysis and human analysis than the variability between the human analyses (i.e. between the human-based usual care clinical and core laboratory TTE reports), and a value greater than 0 (i.e. a positive value) indicates greater variability between automated analysis and humans than between the human analyses. As per previous reports, we defined interchangeability between the imaging reporting modalities as an upper bound of the 95% confidence interval (CI) of the IEC of  $<0.25$ .<sup>5</sup> Agreement between AI-automated (both handheld and cart-based scans) and human core laboratory analyses of the cart-based scans was assessed using Pearson correlation coefficients and presented graphically as Bland–Altman plots.

### Image quality

All TTEs (both handheld and cart-based scans) were assessed for image quality by the core laboratory sonographers. Image quality was categorized into five groups (excellent, good, fair, poor, technically difficult) (online supplementary Table S1). IEC for both handheld and cart scans were presented in the five categories of image quality.

### Sample size calculation

Clinical audits of the diagnostic pathway have reported a prevalence of  $LVEF \leq 40\%$  between 7–30% over the preceding 10 years. Therefore, assuming that the prevalence would be 10% and the marginal error of this estimate does not exceed 2%, with 95% confidence level, a population size of 864 participants would be adequately powered to accurately predict the risk of  $LVEF \leq 40\%$  in individuals referred from the community for investigation of HF.<sup>8</sup> From prior studies, sample sizes of 600 participants have been adequately powered to evaluate the accuracy of automatic echocardiogram measurements compared to manual human measurements.<sup>9,10</sup> Our sample size was expected to provide a population size with enough power to additionally evaluate the accuracy of AI compared to human echocardiogram measurements.

### Study and data management

The Clinical Trials Unit at the Robertson Centre for Biostatistics (University of Glasgow) was responsible for data management and statistical analysis. All analyses were performed using R (version 4.0.0). This study is reported in accordance with the Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (DECIDE-AI) criteria (online supplementary Table S2).<sup>11</sup>

### Patient involvement

We co-developed the study with patients with lived experience of HF. This included the design, management, and conduct of the study. A member of a patient-led charity (Pumping Marvellous) acted as a member of the study steering committee.

## Results

### Baseline characteristics

Between 5 January and 26 August 2021, 1238 patients attended the clinical HF diagnostic pathway, and 867 were enrolled in the study. Baseline characteristics are presented in Table 1 and the Graphical Abstract. The median age was 77 years and 49% were male. The median body mass index was  $31 \text{ kg/m}^2$ . One-third of patients had a history of atrial fibrillation and 14% had a known diagnosis of chronic obstructive pulmonary disease. Median NT-proBNP was 613 (288–1159) pg/ml. A diagnosis of HF with reduced ejection fraction (HFrEF) was made in 44 patients (5.6%) (on the human-reported clinical echocardiogram 41 had an  $LVEF \leq 40\%$  by Simpson's biplane and three had an  $LVEF \leq 40\%$  based on visual assessment of  $LVEF$  where a Simpson's biplane was not possible).

**Table 1** Baseline characteristics

| Characteristic                                  |                  |
|---|------------------|
| <i>n</i>  | 867              |
| Age, years                                      | 77 (69, 83)      |
| Male sex, <i>n</i> (%)                          | 428 (49)         |
| White race, <i>n</i> (%)                        | 823 (98)         |
| BMI, $\text{kg/m}^2$                            | 31 (27, 36)      |
| SBP, mmHg                                       | 148 (131, 165)   |
| Heart rate, bpm                                 | 73 (64, 85)      |
| Past medical history, <i>n</i> (%)              |                  |
| Hypertension                                    | 501 (59)         |
| Type 2 diabetes                                 | 169 (20)         |
| Myocardial infarction                           | 106 (13)         |
| Chronic kidney disease                          | 167 (20)         |
| Atrial fibrillation                             | 275 (33)         |
| Chronic obstructive pulmonary disease           | 117 (14)         |
| Physical examination, <i>n</i> (%)              |                  |
| Elevated JVP ( $>4 \text{ cm}$ )                | 39 (6)           |
| Pulmonary crepitations                          | 83 (17)          |
| Peripheral oedema                               | 571 (69)         |
| Baseline blood tests                            |                  |
| NT-proBNP, pg/ml                                | 613 (288, 1159)  |
| eGFR, $\text{ml/min/1.73 m}^2$                  | 74 (58, 88)      |
| HbA1c, mmol/mol                                 | 41 (38, 45)      |
| HF with reduced ejection fraction, <i>n</i> (%) | 44 (6)           |
| Clinical cart-based echocardiogram              |                  |
| LVEF, %   | 62 (54, 66)      |
| LVIDd, cm                                       | 4.6 (4.2, 5.0)   |
| HFpEF parameters                                |                  |
| E/e'  | 10.2 (8.2, 12.9) |
| LAVI, $\text{ml/m}^2$                           | 37 (29, 49)      |
| LVMI, $\text{g/m}^2$                            | 94 (79, 114)     |
| TR velocity, m/s                                | 2.6 (2.3, 2.8)   |

Values expressed as *n* (%) or median (quartile 1, quartile 3).

