Doctoral theses at NTNU, 2023:288

Anna Katarina Hjorth-Hansen

Limited echocardiography and focused hand-held ultrasound by non-experts with decision-support for diagnosis and follow-up of heart failure patients.

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of Science and Technology

Anna Katarina Hjorth-Hansen

Limited echocardiography and focused hand-held ultrasound by non-experts with decision-support for diagnosis and follow-up of heart failure patients.

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2023

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

ISBN 978-82-326-7276-9 (printed ver.)
ISBN 978-82-326-7275-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:288

Printed by NTNU Grafisk senter

Acknowledgments

This thesis was undertaken at the Department of Internal Medicine at Levanger Hospital and the Department of Circulation and Medical imaging (ISB), Faculty of Medicine and Health Science, Norwegian University of Science and Technology (NTNU). The funding was provided by the EU initiative Interreg and Nord-Trøndelag Hospital Trust, Levanger, Norway.

Main supervisor Håvard Dalen is a creative researcher with a rare capacity for multitasking and focus, while always providing friendly and constructive criticism. Thank you for introducing me to the world of ultrasound research, letting me channel frustrations and never raising an eyebrow at family emergencies mid-project. I have been lucky to learn from your technical insight and research talent. With a great dose of humour plenty of patience, and with a feeling of mutuality you have formed a safe place for me to fail, ask and accomplish.

Co-supervisor Ole Christian Mjølstad has shared his valuable clinical experience, feedback, and notorious optimism throughout this process. Your kind and friendly approach to everyone around you is inspiring. I truly appreciate every effort.

Co-supervisor Bodil Landstad is one of the most inspiring and enthusiastic people I have had the pleasure to get to know. You have a great eye for research and can turn any situation into an advantage. With your unique network, you have introduced me to several new and interesting opportunities for which I am very grateful.

The Cardiology Department at Levanger Hospital deserves special thanks for all the patience with this time-consuming project while also providing an excellent clinical teaching environment. In particular Garrett Andersen, Torbjørn Graven, Olaf Kleinau, Kyrre Skjetne, and Stian Stølen have contributed with significant efforts and valuable inputs. Thanks to Margareta Israelsson, Hilde Haugberg Haug, and Guri Holmen Gundersen for the endurance during these projects. Without the three of you and the nursing staff at the ECG laboratory the project would not have been able to run as smoothly as it did. The general practitioners (Håvard Dybdahl, Anforea Getz, Håkon Gullsvåg, Hermann Leithe, and Ann Kristin Lyngsaunet) participating in the project deserves our gratitude for their significant efforts in the project. Getting to know all of you has shown me the importance of knowing colleagues outside of the hospital walls to be able to provide the best possible care for our patients.

Thanks to the Head of the Department of Internal Medicine, Solfrid Romundstad, for many years of sound advice and flexibility while finishing this project. Thanks to the head of the Clinic of Medicine, Carl Platou, for all your support and for providing a solution to any problem that occurred throughout this period.

This project would have been a struggle without my closest PhD-colleagues and friends Malgorzata "Maggan" Magelssen, Lars Mølgaard Saxhaug, and Sigrid Anna Aalberg Vikjord. Maggan never lets me mope and always sees positive opportunities. We have always worked well together and wanted the best for each other, for which I am very thankful. Lars has a unique ability to think outside the box, while Sigrid has given invaluable inputs and taught me many aspects within "the world of research".

Thank you, Rob, for all your insight, truth, and all the love I could ever ask for. I am appreciative for the time and patience while working long hours, while still having fun attempting something resembling work-life balance with our three incredible children. I am eternally grateful to my parents, Eszter and Henrik, for being only a phone call away to offer endless support, love, motivation, and for introducing me to the academic world before I even started school. Be it either tagging along for "just one more thing" at the hospital or conversations at the dinner table, you have been my supportive structure. Thank you, Peter, and Kristiane, for being so kind and loyal, and for all our wonderful conversations. Thanks to my most generous parents-in-law, Carol, and Jim. You have my deepest gratitude for babysitting, all help indoors and outdoors, advice, and positively refreshing look on life.

I also want to thank Karoline and Martin, Karin and Erik, Hanne Marie, Silje, Harald, Anna, and Maren for all the fun times and distractions to break up the writing of a thesis and the routine of everyday work.

List of papers

Paper I:

Hjorth-Hansen AK, Andersen GN, Graven T, Gundersen GH, Kleinau JO, Mjølstad OC, Skjetne K, Stølen S, Torp H, Dalen H. Feasibility and Accuracy of Tele-echocardiography with Examinations by Nurses and Interpretation by and Expert via Telemedicine in an outpatient heart failure clinic. Journal of Ultrasound in Medicine 2020; 39:2313-2323.

Paper II:

Hjorth-Hansen AK, Magelssen MI, Andersen GN, Graven T, Kleinau JO, Landstad BJ, Løvstakken L, Skjetne K, Mjølstad OC, Dalen H. 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.

Paper III:

Hjorth-Hansen AK, Magelssen MI, Andersen GN, Graven T, Kleinau JO, Landstad BJ, Løvstakken L, Skjetne K, Mjølstad OC, Dalen H. User experience and image quality influence on performance of automated real-time quantification of left ventricular function by hand-held ultrasound devices: A diagnostic accuracy study with data from general practitioners, nurses, and cardiologists. Open Heart 2022; 9:e002083.

Paper IV:

Magelssen MI, Hjorth-Hansen AK, Andersen GN, Graven T, Kleinau JO, Løvstakken L, Skjetne K, Dalen H, Mjølstad OC. Clinical influence of hand-held ultrasound, supported by automatic quantification and telemedicine, in suspected heart failure. Ultrasound in Medicine and Biology Feb 17;S0301-5629(22)00680-9.

Contribution to studies

Study I

Performance of statistical analyses, drafting and revisions of manuscript. Participation in applications for research funding from Nord-Trøndelag Hospital Trust.

Study II

Participation in conception and design, protocol description, data collection, and applications for research funding. Performance of data analyses, drafting and revisions of manuscript.

Study III

Participation in conception and design, protocol description, data collection, and applications for research funding. Performance of data analyses, drafting and revisions of manuscript.

Study IV

Participation in conception and design, protocol description, data collection, data analyses, applications for research funding, and manuscript revisions.

List of acronyms and abbreviations

ACEI	Angiotensin Converting Enzyme Inhibitor
ADSL	Asymmetric digital subscriber line
ARB	Angiotensin II receptor blocker
AutoEF	Automatic measurement of left ventricular ejection fraction
AutoMAPSE	Automatic measurement of mitral annular plane systolic excursion
BMI	Body mass index
COPD	Chronic obstructive pulmonary disease
ECG	Electrocardiogram
EDV	(Left ventricular) end-diastolic volume
eGFR	estimated glomerular filtration rate
ESV	(Left ventricular) end-systolic volume
FoCUS	Focused cardiac ultrasound
EF	Ejection fraction
GP	General practitioner
HF	Heart failure
HFmrEF	Heart failure with mildly reduced ejection fraction
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
HUD	Hand-held ultrasound device
ICC	Interclass correlation coefficient
ICV	Inferior caval vein
IQR	Interquartile range
LA	Left atrium
LAVI	Left atrial volume index

LV	Left ventricle
MAPSE	Mitral annular plane systolic excursion
NRI	Net reclassification index
NTNU	Norwegian University of Technology and Science (Norges teknisk-naturvitenskapelige universitetet)
NT-Pro-BNP	N-terminal pro-Brain Natriuretic Peptide
NYHA	New York Heart Association
PoCUS	Point of care ultrasound

List of figures and tables

Table 5. Echocardiographic indices by the telemedical approach and reference cardiologist.

Table 6. Feasibility of image recording with automatic decision-support software on hand-held ultrasound devices.

Table 7. Agreement of left ventricular function measurements by automatic decision-support software and reference by user groups.

Table 8. Intra- and inter-rater reliability of automatic measurements of left ventricular function by hand-held ultrasound devices according to operator.

Table 9. All recordings of both automatic decision-support software per user and the average scores per parameter.

Table 10. The importance of components of the image quality score for feasibility of automatic decision-support software.

Table 11. General practitioners' diagnostics decision compared to reference.

Table of Contents

Sammendrag på norsk

Dette doktorgradsarbeidet har satt søkelys på begrenset ekkokardiografi og håndholdt ultralyd av ikke-eksperter for å forbedre diagnostikk og oppfølging av pasienter med hjertesvikt. Forekomsten av hjertesvikt er stadig stigende med befolkningsvekst og økende alder. Byrden for enkeltpasienter og helsevesenet relaterer til mange av sykehusinnleggelser, høye kostnader og økt dødelighet. Det er behov for å gjøre utredning og behandling mer strømlinjeformet og å utnytte helsepersonell som ikke er kardiologer slik at helsehjelpen blir mer tilgjengelig. Målet på sikt er å redusere belastende og lange reiser til sykehusene, senke kostnader og å gi trygg og god helsehjelp uavhengig av bosted.

Den første delen av arbeidet har benyttet begrenset ekkokardiografisk undersøkelse gjennomført av kardiologiske sykepleiere med telemedisinsk tolkning fra ekstern kardiolog i en pasientgruppe med etablert hjertesvikt (studie I). Den andre delen har undersøkt fastlegers og sykepleiers nytte av fokusert, orienterende ultralydundersøkelse av hjertet med håndholdt ultralydutstyr, i tillegg til automatiske beslutningsverktøy i en pasientgruppe uten tidligere kjent hjertesvikt som er henvist for vurdering med tanke på hjertesvikt (studie II-IV). Beslutningsstøtten har inkludert automatiske målinger av venstre ventrikkels ejeksjonsfraksjon (autoEF) og mitralplanets langaksebevegelse (autoMAPSE), samt telemedisinsk overføring av bilder for vurdering av ekstern kardiolog med skriftlig tilbakemelding samme dag.

Resultatene viser at slike metoder er gjennomførbare. Studie I viste sammenliknbare måleresultater med referanseundersøkelsene og at slike ultralydundersøkelser var gjennomførbare med rimelig tidsbruk hos selekterte pasienter. Studie II-IV viste at brukerstøtte med kunstig intelligens er gjennomførbart, samt at brukerne har klar diagnostisk nytte av å legge til ultralydundersøkelse av hjertet i vurdering av pasienter med mulig hjertesvikt og også av den telemedisinske vurderingen og tilbakemeldingen de fikk fra kardiologen. Studiene viste også gjennomgående at automatisk analyse av hjertefunksjonen med autoEF og autoMAPSE sviktet i mange tilfeller. Det manglende samsvaret med referanseundersøkelser kunne i liten grad forklares av dårlig bildekvalitet eller brukernes erfaring. Selv når disse målingene ble utført i ekspertenes opptak var måleresultatene upålitelige sammenliknet med referanseundersøkelse.

Studiene er et skritt i rett retning for å øke tilgjengeligheten av mer nøyaktig og pålitelig hjertesviktdiagnostikk og oppfølging. Funnet av begrenset presisjon og nytte av de nyutviklede verktøyene for automatisk måling av hjertefunksjonen synliggjør videre behovet for grundig klinisk validering av alle nyskapende diagnostiske verktøy før de innføres rutinemessig inn i kliniske praksis.

Summary in English

This thesis focused on the use of limited echocardiography and focused hand-held ultrasound examinations performed by non-experts to improve the diagnostics and follow-up of patients with heart failure. Increasing age in the population leads to a higher number of patients suffering from heart failure. The burden for the patients and the health care system includes frequent hospital admissions, increasing costs, as well as a high morbidity and high mortality. There is a need to improve the diagnostic and follow-up workflow, and utilizing non-expert health care personnel supported by technological advances may increase accessibility. The long-term goal is to provide high quality diagnostics at the patients' point-of-care and reduce hospital admissions, patient transportation, and costs.

In the study I, the agreement of limited echocardiographic ultrasound protocols performed by cardiac nurses in combination with telemedical interpretation by an external cardiologist compared to reference exams by an in-hospital cardiologists were evaluated on patients followed up due to heart failure. In the study II-IV, different aspects of focused cardiac ultrasound examinations by general practitioners and cardiac nurses using hand-held ultrasound devices combined with decision-support software was explored. The latter studies evaluate the performance and clinical influence of hand-held ultrasound examinations by inexperienced users supported by automatic decision support software assessing left ventricular ejection fraction (autoEF) and the mitral annular plane systolic excursion (autoMAPSE) as well as telemedical image evaluation by an external cardiologist.

Study I showed that limited echocardiography by nurses supported by telemedical interpretation could provide comparable results to the reference examination and was considered feasible with reasonable time spent on selected patients. Study II-IV concludes that the feasibility of hand-held ultrasound examinations and automated decision-support software was user dependent. The general practitioners can benefit from adding a focused ultrasound examination and from telemedical support from an expert when evaluating patients with potential heart failure. However, the automatic decision-support tools failed to give reliable outputs and was only to a small degree explained by image quality and the experience of the users. Even experienced cardiologist struggled to achieve reliable results.

The results constitute a step in the right direction to increase accessibility to more precise and reliable heart failure diagnostics. The findings of modest feasibility, agreement, reliability, and clinical influence highlight the need for clinical validation of all novel technology before implementation into clinical practice, as well as refinement of the automated decision-support tools studied in this work.

Introduction

Heart failure

Definition and prevalence

Heart failure (HF) is defined as a clinical syndrome, and not as a specific disease. The syndrome consists of symptoms and signs (shortness of breath, peripheral oedemas, fatigue, increased jugular venous pressure) caused by a structural or functional pathology of the heart, increased intracardiac pressure or reduced cardiac output. The objective findings are most often evaluated by echocardiography (1). Importantly HF can occur with or without reduced left ventricular (LV) ejection fraction (EF). The overall incidence of HF is increasing and is estimated to affect over 23 million people worldwide (2). A large population based study from the United Kingdom concluded that the burden of HF is as high as the four most common cancer types combined (3). HF is more common in the elderly population with a prevalence of about 1% below the age of 55 years increasing to >10% in patients over 70 years (4). A study from 2013 states that over 1 million annual hospitalizations in the United States are related to HF (5). Despite current treatment options, morbidity and mortality are still high, and 25-50% of all HF patients hospitalized with decompensation are readmitted within six months (6, 7).

Classification

The common classification of HF in both guidelines and everyday clinic is to group the patients according to LV EF: HF with preserved EF corresponds to LV EF ≥50% (HFpEF), HF with mildly reduced EF (HFmrEF; LV EF 41-49%), and HF with reduced EF (HFrEF; LV EF ≤40%) (1, 8). The New York Heart Association (NYHA) functional score serves as a systematic approach to clinical signs and symptoms and is classified into four groups (Table 1). However, the NYHA classification has limitations and relies only on clinical symptoms which makes it a poor predictor of HF prognosis (9).

Table 1. New York Heart Association functional classification of clinical symptom severity.

Class I	No limitation of physical activity. Ordinary physical activity does not cause undue breathlessness, fatigue, or palpitations.
Class II	Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in undue breathlessness, fatigue, or palpitations.
Class III	Marked limitation of physical activity. Comfortable at rest, but less than ordinary activity results undue breathlessness, fatigue, or palpitations.
Class IV	Unable to carry on any physical activity without discomfort. Symptoms at rest can be present. If any physical activity is undertaken, discomfort is increased.

Table 1. The New York Heart Association functional classification of clinical symptom severity scale is the most used scale to evaluate heart failure patients' severity of symptoms both at diagnosis and during follow-up (1).

Aetiology and pathophysiology

In developed countries HF is primarily caused by cardiovascular disease (e.g., coronary artery disease, hypertension, valvular heart disease, and arrhythmias) (10). Beyond structural changes of the myocardium related to these four specific comorbidities, cardiomyopathies (hypertrophic, dilated, arrhythmogenic, Takotsubo, peripartum, and toxic), congenital heart disease, infectious causes (viral or parasitic), drug-induced/toxic injury, infiltrative disease (amyloidosis, sarcoidosis), and storage diseases (hemochromatosis, Fabry's disease, glycogen storage disease) are also important causes of HF. The cause may be extra-myocardial and related to diseases of e.g., the pericardium or neuromuscular disease as well.

Diagnosing heart failure

HF diagnosis can be challenging due to unspecific clinical findings and symptoms (shortness of breath, oedema, fatigue etc.) that can be misinterpreted as pulmonary or renal disorders. In everyday clinical practice non-experts are usually the first to see a patient with potential HF. Clinical signs combined with the traditional physical examination with auscultation, x-ray imaging and electrocardiograms are not sufficient for accurate diagnosis. The cornerstone of HF evaluation and diagnosis is echocardiography. Echocardiography is demands large resources with limited availability outside specialized laboratories. N-terminal pro-Brain Natriuretic Peptide (NT-proBNP) has high sensitivity to detect HF, but only a moderate specificity and can therefore be used to safely rule out HF outside the specialist health care system (1). In a study by van Riet et.al. only 25% of patients >65 years with elevated NT-proBNP and dyspnoea met the criteria for HF (11). Pleural effusion and the inferior caval vein (ICV) dimensions are parameters easily assessable by hand-held ultrasound devices (HUDs) as shown in a study where cardiac nurses performed a simple and very limited ultrasound examination to assess volume status in a HF population (12). Adding focused, point-of-care ultrasound examinations performed by inexperienced users can improve selection and the diagnostic precision of patients for specialist work-up when applied according to recommendations (12-14).

