

BMJ Open Real-time automatic quantification of left ventricular function by hand-held ultrasound devices in patients with suspected heart failure: a feasibility study of a diagnostic test with data from general practitioners, nurses and cardiologists

Anna Katarina Hjorth-Hansen ^{1,2}, Malgorzata Izabela Magelssen,^{2,3} Garrett Newton Andersen,¹ Torbjørn Graven,¹ Jens Olaf Kleinau,¹ Bodil Landstad ^{4,5}, Lasse Løvstakken,² Kyrre Skjetne,^{1,6} Ole Christian Mjølstad,^{2,3} Havard Dalen^{1,2,3}

To cite: Hjorth-Hansen AK, Magelssen MI, Andersen GN, *et al.* Real-time automatic quantification of left ventricular function by hand-held ultrasound devices in patients with suspected heart failure: a feasibility study of a diagnostic test with data from general practitioners, nurses and cardiologists. *BMJ Open* 2022;**12**:e063793. doi:10.1136/bmjopen-2022-063793

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-063793>).

Received 09 May 2022
Accepted 30 September 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Anna Katarina Hjorth-Hansen; annakh@ntnu.no

ABSTRACT

Objectives To evaluate the feasibility and reliability of hand-held ultrasound (HUD) examinations with real-time automatic decision-making software for ejection fraction (autoEF) and mitral annular plane systolic excursion (autoMAPSE) by novices (general practitioners), intermediate users (registered cardiac nurses) and expert users (cardiologists), respectively, compared to reference echocardiography by cardiologists in an outpatient cohort with suspected heart failure (HF).

Design Feasibility study of a diagnostic test.

Setting and participants 166 patients with suspected HF underwent HUD examinations with autoEF and autoMAPSE measurements by five novices, three intermediate-skilled users and five experts. HUD results were compared with a reference echocardiography by experts. A blinded cardiologist scored all HUD recordings with automatic measurements as (1) discard, (2) accept, but adjust the measurement or (3) accept the measurement as it is.

Primary outcome measure The feasibility of automatic decision-making software for quantification of left ventricular function.

Results The users were able to run autoEF and autoMAPSE in most patients. The feasibility for obtaining accepted images (score of ≥ 2) with automatic measurements ranged from 50% to 91%. The feasibility was lowest for novices and highest for experts for both autoEF and autoMAPSE ($p < 0.001$). Large coefficients of variation and wide coefficients of repeatability indicate moderate agreement. The corresponding intraclass correlations (ICC) were moderate to good (ICC 0.51–0.85) for intra-rater and poor (ICC 0.35–0.51) for inter-rater analyses. The findings of modest to poor agreement and reliability were not explained by the experience of the users alone.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ To our knowledge, no study has evaluated automatic real-time quantification of left ventricular function on hand-held ultrasound devices by inexperienced users.
- ⇒ The three user groups in this study had different levels of experience, ranging from no previous experience to expert level.
- ⇒ The inexperienced operators were recruited by their role in the municipality and not based on motivation for attending the study.
- ⇒ Due to the lack of a gold standard for evaluation of left ventricular function, echocardiographic measurements by experienced cardiologists were used as reference.
- ⇒ An error detected in the first software version of the automatic decision-making software for ejection fraction may have affected the results for the revised software as well.
- ⇒ The study sample is expected to provide adequate power for analyses.

Conclusion Novices, intermediate and expert users were able to record four-chamber views for automatic assessment of autoEF and autoMAPSE using HUD devices. The modest feasibility, agreement and reliability suggest this should not be implemented into clinical practice without further refinement and clinical evaluation.
Trial registration number NCT03547076.

INTRODUCTION

Heart failure (HF) is a severe condition with poor prognosis and reduced quality of life which constitutes a burden on the

healthcare system with high costs and 26 million patients affected worldwide.^{1 2} Echocardiography is the cornerstone imaging modality for HF diagnostics and patient follow-up. HF may be challenging to diagnose and it is shown that (in-training) cardiology fellows inaccurately interpret echocardiograms.³ Moreover, it is shown that a delayed HF diagnosis may be present in up to 40% of patients.⁴

Estimation of left ventricular (LV) ejection fraction (EF) is required for classification and treatment of HF.⁵ Another robust and easily obtainable measure of LV function is mitral annular plane systolic excursion (MAPSE), which is quite sensitive for detection of LV dysfunction,⁶⁻⁸ even when EF is preserved. Semi-automatic quantification of LV EF has been available for some time, but automatic quantification of MAPSE is not widely available.⁷

Hand-held ultrasound devices (HUD) have been widely implemented in the medical field over the last decade and are increasingly used by non-experts.⁹ So far, quantification of LV size and function by HUDs has relied on visual evaluation only.¹⁰ Several studies have shown high feasibility and reliability for inexperienced users performing simple tasks by HUDs.¹¹⁻¹⁵ The experience and skill of the operator is essential for more advanced measures such as assessment of LV function.^{15 16} Automatic measurement

of LV EF (autoEF) from apical HUD recordings are now commercially available, and a novel method for real-time automatic measurement of MAPSE (autoMAPSE) is available on the GE Vscan Extend (GE Ultrasound, Horten, Norway) for research purposes. This allows for real-time quantification of LV function by HUDs, and thus there is a need to evaluate the feasibility and reliability in clinical scenarios by different users before implementation into clinical practice.

We aimed to evaluate the feasibility and reliability of HUD examinations including real-time autoEF and autoMAPSE performed by users with different levels of experience in an outpatient cohort with suspected HF. Specifically, the novice, intermediate and expert groups were represented by general practitioners (GPs), registered cardiac nurses (RCNs) and experienced cardiologists, respectively. Comprehensive echocardiography by experienced cardiologists served as reference.