BMI, body mass index; eGFR, estimated glomerular filtration rate; HbA1c, glycated haemoglobin; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; JVP, jugular venous pressure; LAVI, left atrial volume index; LVEF, left ventricular ejection fraction; LVIDd, left ventricular internal diameter in diastole; LVMI, left ventricular mass index; NT-proBNP, N-terminal pro-B-type natriuretic peptide; SBP, systolic blood pressure; TR, tricuspid regurgitation.



## Left ventricular ejection fraction reporting by artificial intelligence-automated and humans

### Handheld transthoracic echocardiography

On the handheld device, an LVEF was reported in 61% of scans when analysed by the AI-automated software. This is compared with an LVEF reported in 83% and 69% of the two human core laboratory readers, respectively, with an average human core laboratory reporting rate of 76% (Figure 1).

### Cart-based transthoracic echocardiography

On the cart-based TTE images, AI-automated reporting resulted in a similar number of LVEF reports (77%) as the human-based usual care clinical report (72%) and the human core laboratory report (81%) (the average of the usual care and core laboratory reports was 77%). The median LVEF across each of the different methods of assessment is shown in online supplementary Figure S4.

## Outcomes

### Primary endpoint

The primary endpoint of diagnostic accuracy of AI-automated analysis of handheld TTE to identify an LVEF  $\leq 40\%$ , compared to the report of the clinical human on a cart echocardiogram, was 0.93 (95% CI 0.90, 0.95). The sensitivity was 0.61 (95% CI 0.42, 0.78), specificity 0.95 (95% CI 0.93, 0.97), positive predictive value 0.50 (95% CI 0.33, 0.67), negative predictive value 0.97 (0.95, 0.99), with an area under the ROC curve of 0.96 (95% CI 0.94–0.98). These results are also presented as a confusion matrix in online supplementary Table S3.

### Secondary endpoints

The LVEF reported by the AI-automated analysis of the handheld TTE was interchangeable with the LVEF reported by the analysis

of cart-based TTE by two human sonographers (usual care clinical and core laboratory reports) (IEC  $-0.40$ , 95% CI  $-0.60$ ,  $-0.16$ ).

The IEC for additional chamber dimensions and echocardiographic measures of HFpEF are shown in Table 2. For cart-based echocardiograms, Doppler measurements were interchangeable between AI-automated reports and two human reports of cart echocardiograms (usual care clinical and core laboratory reports). Although the point estimate for IEC for left atrial volume was under 0.25, 95% CIs were wide, therefore it did not meet our pre-defined definition of interchangeability based on the upper bound of the 95% CI of IEC.

### Exploratory analyses

Artificial intelligence-automated analysis of the LVEF from cart-based TTE images was interchangeable with the LVEF reported by two human sonographers on cart-based TTE images (IEC  $-0.39$ , 95% CI  $-0.60$ ,  $-0.12$ ).

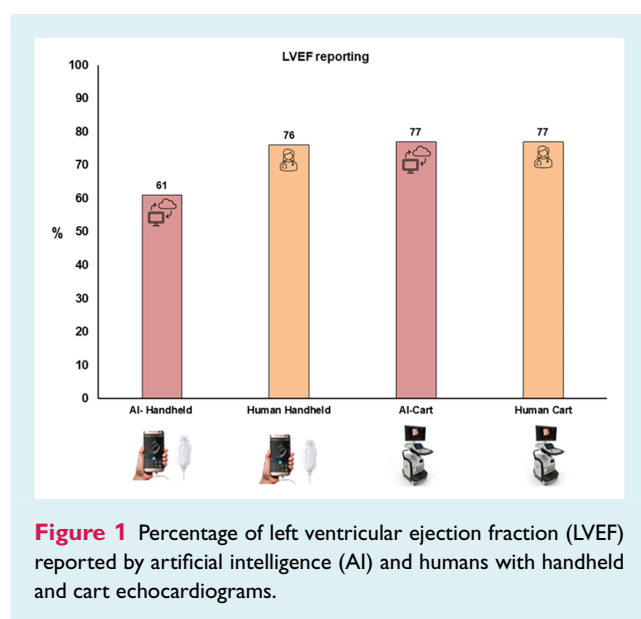
In a *post hoc* analysis, we compared the IEC of AI fully automated analysis of handheld scans to two human core laboratory-blinded analyses of the same handheld scans (rather than comparing to the cart scan). The IEC showed a consistent result regarding LVEF assessment, with an IEC of  $-0.34$  (95% CI  $-0.53$ ,  $-0.11$ ).