Echocardiography and ultrasound in medicine

History

The Curie brothers discovered the piezoelectric effect in 1880, and throughout the 20th century ultrasound was applied to, amongst other, submarine navigation technology, industrial fishing and eventually medicine in the beginning of the 1940 (15). The Swedish cardiologist Inge Edler and nuclear physicist Hellmuth Hertz were the first to successfully image the motion of the heart by ultrasound in 1953 and are considered the “fathers of echocardiography” (16). The first commercially available ultrasound machine for medical purposes, a B-mode scanner, was launched in 1963. Since then, ultrasound has spread to most medical fields. Liv Hatle, working at the Regional Hospital of Trondheim and affiliated with the Norwegian University of Science and Technology (NTNU), was a pioneer within echocardiography and Doppler diagnostics exploring the pulsed wave and continuous wave Doppler modes (17). Several of the principles first shown by Hatle and colleagues are still widely used in the everyday clinic. After the invention of real-time two-dimensional imaging (1982) and colour Doppler imaging of blood blow (1984) echocardiography became a cornerstone in cardiac diagnostics. Since then, a multitude of diverse and refined techniques have been developed, and diagnostic ultrasound is now the most versatile method of imaging in medicine. Constant innovation has led to miniaturization of the apparatus into affordable and highly portable pocked-sized devices which are more easily available to clinicians.

Ultrasound

Ultrasound utilizes high frequency, non-audible sound waves, usually with frequencies in the megahertz-range. It is non-ionizing and considered safe for diagnostic purpose (18). Its entry to the diagnostic armamentarium has provided a non-invasive option that is portable and allows for real-time diagnostics that was previously less accessible. Additionally, ultrasound can be used for treatment by allowing for higher energy output. Thus, to ensure that no harm is put to the patients due to thermal heating or mechanical stress of tissue during diagnostic imaging The Food and Drug Administration (FDA) has allowed the use of a maximal mechanical index of <1.9 and a maximal thermal index for <1.5 for almost all applications. It is widely accepted that there are no known harmful effects associated with properly applied use of diagnostic ultrasound based on clinical use for several decades. However, there is no such thing as *zero risk*, and the ALARA (as low as reasonably achievable) principle relate also to mechanical and thermal indices within diagnostic ultrasound (18).

Echocardiography

Echocardiography is the most comprehensive method for evaluation of cardiac structure and function by ultrasound. Cardiac structures are quantitatively and qualitatively evaluated during a comprehensive echocardiographic examination. The size and function of the (LV), left atrium (LA), right ventricle (RV), and right atrium (RA) are quantified in among others parasternal long- and short-axis views, apical four-chamber, two-chamber, and long-axis views. Additionally, colour Doppler, spectral Doppler, and tissue Doppler are used for evaluation of myocardial and valvular function. Cardiac function can also be assessed qualitatively based on visual evaluation of the recordings.

Hand-held ultrasound devices

The need for easier access to ultrasound has led to the development of smaller, portable, but less sophisticated ultrasound machines called *hand-held ultrasound devices* (HUDs). The increasing availability of HUDs has transformed the way cardiac ultrasound is utilized from the exclusive examination by experts, to the wide range of use by non-experts. The miniaturization and easy-to-use interphase come at a cost of significant compromises with respect to available modalities and image quality compared to state-of-the-art echocardiographic equipment (19). Together, this adds challenges related to user experience, image interpretation, and clinical yield. HUDs allow clinicians immediate access to visualize organs in the acute or bedside setting in addition to the physical examination. The true advantage of the miniaturized devices is the portability and fit to the pocket of a lab coat. Primarily grayscale 2D-images are obtained, while colour Doppler is available on most HUDs. Spectral Doppler is usually not available which limits the quantitative ability to evaluate valvular pathology. Previously, several studies have shown good agreement for e.g., qualitative assessment of left ventricular function between novices and experts (20-23). A weakness of the HUDs is the lack of tools for quantification and objective measures which is desirable to reduce the subjectivity and operator-dependence.

Since introduction of HUDs to the market in 2007 several vendors provide small, easy to use ultrasound scanners. One of the first successful devices was the VScan® developed in collaboration between GE Vingmed Ultrasound (Horten, Norway), NTNU and St. Olavs University hospital in Trondheim, Norway. Several vendors offer HUDs, but the Vscan® has been used in more cardiac studies and by users with more varying levels of experience than any other. Our research group has been one of the central groups in the evaluation of these devices (13, 24). It is currently one of a few HUDs that offer the addition of artificial intelligence to aid the user in assessing LV function by quantification of LV EF (LVivo EF® by DiA Imaging Analysis, Be'er Sheva, Israel) (25).

Image 1. Cardiac nurse performing hand-held ultrasound examination on patient.

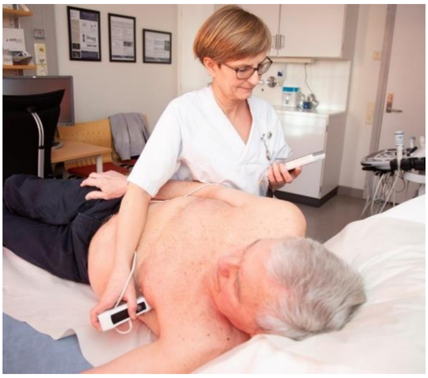


Image 1. A standard hand-held ultrasound examination at the nurse led heart failure clinic at Levanger Hospital. The patient is positioned in the standard left supine position. Both the patient and nurse have consented to the use of the image. ©Karl Jørgen Mathisen/NTNU.

Relevant challenges

HUD technology is inferior to high-end ultrasound equipment and the operators often have limited experience. Examinations are often done in the emergency setting with time constraints or in other unfavourable settings, and complex cardiac pathology is evaluated in an unnuanced manner due to the "pathology/no pathology"-style of reporting (26). Technical hurdles must be overcome with ultrasound diagnostics, but patients also vary greatly in level of echogenicity. Ultrasound does not travel well through air, adipose tissue, and bones which leads to limited fields of view due to costae, air-filled lungs (aggravated by e.g., COPD and emphysema), obesity, and mammary hypertrophy. A moving organ such as the heart is harder to depict, and this is further complicated if cardiac arrythmias are present where the expected duration of a cardiac cycle will appear at random. This leads to beat-to-beat variation in size and functional indices and adds challenges to both recording and interpretation.

Training of personnel

The clinical importance of ultrasound examinations depends on the competence and skill of the user (20, 22, 27). Adequate training of inexperienced personnel is mandatory to benefit from the diagnostic ultrasound examination. The American Society of Echocardiography defines three levels of echocardiographic training. Level I is an introductory level without sufficient experience to independently interpret results, level II allows for performing more specialized care with independent interpretation under certain conditions, and level III requires additional training to acquire specialised skills and knowledge, interpretation and allows for training others (28). Training ultrasound-naïve personnel to perform focused echocardiographic protocols must be planned, limited, and within the scope of the device being used. Any training program should educate the user to a relevant and predetermined extent. It must be stressed that operators must be aware of their limitations and capacity. In the setting where a HUD is used only *point-of-care ultrasonography* (POCUS) and *focused cardiac ultrasound* (FoCUS) can be performed. PoCUS is defined as *a goal-oriented, limited ultrasound examination, extending physical examination performed in any body structure and environment with a predefined limited protocol*, where FoCUS is *a specific type of PoCUS applied to the heart by an operator not necessarily trained in comprehensive echocardiography, but appropriately trained in FoCUS, usually responsible for decision making and/or treatment* (13, 29). These are usually performed on a portable ultrasound system allowing for fast and cost-effective ultrasound assessments of patients by non-experts outside dedicated ultrasound laboratories at the "point-of-care" (30). The recommendations for training with HUDs in various clinical settings have been standardized in a position paper published in 2019 that takes previous experience and the limitation of the HUDs into consideration (13). In summary, their recommendation is to adapt the scope of use to fit the situation, know the limitations, to use it as a supplement to the physical examination, and to train the users appropriately for the task at hand.

Telemedicine

Tele-communicational technology utilized in medicine, i.e., telemedicine, offers possibilities for long-distance contact, advice, education, intervention, and monitoring between patients, primary care physicians and experts. Such technology is particularly useful where the distance to the point-of-care is a critical factor to aid the diagnostic process and therapy. World-wide access through local area networks and wireless mobile telecommunication technology (3G/4G/5G) initiates the potential to utilize telemedicine. Telemedicine has been proven particularly helpful in scarcely populated areas with low accessibility to health care services, with the increasing elderly population being limited for travel, and e.g., in cases of social isolation as experienced during the COVID-19 pandemic (31-33).

Implementing telemedicine for image interpretation is common in the field of radiology to reduce the workload on local radiologists (32). So far, the tele-echocardiographic field has been scarcely explored. Tele-echocardiography was first implemented by paediatric cardiologists supporting physicians in remote areas (33). Telemedicine is not a part of routine HF follow up but if the personnel at site can perform echocardiographic or cardiac ultrasound recordings it could allow for the interpretation to take place or be supported by an expert at another location. The potential benefit for the cardiac patients and society can be a more precise diagnosis as well as reducing the burden of travelling with respect to cost and time.

Dedicated software is necessary to ensure secure transfer of sensitive data in line with the Data Governance Act of 2020 (GDPR) and doctor-patient confidentiality. All software must be approved by the national Data Protection Authority (i.e., “Datatilsynet” in Norway). Several vendors offer various solutions paired and compatible with other analysing software to stream-line workflow and aid the clinicians at the point-of-care.

Aims

The overall aim was to study aspects of feasibility, reproducibility, accuracy, and clinical influence of HUD examinations and limited echocardiography for users with varying level of experience when supported by novel decision-support software as automatic quantification of LV function and telemedical interpretation by experts.

Specific aims

Study I

To examine the feasibility and accuracy of tele-echocardiography in an outpatient HF clinic combining limited echocardiographic recordings by cardiac nurses and telemedical interpretation by an external cardiologist to improve diagnostics and care for HF patients.

Study II

To study the feasibility and reliability of fully automatic quantification of LV function by general practitioners, cardiac nurses, and cardiologists using hand-held ultrasound devices.

Study III

To investigate the influence of operators’ experience on image quality of HUD recordings, and how pre-specified components of image quality influenced feasibility and reliability of the automatic quantification of LV size and function by HUDs.

Study IV

To explore the clinical influence of HUD examinations by general practitioners in patients with suspected heart failure as a stand-alone test and combined with the use of decision-support software for fully automatic quantification of LV size and function as well as telemedical support by expert.

Materials and methods

Study populations

Study I

The population consisted of 50 consecutive patients included from the outpatient HF clinic at Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway (Figure 1). Patients willing to provide informed consent were included between October 2016 and February 2017 given they were ≥18 years, had a known HF diagnosis and planned follow-up at the HF clinic.

Study II-IV

Patients referred for cardiac examination at the outpatient clinic at Levanger Hospital were invited to participate. The referral notes were screened by two cardiology residents (Magelssen M and Hjorth-Hansen A) from June 2018 to June 2020. Inclusion criteria were suspected HF based on the referral note. Exclusion criteria were age <18 years, known HF, and previous cardiac imaging within the last 10 years. Of 185 invited, 170 accepted the invitation, four were later excluded due to withdrawal of consent (n=1), backpain (n=1), no show (n=1), and cognitive dysfunction (n=1). In total, 166 participants were included in the study (Figure 1). The inclusion of patients was paused from March to June 2020 due to the COVID-19 pandemic.

Figure 1. Overview of inclusion and exclusion of patients for both study populations.

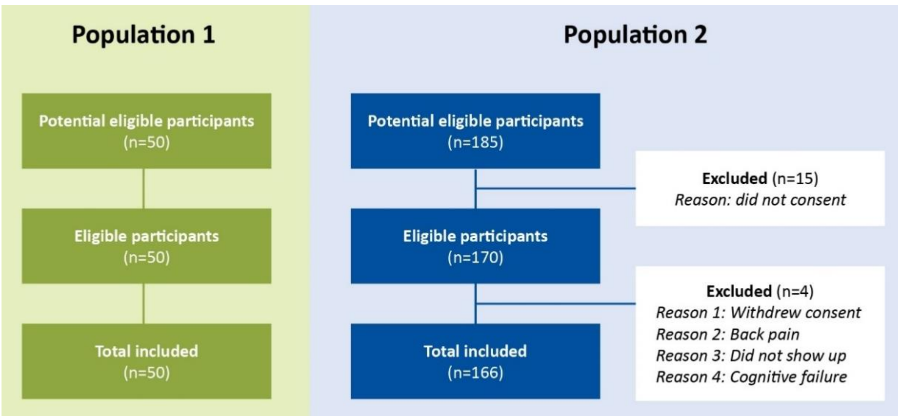


Figure 1. Overview of the study participants in population 1 (study I) and population 2 (study II-IV). All included participants consented to participation in the studies.

Study flow

Figure 2. Overview of the key measurements and outline of all study papers.

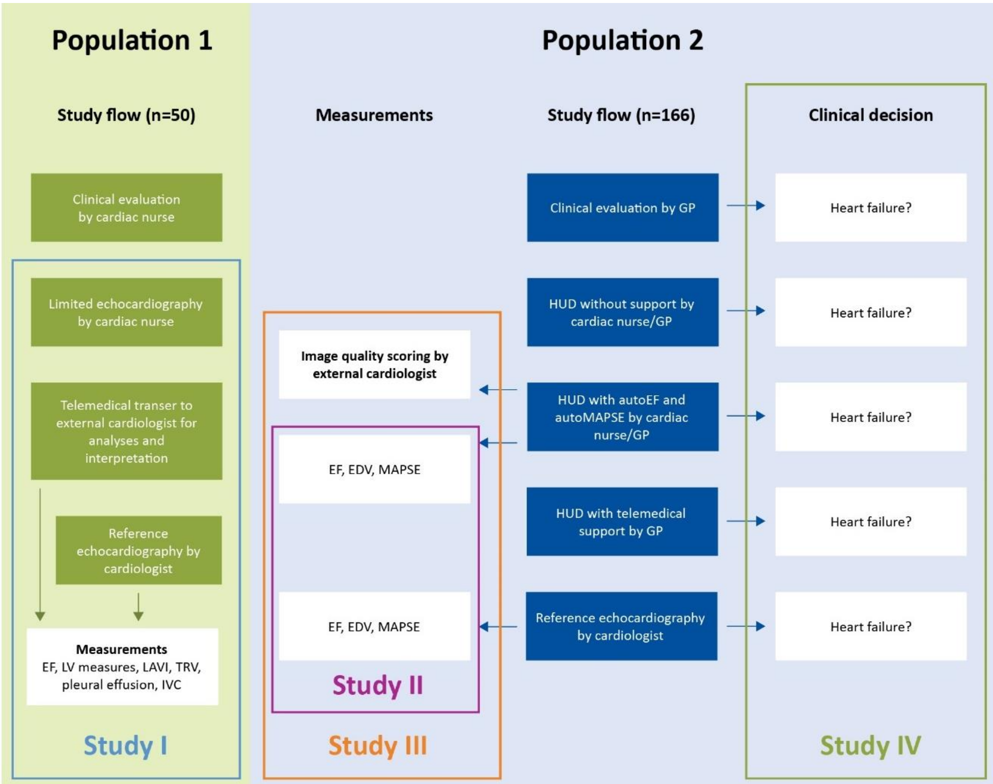


Figure 2. Outline of study flow illustrating the general content of the four studies including key measures, examinations and clinical decision making. *Abbreviations: EDV; LV end-diastolic volume, EF; ejection fraction, GP; general practitioner, HUD; hand-held ultrasound device, ICV; inferior caval vein, LAVI; indexed left atrial end-systolic volume, LV; left ventricle, MAPSE; mitral annular plane systolic excursion, TRV; tricuspid regurgitation peak velocity.*

Study I

The participants were first examined by one of three cardiac nurses performing a limited echocardiographic examination. Immediately after the limited echocardiography the recordings were transferred by telemedicine for further analyses by an out-of-hospital cardiologist (Figure 2). Telemedical interpretation was performed in near real-time categorizing the participants according to whether LV EF was reduced, mildly reduced, or preserved. Key measurements were analysed by the external cardiologist using only the presented recordings. Shortly after the initial limited echocardiography by the cardiac nurses one of four cardiologists or an experienced resident performed reference echocardiography blinded to the telemedical results for comparison.

Study II and III

Patients were examined by one of five GPs and by one of three cardiac nurses, representing the novice and intermediate experienced users, respectively. When examined by GPs the first diagnostic step included only a medical history and a clinical examination. The second step was focused cardiac ultrasound examination of the heart where the GPs visually interpreted their own recordings. For both GPs and cardiac nurses, the next examination steps were addition of the automatic decision-support software for evaluation of LV function, and lastly all images were automatically transferred to one of two out-of-hospital cardiologists by a secured telemedical system and electronic feedback of the cardiologists' interpretation was provided. After the examinations by GPs and the cardiac nurses, a comprehensive reference echocardiography was performed by one of five internal cardiologists, representing the expert group. An additional HUD examination with recordings utilizing the automatic decision-support software was performed by the reference cardiologists for comparison. All examiners were blinded to the results of the others. Four to six patients were included on each of the 30 inclusion days. Due to logistical reasons, the first 29 patients were not examined with the HUD by the cardiologist.

During preliminary analyses of the data, we detected an error in the autoEF software leading to the LVivo EF® applications (DiA Imaging Analysis, Be'er Sheva, Israel) revision by the vendor during the summer of 2019. In total 103 patients were analysed with the first version of the autoEF software and 63 patients with the revised version of the software.

Study IV

As described above and in Figure 2, the GPs considered whether the patients had HF and whether they should be referred to echocardiography after each step of the examination. The GPs did not have the opportunity to answer the question about *further referral or not* at each step of the examination during the first day of inclusion, thus six patients were excluded from those analyses.