METHODS

Study design

Figure 1 indicates the flow of the study participants. The patients were examined by one of five GPs and by one of three RCNs at random order. GPs and RCNs were

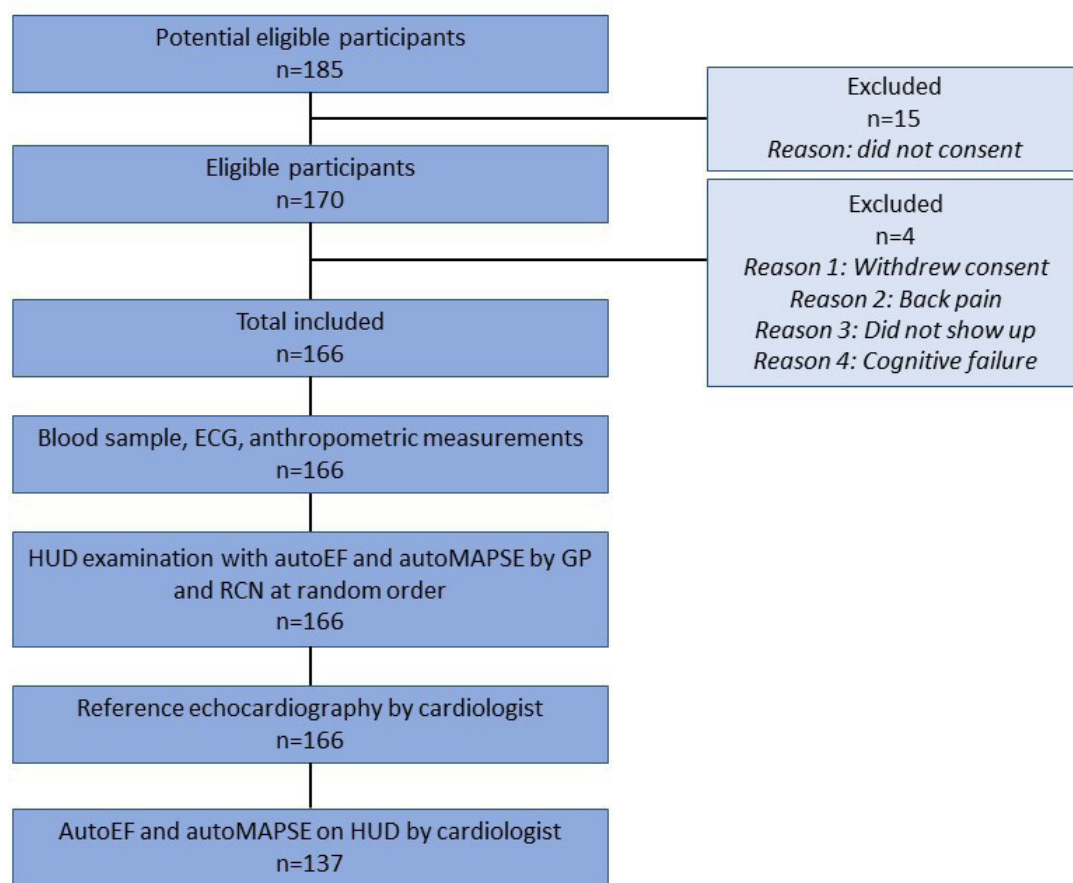


Figure 1 Study flow. AutoEF, automatic measurement of left ventricular ejection fraction; autoMAPSE, automatic measurement of mitral annular plane systolic excursion; GP, general practitioner; HUD, hand-held ultrasound device; RCN, registered cardiac nurse.

blinded to each other's results. Reference echocardiography was performed by one of five cardiologists blinded to preceding examinations. An additional HUD examination was performed by the cardiologists (expert group). Due to logistic reasons, the first 29 patients were not examined by HUD by the cardiologist. No additional follow-up or ultrasound examinations of the participants were performed related to the study. The study was registered in the ClinicalTrials.gov database (NCT03547076).

Participants

Patients referred to Levanger Hospital, Norway, with suspected HF were available for inclusion. Exclusion criteria were age <18 years, known HF and previous cardiac imaging within the last decade. Eligible patients were consecutively included from June 2018 to June 2020. Inclusion was paused from March to June 2020 due to the COVID-19 pandemic.

Training and education of personnel

The conductors of the study had no influence on the selection of GPs for the study who were selected by the municipality administration based on their position in the municipalities of Levanger and Verdal.

A total of six GPs underwent training in focused cardiac ultrasound by HUDs in accordance with the European recommendations.¹⁰ One dropped out due to change of occupation, and thus, five GPs participated in the study. All GPs underwent six in-hospital training days with one-to-one supervision by one of two residents experienced in focused cardiac ultrasound, in addition to two evening lectures provided by experts in diagnostic ultrasound and echocardiography. The GPs had the opportunity to use a personal HUD without supervision from the first day of training, but for no longer than three months prior to inclusion. None of them received additional training prior to study start. On direct request, the GPs considered themselves prepared to start inclusion. Only one of the six had performed focused ultrasound examinations prior to training (n=7 examinations), and thus, the group represents inexperienced users. They performed in total median (range) 46 (45–68) examinations prior to the first inclusion, where median (range) 10 (9–20) examinations were unsupervised and 36 (31–43) supervised, respectively.

Three RCNs with experience from a nurse-led outpatient HF clinic represented intermediate experienced users. They had experience in evaluation of pleural effusion, the inferior caval vein and evaluation of clinical signs in patients with HF. Moreover, they had previously participated in studies with limited ultrasound examinations of the heart.¹⁷ The RCNs had completed a total of median (range) 118 (74–221) limited echocardiographic examinations before patient inclusion, and therefore, they did not undergo the same systematic training as the GPs. They were instructed on how to use the HUD and initialise the autoEF and autoMAPSE software approximately four weeks prior to inclusion.

Five cardiologists experienced in echocardiography (median 18 (6–43) years of experience) were only instructed in how to initialise the automatic decision-support software on the HUDs and were not provided any additional training. All cardiologists were certified by the national authorities.