In a second *post hoc* analysis, we compared the agreement between AI-automated analysis of LVEF on both handheld and cart scans with the human core laboratory analysis of the cart scans (Figure 2). We also compared the agreement between core laboratory cart and other automated echocardiographic measures including those used in the assessment of HFpEF (on cart-based scans only, online supplementary Figures S5–S21). The results of these analyses are presented in the online Supporting Information. There was a strong correlation between AI-automated and human core laboratory LVEF (Pearson  $r=0.61$ ,  $p<0.0001$ ) (Figure 2), and between AI-automated handheld and human core laboratory LVEF ( $r=0.55$ ,  $p<0.0001$ ). The lowest correlation coefficient was seen with mitral valve deceleration time ( $r=0.34$ ,  $p<0.001$ ), and the highest with peak mitral valve E velocity ( $r=0.94$ ,  $p<0.0001$ ). The mean difference between AI-automated LVEF and human core laboratory analyses for handheld scans was 5.2% (95% CI 4.2, 6.2;  $p<0.0001$ ). The mean difference between AI-automated LVEF and human core laboratory analyses for cart scans was 4.2% (95% CI 3.4, 5.0); ( $p<0.0001$ ). Both of these mean differences were smaller than the difference between the human clinical assessment and human core laboratory measurements of cart scans (6.5%, 95% CI 5.7, 7.3;  $p<0.0001$ ).

In a third *post hoc* analysis, we compared the diagnostic accuracy of AI-automated analysis of handheld TTE to identify an LVEF  $<50\%$ , compared to the report of the clinical human on a cart echocardiogram. The results (shown in online supplementary Tables S4 and S5) were consistent with the primary pre-specified analysis.

### Image quality

The interchangeability of LVEF reporting with AI-automated reporting and two humans (usual care clinical and core laboratory



**Table 2** Individual equivalence coefficient values

| Measurement                        | Automated handheld analysis               |                      |                 | Automated cart analysis                   |                      |                 |
|------------------------------------|---|----------------------|-----------------|---|----------------------|-----------------|
|                                    | Scans with 3 assessments (n) <sup>a</sup> | IEC (95% CI)         | Interchangeable | Scans with 3 assessments (n) <sup>a</sup> | IEC (95% CI)         | Interchangeable |
| Ventricular volumes and dimensions |   |                      |                 |   |                      |                 |
| LVEF                               | 410                                       | −0.40 (−0.60, −0.16) | Yes             | 521                                       | −0.39 (−0.60, −0.12) | Yes             |
| LVEDV                              | 410                                       | 0.61 (0.16, 1.21)    | No              | 522                                       | 0.04 (−0.30, 0.49)   | No              |
| LVESV                              | 411                                       | 0.18 (−0.18, 0.63)   | No              | 524                                       | −0.12 (−0.48, 0.35)  | Yes             |
| LVIDd                              | 640                                       | 1.73 (1.19, 2.41)    | No              | 712                                       | −0.01 (−0.25, 0.29)  | No              |
| LVIDs                              | 553                                       | 0.04 (−0.28, 0.45)   | No              | 620                                       | −0.32 (−0.51, −0.10) | Yes             |
| IVSDd                              | 589                                       | −0.06 (−0.30, 0.25)  | Yes             | 672                                       | 0.24 (−0.03, 0.56)   | No              |
| LVPWd                              | 566                                       | −0.70 (−0.84, −0.52) | Yes             | 668                                       | −0.44 (−0.60, −0.25) | Yes             |
| RAA                                | 325                                       | 0.39 (0.05, 0.82)    | No              | 385                                       | −0.14 (−0.39, 0.20)  | Yes             |
| RVIDd                              | 435                                       | 0.00 (−0.29, 0.35)   | No              | 551                                       | −0.26 (−0.49, −0.02) | Yes             |
| TAPSE                              | —   | —                    | —               | 316                                       | −0.12 (−0.51, 0.41)  | No              |
| HFpEF echocardiographic markers    |   |                      |                 |   |                      |                 |
| LA volume                          | 244                                       | 2.23 (1.31, 3.54)    | No              | 322                                       | 0.19 (−0.12, 0.57)   | No              |
| LV mass                            | 541                                       | 0.61 (0.28, 1.03)    | No              | 638                                       | 0.91 (0.53, 1.37)    | No              |
| MV E                               | —   | —                    | —               | 410                                       | −0.86 (−0.98, −0.69) | Yes             |
| MV A                               | —   | —                    | —               | 262                                       | −0.81 (−1.03, −0.52) | Yes             |
| MV E/A                             | —   | —                    | —               | 252                                       | −0.95 (−1.12, −0.45) | Yes             |
| MV Dec T                           | —   | —                    | —               | 324                                       | −0.39 (−0.68, −0.01) | Yes             |
| E/e' average                       | —   | —                    | —               | 319                                       | −0.82 (−0.98, −0.58) | Yes             |
| TV peak velocity                   | —   | —                    | —               | 287                                       | −0.66 (−0.83, −0.41) | Yes             |

CI, confidence interval; Dec T, deceleration time; HFpEF, heart failure with preserved ejection fraction; IEC, individual equivalence coefficient; IVSDd, interventricular septal diameter in diastole; LA, left atrial; LV, left ventricular; LVEDV, left ventricular end-diastolic volume; LVEF, left ventricular ejection fraction; LVESV, left ventricular end-systolic volume; LVIDd, left ventricular internal diameter in diastole; LVIDs, left ventricular internal diameter in systole; LVPWd, left ventricular posterior wall in diastole; MV, mitral valve; RAA, right atrial area; RVIDd, right ventricular internal diameter in diastole; TAPSE, tricuspid annular plane systolic excursion; TV, tricuspid valve.