Blood samples were drawn at the day of inclusion and analysed at the in-hospital accredited laboratory prior to examinations. Serum N-terminal pro-Brain Natriuretic Peptide (NT-pro-BNP), serum creatinine and estimated glomerular filtration rate (eGFR), basic electrolytes and haemoglobin (Hb) were measured. New York Heart Association (NYHA) functional classification was scored, and anthropometric measures were collected (body weight (kg), body height (cm)) and blood pressure (mmHg) were measured. Electrocardiograms (ECGs) were taken. No further follow-up or ultrasound examinations of the participants was performed for research purposes during all studies. Due to regular delays at the laboratory not all laboratory test results were available prior to the GPs assessment.

Training and education of personnel

Study I

The three specialized cardiac nurses who performed the limited echocardiographic examinations had 6-12 years of clinical experience from the HF outpatient clinic at Levanger Hospital and were previously trained in evaluating fluid status by ultrasound. The nurses were systematically trained one-on-one by specialists in cardiology with extensive experience in echocardiography and performed mean (range) 67 (47-97) limited echocardiographic examinations before start of inclusion.

Three cardiologists and one experienced resident performed the reference echocardiographic examinations. They were all previously trained and skilled in echocardiography. No additional training was provided.

Study II-IV

The GPs included in the study were selected by the respective administrators of the municipalities of Verdal and Levanger. The study conductors therefore had no influence on the selection of GPs. In total, six GPs underwent training in focused cardiac ultrasound by HUDs. Only five participated in the study as one GP changed occupation during the study period. Only one of the six had performed focused ultrasound examinations prior to training (n=7 examinations), and thus, the GPs represented the inexperienced users. Six in-hospital training days with one-to-one supervision by one of two residents experienced in focused cardiac ultrasound, and two evening lectures provided by cardiologists under a three-month training period was the entirety of the training protocol. In addition, they had the opportunity to use a personal HUD without supervision from the first day of training. They performed in total median (range) 46 (45-68) examinations prior to first inclusion whereof median (range) 10 (9-20) were unsupervised and 36 (31-43) were supervised.

The same three cardiac nurses (study I) from the outpatient HF clinic represented the intermediate level users. They had experience from study I (14), as well as years of hands-on ultrasound experience from the same HF outpatient clinic (evaluation of pleural effusion, ICV, and clinical signs and symptoms in a HF patients). They had completed a total of median (range) 118 (74-221) limited echocardiographic examinations before patient inclusion in study II-IV. There was therefore no need for the cardiac nurses to undergo the same systematic training as the GPs. Approximately four weeks prior to inclusion they were instructed in how to initialize the automatic decision-support software for quantification of LV function on the HUD.

Five cardiologists experienced in echocardiography represented the expert group. They were only instructed in how to initialize the automatic tools on the HUDs at the day of inclusion. No additional training was provided.

Details of the ultrasound examinations

Study I

The cardiac nurses had access to medical history, keywords of previous echocardiographic exams, but no access to previous echocardiographic recordings. A Vivid 7 scanner (GE Ultrasound AS, Horten, Norway) was used. All recordings contained at least three cardiac cycles when appropriate. The views recorded were: parasternal long- and short-axis (with and without colour Doppler), three standard apical views (four-chamber, two-chamber and long-axis) with and without colour Doppler focusing on left ventricle (LV), left atrium and right ventricle, respectively, pulsed-wave tissue Doppler with sample volume in the basal part of the septal and lateral walls (four-chamber), pulsed-wave Doppler recordings with the sampled volume at the tip of the mitral leaflets for mitral inflow and continuous wave Doppler through the aortic and tricuspid valves . The sub-costal view allowed for assessment of the ICV, including both the maximal and minimal dimensions by respiratory variation. Pleural cavities were assessed in the sitting position in the mid-clavicular and mid-axillary lines. If pleural effusion was present longitudinal and transverse images were recorded.

One of the four experienced echocardiographers (three cardiologists and one experienced resident) performed the reference echocardiography immediately after the cardiac nurse's examination. High-end ultrasound scanners were used (Vivid E9 or Vivid E95, both GE Ultrasound AS, Horten, Norway). The exams were comprehensive and included all the same recordings specified above for the nurses. At least three cardiac cycles were recorded for each specified recording. The sub-costal recording for assessment of the ICV (including both the maximal and minimal dimensions) included a quick inspiration (*sniff*). The pleural cavities were assessed in the mid-axillary and mid-clavicular lines as described above (34).

The external cardiologists analysed the nurses' recordings and the internal echocardiographers analysed their own recordings. The LV EF was calculated according to the biplane Simpsons method where the endocardial LV border were traced in end-diastole (LV EDV) and end-systole (LV ESV) in four- and two-chamber views. LV internal length was measured as the midline of the traces. LV EF was calculated. LV internal diameter and wall thickness were measured at the level of the tip of the mitral leaflets in two-dimensional grey scale parasternal long-axis view. The left atrial endocardial border was traced in end-systole in four-chamber and two-chamber views, the volume was calculated by the area-length (A-L) method, and subsequently indexed per square meter body surface area (LAVI). The left atrial appendage and pulmonary veins were not included in the traces. Mitral inflow peak early (E) and late (A) velocity, and early filling mitral deceleration time were measured in pulsed-wave Doppler recordings at the tip of the leaflets in the apical four-chamber view. Pulsed-wave tissue Doppler was used to assess mitral annular peak systolic (S') and peak early diastolic (e') longitudinal velocities. The early mitral inflow to the early diastolic mitral annular velocity ratio (E/e') was calculated. Peak velocity through the tricuspid regurgitation (TR) was measured by continuous Doppler. Based on the European Association of Cardiovascular Imaging and the American Society of Echocardiography recommendations LV filling pressure was estimated based on left atrial volume index (LAVI), e’, E/e’, and peak velocity of the TR as normal or elevated (35). In the case of pleural effusion, the amount of fluid was quantified as the distance from the diaphragm to the basal part of the lung.

Study II-IV

All patients underwent HUD examinations with VScan Extend by the three user groups, except for the first 29 patients where the cardiologist did not perform this exam. Reference echocardiography was performed using Vivid E9 or E95 scanners.

The GPs had access to the referral, but not to hospital records or previous cardiac imaging. The scan-protocol included parasternal long- and short-axis views, apical four-chamber view, subcostal four-chamber view, evaluation of the ICV in the subcostal view with assessment of minimal and maximal dimensions, and assessment of the pleural cavities.
The cardiac nurses had access to the same patient information as the GPs. They recorded parasternal long- and short-axis views, apical four-chamber view, apical two-chamber, apical long-axis views, right ventricular focused view, atrial focused recordings, subcostal four-chamber view, and subcostal view for evaluation of the ICV in addition to assessment of the pleural cavities in the sitting position. The recording of the ICV included both the maximal and minimal dimension by including inspiration. Colour Doppler images of the mitral, aortic, and tricuspid valves was recorded.

Reference cardiologists had full access to hospital records, referral notes, and previous imaging, and performed the reference echocardiography as previously described in study I (36). In addition, the reference cardiologist used the HUD for recording of four-chamber view with quantification by the automatic decisions-support software.

Details of the telemedical technology

Study I

Immediately after completion of the limited echocardiographic examination by the nurses a commercial software system, PaCentric® (Fimreite Software, Stavanger, Norway), installed on the Vivid 7 scanner and connected to the hospitals local area network was used to transfer the images. This software allowed for secure transmission of pseudonymized data for subsequent interpretation and reporting of the medical DICOM images per internet. PaCentric® is accredited (ISO 13485:2003) and certified (CE 0434), as well as FDA approved (FDA 510 k100837). Data exported to the PaCentric® server was downloaded by the external cardiologist and stored on a password protected laptop computer for analyses. Depending on the external cardiologist's actual location fibreoptic cables, Asymmetric Digital Subscriber Line (ADSL) and wireless mobile network (3G and 4G) were used for data transfer.

Study II-IV

Pseudonymized HUD images were consecutively transferred without delay via the Tricefy™ cloud-based server (Trice Imaging Inc., CA, USA) software installed on the VScan Extend (GE Ultrasound AS, Horten, Norway). Tricefy is accredited (ISO 13485) and FDA approved (FDA 3009831823) The images were downloaded to a password protected laptop computer and/or a stationary computer outside the outpatient clinic. All measurements were done in the EchoPAC, version 202 and 203 (GE Ultrasound). Feedback to the GPs were provided by pseudoanonymized standardized forms containing semi-quantitative grading and a written evaluation of key messages sent by e-mail consecutively throughout the day of inclusion.

Details of the automatic decision-support tools

For all HUD examinations live cine-loops of at least one cardiac cycle were recorded. Before storing of the four-chamber view cine-loop, the specific automatic decision-support software (autoEF or autoMAPSE) was launched on the HUD and performed the automatic analyses. The recording was repeated aiming for six recordings in total, three with autoEF and three with autoMAPSE. For both autoEF and autoMAPSE the four-chamber view cine-loop with the overlay of the tracking and results from the automatic algorithm was stored on the HUD and transferred without delay to the cloud-based server.

AutoEF

The automatic measurements of LV volumes and subsequent calculation of LV EF was done using the commercially available LVivo EF® application (DiA Imaging Analysis). AutoEF provided fully automatic tracing of the endocardial border in the apical four-chamber view throughout the cardiac cycle. LV volumes were estimated in end-diastole and end-systole. LV EF was calculated from the volume estimates. More specific details on the software was not made available by the vendor.

AutoMAPSE

MAPSE was estimated by an automated algorithm tracking the mitral annular septal and lateral points (autoMAPSE) using an LV model. Briefly, a Real-time Contour Tracking Library (RCTL) processed and tracked the LV movement in images (GE, Vingmed, Norway) using a non-uniform rational B-spline model (37). Further, the septal and lateral points of the mitral annulus were retrieved from the RCTL. The array of points located the maximum displacement of the mitral annular plane. MAPSE was calculated automatically as the mean and specific displacement of the septal and lateral mitral annular points. More extensive technical details of the method is described in a previous paper (38).

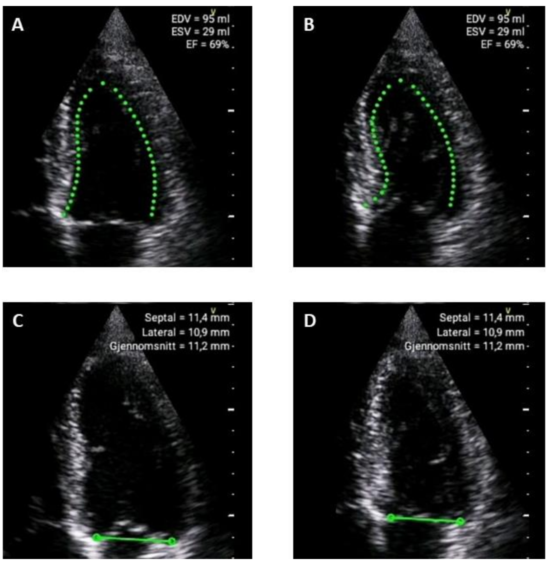


Image 2. Recordings with automatic left ventricular ejection fraction and automatic mitral annular plane systolic excursion measurement overlay on hand-held ultrasound device.

Image 2. Examples of autoEF in A) end-diastole and B) end-systole, and autoMAPSE in C) end-diastole and D) end-systole. *Abbreviations: autoEF; automatic measurement of left ventricular ejection fraction, autoMAPSE; automatic measurement of mitral annular plane systolic excursion, EDV; end-diastolic volume, ESV; end-systolic volume, EF; left ventricular ejection fraction.*

Outcome measures

Table 2. Central outcome measures for study I-IV.

	Central outcomes	Comment
Study I	LV/LA size, functional indices, Doppler, IVC, HF category	Tele-echocardiography vs. reference
Study II	LV EF, EDV, MAPSE, image score	Non-expert users using HUD with decision-support vs. reference
Study III	LV EF, EDV, MAPSE, image score	The effect of user experience and image quality on autoEF and autoMAPSE performance by all users
Study IV	HF diagnosis	Influence of HUD with decision support on HF diagnosis

Table 2. A brief overview of central outcome measures for the four studies. Supplementary table to Figure 2. *Abbreviations: autoEF; automatic measurement of left ventricular ejection fraction, autoMAPSE; automatic measurement of mitral annular plane systolic excursion, EDV, left ventricular end-diastolic volume; EF, ejection fraction; HF, heart failure; HUD, hand-held ultrasound; ICV, inferior caval vein; LA, left atrium; LV, left ventricle; MAPSE, mitral annular plane systolic excursion.*

Image quality was evaluated by one external cardiologist on all HUD recordings with applied automatic decision-support software. The cardiologist was blinded to details of the operator and patient but had access to the referral prior to the evaluation (39-41). The evaluation took the image quality and performance of the software into account. Image quality was quantified by a predefined five category score (Table 3). The mean score was applied to represent the average image quality per recording. In addition, both automatic decision-support software were given an individual evaluation whether they could be recommended for clinical use or not. This was based on 1) the image quality scores, 2) the tracking of the endocardial border (autoEF) or the mitral annular points (autoMAPSE) and 3) the overall performance combined with the numerical output of both algorithms. Further, this was scored as: 1, discard (not for clinical use); 2, accept, but needs adjustment to fit the clinical context due to suboptimal performance of the software; 3, accept result as it is.

Table 3. Image quality score for recordings with automatic decision support software.

Score value	0	1	2	3	4	5	6
4Ch		2Ch, ALAX or others		5Ch or posterior 4Ch			4Ch
LV alignment	≥45°		30-44°		15-29°		<15°
Mispositioning of apex			≥15 mm		<15mm		None
Mitral annular assessment		Not judgeable	Poor	Fair	Good	Near excellent	Excellent
Number of visible endocardial segments		One	Two	Three	Four	Five	Six

Table 3. The prespecified image quality score parameters and the numerical score values in table format as previously described. *Abbreviations: 2Ch, two-chamber view; 4Ch, four-chamber view; 5Ch, five-chamber view; ALAX, apical long-axis view; LV, left ventricular,*

Statistical analyses

Sample size was estimated by Sample Power (SPSS, Inc., Chicago, IL, USA). In study I, a relative error of 15% (absolute 5%) in LV EF provided a sample size of 48. For all the described tests 50 to 70 participants provided adequate power to detect a clinically meaningful difference (relative error of 15%). In study II-IV the power calculations provided a need for 104 patients, but the sample size was expanded to 150 to account for the likelihood of few patients with pathological findings. During study II-IV, preliminary analyses indicated an error in the autoEF algorithm leading to adjustment of the sample size to 170 patients to account for the software upgrade.

Categorical data were reported as frequencies and percentages. Normality was evaluated by histograms and normality plots. Continuous variables are presented as mean ± SD or as median (interquartile range) in case of skewed data. Proportions were analysed by Chi-square test and Fisher's exact test when appropriate. Comparison of means were tested by paired t-tests, Student's t-test, or Wilcoxon test when appropriate. A p-value <0.05 was considered statistically significant. All statistical analyses were performed using IBM SPSS Statistics, version 26, 27 and 28 (SPSS Inc, Chicago, IL, USA) and Microsoft® ExCel®.

Study I

The agreement of the telemedical approach and reference was tested by Bland-Altman statistics, coefficient of variation and Pearson's correlation coefficient. Weighted kappa was used to classify agreement with respect to correct HF. Sensitivity and specificity, negative and positive predictive values was used to assess semi-quantitative data. Logistic regression was used to evaluate the association of HF type with correct classification of LV filling pressures.

Study II

ANOVA with post-hoc LSD correction was used for comparison between groups. The performance was judged as feasible if the user was able to acquire the image and run the algorithm combined with score ≥2 for automatic decision-support software's performance (accept results with or without changes) by the blinded cardiologist. Reliability 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 (38). The intra-rater reliability of the automated decision-support software was calculated in the single-measurement dataset by a two-way mixed model defined by absolute agreement. The inter-rater reliability was calculated with a two-way random model defined by absolute agreement in the dataset of average measurements by HUDs analysed by all groups (GPs, cardiac nurses, and cardiologists) compared to reference. Agreement with the reference examination was evaluated by coefficients of variation, coefficient of repeatability indicating the minimal detectable change, and Bland-Altman statistics.

Study III

Comparison of paired nominal data was done by McNemar's test. Differences between groups was analysed by repeated measure ANOVA with post-hoc Bonferroni correction. Logistic regression and general linear models were used as appropriate to evaluate the influence of the different image quality parameters for performance of the automatic decision-making software. The importance of image quality on the feasibility and agreement with reference was evaluated first on the whole dataset of images from all users, and subsequently within the separate user groups. Agreement with reference was tested at the level of all available automatic measurements. The maximal difference in measurements of autoEF, autoMAPSE, and image quality scores were used to test the importance of the different image quality categories for the within-patient reliability. These analyses were also performed for the dataset as a whole and within user groups.

Study IV

Net reclassification improvement (NRI) was used as an indicator of the clinical gain of HUD examinations with and without decision-support software. NRI was evaluated at each step from HUD examination without user support through the different modes of support. In other words, NRI quantified whether the use of HUDs with or without supporting tools improved the correct classification of patients to disease (heart failure) or no disease (not heart failure) (42). NRI combines *NRI event* and *NRI non-event. NRI event* shows the net proportion of correct HF diagnosis reclassification in those with the event (i.e., HF stated by reference examination) and *NRI non-event* shows the net proportion of correct no-HF diagnosis reclassification (in those without events). The sum of the net proportions of correct reclassification with and without the event is the *overall NRI.* The agreement between the stages of evaluation was treated as paired nominal data and calculated by McNemar's test. The positive predictive value (PPV) was calculated as the true positive test results divided by the true positive in addition to the false positive test results. Similarly, the negative predictive value (NPV) was calculated as the true negative results divided by the true negative in addition to the false negative test results, respectively.