Test method

Each patient underwent three HUD examinations in addition to the reference imaging. All HUD examinations were performed by a Vscan Extend with a sector probe, and similarly, reference echocardiography by a Vivid E9 or E95 scanner (GE Ultrasound) with a 1.4–4.6 MHz phased array transducer. All examinations were performed according to standard operating procedures and included four-chamber recordings of the LV. The protocol for the GPs included parasternal long-axis and short-axis views, apical four-chamber view, subcostal four-chamber view and evaluation of the inferior caval vein and the pleural cavities. The recording of the inferior caval vein included both maximum and minimum dimension during normal breathing. Pleural cavities were assessed in the sitting position, and in case of pleural effusion craniocaudal images were recorded. RCNs recorded the same above-mentioned views, as well as apical two-chamber and apical long-axis views, right ventricular focused four-chamber view and atrial focused recordings. Additionally, RCNs recorded colour Doppler images of the mitral, aortic and tricuspid valve not related to the objectives of the current study. Cardiologists recorded the four-chamber view only by the HUD, but the reference echocardiography was comprehensive.¹⁸

For all HUD examinations, live cine-loops of at least one cardiac cycle were recorded. The software for autoEF or autoMAPSE implemented on the HUD was initialized by the user and the automatically analysed recordings were subsequently stored on the HUD. This was repeated aiming for six separate recordings for automatic analyses by autoEF (three recordings) and autoMAPSE (three recordings). All recorded views and analyses were stored and transferred without delay to a cloud-based server (Tricefy, Trice Imaging, California, USA).

Reference echocardiographic examinations were performed according to recommendations¹⁸ in a separate room immediately after the examinations by the GPs and RCNs. All measurements reflect the average of at least three (five in the case of arrhythmia) cardiac cycles. Central methodology follows: all measurements were performed using EchoPAC, V.202 and V.203 (GE Ultrasound). The LV endocardial borders were traced in end-diastole and end-systole in four-chamber and two-chamber view. LV volumes (end-diastolic and end-systolic) and EF was calculated based on the traces using the biplane Simpson's method. MAPSE was measured as the longitudinal displacement of the mitral annular septal and lateral points in reconstructed motion mode.

Details of the automatic tools for quantification of LV function and image analyses

Before storing of the four-chamber view recording, the specific application (autoEF or autoMAPSE) was initialized on the HUD. The automatic measurements of LV volumes and EF were done by the commercially available LVivo application (DiA Imaging Analysis, Be'er Sheva, Israel). The software provides fully automatic edge detection and tracing of the endocardial border in standard apical four-chamber views throughout the cardiac cycle. LV volume was estimated at end-diastole and end-systole and EF was calculated from the volume estimates. MAPSE was estimated by an automated algorithm tracking the mitral annular septal and lateral points using a LV model. Technical details of the method are described in a previous paper.¹⁹ Shortly, a Real-time Contour Tracking Library (RCTL) was used to process and track the LV movement and images (GE, Vingmed, Norway) using a non-uniform rational B-spline model.²⁰ The mitral annular septal and lateral points of the model were returned from the RCTL. The array of points were evaluated to locate the maximum mitral annular plane displacement. MAPSE was calculated at the septal and lateral mitral annular points and as averaged values. For both autoEF and autoMAPSE the four-chamber view recording with the overlay of the results from the automatic algorithm was stored as described above.

All HUD recordings were made available for blinded analyses by external cardiologists experienced in echocardiography. These cardiologists scored all recordings with the automatic measurement overlay as one of the following categories: (1) discard (not for clinical use), (2) accept, but adjust the result according to suboptimal performance or (3) accept the result as it is. The scoring took both the quality of the recordings and the performance of the application used into account. Thus, if the recording was not representative for a four-chamber view, the score was lower. The latter part of the scoring was based on identification and tracking of the endocardial border (autoEF), or mitral annular points (autoMAPSE) combined with the numerical output.

During the study we detected an error in the autoEF software, so the LVivo app was revised by the vendor during the summer of 2019. In total, 103 were analysed with the first version of the autoEF software and 63 patients with the revised software.

Other measurements

Blood samples were drawn the same day and analysed at the in-hospital accredited laboratory. Serum N-terminal pro-brain natriuretic peptide (NT-pro-BNP), serum creatinine and estimated glomerular filtration rate (calculated by the Cockcroft-Gault equation), as well as serum electrolyte (sodium and potassium) and haemoglobin (g/L) were measured. New York Heart Association (NYHA) functional classification was scored by the nurses and body weight (kg), body height (cm)

and blood pressure (mm Hg) were measured. Anthropometric measurements were rounded up to the nearest multiple of one.

Patient and public involvement

Patients were not involved in decisions regarding the research question or the outcome measures. However, the patient user group was involved in planning of the study period as well as the ways of informing the patients and the society of the study results.

Analyses

Continuous variables were expressed as mean and SD or as median and interquartile range (IQR) as appropriate. Evaluation of normality was done by evaluation of histograms and normality plots. Categorical variables are presented as frequencies and proportions. Student's t-test and Wilcoxon test were used for comparison of groups when appropriate, analysis of variance with post-hoc least significant difference correction was used to compare the three user groups. A study was judged as feasible if the following two criteria were present: first, the user was able to acquire data with the fully automatic decision-support software; second, the cardiologists blinded score of the recordings with the automatic measurement overlay was at least 2 (indicating that the recording and automatic measurement was accepted for clinical use). Proportions were compared using the χ^2 test and Fisher's exact test when appropriate. Reliability of the measurements was evaluated by intraclass correlations (ICC), where values <0.5 were considered poor, 0.5–0.75 moderate, 0.75–0.9 good and >0.9 excellent.¹⁹ The intra-rater reliability was calculated by a two-way mixed-effect model defined by absolute agreement in the dataset of single measurements analysed by the automatic methods as repeated measurements from the same patient are assumed to be more similar to each other than measurements between patients.²¹ The inter-rater reliability was calculated with a two-way random model defined by absolute agreement in the dataset of average measurements analysed by the GPs, nurses and cardiologists by HUDs compared with reference. The agreement with reference echocardiography was evaluated by coefficients of variation, coefficient of repeatability indicating the minimal detectable change and Bland-Altman statistics. A p-value <0.05 was considered statistically significant. Sample size was calculated based on estimates of diagnostic precision using Sample Power (SPSS, Chicago, Illinois, USA). A sample size of 104 was needed to detect a difference of <15% of correctly diagnosed patients with HF compared with reference. As the proportion of patients with HF was expected to be small, we adjusted to a sample size of 150. Due to the revision of the autoEF software, the sample size was further adjusted to 170 to account for the new software version. All statistical analyses were performed using IBM SPSS Statistics, V.27 (SPSS).