<sup>a</sup>Two human cart analyses and one automated AI analysis.

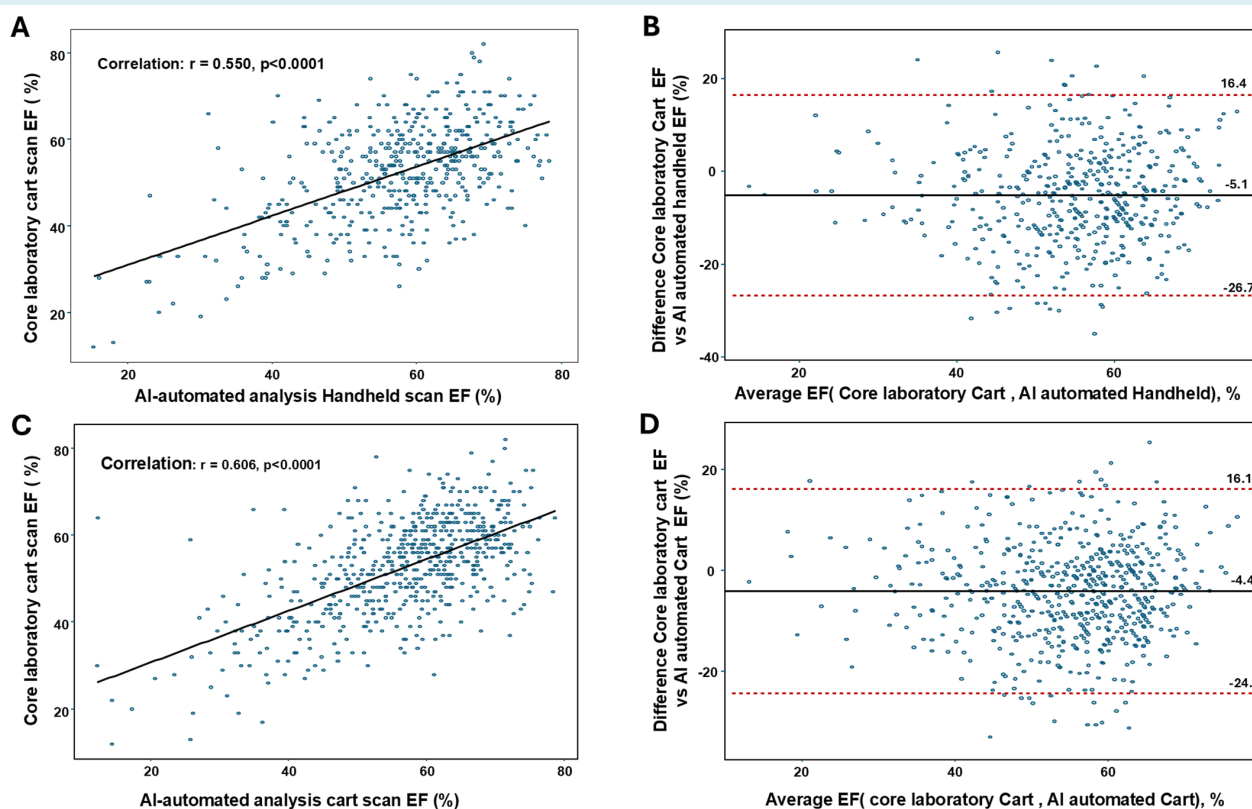
reports) was seen regardless of image quality on both handheld and cart scans (Figure 3). Image quality was categorized as substantially worse on handheld than on cart echocardiograms (online supplementary Figure S3).

## Discussion

We conducted a large prospective, real-world assessment of AI-automated reporting of handheld echocardiograms compared to human reporting of cart-based TTE. We have shown that the AI-automated analysis of images acquired with a handheld scanner had good diagnostic performance for detecting an LVEF  $\leq 40\%$  (when compared to humans in daily clinical practice). We have also shown that AI-automated LVEF was interchangeable with LVEF reported by humans on a cart-based TTE in a clinical setting. AI-automated LVEF reporting on a cart is also interchangeable with human reporting of cart-based TTE. On cart scans, AI-automated reporting met criteria for interchangeability with human reporting of most, but not all, echocardiographic parameters suggested in guidelines to be used in the diagnosis of HFpEF.<sup>1</sup> The interchangeability of LVEF reporting between AI-automated reporting and humans was seen regardless of image quality on both handheld and cart scans.