Ethics

There are no harmful effects of diagnostic ultrasound when properly applied for medical diagnostic purposes. When inexperienced users perform complex examinations with limited diagnostics training errors and overdiagnosis is inevitable. However, the overall results of adding ultrasound to the inexperienced clinician's assessment potentially leads to less errors in diagnosis, faster and more precise treatment (12, 21, 23, 43). No compromises on behalf of the patients during diagnostic work-up were made as all were examined with a thorough clinical evaluation and comprehensive echocardiography by experts (reference examination).

All participants gave their informed, written consent prior to inclusion. All data were handled confidentially and according to data protection regulations utilizing data minimizing strategies. Cardiac images were stored in the in-hospital imaging archive according to health care regulations. All participants were given a study identification number as ultrasound recordings were transferred and analysed on a limited number of computers secured by passwords. The coupling code was stored on the hospitals secure server separated from the data material. At the end of December 2027, the key to personal identification will be deleted and data will be anonymized. The study was performed in conformity with the Declaration of Helsinki. The studies were approved by the Regional Committee for Medical and Health Research Ethics (study I; REK 2015/2312, study II-IV; REK 2017/2054), by Nord-Trøndelag Health Trust, and registered in the ClinicalTrial.gov database (study I; NCT02936050, study II-IV; NCT03547076).

Results

Patient population 1 and 2

Population 1 consisted of previously diagnosed HF patients, while *Population 2* was referred due to suspected HF. For both populations the median age was approximately 70 years and women and men were equally represented. Overweight/obesity, reduced renal function, and atrial fibrillation was common in both populations. Table 4 shows the baseline characteristics of both study populations.

Table 4. Baseline characteristics of the study populations.

	Population 1	Population 2
	Study I	Study II-IV
Number	50	166
Age (years)	70 (74-85) [#]	73 (63-78)
Women, n (%)	23 (46)	78 (47)
BMI (kg/m^2)	25.5 ± 5.3	28.7 ± 5.3
Systolic blood pressure (mmHg)	131 ± 22	150 ± 22
Diastolic blood pressure (mmHg)	76 ± 11	83 ± 11
NT-proBNP (ng/L)	1881 (918-6090) [#]	295 (66-864)
eGFR (mL/min/1.73m^2) *	40 (25-54) [#]	89 (68-109)
NYHA functional class		
I, n (%)	12 (24)	63 (37)
II, n (%)	25 (50)	80 (47)
III, n (%)	7 (14)	12 (7)
IV, n (%)	1 (2)	1 (1)
Atrial fibrillation, n (%)	24 (48)	49 (29)
Diuretics, n (%)	42 (84)	41 (25)
Beta-blocker, n (%)	40 (80)	51 (31)
ACEI/ARB, n (%)	18 (36)	32 (19)
COPD/asthma, n (%)		26 (16)
Diabetes mellitus type 2, n (%)		23 (14)
Coronary artery disease, n (%)		19 (11)
Haemoglobin (g/dL)		14.4 ± 1.5

Table 4. Normally distributed data are expressed as mean ± SD, non-normally distributed data as median (interquartile range), and n (%) where appropriate. *Cockcroft-Gault equation. [#] Values have been recalculated compared to original publication as they deviated somewhat from normal distribution and is now shown as median (interquartile range). A comment to the editor was sent regarding the correction. *Abbreviations: ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; BMI, body mass index; NT-Pro-BNP, N-terminal pro B-type natriuretic peptide; eGFR, estimated glomerular filtration rate; COPD, chronic obstructive pulmonary disease; NYHA, New York Heart Association.*

Study I

Feasibility of tele-echocardiography

The time used for the complete examination by the cardiac nurses, transfer of data, and finalizing of the report was mean ± SD 1.32 ± 0.36 hours. The corresponding time used only for the acquisition of the recordings was 0.48 ± 0.25 hours. Complications in transfer was mainly due to computer crashes and downloading issues. The feasibility was high for most measurements by both the telemedical approach and the reference examination (Figure 3).

Figure 3. Feasibility of echocardiographic parameters in study I

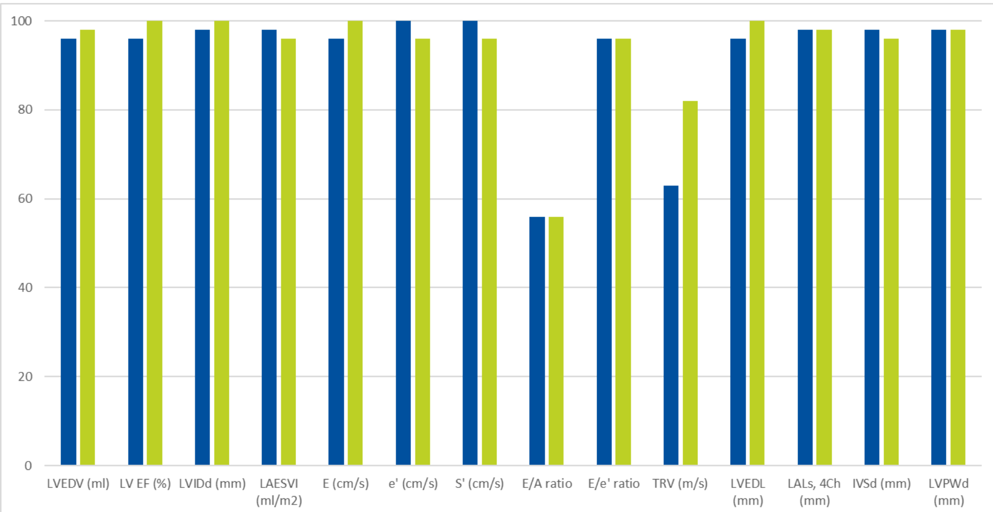


Figure 3. Feasibility of key measurements comparing the tele-echocardiographic method with recordings of cardiac nurse recordings interpreted by external cardiologist (blue), versus results from reference echocardiography (green). *Abbreviations: LVEDV, left ventricular end-diastolic volume; LV EF, left ventricular ejection fraction; LVIDd, left ventricular end-diastolic internal diameter; LAESVI, indexed left atrial end-systolic volume index; E, mitral early diastolic velocity; e', mitral annular early diastolic velocity; S', mitral annular systolic velocity; E/A ratio, mitral E/A ratio; E/e' ratio, mitral E/e' ratio; TRV, tricuspid regurgitation peak velocity; LVEDL, Left ventricular end-diastolic length; LALs, Left atrial end-systolic length; IVSd, Interventricular septum end-diastolic length; LVPWs, Left ventricular end-systolic posterior wall thickness.*

Agreement with reference

Table 5 shows the comparison of the different echocardiographic measurements by the telemedical approach and the reference cardiologists. Figure 4 illustrates the agreement between the telemedical approach and reference for selected measurements.

Table 5. Echocardiographic indices by the telemedical approach and reference cardiologist.

	N_{tele} / N_{ref}	Telemedical approach, mean (SD)	Reference echocardiography, mean (SD)	p-value for difference
LV end-diastolic volume (ml)	48 / 49	113 (39)	115 (49)	0.382
LV ejection fraction (%)	48 / 50	42 (12)	43 (11)	0.395
LV internal end-diastolic diameter (mm)	49 / 50	49 (8)	52 (10)	<0.001
LA end-systolic volume index (ml/m^2)	48# / 48	62 (19)	61 (22)	0.689
Mitral early diastolic velocity (cm/s)	48 / 50	85 (41)	84 (41)	0.544
Mitral annular early diastolic velocity (cm/s) *	50 / 48	6.1 (2.3)	5.7 (2.3)	0.052
Mitral annular systolic velocity (cm/s) *	50 / 48	4.8 (1.3)	4.8 (1.3)	0.621
Mitral E/A ratio	28 / 28	1.67 (1.22)	1.43 (0.91)	0.177
E/e' ratio *	48 / 48	16.9 (12.3)	15.9 (11.4)	0.562
Tricuspid regurgitation peak velocity (m/s)	32 / 41	2.65 (0.57)	2.79 (0.46)	0.018
LV end-diastolic length (mm)	48 / 50	84 (8)	79 (9)	<0.001
LA end-systolic length, 4Ch (mm)	49/ 49	6.2 (0.9)	6.2 (0.9)	0.641
IVS end-diastolic thickness (mm)	49 / 48	11.1 (3.0)	10.3 (2.7)	0.164
LV posterior wall end-diastolic thickness (mm)	49 / 49	11.1 (2.9)	10.2 (2.9)	0.013

Table 5. Comparison of echocardiographic measurements by the telemedical approach compared to reference measurements. *Mean of peak septal and lateral mitral annular velocities by pulsed-wave tissue Doppler. LV and LA lengths are mean of measurements from four- and two-chamber views. #The number differs from the original publication due to

typing error and notice is sent to the Editor. *Abbreviations: 4Ch, four-chamber view; E/A, early to late diastolic mitral inflow velocity; E/e', early diastolic mitral inflow velocity to mitral annular velocity.* N_{tele}; number of recordings available for telemedical evaluation, $N_{ref;}$ number of recordings available for reference evaluation. With permission from Journal of Ultrasound in Medicine under CC BY-NC license.

Figure 4. Agreement between methods illustrated in Bland-Altman plots of selected echocardiographic measures.

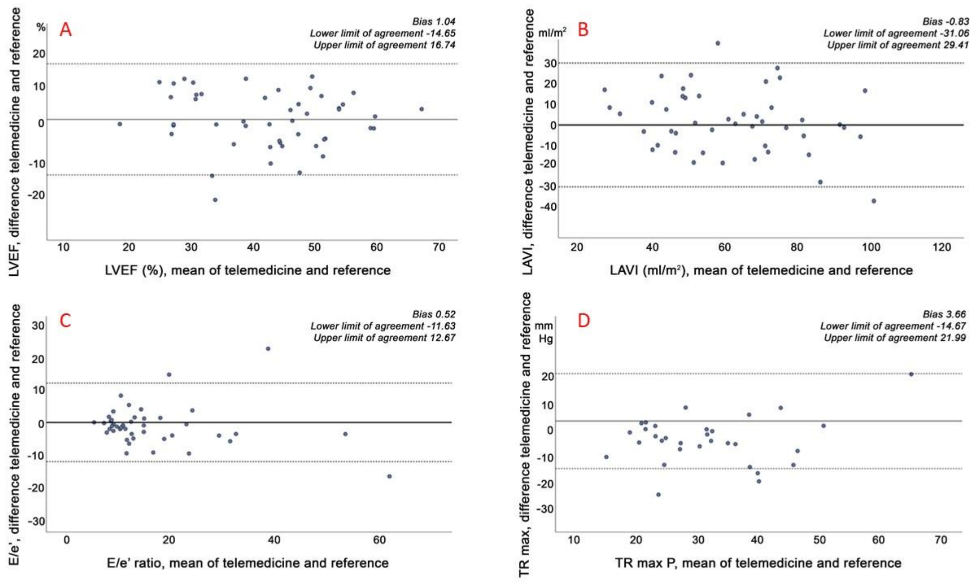


Figure 4. Agreement between the telemedical approach and reference echocardiography visualized by Bland-Altman plots of A) LV EF (%), B) LAVI (ml/m^2), C) E/e'-ratio and D) TR max P. The differences of the respective measures are plotted against the difference in the means of the measures. *Abbreviations: LVEF, left ventricular ejection fraction; LAVI, left atrial volume index; E/e', mitral early diastolic inflow velocity to mitral annular early; TR max P, tricuspid regurgitation maximal pressure gradient.* With permission from Journal of Ultrasound in Medicine under CC BY-NC license.

For the central measurements highlighted in Figure 4, the coefficients of variation ranged from to 6% to 15% and the biases were close to zero. There was no significant relation of the errors and the magnitude of the measurements. All echocardiographic measurements (Table 5) showed high correlations (r ≥0.71, p ≤0.007), except for wall thickness (interventricular septum and the posterior wall, with r 0.60 to 0.62, both p ≤0.03). The outliers shown in Figure 4 were mainly found in patients with severe cardiac pathology. Whether sub-optimal recordings by either group explain the differences within these cases are unknown.

Classification of HF according to LV EF showed substantial agreement between tele-echocardiography and reference (weighted kappa 0.73, p<0.001). No patients were misclassified between HFrEF and HFpEF. Correspondingly, 17 misclassifications between HFmrEF and HFrEF or HFpEF was found.

LV filling pressure was determined by tele-echocardiography in 39 cases and reference echocardiography in 41 cases. Figure 5 shows details of the 31 cases where the LV filling pressures were determined by both approaches. Disagreement between tele-echocardiography and reference was found in 7 cases, and among these HFpEF and HFmrEF were more prevalent.

Figure 5. Agreement between telemedical approach and reference for estimation of left ventricular filling pressures.

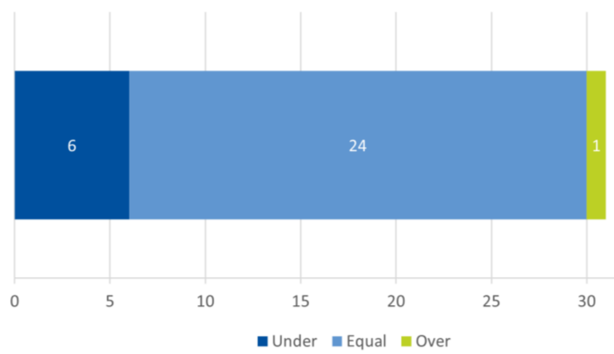


Figure 5. *Under*, *equal* and *over* refers to the telemedical classification of LV filling pressure compared to the reference echocardiography, where *under* indicates normal filling pressure by the telemedical approach and elevated filling pressure by reference, etc. With permission from Journal of Ultrasound in Medicine under CC BY-NC license.

Estimation of valvular pathology was done semi-quantitatively in a sub-analysis and was not a specific aim of the study. The telemedical approach detected at least moderate mitral stenosis, mitral regurgitation, and tricuspid regurgitation with a sensitivity of 100% and specificity of ≥95%. For detection of at least moderate aortic stenosis the sensitivity was lower (43%) but still with excellent specificity (97%). Moderate to severe valvular pathology was present only in a limited proportion of patients (mitral stenosis 4%, aortic stenosis 8%, mitral regurgitation 2%, tricuspid regurgitation 6%).

Study II

Feasibility

The time used for the focused cardiac ultrasound examinations was on average 18 ± 7 min for novices and 23 ± 7 min for the intermediate group. The time used for the six recordings with the automatic decision-support software was 4.6 ± 2.3 min, 3.3 min ± 1.9 min and 2.3 ± 1.3 min for the novices, intermediate group, and experts, respectively.

Figure 6 shows the proportion of successful recordings of four-chamber images with autoEF and autoMAPSE overlays by the three user groups and the corresponding feasibility defined as successful recordings of a four-chamber image with the automatic decision-support software overlay scored as acceptable for use (score ≥2) in the image analyses by external blinded cardiologist. There were significant differences between the groups where the best results were found in the most experienced users, and the differences were more pronounced when only considering the recordings accepted for use (score ≥2). These findings were consistent for both autoEF and autoMAPSE (all $p \leq 0.001$). With respect to the two versions of the autoEF algorithm the feasibility improved after the revision for all groups ranging from 41% to 68% for novices and from 77% to 91% for experts (Table 6).

Figure 6. Percentage of exams with autoEF and autoMAPSE recordings in the whole population and the sub-population with measurements accepted for use.

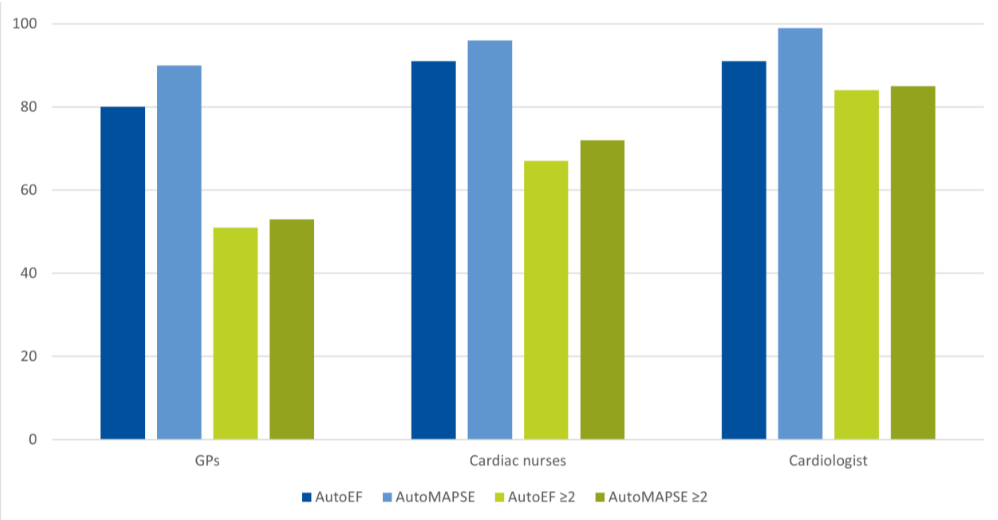


Figure 6. Blue coloured columns refer to autoEF and autoMAPSE recordings irrespective of whether they were recommended for use by the blinded analyses by the external cardiologist. Similarly, green coloured columns refer to the recordings accepted for use (score ≥2 in the blinded analyses). *Abbreviations: AutoEF, automatic measurement of left ventricular ejection fraction; AutoMAPSE, automatic measurement of mitral annular plane systolic excursion.*

Table 6. Feasibility of image recording with automatic decision-support software on hand-held ultrasound devices.