Table 1 Baseline data, medications and comorbidities of the study population

Variable	
Age, years	73 (63–78)
Women, n (%)	78 (47)
Body mass index (kg/m ²)	28.7±5.3
Systolic blood pressure (mm Hg)	150±22
Diastolic blood pressure (mm Hg)	83±11
Glomerular filtration rate (mL/min)*	89 (68–109)
Haemoglobin (g/L)	144±15
N-terminal pro-brain natriuretic peptide (ng/L)	295 (66–864)
NYHA functional class	
I, n (%)	63 (37)
II, n (%)	80 (47)
III, n (%)	12 (7)
IV, n (%)	1 [†]
Diuretics, n (%)	41 (25)
Beta-blockers, n (%)	51 (31)
ACE inhibitor or angiotensin receptor blocker, n (%)	32 (19)
Atrial fibrillation, n (%)	49 (29)
Chronic obstructive pulmonary disease/asthma, n (%)	26 (16)
Diabetes mellitus type 2, n (%)	23 (14)
Coronary artery disease, n (%)	19 (11)

Normally distributed data are expressed as mean±SD. Skewed data are presented as median (IQR). Proportions are presented as n (%). Medications refer to the current use.
*Calculated by the Cockcroft–Gault equation.
NYHA, New York Heart Association.

RESULTS

Participants

Baseline characteristics are shown in [table 1](#). In total, 185 patients were invited to participate, 170 were included and four (n=4) were excluded (did not show up (n=1), cognitive failure (n=1), withdrawal of consent (n=2)). The 166 participants included (47% women), median

(IQR) age 70 (63–78) years. NT-pro-BNP was above 125 ng/L in 101 (61%) with an overall median (IQR) of 295 (66–864)ng/L. More than half the population was in NYHA class ≥II (93 (55%)) and were obese or overweight (123 (74%)). Chronic pulmonary diseases were relatively rare (24 (15%)). Atrial fibrillation was known in 49 (29%) patients, and present at inclusion in 40 (23%).

Test results

Feasibility

The novices were able to record at least one four-chamber image with autoEF and autoMAPSE in 134 (80%) and 153 (92%) patients, respectively. The corresponding numbers for the intermediate group were 151 (90%) and 161 (96%), respectively (difference vs novices, both p<0.001). The experts were able to obtain the same views using the HUD for autoEF in 91% of the cases and autoMAPSE in 99% (difference vs the intermediate group, both p<0.001).

The proportion of images judged as feasible (score of ≥2) by the blinded cardiologist was lowest for novices, higher for the intermediate group and highest for experts for both autoEF and autoMAPSE (all p≤0.001, [table 2](#)). Overall, ≤53% of images with autoEF or autoMAPSE by novices were judged as feasible, compared with 84% and 85% for autoEF and autoMAPSE by experts, respectively. In analyses taking the two versions of the autoEF algorithm into account, the feasibility for autoEF improved after the revision for all examiners ranging from 68% for novices to 91% for experts ([table 2](#)). Only very few recordings with the automatic algorithm overlays were scored as 3: ‘accept the result as it is’. In total, the numbers (%) for autoEF and autoMAPSE were 7 (2%) and 23 (5%) for novices, 13 (3%) and 52 (11%) for the intermediate group and 25 (7%) and 67 (17%) for experts. The proportion of recordings scored as 3 (‘result accepted as it is’) using autoEF was lower using the revised autoEF algorithm in novices and experts.

The time used for the focused cardiac ultrasound examination was mean (SD) 18 (7) min for novices and 23 (7) min for the intermediate group. The time used for the six recordings with the automatic measurements were mean (SD) 4 min 34 s (2 min 20 s) for novices, 3 min 21 s

Table 2 Feasibility (ie, score ≥2) for the combinations of image recording and the use of automatic applications

	Hand-held ultrasound operator		
	GP (novice)	RCN (intermediate)	Cardiologist (expert)
AutoEF, all patients	205/400 (51%)	296/442 (67%)	298/357 (84%)
AutoEF, first software version	100/246 (41%)	149/270 (55%)	148/193 (77%)
AutoEF, revised software version	105/154 (68%)	147/172 (85%)	150/164 (91%)
AutoMAPSE, all patients	248/471 (53%)	335/467 (72%)	333/391 (85%)

Data are presented as number of feasible/available recordings (%). Feasible recordings were defined as score of ≥2 (ie, accepted with or without need for adjustments by the blinded cardiologist).

AutoEF, automatic measurement of left ventricular ejection fraction; AutoMAPSE, automatic measurement of mitral annular plane systolic excursion; GP, general practitioner; RCN, registered cardiac nurse.