Our study was large and prospective, with 867 near-consecutive patients evaluated as part of a routine clinical diagnostic pathway. This is in contrast to most previous studies assessing the utility of semi-automated and fully automated AI algorithms for analysis of echocardiograms which were mostly retrospective. In contrast to previous validation cohorts, the population included in the present study had substantial comorbidity (e.g.  $>50\%$  were obese). Many previous studies used core laboratory echocardiography scans from selected populations (usually from clinical trial or research study populations). We did not exclude patients with poor acoustic windows, although we were able to only compare scans where an ejection fraction was produced by both human and automated analyses. Although prior studies have been very useful for the development and early validation of AI-automated algorithms, questions remained regarding their utility in 'real-world' patients assessed in everyday clinical practice. One of the key strengths of our study was the focus on 'real-world', consecutive patients referred by their primary care physician for a clinically indicated diagnostic TTE.

Our primary endpoint focused on the ability of AI-automated reporting of handheld scans to diagnose those with an LVEF  $\leq 40\%$  (i.e. the echocardiographic biomarker of the syndrome of HFpEF). We compared its performance to sonographers' reporting on a cart-based TTE in a routine clinical setting. This was deemed to



**Figure 2** Correlation and Bland–Altman plots of artificial intelligence (AI)-automated measurement (handheld and cart) with cart core laboratory ejection fraction (EF). (A) Scatter plot of AI-automated handheld EF and core laboratory cart EF. (B) Bland–Altman plot comparing AI-automated handheld EF and core laboratory cart EF. (C) Scatter plot of AI-automated cart EF and core laboratory cart EF. (D) Bland–Altman plot comparing AI-automated cart EF and core laboratory cart EF.

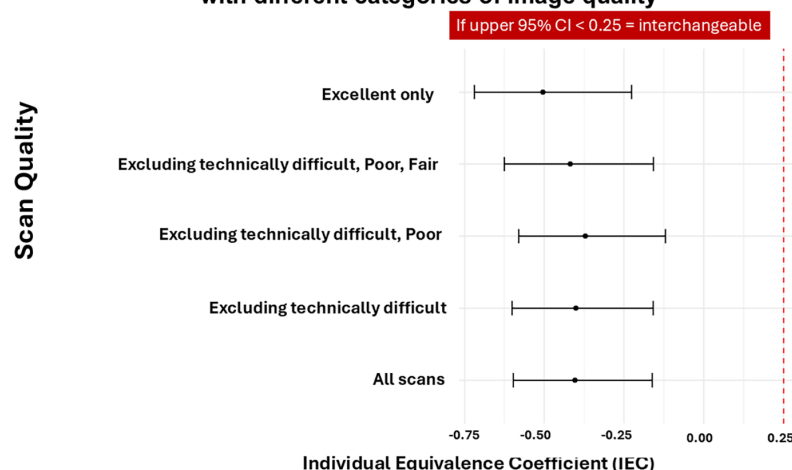
be the most clinically meaningful standard of comparison for the combined AI-handheld TTE approach. A key goal of HF diagnostic pathways is to identify very high-risk patients where there is a major treatment opportunity. In patients with HFrEF, comprehensive medical therapy can reduce mortality by around 50% and hospitalizations for HF by 60–70%.<sup>12</sup> The diagnostic accuracy of AI-automated reporting of LVEF  $\leq 40\%$  was high. The negative predictive value was 0.97. The positive predictive value (0.50) suggests that those who are thought to have LVEF  $\leq 40\%$  by AI handheld assessment require a confirmatory cart-based scan, although this lower positive predictive value will be influenced by the relatively low prevalence of HFrEF in the cohort. The high negative predictive value is of importance, excluding patients who do not have LVEF  $\leq 40\%$ . By using AI-automated handheld TTE, HFrEF could be excluded efficiently and those with potential HFrEF identified. The use of a handheld TTE potentially allows the diagnosis of HFrEF to be moved from secondary care into primary or community care.