	Hand-held ultrasound operators		
	General practitioners	**Cardiac nurses**	**Cardiologists**
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%)

Table 6. Data are presented as numbers of feasible/available recordings (percent feasibility) according to user groups and decision-support software. A feasible recording was defined as score of ≥2 (i.e., accepted with or without need for adjustments in blinded evaluation by the external cardiologist). *Abbreviations: AutoEF, automatic measurement of left ventricular ejection fraction; AutoMAPSE, automatic measurement of mitral annular plane systolic excursion.* With permission from BMJ Open under CC BY-NC license.

Agreement and reliability

Table 7 and figures 7 and 8 show the agreement of the automatic decision-support software by the different users compared to reference. The agreement between the automatic decision-support software and reference was poor too modest for all user groups, even though, as for feasibility, a similar gradient across user groups was found. This is shown by the large coefficients of variation and coefficients of repeatability. The differences between operators were modest. For autoEF the minimal detectable change (calculated from the coefficients of repeatability) ranged 24.2-21.5%, and the corresponding values for autoMAPSE were 5.0-4.1 mm, respectively. After revision of the autoEF software the minimal detectable change improved slightly but remained around 20%.

The agreement was poor to moderate overall. There was no association of the size of the measurement, but the limits of agreement were narrower for the experts.

Table 7. Agreement of left ventricular function measurements by automatic decision-support software and reference by user groups.

	Hand-held ultrasound operator			
	General practitioners	Cardiac nurses	Cardiologist	Reference echocardiography
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)				
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	-

Table 7. The presented data relate to analyses utilizing all available feasible measurements in images recorded by GPs, cardiac nurses, and cardiologists compared to the reference measurements. *%-points. *Abbreviations: autoEF, automatic measurement of left ventricular ejection fraction; autoMAPSE, automatic measurement of mitral annular plane systolic excursion; GP, general practitioner.*

Figure 7. Bland-Altman plots of the agreement between HUD and reference by user groups in all available measurements.

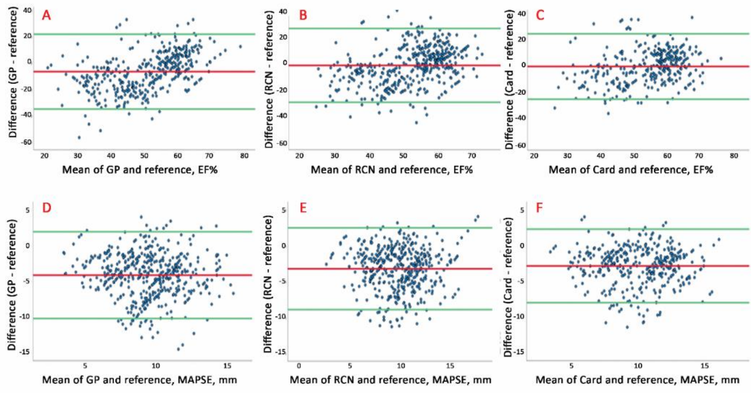


Figure 7. Bland-Altman plots for autoEF (upper plots) and autoMAPSE (lower plots) by GPs (left), RCNs (middle), and cardiologists (right) with no restriction to whether the automatic measurement was judged as feasible. *Abbreviations; Card, cardiologist; EF, left ventricular ejection fraction; GP, general practitioner; MAPSE, mitral annular plane systolic excursion; RCN, registered cardiac nurse.* With permission from BMJ Open under CC BY-NC license.

Figure 8. Bland-Altman plots of the agreement between HUD and reference by user groups in measurements judged as feasible.

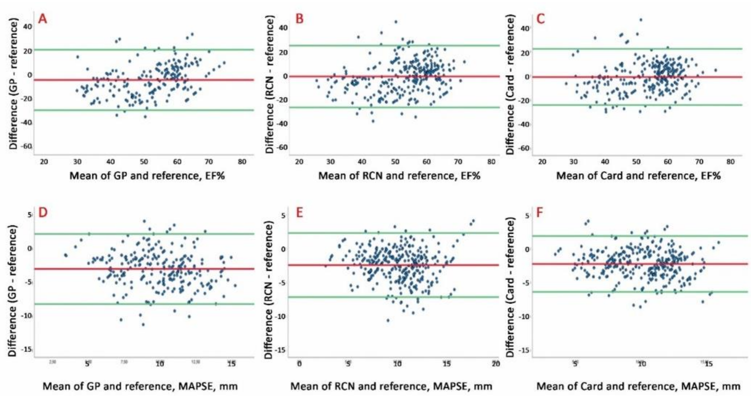


Figure 8. Presentation and abbreviations as in Figure 7 except that only measurements classified as feasible. With permission from BMJ Open under CC BY-NC license.

Intra-rater ICCs were moderate for all user groups (<0.75) except for autoMAPSE by the intermediate group and experts (≥0.83) shown in Table 8. The intra-rater ICC was as expected highest for experts for both automatic decision-support software. The inter-rater ICC was poor (≤0.51) for all users for both autoEF and autoMAPSE, ranging 0.43 to 0.51 for autoEF, and 0.35 to 0.51 for autoMAPSE for all three user groups.

Table 8. Intra- and inter-rater reliability of automatic measurements of left ventricular function by hand-held ultrasound devices according to operator.

	Hand-held ultrasound operator		
	General practitioners	**Cardiac nurses**	**Cardiologist**
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

Table 8. *Intraclass correlation was calculated with only two repeated measures as only 38 and 50 patients examined by the GPs had three repeated measures of autoEF and autoMAPSE, respectively. *Abbreviations: autoEF; automatic measurement of left ventricular ejection fraction, autoMAPSE; automatic measurement of mitral annular plane systolic excursion, GP; general practitioner, ICC; intraclass correlation.* With permission from BMJ Open under CC BY-NC license.

Agreement between two external cardiologists' evaluation of images

A second blinded external cardiologist evaluated a random sample of 771 images (548 with autoEF and 174 with autoMAPSE) and scored them as described previously. The agreement between the cardiologist was good with mean ± SD average scores by the second/first cardiologists 3.5 ± 0.9/3.3 ± 1.2 in GPs' recordings, 4.0 ± 0.9/4.0 ± 1.6 in nurses' recordings, and 4.5 ± 0.8/4.5 ± 1.4 in the cardiologists' recordings, respectively. Using image quality scores by the second cardiologist did not alter the results (data not shown).

Study III

Feasibility and image quality

Image quality was assessed in 2543 images (Table 9). There was a gradient across user groups for averaged and specific image quality scores being highest for experts and lowest for the novices. This finding was constituent across images recorded for both automatic decision-support software. The subcategories of *LV alignment* showed the highest image quality score, while the *mitral annular assessment* showed lowest, with the same gradient across user groups for both autoEF and autoMAPSE recordings.

In univariate logistic regression analyses including all recordings independent of user groups we found that the five prespecified image quality parameter scores and the average score were associated with the feasibility for both automatic decision-support software (all p<0.001). The feasibility of autoEF was significantly associated with all five prespecified image score parameters in multivariate analyses (all p <0.001, except four chamber view (p=0.02)). For autoMAPSE two prespecified parameters were not associated with the feasibility (*mispositioning of apex* (p=0.94) and *number of visible LV endocardial segments* (p=0.06)), while the others showed significant association (*four-chamber view*; p=0.046, *others* p <0.001).

Table 9. All recordings of both automatic decision-support software per user and the average scores per parameter.

AutoEF			
	General practitioners	Cardiac nurses	Cardiologists
Images, n	403	445	360
Total average score	3.7 ± 0.9	4.2 ± 0.9	4.8 ± 0.8
Four-chamber view	3.9 ± 1.7	4.7 ± 1.7	5.1 ± 1.5
LV alignment	4.7 ± 1.3	5.5 ± 1.0	5.6 ± 0.8
Mispositioning of apex	3.6 ± 1.5	3.9 ± 1.5	4.5 ± 1.4
Mitral annular assessment	2.8 ± 1.0	3.1 ± 1.1	3.5 ± 1.0
Number of visible LV endocardial segments	3.5 ± 1.3	3.8 ± 1.2	4.2 ± 1.2

AutoMAPSE			
	General practitioners	Cardiac nurses	Cardiologists
Images, n	476	470	389
Total average score	3.2 ± 1.9	3.7 ± 0.9	4.4 ± 0.8
Four chamber view	3.3 ± 1.6	4.1 ± 1.8	4.9 ± 1.5
LV alignment	4.4 ± 1.5	5.1 ± 1.2	5.5 ± 0.9
Mispositioning of apex	3.0 ± 1.3	3.3 ± 1.4	4.0 ± 1.4
Mitral annular assessment	2.5 ± 0.9	2.9 ± 1.0	3.5 ± 1.0
Number of visible LV endocardial segments	2.8 ± 1.3	3.2 ± 1.3	3.9 ± 1.3

Table 9. Numbers of evaluated images and mean ± SD image quality scores of the recordings with decision-support software overlay according to user groups. *Abbreviations: autoEF; automatic measurement of LV ejection fraction, autoMAPSE; automatic measurement of mitral annular plane systolic excursion, LV; left ventricular.* With permission from Open Heart under CC BY-NC license.

The feasibility of autoMAPSE was more affected by image quality than autoEF (Table 10). The explained variances (R^2) in feasibility of autoMAPSE by image quality ranged from 41% for novices to 22% for experts. A similar pattern was not seen for autoEF. In the intermediate and expert groups, the feasibility of autoEF was significantly explained by the *numbers of visible LV endocardial segments* (R^2 13-14%). Similarly, for autoMAPSE *mitral annular assessment* explained most of the variability with R^2 33%, 32%, and 20% in the novice, intermediate, and expert group, respectively. Comparison of the first and second version of the autoEF software did not reveal clinically meaningful changes.

In analyses merging all user groups patient characteristics were only weakly associated with image quality score of both autoEF and autoMAPSE recordings, shown by explained variances by body mass index (BMI) and systolic blood pressure (both $R^2 \leq 4\%$, $p \leq 0.04$). The association of BMI with image quality was stronger in the expert group (R^2 12% and 9% for autoEF and autoMAPSE, both $p < 0.001$), while systolic blood pressure was not associated with image quality among the experts (p ≥0.42). The associations of systolic blood pressure with image quality for novices and intermediate experienced users were weak (R^2 ≤2%, p ≤0.06) and known hypertension and COPD showed no significant associations with image quality (R^2 <1%, p >0.09).

Table 10. The importance of components of the image quality score for feasibility of automatic decision-support software.

	Novice	Intermediate	Expert
Average image quality	autoEF 20% autoMAPSE 41%	autoEF 18% autoMAPSE 37%	autoEF 24% autoMAPSE 22%
Four chamber view	autoMAPSE 13%		
LV alignment	autoEF 13% autoMAPSE 17%	autoEF 8% autoMAPSE 10%	
Mispositioning of apex			autoEF 2%
Mitral annular assessment	autoEF 10% autoMAPSE 33%	autoMAPSE 32%	autoMAPSE 20%
Number of visible LV endocardial segments		autoEF 13%	autoEF 14%

Table 10. The explained variance (R^2) of the importance of different image score parameters for the feasibility of autoEF and autoMAPSE is presented. Only significant associations are shown. In multivariate regression analyses including all five prespecified image quality scores, all the presented associations were still significant with p ≤0.02. *Abbreviations: autoEF, automatic measurement of LV ejection fraction; LV, left ventricle; autoMAPSE, automatic measurement of mitral annular plane systolic excursion.* With permission from Open Heart under CC BY-NC license.

Image quality and agreement with reference for autoEF and autoMAPSE

Image quality was only weakly associated with the agreement of measurements by automatic decision-support software on HUDs and reference imaging (Figure 9). Only 2% of the variability (R^2) between methods was explained by image quality when comparing all autoEF measurements with reference. The findings were quite similar across all user groups, but as shown by the regression lines the effect was even less in experienced users. For autoMAPSE recordings only 7% of the variance compared to reference was explained by image quality. The distribution across groups was more pronounced compared to autoEF, with R^2 being 7%, 5%, and 1% for the novices, intermediate group, and experts, respectively.

Figure 9. Agreement between automatic measurements on hand-held ultrasound devices and reference according to image quality.

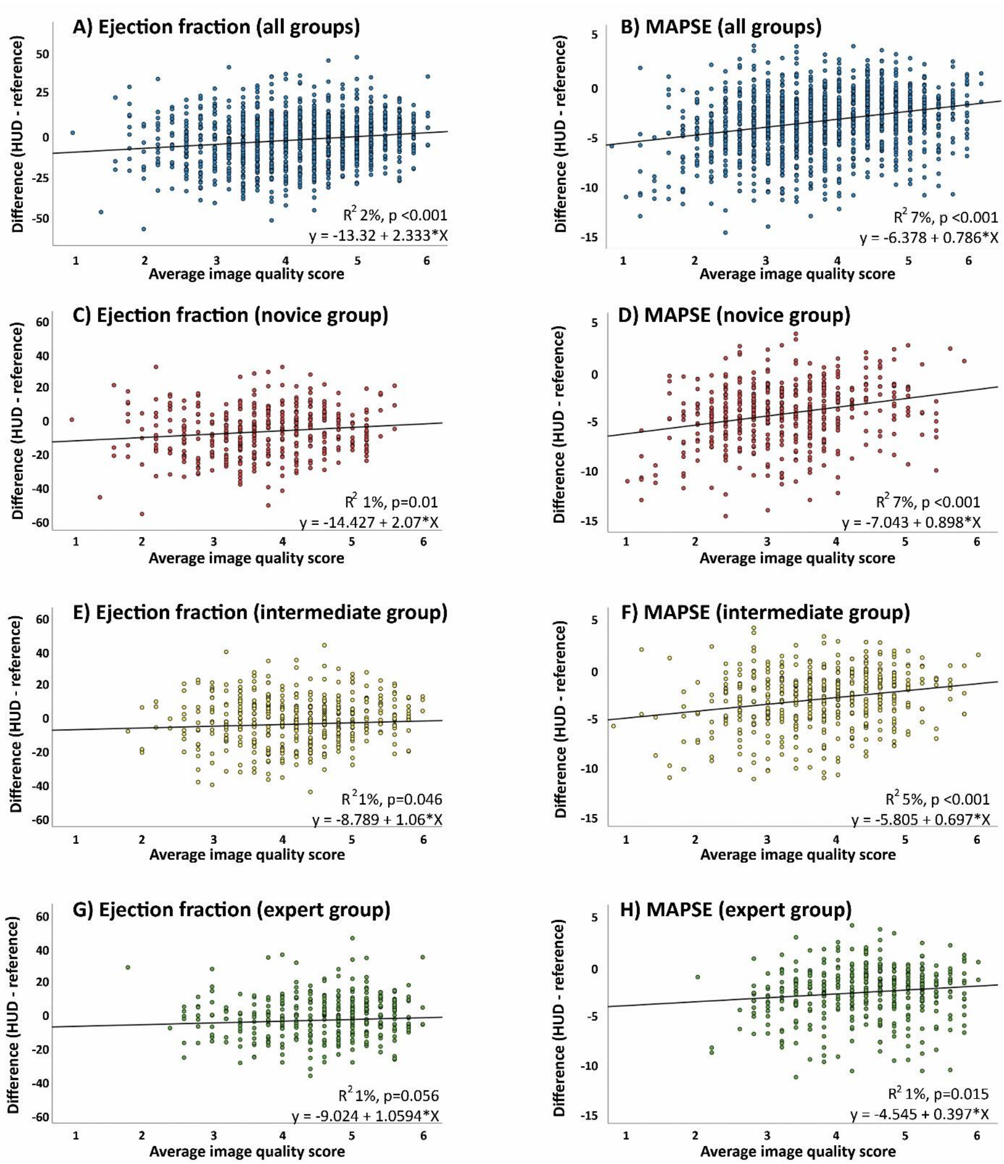
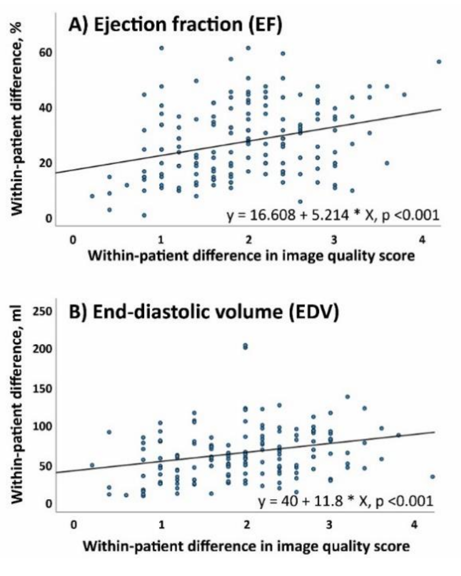


Figure 9. Scatter plots with *line of best fit* for the agreement of autoEF and autoMAPSE measurements with reference according to user groups (y-axis) plotted against the average image quality score (x-axis). Explained variances are shown by the R^2. The equation for the *line of best fit* is showed as well. *Abbreviations: HUD, hand-held ultrasound device; MAPSE, mitral annular plane systolic excursion.* With permission from Open Heart under CC BY-NC license.

Image quality and the reliability of automatic decision-support software measurements

There were significant associations of lower reliability with larger difference in image quality scores of the repeated recordings (all p ≤0.005). Figure 10 shows data when all three user groups were merged on the level of the individual participants. Analysing the three user groups separately, the reliability was not significantly associated with image quality for either LV EF, EDV or MAPSE in experts (all p ≥0.16). For the novices and the intermediate group, the association of reliability with within-patient differences in image quality was significant for LV EF, but not for EDV and MAPSE (both p ≥0.051).