Table 3 Mean values and the agreement of automatic hand-held ultrasound measurements of left ventricular function compared with reference

	Hand-held ultrasound operator			Reference echocardiography
	GP (novice)	RCN (intermediate)	Cardiologist (expert)	
Mean and agreement, autoEF (all recordings)				
Mean (SD), %*	51.7 (10.1)	52.9 (9.6)	53.3 (9.5)	53.4 (10.1)
Coefficient of variation, %	15.4	13.3	12.0	–
Coefficient of repeatability, %*	24.0	24.2	21.5	–
Mean and agreement, autoEF (first software version, n=107)				
Mean (SD), %*	52.6 (11.6)	54.2 (10.3)	55.0 (10.4)	53.5 (10.0)
Coefficient of variation, %	14.8	13.5	11.2	–
Coefficient of repeatability, %*	24.7	24.6	21.4	–
Mean and agreement, autoEF (revised software version, n=63)				
Mean (SD), %*	50.8 (8.4)	51.0 (8.3)	51.6 (8.1)	54.7 (9.6)
Coefficient of variation, %	16.0	13.1	12.9	–
Coefficient of repeatability, %*	20.6	20.6	19.8	–
Mean and agreement, autoMAPSE (all patients)				
<i>Mean of septal and lateral position</i>				
Mean (SD), mm	9.8 (2.4)	10.1 (2.6)	10.2 (2.5)	11.4 (2.9)
Coefficient of variation, %	24.3	20.5	18.9	–
Coefficient of repeatability, mm	5.0	4.8	4.1	–

Comprehensive echocardiography by experienced cardiologists used as reference.
*% points.
AutoEF, automatic measurement of left ventricular ejection fraction; AutoMAPSE, automatic measurement of mitral annular plane systolic excursion; GP, general practitioner; RCN, registered cardiac nurse.

(1 min 52s) in the intermediate group and 2 min 21s (1 min 19s) for experts, respectively.

Reliability

Table 3 shows the agreement of autoEF and autoMAPSE by the different users with reference. In short, the large coefficients of variability and large coefficients of repeatability for all three user groups indicate poor agreement of the automatic applications compared with reference. There was only a modest difference with respect to agreement between the operators. The minimal detectable change estimated from the coefficient of repeatability for autoEF and autoMAPSE ranged 24.2%–21.5% points and 5.0–4.1 mm, respectively. After revision of the autoEF software, the minimal detectable change was somewhat improved but was still approximately 20% points.

Table 4 shows that intra-rater ICCs were moderate for all user groups with values <0.75 for all except for autoMAPSE by the intermediate group (0.85) and experts (ICC 0.83). The intra-rater ICC for autoEF was highest for experts, with ICCs for the three groups ranging 0.51–0.72. The intra-rater ICC for autoMAPSE was lowest for novices and highest for experts, with ICC ranging 0.70–0.85, respectively.

The inter-rater ICCs were poor (≤ 0.51) for both automatic decision support software and all users. Inter-rater ICC for autoEF was highest for experts, with ICCs for

Table 4 Intra-rater and inter-rater reliability of automatic measurements of left ventricular function by HUD according to operators

	HUD measurements by		
	GP (novice)	RCN (intermediate)	Cardiologist (expert)
Intra-rater ICC			
AutoEF	0.58*	0.51	0.72
AutoMAPSE	0.70*	0.85	0.83
Inter-rater ICC			
AutoEF	0.44	0.43	0.51
AutoMAPSE	0.35	0.44	0.51

ICC calculated from single recordings per patient with automatic quantification of left ventricular function. Inter-rater ICC based on average values per patient and operator.

ICC of two repeated measures as only few patients had three repeated measures of autoEF (n=38) and autoMAPSE (n=50), respectively.

AutoEF, automatic measurement of left ventricular ejection fraction; AutoMAPSE, automatic measurement of mitral annular plane systolic excursion; GP, general practitioner; HUD, hand-held ultrasound device; ICC, intraclass correlation; RCN, registered cardiac nurse.

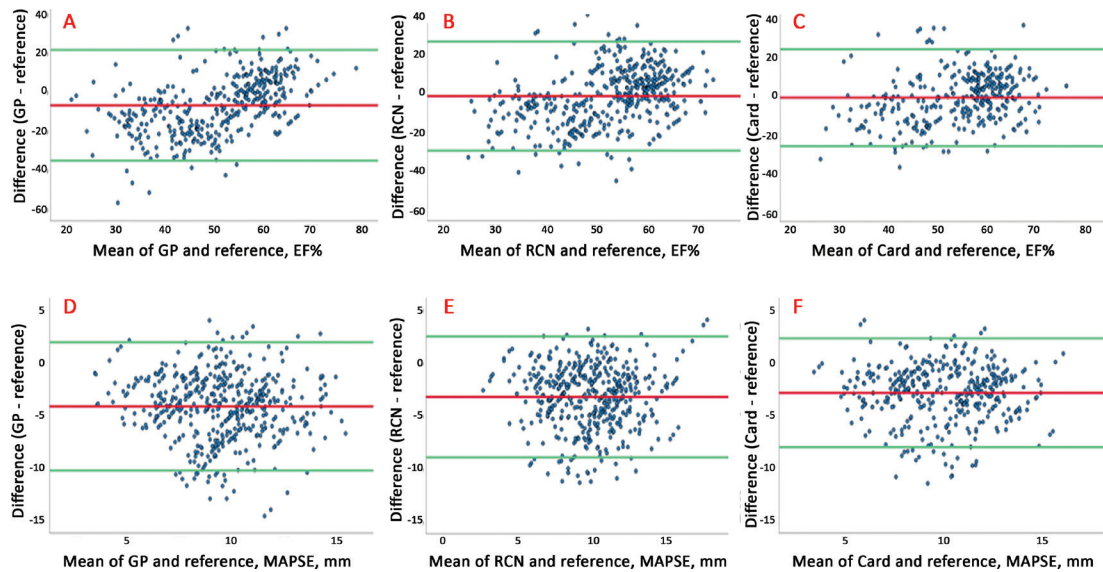


Figure 2 Bland-Altman plots illustrating the agreement between all autoEF and autoMAPSE recordings taken by GPs, RCNs and cardiologists compared to reference echocardiography for all recordings with automatic decision-support software irrespective of image score. Upper panel: autoEF by (A) GPs, (B) RCNs and (C) Card compared with reference. Lower panel: autoMAPSE by (D) GPs, (E) RCNs and (F) Card compared with reference. AutoEF, automatic measurement of left ventricular ejection fraction; autoMAPSE, automatic measurement of mitral annular plane systolic excursion; Card, cardiologist; GP, general practitioner; RCN, registered cardiac nurse.

the three groups ranging 0.43–0.51. The inter-rater ICC for autoMAPSE was lowest for novices and highest for experts, with ICC ranging 0.35–0.51, respectively.