We also found that LVEF reported by AI on handheld devices was interchangeable with LVEF reported by humans reporting cart images in a clinical environment, although this was only where both AI-automated analysis and two human analyses were possible. This finding has major clinical implications. For example, many echocardiograms are requested to investigate recovery of myocardial

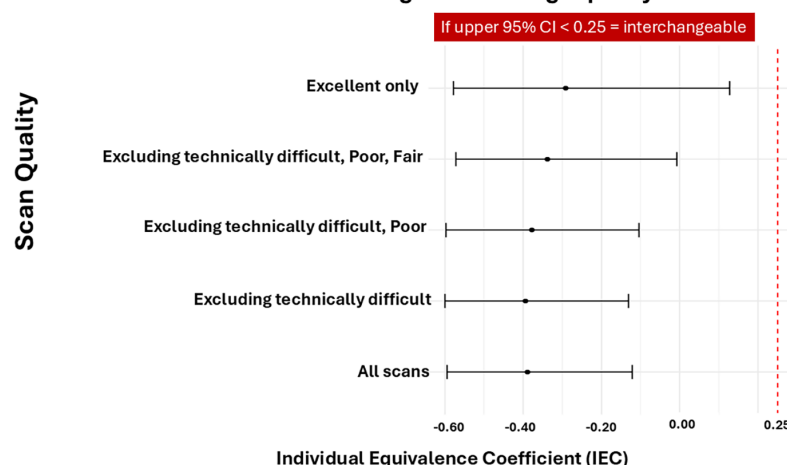
function after HF therapy has been instituted. Among other clinical implications, the degree of myocardial recovery determines eligibility for cardiac resynchronization therapy or implantable cardioverter-defibrillators. A focused, rapid, accurate assessment of LVEF by AI reporting on a handheld device in this setting is potentially of great clinical value, negating the need for a time-consuming cart-based scan and human reporting. The value of AI reporting of LVEF on handheld machines has major potential, especially in regions of the world where there is a lack of cart machines and a dearth of expert reporting sonographers. The impact on many aspects of HF diagnosis and management could be profound.

Artificial intelligence-automated reporting of LVEF on cart images was also interchangeable with human reporting of cart images. Obtaining an AI-automated LVEF report, which is equivalent to a human sonographer, would result in major time-saving and efficiencies, through workflow optimization. Importantly AI reporting of cart scans resulted in similar numbers of LVEF reports as cart scans reported by human clinical sonographers. Workflow acceleration has recently been demonstrated using this approach in clinical echo labs where AI-automated echo reporting was shown to reduce echo analysis time by 70% without compromising accuracy.<sup>13</sup> Such efficiency benefits are similar to other areas of diagnostic imaging where AI-automated algorithms are being deployed

### A HANHDELD LVEF- Interchangeability of AI and human reporting with different categories of image quality



### B CART LVEF- Interchangeability of AI and human reporting with different categories of image quality



**Figure 3** Interchangeability of of artificial intelligence (AI) and human reporting of left ventricular ejection fraction (LVEF) on handheld (A) and cart (B) echocardiograms according to image quality. CI, confidence interval.

in clinical care.<sup>14</sup> This workflow acceleration could increase the number of scans performed per sonographer/scanner and therefore increase capacity in the systems which are often working beyond capacity. By increasing the throughput of an echocardiogram department, the time to diagnostic scan for HF would likely reduce, and this in turn has the potential to allow patients to be started on HFrEF disease-modifying therapies sooner. As AI-automated analysis models and handheld echocardiographic equipment continue to improve, it is anticipated that there are likely to be further gains in diagnostic accuracy and in the proportion of patients where automated analysis is possible.

## Limitations

A key strength of our study is that every TTE, both handheld and cart-based, was performed by accredited sonographers/

cardiologists. All of the blinded analyses were also performed by expert/accredited sonographers. This could limit the potential generalizability, with our results only applicable to scanning performed by accredited sonographers and cardiologists. It is probable, although we did not formally test this, that a novice acquisition of images would result in a lower yield of LVEF assessment. A potential source of bias in our study was the timing of the handheld scan in relation to the cart scan. In our study the same sonographer performed both handheld and cart scans with the handheld scan being performed first in each case. It may be that the sonographer was able to acquire better images on the second scan. It was not possible to calculate an LVEF in every patient on every scan. The assessment of the primary endpoint was only possible in 445 of 867 (51%) participants. This was anticipated and expected, with estimates of 10–15% of



patients having a TTE requiring contrast agents to allow accurate endocardial definition and LVEF calculation.<sup>15</sup> To obtain an LVEF in every patient, ultrasound enhancing contrast agents would be required for the patients with limited acoustic windowing. We elected not to administer contrast agents for those with poor acoustic windows as the AI-automated algorithm had not been trained on datasets utilizing contrast echocardiography. Although rates of LVEF reporting on the handheld scans were lower than cart-based scans, AI reporting could still represent a potential major step forward in a triage setting, particularly when used by a trained sonographer. Very few non-White patients were enrolled and the study was performed in one country. Our results require validation in other populations.

## Conclusions

Artificial intelligence-automated reporting of LVEF  $\leq 40\%$  on handheld devices has good diagnostic accuracy compared to humans reporting cart LVEF. AI reporting of LVEF on handheld devices and cart scanners is interchangeable with human cart reporting. Incorporating AI-automated reporting of TTE has the potential to accelerate exclusion, identification and treatment of HFrEF.