Figure 10. Within-patient differences in automatic measurements plotted against the within-patient differences in image quality on hand-held ultrasound recordings.



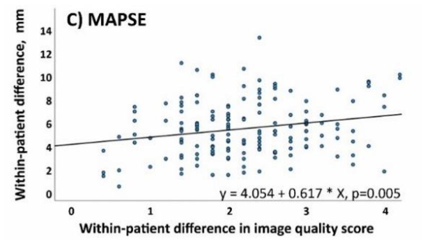


Figure 10. Within-patient differences in image quality score (x-axis) plotted against the within-patient difference in automatic measurements of A) EF, B) EDV, and C) MAPSE (y-axis). The lines of best fit have been applied and calculated by linear regression analyses. Abbreviations: EF, ejection fraction; EDV, end-diastolic volume; MAPSE, mitral annular plane systolic excursion.

Study IV

Of the 166 patients included, 28 patients (13 with HFpEF) were diagnosed with HF and 130 had HF excluded by reference examination. In eight patients the diagnosis remained uncertain. Figure 11 illustrates the distribution of correct and incorrect diagnosed patients by the different steps of the GPs' evaluation of the patients. As shown by the heights of the columns the numbers of uncertain cases were lowest when the diagnoses were based on HUD examinations without automatic measurements and by HUD examinations with telemedical support. The GPs correctly classified 92 (55%) patients (15 with and 77 without HF) by clinical evaluation only, increasing to 118 (71%) after HUD examinations (19 with and 99 without HF), and further improving slightly to 123 (74%) correctly classified after feedback from telemedical evaluation (20 with, 103 without HF) (difference from clinical examination; both p<0.001, non-significant difference for HUD without support and telemedicine; p=0.44). The GPs failed to run autoEF and autoMAPSE due to suboptimal recordings in 34 and 16 cases, respectively. Using the automatic decision-support software did not improve the diagnostic precision compared to clinical examinations (55% and 57% correctly classified after adding autoEF and autoMAPSE, respectively).

Figure 11. Diagnostic precision of ruling in and out heart failure by general practitioners.

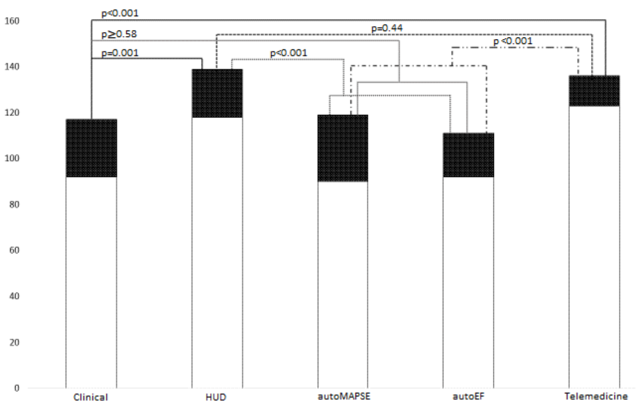


Figure 11. The numbers of patients with correct and incorrect classification according to the stages of the GPs examination are shown. Uncertain cases are shown by the total number of 166 minus the numbers each column. Correct classifications of *HF* or *no HF* (white), incorrect classifications (black), respectively. *Abbreviations: autoEF; automatic quantification of left ventricular ejection fraction, autoMAPSE automatic quantification of*

mitral annular plane systolic excursion; HUD, hand-held ultrasound device. With permission from Ultrasound in Medicine and Biology under CC BY-NC license.

After the clinical assessment the GPs were uncertain about the diagnosis of 43 patients. Using HUD (without decision-support software) and HUD examinations with telemedicine the number of uncertain cases were significantly decreased (20 and 24, respectively, both $p<0.05$), but the difference was not significant between the two ($p=0.44$). With use of the automatic decision-support software there was no significant reduction of uncertain cases (autoEF and autoMAPSE with 40 and 36 uncertain cases, respectively).

The diagnostic yield of the stages including HUD alone and combined with telemedical approach showed NRI of 0.10 and 0.19 respectively. The data were based on 3.6% and 12.0% correct reclassification of patients with HF by HUD and telemedicine, respectively (NRI_{events}), and 6.5% (both HUD and telemedicine) correct reclassification of non-events (Figure 12).

Figure 12. Reclassification to heart failure or healthy according to diagnostic stages.

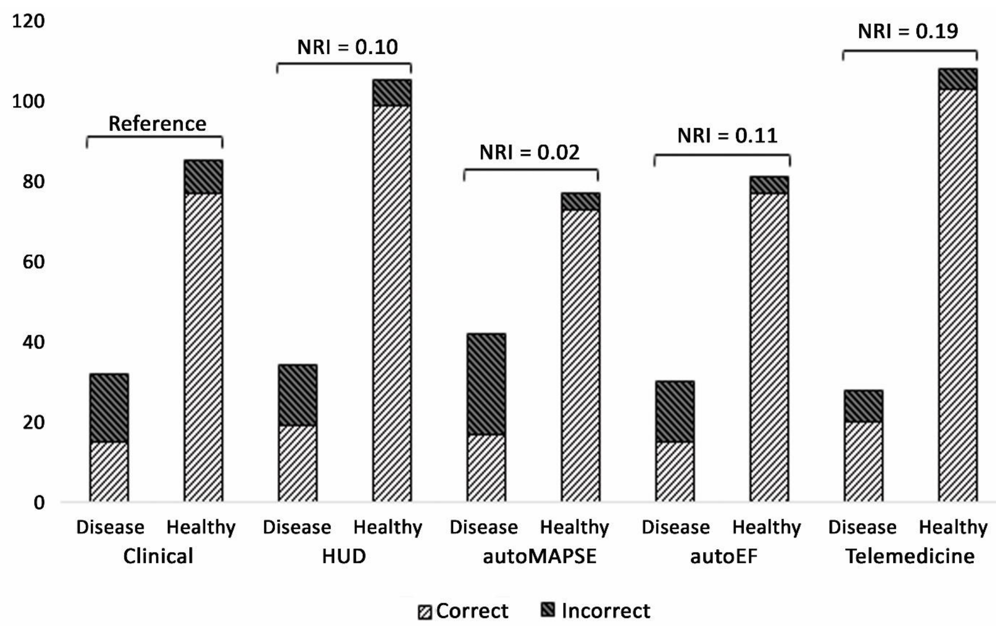


Figure 12. Numbers of correct and incorrect classification of patients with heart failure diagnosis dichotomized to "yes" or "no" according to the different stages of the general practitioners' examination. Net reclassification improvement is shown compared to clinical examination per stage. *Abbreviations: autoEF; automatic quantification of left ventricular ejection fraction, autoMAPSE automatic quantification of mitral annular plane systolic excursion, HUD; hand-held ultrasound device. NRI, net reclassification improvement.* With permission from Ultrasound in Medicine and Biology under CC BY-NC license.

The diagnostic decision made by the GPs compared to the reference cardiologist is shown in Table 11. PPV was highest for telemedicine and lowest for autoMAPSE (0.71 and 0.40, respectively). The NPV was >0.91 at all diagnostic stages. The revision of the autoEF software did not significantly alter the results (PPV 0.58 after revision vs. 0.38 and NPV 0.96 after revision vs. 0.86, respectively).

Table 11. General practitioners' clinical decisions compared to reference.

	HF positive	HF negative	False positive	False negative	PPV	NPV
Diagnostic stages						
Clinical	35 (21)	88 (53)	17 (10)	8 (5)	0.47	0.91
HUD	36 (22)	110 (66)	15 (9)	6 (4)	0.56	0.94
AutoMAPSE	45 (27)	81 (49)	25 (15)	4 (2)	0.40	0.95
AutoEF	31 (19)	85 (51)	15 (9)	4 (2)	0.50	0.95
Telemedicine	30 (18)	112 (67)	8 (5)	6 (4)	0.71	0.94

Table 11. Data are provided as numbers (percentages), except for PPV and NPV which are presented as numbers only. Uncertain cases were not included. *Abbreviations: autoEF, automatic ejection fraction; autoMAPSE, automatic mitral annular plane systolic excursion; HF, heart failure; HUD, hand-held ultrasound device. PPV, positive predictive value; NPV, negative predictive value.* With permission from Ultrasound in Medicine and Biology under CC BY-NC license.

Referral of patients

The GPs were asked to state whether they would refer the patients for a cardiac examination by a specialist or not after each stage of the examination, except for the first day of inclusion due to an error in the questionnaire. The GPs intended to refer 113 (68%) patients after clinical examination, even though suspecting HF in only 35 patients. In the remaining 78 (69%) of referred patients, referral was due to other reasons (e.g., atrial fibrillation, heart murmurs, suspected coronary artery disease, etc). The number of referrals in total decreased significantly after addition of HUD, analyses by the automatic decision-support software, and telemedical feedback by expert (all $p < 0.02$). Further, the proportions of referrals based on suspected HF were non-significantly reduced after adding HUD alone, HUD combined with autoEF, and HUD combined with autoMAPSE (31, 35, and 26 patients, respectively). Adding telemedical support reduced the number of referrals with respect to HF (referral of 23, $p=0.02$). Interestingly, out of the latter 23 patients the reference cardiologist diagnosed HF in 21 (91%). Seven of a total of eight patients being false negative after clinical evaluation were referred for cardiac evaluation of other reasons than HF.

Discussion

The main findings this PhD work showed the following: 1) The telemedical approach with limited echocardiographic recordings by non-experts and subsequent interpretation by a remote cardiologist was feasible and accurate with results in line with the in-hospital cardiologists' reference measurements. 2) The feasibility of GPs, cardiac nurses, and cardiologists with respect to record four-chamber views and utilize automatic decision-support software on HUDs was only modest, with a gradient from the least to the most experienced users. Further, the poor to modest agreement and reliability of the automatic measurements suggest a need for further refinement and clinical evaluation before implementation into clinical practice. 3) The operators experience influenced image quality and the feasibility of HUD recordings. Image quality explained approximately 20% of the variability in feasibility for autoEF across user groups, while for autoMAPSE there was a difference across groups with explained variability 41% in GPs and 22% in cardiologist. Additionally, image quality did not explain the low test-retest reliability of the decision-support software measurements among GPs, specialized nurses, and cardiologists. 4) In patients referred to cardiac consultation with suspected HF, GPs improved their diagnostic precision by adding HUD examination to the clinical examination while the highest NRI was found when the HUD examinations were supported by an external cardiologist via telemedicine. The results also show that the GPs had no significant benefit of the automatic decision-support software.

General discussion

Training and the use of cardiac diagnostic ultrasound

Training of inexperienced users within cardiac ultrasound must be focused and with a clear and predetermined goal to achieve reliable results. In study I, the cardiac nurses had not performed echocardiography until training for this study but had years of experience evaluating volume status with HUDs and mobile ultrasound scanners (12, 44). Their training involved one-to-one teaching by cardiologists experienced in echocardiography. No formally recommended training protocol was followed, but all views and parameters were predetermined. The cardiac nurses performed only dedicated recordings without interpretation or quantitative analyses, which differs from the training of cardiology fellows and sonographers. The training in limited echocardiography was beyond the scope of *focused cardiac ultrasound*, but not as extensive as for standard echocardiography (13). In study II-IV, the GPs time for training was adequate lasting three months prior to inclusion and included six one-to-one training days in addition to evening lectures. This training protocol was in line with previous studies (21, 22, 43), and as expected, asking the GPs directly they considered themselves prepared to start inclusion in the study. We believe that the training in these studies was adequate, providing a relevant level of skills which makes the results generalisable and realistic for an average clinical setting.

The aspect of personal enthusiasm for learning new skills is a common bias in research. Thus, the GPs were selected by the municipality administration based on their positions, and not by the research group or their motivation to learn cardiac diagnostic ultrasound. In our opinion this provides a realistic setting and is representable when discussing broad implementation into practice, even though the performance may have been improved by including enthusiasts only. On the other hand, the cardiac nurses had an above average interest in diagnostic ultrasound and had years of experience with focused ultrasound of the ICV and pleural cavities. The nurses' participation in study I, practicing a limited echocardiographic protocol at a HF clinic was a strong motivator for study II-III and is to our knowledge unique. Nurses in general, and most cardiac nurses, will most likely fall into the *novice* category in most cases, while the nurses included in this work were due to the training and years of experience categorized as intermediate experienced at start of study II-III.

All training and personal motivation aside, it takes years to become an experienced echocardiographer. However, evaluating experienced echocardiographers was not the purpose of this work. The need to support inexperienced and intermediate experienced users by decision-support software is increasing along with the ever-increasing availability of HUDs. The type of support offered in everyday clinical work depends on the access to specialists, technological development, innovations, and creativity. The presented studies show that the telemedical support provided useful support for the non-experts to make solid clinical evaluations quite well aligned with the reference examinations, while the benefit of the automatic decision-support software was negligible. The latter could not be fully explained by poor image quality and inexperienced users.

Study populations

Population 1 (study I) consisted of established HF patients enrolled from an outpatient HF clinic, while Population 2 (study II-IV) was selected to represent the typical patient group referred to cardiac consultation to rule in or out HF as seen in everyday clinical practice. The high rate of comorbidities such as atrial fibrillation, hypertension, COPD, and overweight and obesity was expected, as these comorbidities are commonly seen in HF patients and constitute relevant differential diagnoses and risk factors for HF (1). These comorbidities and patient characteristics may negatively affect the acoustic window leading to challenges in image acquisition as well as manual and automatic image analyses. Image quality is essential and poor image quality constitutes a major challenge within ultrasound diagnostics. BMI and systolic blood pressure were of limited clinical importance with respect to feasibility, agreement with reference, and reliability of measurements in this work.

The study populations were heterogeneous with a wide range of age leading to a broader distribution of relevant patient characteristics compared to more strictly selected patient samples (14, 38, 43). Arrhythmias, and in particular atrial fibrillation, may lead to unpredictable duration of cardiac cycles, which challenged the automatic decision-support software in study II-IV. Only 28 (17%) of the population included in study II-IV were diagnosed with HF. This highlights the indeterminate nature of signs and symptoms of HF, as well as the need to improve HF diagnostics outside the hospital setting. As seen by the results of study IV atrial fibrillation was common among misclassified patients (41). Even among experts, differentiating between HF and atrial fibrillation may be challenging, as the presence of tachycardia may alter LV volumes and functional indices even when HF is not present.

Study specific discussion

Study I

The telemedical approach with limited echocardiography by specialized nurses and telemedical interpretation by external cardiologists was feasible and reliable. The total examination time was approximately 1.5 hours from start to finish, including recording, transfer, interpretation, and feedback. The echocardiographic measurements and clinical conclusions of the telemedical approach were comparable to reference echocardiography. Only small differences in dimensions and blood flow measurements compared to reference measurements were seen, and the results were well in line with known test-retest variation for echocardiographic indices (45-47). The agreement for estimation of LV filling pressure was good with misclassification by the telemedical approach of 17% compared to reference. Outliers represented severe cardiac pathology which can be difficult to assess for less experienced users. Severe pulmonary hypertension, a degenerated aortic bioprosthesis, and hypertrophic cardiomyopathy may be overlooked during echocardiographic examination if the operator is not aware of the condition or the imaging details important for the diagnosis. The findings were also in line with a previous study evaluating the agreement in classification of diastolic function between echocardiographers in 105 participants (48). The evaluation of valvular pathology by HUDs has been found to be more challenging than evaluation of ventricular size and function, and the lack of quantitative tools on HUDs may partly explain the reduced diagnostic performance within valvular diseases (49). The latter is also shown in a previous publication by our group, where inexperienced users perform poorly when evaluating valvular pathology compared to evaluation of global LV function (21). A study comparing the accuracy for diagnosing cardiac pathology by medical students after brief training with HUDs compared to cardiologists without access to ultrasound (physical examination and auscultation only), showed superior accuracy for identifying valvular pathology by the medical students. This indicates the usefulness of FoCUS even when assessing the cardiac valves which is considered particularly difficult for novices (50). However, as the level of training needed for proper evaluation of valvular pathology is too extensive this was not included as an aim for this work.

The agreement with respect to correct classification of HF was substantial (kappa 0.73) between the telemedical approach and reference, and all inconsistencies were related to misclassification of HFmrEF vs. HFpEF or HFrEF. The HFmrEF phenotype is controversial, and whether HF should be categorized based on LV EF is debated (51-53). Moreover, it has previously been shown that 40% of patients will move from HFmrEF to HFpEF or HFrEF phenotypes in repeated analyses (54).

The limited echocardiographic protocol included the views and parameters needed for assessment of a HF patient, allowing for HF subtype classification and estimation of LV filling pressure. The protocol could be categorized somewhere between recommendations for FoCUS (more thorough evaluation) and comprehensive echocardiography (less extensive) (13, 55). There are to the best of our knowledge no similar studies where non-experts perform limited echocardiography with expert's interpretation by telemedicine. In general, there are few studies where inexperienced personnel perform echocardiographic recordings with interpretation by experts (55, 56). The presented results confirmed previous studies aiming to expand telemedical solutions. Large efforts and resources are related to transportation costs vs. training of e.g., cardiac nurses. In our opinion, the presented results justify implementation of telemedical tools into everyday practice to support users and to overcome geographical challenges. The use of telemedical solutions accelerated during the COVID-19 pandemic and changed the way health care services can be offered (57). Reducing the time to a correct diagnosis is an urgent issue with increasing numbers of patients in need of specialized care. HF patients may benefit from diagnostic ultrasound even without seeing an expert, and this may improve both diagnostics and follow-up, and has the potential to shorten waiting times. It is important to highlight that the long-term aim for these studies was to expand the use of diagnostic ultrasound and not to replace comprehensive echocardiography so more patients may benefit from improved diagnostics in the future.