Figure 2 shows the Bland-Altman plots for HUD recordings with autoEF and autoMAPSE compared with reference according to user groups. Similarly, figure 3 shows images accepted (score 2 or 3) by the blinded cardiologist. Overall, the agreement was poor to moderate. We found no association of size of the measurement with agreement, but the limits of agreement were lower for

the most experienced users (also shown in table 3) and after excluding the images deemed too poor for clinical use (figure 3).

DISCUSSION

This is to our knowledge the first study to evaluate the feasibility and reliability of real-time automatic decision-support software for quantification of LV function by HUDs across novices, intermediate experienced users

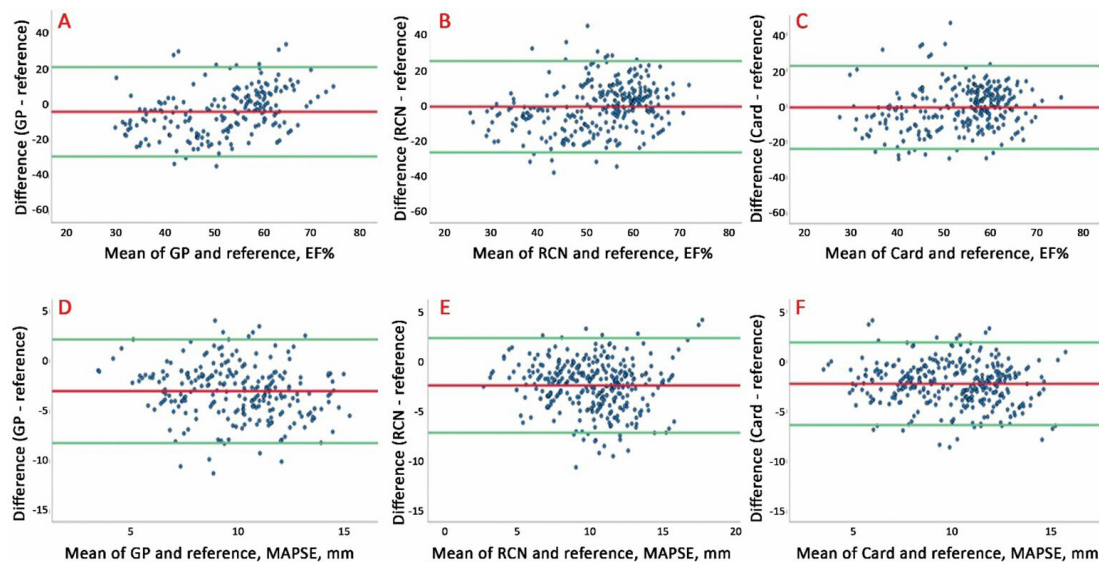


Figure 3 Bland-Altman plots illustrating agreement between the autoEF and autoMAPSE in recordings deemed acceptable for clinical use by evaluation of the blinded cardiologist (blinded image score ≥ 2). Upper panel: autoEF recorded by (A) GPs, (B) RCNs and (C) Card. Lower panel: autoMAPSE by (D) GPs, (E) RCNs and (F) Card. AutoEF, automatic measurement of left ventricular ejection fraction; autoMAPSE, automatic measurement of mitral annular plane systolic excursion; Card, cardiologist; GP, general practitioner; RCN, registered cardiac nurse.



and experts. The main findings were: first, that the feasibility of the applications was acceptable, even though being highest among experts and second, the agreement with reference was poor to moderate, and even for the experts the agreement and reliability were barely within the ranges recommended for clinical use.

Participants

The study population represents patients referred for cardiac examination to rule-in or rule-out HF in everyday clinical practice. The novices underwent limited, but dedicated training. The intermediate group used focused cardiac ultrasound in their clinical practice, and the experts were experienced in echocardiography and the use of HUDs. The training of novices, as well as lack of additional training for the more advanced user groups, was in line with comparable studies and present recommendations.^{10 22 23} Most of the patients were overweight or obese and comorbidities such as atrial fibrillation and hypertension were common. Thus, both poor acoustics and atrial fibrillation (present at examination in 24%) could interfere with image acquisition and the precision of the automatic measurements.

Feasibility

The ability to run the automatic decision-support software was high for autoEF and autoMAPSE with >80% and >92% success rate for performance by all user groups when no quality assessment of the recorded image or performance of the applications was performed. The proportions were lowest for the novices and highest for the experts. The feasibility of the autoEF application significantly improved after revision. However, after blinded quality assessment by the external cardiologist the feasibility was markedly impaired for both applications. In novices, 35%–40% of the automatic decision-support software recordings were not recommended for clinical use. In the intermediate group and experts, the corresponding proportions were approximately 20% and 10%, respectively. Additionally, the proportion of images where the operators were able to run the autoEF software was somewhat lower with the second version of the software, which may be caused by stricter rules for when the algorithm succeeded. Recently, automatic quantification of LV EF has been evaluated in a couple of studies by experienced users.^{15 24} One study evaluated the same autoEF software operated by a cardiology fellow trained in advanced echocardiography for six months prior to study start. There the automatic LV quantification succeeded in 76 of 112 patients (68%).²⁴ In our study, the feasibility of the autoEF application significantly improved after revision for all user groups. This finding indicates that the training effect was minimal. Our findings also highlight the importance of comprehensive evaluation of diagnostic decision-support software before implementation into clinical practise. This also applies to revised versions of the decision-support software and not only before introduction to the market. Additionally, the proportion of recordings with the highest possible

score in blinded evaluation by the external cardiologist was somewhat lower after revision of the autoEF software. The time consumption for the complete HUD examinations was on average 18–23 min for novices and the intermediate group, which we believe is acceptable in selected cases in the everyday practice with significant potential for clinical benefit. However, the time used was higher than in previous publications evaluating focused cardiac ultrasound by HUDs performed by more experienced users.^{11 15 25}

The intra-rater and inter-rater ICCs for novices and the intermediate group were mainly lower than what would be recommended for clinical use (commonly used cut-off of 0.75).²⁶ For experts the ICCs were somewhat higher, but compared with reference only 0.51, and in intra-rater analyses 0.72–0.83, respectively. In a recent publication using another HUD platform by a single cardiologist for automatic quantification of LV EF the ICC was 0.91.¹⁵ Even though the presented data are not directly comparable, they may indicate that reliability was somewhat lower in the present study, even when the autoEF software was used by experienced cardiologists in the current study. Furthermore, we find that image quality and operator experience alone cannot fully explain the moderate intra-operator reliability among the experienced cardiologists. Future studies must address how the next-generation automatic analyses of LV function will perform across users of varying level of experience.