## Acknowledgements

We would like to acknowledge and thank the support given to the study by the NHS Greater Glasgow and Clyde & University of Glasgow Sponsor teams. We would like to acknowledge and thank the contributions made by Ruairidh Davison, Barbara Meyer, and Joanne Cusack. We would like to acknowledge and pay particular thanks to all the members of the NHS clinical teams who contributed to delivering this project, and most of all to the patients who were willing to participate in clinical research during the COVID-19 pandemic.

## Supplementary Information

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

## Funding

This study was funded by investigator-initiated unrestricted research grants from AstraZeneca and Roche Diagnostics. M.C.P./J.J.V.M. are supported by a Centre of Research Excellence Grant from the British Heart Foundation (RE/18/6/34217). Handheld (Philips Lumify) scanners were provided by Phillips®, 'cart' scanners (GE Vivid E95) were provided by GE®, automated analysis software was provided by Us2.Ai®. Reagents and assays were provided free of charge from Roche Diagnostics.

**Conflict of interest:** R.T.C. has received consultancy honoraria from Bayer AG and speaking honoraria from AstraZeneca and Circle CVI, and has received research grant support from SQ Innovations, Boehringer Ingelheim, Roche Diagnostics, and AstraZeneca (paid to his institution). M.C.P. reports research funding from Boehringer Ingelheim, Roche, SQ Innovations, AstraZeneca, Novartis, Novo Nordisk, Medtronic, Boston Scientific, Pharmacosmos, and consultancy and trial committee membership for Abbott, Akero, Applied Therapeutics, Amgen, AnaCardio,

Biosensors, Boehringer Ingelheim, Corteria, Novartis, AstraZeneca, Novo Nordisk, Abbvie, Bayer, Horizon Therapeutics, Foundry, Takeda, Cardioentis, Pharmacosmos, Siemens, Eli Lilly, Vifor, New Amsterdam, Moderna, Teikoku, LIB Therapeutics, 3R Lifesciences, Reprieve, FIRE 1, Corvia, Regeneron. M.C.P. is a director of Global Clinical Trial Partners Ltd. K.F.D. reports that his employer, the University of Glasgow, has been remunerated by AstraZeneca and FIRE-1 for work related to clinical trials; has received speaker honoraria from AstraZeneca, Boehringer Ingelheim, Pharmacosmos, Translational Medicine Academy and Radcliffe Cardiology; has served on an advisory board for Us2.ai and Bayer AG; has served on a clinical endpoint committee for Bayer AG; and has received grant support from Boehringer Ingelheim, Roche Diagnostics, Novartis and AstraZeneca (paid to his institution). C.S.P.L. reports service as Co-founder & non-executive director of Us2.ai and as consultant/advisor for Alleviant Medical, Allysta Pharma, AnaCardio AB, Applied Therapeutics, AstraZeneca, Bayer, Biopeutics, Boehringer Ingelheim, Boston Scientific, Bristol Myers Squibb, CardioRenal, CPC Clinical Research, Eli Lilly, Hanmi, Impulse Dynamics, Intellia Therapeutics, Ionis Pharmaceutical, Janssen Research & Development LLC, Medscape/WebMD Global LLC, Merck, Novartis, Novo Nordisk, Prosciento Inc, Quidel Corporation, Radcliffe Group Ltd., Recardio Inc, ReCor Medical, Roche Diagnostics, Sanofi, Siemens Healthcare Diagnostics and Us2.ai; research support from Novo Nordisk and Roche Diagnostics; and grant funding by a Clinician Scientist Award from the National Medical Research Council of Singapore. M.M.Y.L. has received research grants through his institution, the University of Glasgow, from AstraZeneca, Boehringer Ingelheim, and Roche Diagnostics; serves on Trial Steering Committees for Amgen and Cytokinetics; and Clinical Endpoint Committees for Bayer and GlaxoSmithKline. J.P.C. reports consultancy fees from FIRE1. N.H.E. reports speaker honorarium from AstraZeneca and Novartis. F.J.G. reports travel support from Pharmacosmos to attend a meeting, and consultancy fees from Vifor. L.M. reports speaker honorarium from AstraZeneca. K.M. reports honoraria from Amarin outside this submitted work. P.P. has received consultancy honoraria and/or sponsorship support from Boehringer Ingelheim, Pharmacosmos, Novartis, Vifor, AstraZeneca, and Caption Health in the past 3 years, not connected with this manuscript. J.S. reports speaker honorarium from Pfizer and Alnylan. P.W. reports grant income from Roche Diagnostics, AstraZeneca, Boehringer Ingelheim, and Novartis, and speaker fees from Novo Nordisk and Raisio, outside the submitted work. J.J.V.M. reports payments to Glasgow University for clinical trials and other research projects from the British Heart Foundation, National Institute for Health – National Heart Lung and Blood Institute (NIH-NHLBI), Alnylan Pharmaceuticals, AstraZeneca, Bayer, Cardurion, Cytokinetics, Novartis, Roche; personal consultancy fees from Alnylan Pharmaceuticals, AnaCardio, AstraZeneca, Bayer, Cardurion, Cytokinetics, Novartis, River BioMedics, Biohaven Pharmaceuticals, Chugai Pharmaceuticals, Protherics Medicine Developments Ltd., DalCor Pharmaceuticals; personal lecture fees from Alkem Metabolics, AstraZeneca, Canadian Medical and Surgical Knowledge, Centrix Healthcare, Emcure Pharmaceuticals, Eris Lifesciences, Hikma Pharmaceuticals, Imagica Health, Intas Pharmaceuticals, J.B. Chemicals & Pharmaceuticals, Lupin Pharmaceuticals, Medscape/Heart.Org., ProAdWise Communications, Radcliffe Cardiology, Sun Pharmaceuticals, Translational Medicine Academy, Regeneron, MCI India, Hilton Pharmaceuticals, IMEDIC Pharmaceuticals Micro Labs Ltd., At the Limits Ltd., ARMGO Pharmaceuticals; data safety monitoring boards: WCG Clinical Services. J.J.V.M. is a director of Global Clinical Trial Partners Ltd (which provides clinical trial services such as endpoint committees and educational programmes). C.L.M. reports speaker honorarium from AstraZeneca and Novartis. D.J.L. has received consultancy honoraria from Boehringer Ingelheim, AstraZeneca (paid to his institution) and research grants from AstraZeneca, Qure.ai, Roche. All other authors have nothing to disclose.