Study II

Feasibility and reliability of the automatic decision-support software

The time consumption of approximately 20 minutes for GPs and cardiac nurses to complete HUD examinations is considered acceptable for selected cases in everyday clinical practice. The feasibility of the automatic decision-support software was acceptable for the most experienced users, but still below what would be clinically acceptable. This even applied for images of decent quality looking at the blinded image quality score. The agreement with reference was poor to moderate and even for experts the agreement and reliability were barely within the ranges recommended for clinical use.

The intra- and inter-rater ICCs for novices and the intermediate experienced group were lower than what is recommended for clinical use, and even for experts the inter-rater ICCs were only 0.51 for autoEF and autoMAPSE. This indicates that user experience alone could not explain the modest inter- and intra-operator reliability. The agreement with reference was poor for autoEF and autoMAPSE for all users. Suboptimal image acquisitions by inexperienced users partially explains the poor agreement. It is important to emphasize that the agreement was suboptimal also in experts. The automatic quantification of LV EF has been evaluated in a few studies by experienced users (25, 58, 59). Interestingly, two of these studies published somewhat better agreement for autoEF compared with the presented work, but the results are not directly comparable. In these studies, the users were either cardiologist or a cardiology fellow trained in advanced echocardiography (25, 58). A third study evaluated the autoEF software on Vscan Extend HUDs by intensive care physicians.(59) They presented a kappa value of 1.0 for the agreement of LV function assessment! However, this study has several limitations and potential errors including poor methodological description, lack of presentation of the results, no difference between end-diastolic and end-systolic LV volumes, and for different echocardiographic tasks the kappa values ranged from 0.01 to 1.0.

For the autoMAPSE application, we know from recent studies by our group that the software underestimates MAPSE in recordings by HUDs compared to reference echocardiography (38, 60). Supplying motion-mode MAPSE with tissue Doppler mode has been shown to reduce the underestimation, but tissue Doppler mode is not available on most HUDs (37). These findings also highlight that cut-offs and references for pathology is not interchangeable between different methods. With respect to the GPs interpretation of autoMAPSE, they were presented with these results prior to study start and were told that due to the underestimation by the algorithm autoMAPSE would lead to a lower normal limit compared to the commonly used cut-off of ≤10mm (61).

The findings indicate that the automatic decision-support software cannot be implemented as reliable tools independently of the skills of the users. These findings should be taken into consideration when the next generation of automatic LV quantification software is developed.

Study III

When evaluating the importance of image quality and operators experience for automatic decision-support software measurements on HUD examinations by three user groups, the main findings were as follows: 1) Image quality was significantly associated with the feasibility of autoEF and autoMAPSE, and image quality explained more of the variance of feasibility among the less experienced compared to the more experienced user groups. Across the user groups, image quality explained 18-24% of the variability in the feasibility of autoEF vs. 22-41% of the variability for autoMAPSE. 2) The most important prespecified image quality category (out of five) for the feasibility of autoEF and volumes was the *numbers of visible LV endocardial segments*, while *mitral annular assessment* was most important for autoMAPSE. 3) The reliability of the automatic decision-support software was less influenced by image quality, shown by the very low R^2 and the finding of significant associations only for autoEF within the less experienced user groups.

Feasibility of the decision-support software and image quality

As expected, both automatic decision-support software showed higher feasibility when the image quality score was higher. Moreover, the less experience the users had, the more image quality influenced the feasibility of both autoEF and autoMAPSE. The feasibility of autoMAPSE was more closely related to image quality than autoEF, as seen by the explained variances of the feasibility which was almost twice as hight for autoMAPSE compared to autoEF. The differences found across user groups can partly be explained by higher image quality with less variation among the expert users. Importantly, the performance of decision-support software performance was not dependent on image quality alone.

As shown above, the image quality parameter explaining most of the variance in feasibility for autoEF was the *number of visible LV endocardial segments*, while the *mitral annular assessment* was most important for autoMAPSE. This corresponds to the requirements for LV EF and MAPSE in echocardiography.

Technological aspects of the automatic decision-support software may influence their feasibility. As the autoEF software is assisted by artificial intelligence the training of the algorithm seems not to have been adequate for the HUD recordings of this study (58). Accordingly, the autoMAPSE software utilized an underlying model of the left ventricle, which seems inadequate when the inexperienced users recorded the images. The well-known robustness of MAPSE has been shown for other methods than the one currently used, and the results are not directly transferrable across methods (62).

Image quality, agreement, and reliability

Compared to previous studies by experts using high-end echocardiography the coefficients of repeatability for the automatic decision-support software were much higher (ranging 19-24%) (39, 45). In a recent publication utilizing decision-support software for assessment of LV EF by a novel HUD the limits of agreement was 14.5% (25). In the latter study, image quality did not seem to significantly influence the agreement with reference (25). Exploring the unacceptably high variability in the present study we found that image quality could explain only a small part of the poor agreement. However, both autoEF and autoMAPSE underestimated LV measurements compared to reference when image quality was low.

It is important to highlight that even for experts where image quality was highest, the agreement with reference for autoMAPSE was below recommendations for clinical use. In addition, compared to reference autoMAPSE showed a linear relation of more underestimation with lower image quality for all user groups. Only 7% of the variability was explained by the image quality for autoMAPSE. In a recent publication by our group we showed a slight underestimation of autoMAPSE compared to reference, but it is important to notice that even if agreement was good and intra- and interrater ICCs were excellent the experts usually do not seek automatic decision-support tools when using HUDs (38). Further studies evaluating how less experienced users can record echocardiographic images of decent quality for the automatic decision-support software to function as intended are needed. The use of real-time guiding systems for probe positioning, with and without automatic decision-support tools, could be a step in the right direction (63).

As discussed in relation to study II, only a minor part of the test-retest reliability of the measurements by the automatic decision-support software was caused by image quality variations. These findings are in disfavour of clinical implementation of both the autoEF and autoMAPSE software before refinement and proper clinical evaluation has been performed.

Such a comprehensive evaluation of blinded image quality assessment of repeated HUD recordings and the influence of different aspects of image quality on agreement and reliability of automatic measurements of LV function has to the best of our knowledge not yet been performed. The method was further strengthened by testing the reproducibility of the blinded image quality assessment in a large sample of 771 recordings by another blinded cardiologist. Direct comparison to other studies (e.g., Papadopoulou et al. and Filipiak-Strzecka et al.) is difficult (25, 58). The mentioned two studies provide little information about image quality scoring, and they have only evaluated data from single operators with ultrasound experience (25, 58). In comparison, our data were based on findings from eight novices or intermediate experienced users, and five experts. Additional refinement and better training of the artificial intelligence algorithms used in the autoEF software may at least partly explain the differences between these studies, but this information is hidden for the readers (25, 58). However, more advanced supportive software is already on the horizon, including real-time feedback to the user on how to improve examination techniques with HUDs as well as automatic real-time measurements of LV volumes, LV EF, and deformation imaging (64, 65).

Study IV

The precision of a HF diagnosis based on a standard clinical examination is imprecise with low accuracy and high false positive rates (66, 67). Correspondingly, we found a PPV of only 0.47 for diagnosis of HF based on the clinical examination. The proportion of patients correctly diagnosed by GPs improved with ≥25% (with less false positive results) and uncertain cases was reduced by approximately 50% after adding focused cardiac ultrasound by HUD and telemedical support (41). Improving patient selection for referrals to cardiological evaluation allows for better utilization of health care resources. Relying solely NT-Pro-BNP for HF diagnosis is imprecise as the PPV is low (68). In a study of more than 500 patients presenting with shortness of breath over 65 years of age, only a quarter of those with pathologically elevated NT-pro-BNP had HF (11).

Adding FoCUS to the initial clinical evaluation by GPs in patients with suspected HF had diagnostic value in this study. A previous study from our research group showed that GPs could assess LV function by HUDs, and other studies evaluating users of varying experience have shown the benefits of adding HUD examinations to the clinical evaluations (27, 49, 69, 70). In contrast, adding automatic decision-support software for LV function did not improve the diagnostic precision. The automatic decision-support software seemed to somewhat confuse the GPs making the interpretation difficult. After adding telemedical support the referrals were reduced by 35%, and for those selected for further diagnostic workup >90% were diagnosed with HF by the reference cardiologist. Evangelista et al. showed the value of adding HUD examinations by GPs combined with telemedical interpretation and feedback from external experts in a non-selected population (49). The NPV for non-pathological findings was excellent (>97% for valvular pathology and LV dysfunction) and the reduction of referrals to comprehensive echocardiographic examination was reduced by 32% after HUD examination and telemedical feedback. However, the study lacks a systematic evaluation of image quality and only some of the HUD recordings were compared to comprehensive echocardiography, which can obscure the true results for those with actual HF.

Clinical influence of HUD examinations and diagnostic supportive tools

Previously, autoEF has been proven feasible for experienced users (25, 58) and autoMAPSE has been shown to be reliable on high-end equipment by experts (70, 71). To the best of our knowledge, automatic quantification of LV function on HUDs performed by inexperienced users have so far not been evaluated to the same extent as in study II-IV. The lack of diagnostic improvement by autoEF and autoMAPSE was explained by an increase in false positive results after autoMAPSE, a high proportion of uncertain cases for both autoEF and autoMAPSE, and no significant difference between the two methods was shown. After the revision of the autoEF software, we found no improvement in false positive or false negative cases. For autoMAPSE, the underestimation of MAPSE seems to be the main cause for the high false positive rate. As discussed previously related to study II this may be compensated by using tissue Doppler mode which is not yet available on HUDs (37, 70, 72). Both autoEF and autoMAPSE was underestimated compared to reference in this study (mean autoEF vs mean LV EF; 48% vs 53% and mean autoMAPSE vs mean MAPSE; 8 mm vs 12 mm). The underestimation by autoMAPSE was larger than previously shown (70, 72). This can partly explain the uncertainty and overdiagnosis by the GPs.

Further, atrial fibrillation makes HF diagnostics challenging in general when performing standard echocardiography due to the inconsistent cardiac cycle length. HFpEF and atrial fibrillation present clinical challenges alone, but the combination is particularly difficult due to overlapping symptoms and technical difficulties during echocardiography. Atrial fibrillation was present in 54% of all patients diagnosed with heart failure, and 62% in patients diagnosed with HFpEF. Atrial fibrillation may also have influenced the success of the automatic decision-support software.

Telemedical solutions in combination with echocardiography can make HF diagnostics more accessible and time to treatment can be reduced (73, 74). In this study the number of false positive and uncertain cases were reduced after feedback from the external cardiologist. The PPV (0.71) was highest for telemedical support, and the proportion of correct reclassification (NRI 0.19) compared to clinical evaluation and automatic quantification was significant. The small difference between the HUD examination interpreted by the GPs themselves and telemedicine is somewhat unexpected. However, not supported by the collected data it may seem like it was difficult to put expert report into clinical context. As shown, 13 of 28 patients were diagnosed with HFpEF. Further, most of the misclassified HF patients in the present study had HFpEF, which is a known pitfall in HF diagnosis even for experienced echocardiographers (75). Giving the external cardiologist access to the clinical information from the present exam could have improved the precision of the telemedical approach. Furthermore, improved image quality would likely improve the success of telemedical support as the feedback from the external cardiologists were highly dependent of the material interpreted.

Strengths and limitations

Study I

The main strength of the study was the complete blinding of the telemedical approach and the reference method. The number of patients included, and the number of cardiac nurses involved were modest. However, for the aim of the study the sample size was adequate. Still, this reduces the generalisability of the results and should be taken into consideration before applying the results beyond the aims of the study. Since echocardiography served as reference for the estimation of filling pressures and no invasive measurements were performed, we were only able to evaluate the agreement of the telemedical approach compared to reference as a surrogate for the actual filling pressures. The limited echocardiography by cardiac nurses is not equal to comprehensive echocardiography but was shown to add valuable information to clinical decision making. Since only the cardiologists performing the reference echocardiography (and not the cardiac nurses or the out-of-hospital cardiologist performing the interpretation by telemedicine) had access to previous echocardiograms it is unlikely that the previous examinations have influenced the results of the study.

Study II-IV

The main strength of these studies was the rigid design with complete blinding between all users. Another strength was the comprehensive blinded evaluation of image quality by a predefined five parameter score in attempt to better understand the performance of the automatic decision-support software. The verification of the robustness of the blinded evaluation of image quality by re-evaluating almost 800 images by another external cardiologist confirmed the consistency of the presented results across user groups. The sample size was adequate and based on power calculations, and further expanded to account for an expected low proportion with disease. Additionally, the sample size was further adjusted to compensate for the revision of the autoEF software. We believe the total sample size of patients and operators provide adequate power for all analyses.

For broad generalisability of the study results an even larger number of operators would have been optimal. Still, the number of examiners (five GPs, three cardiac nurses, and five cardiologists) was higher than in comparable studies (25, 58). Further, the users' level of experience ranged from inexperienced to expert level (most GPs were introduced to practical ultrasound diagnostic in preparation for the study, while all cardiologist had level III experience according to the American Society of Echocardiography (28)). However, with respect to the user specific results a larger group of operators would have been preferable to reduce potential bias. We believe the studies paint a realistic image of the performance of decision-support software across different levels of user experience. All available cardiac nurses previously trained in limited echocardiography and all cardiologists at Levanger Hospital were included as users while GPs were recruited by their current position in the municipality. No user was recruited based on motivation for attending the study, making the study representable and applicable for the average GPs experience level, and removed personal motivation and interest as a bias.

A current limitation when evaluating cardiac function and HF is that there is no gold standard for evaluation of LV function. Echocardiography performed by experienced cardiologists was therefore used as reference method, and this introduces a possibility for some variation also for the reference method. The diagnosis of HF includes signs and symptoms in addition to more objective findings of cardiac pathology which challenges the use of HF as an endpoint. The HF diagnosis may be particularly challenging when LV EF is preserved and when patients have high-rate atrial fibrillation combined with a mild or moderate reduction of LV EF (41). This remains a challenge in clinical work.

An important limitation regarding the evaluation of image quality was that only images with autoEF and autoMAPSE overlays were reviewed, meaning cases where image quality or cardiac cycle inconsistency did not allow for the software to run were excluded from the analyses. As a lower proportion of recordings were able to run the decision-support software for the less experienced groups, it can potentially influence the results between the groups. However, it may be expected that the image quality for the images not analysed had even lower quality than those included, and that the presented results may be biased towards underestimation of the impact of image quality particularly for the GPs. Due to logistic reasons the cardiologists did not perform HUD examinations on the first 29 participants. However, the feasibility in this groups was high and over 360 images by cardiologists were reviewed for both autoEF and autoMAPSE. As we detected an error in the first autoEF version during the blinded image quality analyses, this experience may have affected the results for the revised software as well. This issue is particularly relevant for the GPs with respect to placing *HF diagnosis* or *no HF diagnosis* in study IV. The error detected caused an irregularity between the shown endocardial traces and the measured LV EF, which was only detected by the blinded analyses and not by performing operators in the study. This caused additional challenges to the GPs when interpreting the results of the automatic decision-support software as their clinical experience led them to mistrust the results often leading to overdiagnosis rather than to trust the initial assessment with HUD without automatic decision-support software. One could argue that due to the autoEF software revision it would be beneficial to reanalyse all recordings with the revised version, but the decision was made to stay true to the original aim of the study. We did not reanalyse the data, and for the evaluation of the clinical impact of the autoEF software in study IV reanalysing LV EF months after the GPs evaluation of the patients would not represent clinical practice.

Other factors may have influenced the clinical decisions regarding HF diagnostics in study IV. First, there was a delay in the reporting of the in-hospital laboratory tests. This made the results from the blood samples unavailable to the GPs in most cases and NT-proBNP values of the present day were rarely available at the time of the GPs consultation. It is worth noticing that the rule-out threshold of <125 ng/ml has excellent reliability, and some HF referrals may have been avoided if the NT-pro-BNP values were available for the GPs to rule out HF (76). However, not having NT-pro-BNP available the same day reflects the everyday practice of GPs in Norway, and we do not expect significant changes if the laboratory results were available in time. Most patients in *population 2* (study II-IV) had moderately increased NT-pro-BNP which corresponds to higher uncertainty with respect to HF diagnosis (11, 77).

Secondly, more training in clinical decision making and image analyses, shorter and more consistent time intervals between the training period and inclusion days could potentially improve image quality and the gain from the telemedical feedback option.

Future aspects

Image analyses in general, and particularly echocardiography, is well suited for artificial intelligence (machine learning and deep learning algorithms) due to standardized views and multiple measurements. Normal reference ranges for most echocardiographic parameters exists. This in combination of enormous databases of images collected over time allows for large scale training (78-82). However, the quality of the databases and the reference annotations are mandatory for the development of reliable software. The technological developments discussed in the current work add valuable experience to the future development of HF diagnostics and follow-up. Recent development at Centre for Innovative Ultrasound Solutions, NTNU, has shown that real-time automatic image analyses by deep learning algorithms with continuous feedback to the user reduced the common error of *LV foreshortening* (where the imaging plane does not transect the centre of the LV) (63). Further, the use of automatic image analyses for measurements of global longitudinal strain reduces inter-observer variability (65). The UltraSight artificial intelligence-driven guidance software (UltraSight Inc.©, collaboration with GE ultrasound) and Caption Guidance (Caption Health Inc. ©, Brisbane, CA, USA) software allow inexperienced users to record images while the software gives real-time feedback on probe position utilizing deformation analysis and real-time feedback to the user on how to achieve images of better quality (83, 84). However, such feedback is not yet commercially available on HUDs and applications are currently available for research purposes only. Real-time feedback to the examiner on how to improve the image quality is thought to be of great value with these portable HUDs. This may also reduce common errors in image acquisition and improve the databases so that automatic image analyses could be optimized. Importantly, there is a continuous need to scientifically test and clinically validate all new methods before implementation into clinical practice. Even software revisions of different image analysis algorithms should be scientifically evaluated to ensure the quality of the software. This is not yet the common practice, but to it should be mandatory to guarantee that all tools used in clinical diagnostics and treatment of patients have optimal quality.