The agreement was poor for automatic measurements of EF and MAPSE for all users. Even though the bias for autoEF was lower for the most experienced users, the agreement was poor to moderate for all user groups. In the recent publications by Filipiak-Strzecka and Papadopoulou, the lower–upper limits of agreement with reference were –10–12 (EF %) and –16–13 (EF %), respectively.^{24 27} Thus, both studies found somewhat better agreement for LV EF compared with the presented limits of agreement as shown in figures 2 and 3, but neither the design nor the presented data are directly comparable. For autoMAPSE, the underestimation compared with reference was consistent and replicates the findings from a previous study by our group.¹⁹ This highlights that the cut-off for pathology is not interchangeable between different methods. Suboptimal image acquisition by less experienced users partially explains the difference across user groups. Importantly, the agreement and reliability were suboptimal also in experts which indicate that the decision-support software needs refinement before incorporation as a reliable tool in everyday clinical practice. The latter is of special importance before implementation by less experienced operators.

The patients' perspective

From the patients' perspective it is important to provide correct diagnosis, and thus, treatment as soon as possible. Fast and precise diagnostics may reduce patient suffering and improve the quality of care. Moving advanced diagnostics to the patients' point-of-care may shorten time to

diagnosis and improve care. As indicated by this study, it is of utmost importance to thoroughly evaluate novel methodology before implementation into clinical practice, since further diagnostic workup may be delayed in case of false negative findings.

Strengths and limitations

The main strengths of this study design is the use of blinded examinations of the consecutive patients by three different user groups ranging from trained novices to experts, blinded review of the feasibility of the automatic algorithms' performance and the use of similar HUDs equipped with two relevant automatic decision-support software. The real-time automatic quantification of LV function on HUDs by inexperienced users with real-time feedback has to our knowledge not been done before. Furthermore, the novices were recruited by the municipality based on their role at various healthcare institutions and not on personal motivation to attend the study. This improves the generalisability but may have impaired the performance of the novices compared with the more experienced user groups. The adequate power of the study is another strength.

The most important limitation relates to the lack of a gold standard for evaluation of LV function. Thus, measurements of LV function by HUDs were compared with the experts' comprehensive echocardiographic measurements. However, the feasibility and reliability across groups are less influenced by the lack of a gold standard. Further, we believe that the blinded evaluation of all recordings with the automatic decision-support overlay provides valuable insight into the performance of the HUD and the automatic decision-support software across user groups. Another limitation which may have influenced the performance of the autoEF software is related to internal error of the first software version which was detected during blinded image analyses. The reduced performance of the first version may particularly have challenged the less experienced users and may also be of importance after software revision. However, the performance of the revised software (among experts) indicates that the automatic decision-support software needs further refinement before broad clinical implementation.

CONCLUSION

Novice GPs, intermediate experienced RCNs and expert cardiologists were able to perform automatic analyses of LV function by automatic decision-support software implemented on HUDs. However, these automatic measurements showed poor to moderate agreement with reference and modest reliability. While this study is a step in the right direction using novel technology to aid healthcare providers in diagnostic decision-making, there is a need for more reliable methods before large-scale implementation into clinical practice.

Author affiliations

¹Department of Internal medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

²Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

³Clinic of Cardiology, St. Olavs University Hospital, St. Olavs University Hospital, Trondheim, Norway

⁴Department of Research, Nord-Trøndelag Hospital Trust, Levanger, Norway

⁵Department of Health Sciences, Mid Sweden University, Östersund, Sweden

⁶Innherred Heart Clinic, Levanger, Norway

Acknowledgements We want to thank all clinicians and other employees involved at Nord-Trøndelag Hospital Trust for their support and for contributing to data collection in this research project.

Contributors AKH-H has contributed to protocol description, data collection, data analyses, manuscript draft and revision. MM has contributed to data collection, manuscript draft and revision. GA, TG, JOK, KS, and OCM have contributed to data acquisition and manuscript revision. BL has contributed to manuscript revision. LL has provided software development, contribution to study design and manuscript revision. HD is the main developer of study design, has contributed to data acquisition, data analyses, manuscript revision, and is responsible for the overall content as the guarantor.

Funding The study was funded by grants from the European Interreg A initiatives (Norwegian-Swedish initiative), Research Council of Norway (Norges Forskningsråd) and Nord-Trøndelag Hospital Trust. The Department of Circulation and Medical Imaging, Norwegian University of Science and Technology hosts a research collaboration between university, hospitals and various vendors funded by the Norwegian Research Council.

Competing interests GE Ultrasound provided the HUD devices for loan through a research contract with the project leader (HD), but GE had no role in performance of the study. MIM, OCM, LL and HD hold positions at Centre for Innovative Ultrasound Solutions (CIUS) where GE Ultrasound is one of the industrial partners. LL acts as part-time consultant for GE Ultrasound.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the 'Methods' section for further details.

Patient consent for publication Not applicable.