## References

- McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail* 2022;**24**:4–131. <https://doi.org/10.1002/ehf.2333>
- Lawson CA, Zaccardi F, Squire I, Ling S, Davies MJ, Lam CSP, et al. 20-year trends in cause-specific heart failure outcomes by sex, socioeconomic status, and place of diagnosis: A population-based study. *Lancet Public Health* 2019;**4**:e406–e420. [https://doi.org/10.1016/S2468-2667\(19\)30108-2](https://doi.org/10.1016/S2468-2667(19)30108-2)
- Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: A comprehensive and updated review of epidemiology. *Cardiovasc Res* 2023;**118**:3272–3287. <https://doi.org/10.1093/cvr/cvac013>
- Robinson S, Rana B, Oxborough D, Steeds R, Monaghan M, Stout M, et al. A practical guideline for performing a comprehensive transthoracic echocardiogram in adults: The British Society of Echocardiography minimum dataset. *Echo Res Pract* 2020;**7**:G59–G93. <https://doi.org/10.1530/ERP-20-0026>
- Tromp J, Bauer D, Claggett BL, Frost M, Iversen MB, Prasad N, et al. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun* 2022;**13**:6776. <https://doi.org/10.1038/s41467-022-34245-1>
- Tromp J, Seekings PJ, Hung CL, Iversen MB, Frost MJ, Ouwerkerk W, et al. Automated interpretation of systolic and diastolic function on the echocardiogram: A multicohort study. *Lancet Digit Health* 2022;**4**:e46–e54. [https://doi.org/10.1016/S2589-7500\(21\)00235-1](https://doi.org/10.1016/S2589-7500(21)00235-1)
- Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, di Salvo G, et al.; 2016–2018 EACVI Scientific Documents Committee. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: An expert consensus document of the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2017;**18**:1301–1310. <https://doi.org/10.1093/ehjci/jex244>
- Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;**48**:193–204. <https://doi.org/10.1016/j.jbi.2014.02.013>
- Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation* 2018;**138**:1623–1635. <https://doi.org/10.1161/CIRCULATIONAHA.118.034338>
- Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circ Cardiovasc Imaging* 2019;**12**:e009303. <https://doi.org/10.1161/CIRCIMAGING.119.009303>
- Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 2022;**377**:e070904. <https://doi.org/10.1136/bmj-2022-070904>
- Burger PM, Savarese G, Tromp J, Adamson C, Jhund PS, Benson L, et al.; European Society of Cardiology's Cardiovascular Risk Collaboration (ESC CRC). Personalized lifetime prediction of survival and treatment benefit in patients with heart failure with reduced ejection fraction: The LIFE-HF model. *Eur J Heart Fail* 2023;**25**:1962–1975. <https://doi.org/10.1002/ehf.3028>
- Hirata Y, Nomura Y, Saijo Y, Sata M, Kusunose K. Reducing echocardiographic examination time through routine use of fully automated software: A comparative study of measurement and report creation time. *J Echocardiogr* 2024;**22**:162–170. <https://doi.org/10.1007/s12574-023-00636-6>
- Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med* 2019;**25**:44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- Porter TR, Mulvagh SL, Abdelmoneim SS, Becher H, Belcik JT, Bierig M, et al. Clinical applications of ultrasonic enhancing agents in echocardiography: 2018 American Society of Echocardiography guidelines update. *J Am Soc Echocardiogr* 2018;**31**:241–274. <https://doi.org/10.1016/j.echo.2017.11.013>