Ethics approval This study was approved by Regional Committee for Medical and Health Research Ethics (REK2017/2054). The study was performed in conformity with the Declaration of Helsinki. All participants gave their informed, written consent prior to inclusion.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Deidentified patient data can be made available from the last author Havard Dalen (ORCID 0000-0003-1192-3663) upon reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Anna Katarina Hjorth-Hansen <http://orcid.org/0000-0003-4626-2113>

Bodil Landstad <http://orcid.org/0000-0001-6558-3129>

REFERENCES

- 1 Ponikowski P, Anker SD, AlHabib KF, *et al*. Heart failure: preventing disease and death worldwide. *ESC Heart Failure* 2014;1:4–25.
- 2 Maggioni AP, Dahlström U, Filippatos G, *et al*. EURObservational research programme: the heart failure pilot survey (ESC-HF pilot). *Eur J Heart Fail* 2010;12:1076–84.
- 3 Zhitny V, Iftekhar N, Alexander L, *et al*. Cardiology fellow diagnostic accuracy and data interpretation outcomes: a review of the current literature. *Vasc Health Risk Manag* 2020;16:429–35.
- 4 Bottle A, Kim D, Aylin P, *et al*. Routes to diagnosis of heart failure: observational study using linked data in England. *Heart* 2018;104:600–5.
- 5 Ponikowski P, Voors AA, Anker SD, *et al*. 2016 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;37:2129–200.
- 6 Willenheimer R, Cline C, Erhardt L, *et al*. Left ventricular atrioventricular plane displacement: an echocardiographic technique for rapid assessment of prognosis in heart failure. *Heart* 1997;78:230–6.
 - 7 Snare SR, Mjølstad OC, Orderud F, *et al*. Fast automatic measurement of mitral annulus excursion using a pocket-sized ultrasound system. *Ultrasound Med Biol* 2011;37:617–31.
 - 8 Vinereanu D, Nicolaidis E, Tweddel AC, *et al*. “Pure” diastolic dysfunction is associated with long-axis systolic dysfunction. Implications for the diagnosis and classification of heart failure. *Eur J Heart Fail* 2005;7:820–8.
 - 9 Díaz-Gómez JL, Mayo PH, Koenig SJ. Point-Of-Care ultrasonography. *N Engl J Med* 2021;385:1593–602.
 - 10 Cardim N, Dalen H, Voigt J-U, *et al*. The use of handheld ultrasound devices: a position statement of the European association of cardiovascular imaging (2018 update). *Eur Heart J Cardiovasc Imaging* 2019;20:245–52.
 - 11 Andersen GN, Graven T, Skjetne K, *et al*. Diagnostic influence of routine point-of-care pocket-size ultrasound examinations performed by medical residents. *J Ultrasound Med* 2015;34:627–36.
 - 12 Andersen GN, Viset A, Mjølstad OC, *et al*. Feasibility and accuracy of point-of-care pocket-size ultrasonography performed by medical students. *BMC Med Educ* 2014;14.
 - 13 Gundersen GH, Norekval TM, Haug HH, *et al*. Adding point of care ultrasound to assess volume status in heart failure patients in a nurse-led outpatient clinic. A randomised study. *Heart* 2016;102:29–34.
 - 14 Lipczyńska M, Szymański P, Klisiewicz A, *et al*. Hand-Carried echocardiography in heart failure and heart failure risk population: a community based prospective study. *J Am Soc Echocardiogr* 2011;24:125–31.
 - 15 Panoulas VF, Daigeler A-L, Malaweera ASN, *et al*. Pocket-size hand-held cardiac ultrasound as an adjunct to clinical examination in the hands of medical students and junior doctors. *Eur Heart J Cardiovasc Imaging* 2013;14:323–30.
 - 16 Mjølstad OC, Andersen GN, Dalen H, *et al*. Feasibility and reliability of point-of-care pocket-size echocardiography performed by medical residents. *Eur Heart J Cardiovasc Imaging* 2013;14:1195–202.
 - 17 Hjorth-Hansen AK, Andersen GN, Graven T, *et al*. Feasibility and accuracy of Tele-Echocardiography, with examinations by nurses and interpretation by an expert via telemedicine, in an outpatient heart failure clinic. *J Ultrasound Med* 2020;39:2313–23.
 - 18 Lang RM, Badano LP, Mor-Avi V, *et al*. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of echocardiography and the European association of cardiovascular imaging. *J Am Soc Echocardiogr* 2015;28:1–39.
 - 19 Magelssen MI, Palmer CL, Hjorth-Hansen A, *et al*. Feasibility and reliability of automatic quantitative analyses of mitral annular plane systolic excursion by handheld ultrasound devices: a pilot study. *J Ultrasound Med* 2021;40:341–50.
 - 20 Storve S, Grue JF, Samstad S, *et al*. Realtime automatic assessment of cardiac function in echocardiography. *IEEE Trans Ultrason Ferroelectr Freq Control* 2016;63:358–68.
 - 21 Detry MA, Ma Y. Analyzing repeated measurements using mixed models. *JAMA* 2016;315:407–8.
 - 22 Spencer KT, Kimura BJ, Korcarz CE, *et al*. Focused cardiac ultrasound: recommendations from the American Society of Echocardiography. *Journal Am Soc Echocardiogr* 2013;26:567–81.
 - 23 Via G, Hussain A, Wells M, *et al*. International evidence-based recommendations for focused cardiac ultrasound. *J Am Soc Echocardiogr* 2014;27:683.e1–683.e33.
 - 24 Filipiak-Strzecka D, Kasprzak JD, Wejner-Mik P, *et al*. Artificial Intelligence-Powered measurement of left ventricular ejection fraction using a handheld ultrasound device. *Ultrasound Med Biol* 2021;47:1120–5.
 - 25 Andersen GN, Haugen BO, Graven T, *et al*. Feasibility and reliability of point-of-care pocket-sized echocardiography. *Eur Heart J Cardiovasc Imaging* 2011;12:665–70.
 - 26 Kleijn SA, Aly MFA, Terwee CB, *et al*. Reliability of left ventricular volumes and function measurements using three-dimensional speckle tracking echocardiography. *Eur Heart J Cardiovasc Imaging* 2012;13:159–68.
 - 27 Papadopoulou S-L, Sachpekidis V, Kantartzi V, *et al*. Clinical validation of an artificial intelligence-assisted algorithm for automated quantification of left ventricular ejection fraction in real time by a novel handheld ultrasound device. *European Heart Journal - Digital Health* 2022;3:29–37.