Conclusion

This work has evaluated the feasibility, agreement with reference measurements, and the clinical influence of novel decision-support software used by operators of varying level of experience to create new knowledge and possibilities to expand the use of ultrasound diagnostics, as well as highlighting important limitations.

Specific conclusion

Tele-echocardiography with limited echocardiographic data collection by inexperienced users and interpretation by experts was feasible and accurate for quantitative measurements of left-sided cardiac chambers size and function, HF classification, and LV filling pressures. The use of tele-echocardiography for assessment of HF patients at remote locations can potentially lead to improve diagnostics and follow-up.

Fully automated analyses of LV function on HUDs by autoEF and autoMAPSE showed modest feasibility with a gradient from the novices to the expert users. The automatic decision-support software showed agreement and reliability below what is commonly recommended for clinical use.

The image quality of HUD recordings was closely related to experience of the users, with the lowest image quality score among the GPs and highest score for cardiologists. However, neither image quality, user experience, nor patients' characteristics could explain the poor performance of the automatic decision-support software. Further refinement of the automatic decision-support software is warranted before clinical use.

HUD examinations alone by GPs and in combination with expert interpretation by telemedicine improved the diagnostic precision for GPs evaluation of patients with suspected HF compared to clinical evaluation alone. However, the use of HUD examinations with automatic decision-support software did not improve the diagnostic precision.

Clinical implications

The current work is a step in the right direction for improving the evaluation of novel technological solutions to simplify assessment of LV size and function in situations where experts are unavailable.

In line with previous studies, adding a focused cardiac ultrasound by GPs to the clinical evaluation increases the diagnostic precision indicating that implementation of ultrasound diagnostics outside the specialized echocardiographic laboratories can allow for more patients to benefit from diagnostic ultrasound.

Tele-echocardiography can be a valuable addition to diagnostic ultrasound if the users are sufficiently trained in recording of ultrasound images but less experienced in image analyses. This type of technology can improve the diagnostic process for patients by overcoming geographical challenges.

The automatic decision-support software evaluated in this work needs further refinement before large scale implementation into the clinical workflow. It is crucial to always evaluate novel medical technology with clinical studies before implementation into practice.

References

1. 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. European Heart Journal. 2021;42(36):3599-726.
2. Dunlay SM, Roger VL. Understanding the epidemic of heart failure: past, present, and future. Curr Heart Fail Rep. 2014;11(4):404-15.
3. Conrad N, Judge A, Tran J, Mohseni H, Hedgecott D, Crespillo AP, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. The Lancet. 2018;391(10120):572-80.
4. van Riet EE, Hoes AW, Wagenaar KP, Limburg A, Landman MA, Rutten FH. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. Eur J Heart Fail. 2016;18(3):242-52.
5. Blecker S, Paul M, Taksler G, Ogedegbe G, Katz S. Heart failure-associated hospitalizations in the United States. J Am Coll Cardiol. 2013;61(12):1259-67.
6. Ambrosy AP, Pang PS, Khan S, Konstam MA, Fonarow GC, Traver B, et al. Clinical course and predictive value of congestion during hospitalization in patients admitted for worsening signs and symptoms of heart failure with reduced ejection fraction: findings from the EVEREST trial. Eur Heart J. 2013;34(11):835-43.
7. Khan H, Greene SJ, Fonarow GC, Kalogeropoulos AP, Ambrosy AP, Maggioni AP, et al. Length of hospital stay and 30-day readmission following heart failure hospitalization: insights from the EVEREST trial. Eur J Heart Fail. 2015;17(10):1022-31.
8. Lam CS, Solomon SD. The middle child in heart failure: heart failure with mid-range ejection fraction (40-50%). Eur J Heart Fail. 2014;16(10):1049-55.
9. Caraballo C, Desai NR, Mulder H, Alhanti B, Wilson FP, Fiuzat M, et al. Clinical Implications of the New York Heart Association Classification. J Am Heart Assoc. 2019;8(23):e014240.
10. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1211-59.
11. van Riet EE, Hoes AW, Limburg A, Landman MA, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. Eur J Heart Fail. 2014;16(7):772-7.
12. Gundersen GH, Norekval TM, Haug HH, Skjetne K, Kleinau JO, Graven T, 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(1):29-34.
13. Cardim N, Dalen H, Voigt JU, Ionescu A, Price S, Neskovic AN, 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(3):245-52.
14. Hjorth-Hansen AK, Andersen GN, Graven T, Gundersen GH, Kleinau JO, Mjølstad OC, 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. Journal of Ultrasound in Medicine. 2020;39(12):2313-23.
15. Curie J, Curie P. Développement par compression de l'électricité polaire dans les cristaux hémièdres à faces inclinées. Bulletin de la Société minéralogique de France. 1880;3(4):90-3.
16. Inge Edler CHH. The use of ultrasonic reflectoscope for the continuous recording of the movements of heart walls. 1954.
17. Liv Hatle BAA, A Tromsdal. Non-invasive assessment of aortic stenosis by Doppler ultrasound. British Heart Journal. 1980;43: 284-92.
18. Kollmann C, Jenderka KV, Moran CM, Draghi F, Jimenez Diaz JF, Sande R. EFSUMB Clinical Safety Statement for Diagnostic Ultrasound - (2019 revision). Ultraschall Med. 2020;41(4):387-9.

19. Liebo MJ, Israel RL, Lillie EO, Smith MR, Rubenson DS, Topol EJ. Is pocket mobile echocardiography the next-generation stethoscope? A cross-sectional comparison of rapidly acquired images with standard transthoracic echocardiography. Ann Intern Med. 2011;155(1):33-8.
20. Andersen GN, Haugen BO, Graven T, Salvesen O, Mjolstad OC, Dalen H. Feasibility and reliability of point-of-care pocket-sized echocardiography. Eur J Echocardiogr. 2011;12(9):665-70.
21. Mjolstad OC, Andersen GN, Dalen H, Graven T, Skjetne K, Kleinau JO, et al. Feasibility and reliability of point-of-care pocket-size echocardiography performed by medical residents. Eur Heart J Cardiovasc Imaging. 2013;14(12):1195-202.
22. Galderisi M, Santoro A, Versiero M, Lomoriello VS, Esposito R, Raia R, et al. Improved cardiovascular diagnostic accuracy by pocket size imaging device in non-cardiologic outpatients: the NaUSiCa (Naples Ultrasound Stethoscope in Cardiology) study. Cardiovasc Ultrasound. 2010;8:51.
23. Andersen GN, Graven T, Skjetne K, Mjolstad OC, Kleinau JO, Olsen O, et al. Diagnostic influence of routine point-of-care pocket-size ultrasound examinations performed by medical residents. J Ultrasound Med. 2015;34(4):627-36.
24. Chamsi-Pasha MA, Sengupta PP, Zoghbi WA. Handheld Echocardiography: Current State and Future Perspectives. Circulation. 2017;136(22):2178-88.
25. Papadopoulou S-L, Sachpekidis V, Kantartzi V, Styliadis I, Nihoyannopoulos P. 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.
26. Neskovic AN, Edvardsen T, Galderisi M, Garbi M, Gullace G, Jurcut R, et al. Focus cardiac ultrasound: the European Association of Cardiovascular Imaging viewpoint. Eur Heart J Cardiovasc Imaging. 2014;15(9):956-60.
27. Mjolstad OC, Snare SR, Folkvord L, Helland F, Grimsmo A, Torp H, et al. Assessment of left ventricular function by GPs using pocket-sized ultrasound. Fam Pract. 2012;29(5):534-40.
28. Wiegers SE, Ryan T, Arrighi JA, Brown SM, Canaday B, Damp JB, et al. 2019 ACC/AHA/ASE Advanced Training Statement on Echocardiography (Revision of the 2003 ACC/AHA Clinical Competence Statement on Echocardiography): A Report of the ACC Competency Management Committee. Circulation: Cardiovascular Imaging. 2019;12(7).
29. Via G, Hussain A, Wells M, Reardon R, ElBarbary M, Noble VE, et al. International evidence-based recommendations for focused cardiac ultrasound. J Am Soc Echocardiogr. 2014;27(7):683.e1-.e33.
30. Moore CL CJ. Point-of-Care Ultrasonography. N Engl J Med. 2011.
31. Raikhelkar J, Raikhelkar JK. The impact of telemedicine in cardiac critical care. Crit Care Clin. 2015;31(2):305-17.
32. Jarvis L, Stanberry B. Teleradiology: threat or opportunity? Clin Radiol. 2005;60(8):840-5.
33. Grant B, Morgan GJ, McCrossan BA, Crealey GE, Sands AJ, Craig B, et al. Remote diagnosis of congenital heart disease: the impact of telemedicine. Arch Dis Child. 2010;95(4):276-80.
34. Graven T, Wahba A, Hammer AM, Sagen O, Olsen O, Skjetne K, et al. Focused ultrasound of the pleural cavities and the pericardium by nurses after cardiac surgery. Scand Cardiovasc J. 2015;49(1):56-63.
35. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF, 3rd, Dokainish H, Edvardsen T, et al. Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging. 2016;17(12):1321-60.
36. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, 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):1-39 e14.
37. Storve S, Grue JF, Samstad S, Dalen H, Haugen BO, Torp H. Realtime Automatic Assessment of Cardiac Function in Echocardiography. IEEE Trans Ultrason Ferroelectr Freq Control. 2016;63(3):358-68.

38. Magelssen MI, Palmer CL, Hjorth-Hansen A, Nilsen HO, Kiss G, Torp H, et al. Feasibility and Reliability of Automatic Quantitative Analyses of Mitral Annular Plane Systolic Excursion by Handheld Ultrasound Devices. Journal of Ultrasound in Medicine. 2020;40(2):341-50.
39. Hjorth-Hansen AK, Magelssen MI, Andersen GN, Graven T, Kleinau JO, Landstad B, 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(10).
40. Hjorth-Hansen AK, Magelssen MI, Andersen GN, Graven T, Kleinau JO, Landstad B, et al. User experience and image quality influence on performance of automated real-time quantification of left ventricular function by handheld ultrasound devices: a diagnostic accuracy study with data from general practitioners, nurses and cardiologists. Open Heart. 2022;9(2).
41. Magelssen MI, Hjorth-Hansen AK, Andersen GN, Graven T, Kleinau JO, Skjetne K, et al. Clinical Influence of Handheld Ultrasound, Supported by Automatic Quantification and Telemedicine, in Suspected Heart Failure. Ultrasound in Medicine & Biology. 2023.
42. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology. 2014;25(1):114-21.
43. Panoulas VF, Daigeler AL, Malaweera AS, Lota AS, Baskaran D, Rahman S, 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(4):323-30.
44. Dalen H, Gundersen GH, Skjetne K, Haug HH, Kleinau JO, Norekval TM, et al. Feasibility and reliability of pocket-size ultrasound examinations of the pleural cavities and vena cava inferior performed by nurses in an outpatient heart failure clinic. Eur J Cardiovasc Nurs. 2015;14(4):286-93.
45. Baron T, Berglund L, Hedin EM, Flachskampf FA. Test-retest reliability of new and conventional echocardiographic parameters of left ventricular systolic function. Clin Res Cardiol. 2019;108(4):355-65.
46. Karlsen S, Dahlslett T, Grenne B, Sjoli B, Smiseth O, Edvardsen T, et al. Global longitudinal strain is a more reproducible measure of left ventricular function than ejection fraction regardless of echocardiographic training. Cardiovasc Ultrasound. 2019;17(1):18.
47. Thorstensen A, Dalen H, Amundsen BH, Aase SA, Stoylen A. Reproducibility in echocardiographic assessment of the left ventricular global and regional function, the HUNT study. Eur J Echocardiogr. 2010;11(2):149-56.
48. Chapman CB, Ewer SM, Kelly AF, Jacobson KM, Leal MA, Rahko PS. Classification of left ventricular diastolic function using American Society of Echocardiography Guidelines: agreement among echocardiographers. Echocardiography. 2013;30(9):1022-31.
49. Evangelista A, Galuppo V, Mendez J, Evangelista L, Arpal L, Rubio C, et al. Hand-held cardiac ultrasound screening performed by family doctors with remote expert support interpretation. Heart. 2016;102(5):376-82.
50. Kobal SL, Trento L, Baharami S, Tolstrup K, Naqvi TZ, Cercek B, et al. Comparison of effectiveness of hand-carried ultrasound to bedside cardiovascular physical examination. Am J Cardiol. 2005;96(7):1002-6.
51. Manisty CH, Francis DP. Ejection fraction: a measure of desperation? Heart. 2008;94(4):400-1.
52. Pellikka PA, She L, Holly TA, Lin G, Varadarajan P, Pai RG, et al. Variability in Ejection Fraction Measured By Echocardiography, Gated Single-Photon Emission Computed Tomography, and Cardiac Magnetic Resonance in Patients With Coronary Artery Disease and Left Ventricular Dysfunction. JAMA Netw Open. 2018;1(4):e181456.
53. Wehner GJ, Jing L, Haggerty CM, Suever JD, Leader JB, Hartzel DN, et al. Routinely reported ejection fraction and mortality in clinical practice: where does the nadir of risk lie? Eur Heart J. 2020;41(12):1249-57.
54. Shah AM, Cheng S, Skali H, Wu J, Mangion JR, Kitzman D, et al. Rationale and design of a multicenter echocardiographic study to assess the relationship between cardiac structure and

function and heart failure risk in a biracial cohort of community-dwelling elderly persons: the Atherosclerosis Risk in Communities study. Circ Cardiovasc Imaging. 2014;7(1):173-81.
55. Sengupta PP, Narula N, Modesto K, Doukky R, Doherty S, Soble J, et al. Feasibility of intercity and trans-Atlantic telerobotic remote ultrasound: assessment facilitated by a nondedicated bandwidth connection. JACC Cardiovasc Imaging. 2014;7(8):804-9.
56. Boman K, Olofsson M, Berggren P, Sengupta PP, Narula J. Robot-assisted remote echocardiographic examination and teleconsultation: a randomized comparison of time to diagnosis with standard of care referral approach. JACC Cardiovasc Imaging. 2014;7(8):799-803.
57. Ghilencea LN, Chiru MR, Stolcova M, Spiridon G, Manea LM, Stanescu AA, et al. Telemedicine: Benefits for Cardiovascular Patients in the COVID-19 Era. Front Cardiovasc Med. 2022;9:868635.
58. Filipiak-Strzecka D, Kasprzak JD, Wejner-Mik P, Szymczyk E, Wdowiak-Okrojek K, Lipiec P. Artificial Intelligence-Powered Measurement of Left Ventricular Ejection Fraction Using a Handheld Ultrasound Device. Ultrasound Med Biol. 2021;47(4):1120-5.
59. Mahneshwarappa HM MS, Kulkarni AV, GUnaseelan V, Kanchi M. Use of Handheld Ultrasound Device with Artificial Intelligence for Evaluation of Cardiorespiratory System in COVID-19. Indian Journal of Critical Care Medicine. 2021;25(5):524-7.
60. Grue JF, Storve S, Dalen H, Salvesen O, Mjolstad OC, Samstad SO, et al. Automatic Measurements of Mitral Annular Plane Systolic Excursion and Velocities to Detect Left Ventricular Dysfunction. Ultrasound Med Biol. 2018;44(1):168-76.
61. Støylen A, Mølmen HE, Dalen H. Relation between Mitral Annular Plane Systolic Excursion and Global longitudinal strain in normal subjects: The HUNT study. Echocardiography. 2018;35(5):603-10.
62. R W, C C, B I. Left ventricular atrioventricular plane displacement: an echocardiographic technique for rapid assessment of prognosis in heart failure. Heart. 1997(78):230-6.
63. Sæbø S HP, Østvik A, Pasdeloup D, Stølen S, Grenne B, Løvstakken L, Holte E, Dalen H. Automated analyses and real-time guiding by deep learning to reduce test-retest variability of global longitudinal strain. ESC Congress 2022. 2022.
64. Leeson P, Fletcher AJ. Let AI Take the Strain. JACC Cardiovasc Imaging. 2021;14(10):1929-31.
65. Salte IM, Ostvik A, Smistad E, Melichova D, Nguyen TM, Karlsen S, et al. Artificial Intelligence for Automatic Measurement of Left Ventricular Strain in Echocardiography. JACC Cardiovasc Imaging. 2021;14(10):1918-28.
66. Wheeldon NM, MacDonald TM, Flucker CJ, McKendrick AD, McDevitt DG, Struthers AD. Echocardiography in chronic heart failure in the community. Q J Med. 1993;86(1):17-23.
67. Sparrow N, Adlam D, Cowley A, Hampton JR. The diagnosis of heart failure in general practice: implications for the UK National Service Framework. Eur J Heart Fail. 2003;5(3):349-54.
68. Januzzi JL, van Kimmenade R, Lainchbury J, Bayes-Genis A, Ordonez-Llanos J, Santalo-Bel M, et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study. Eur Heart J. 2006;27(3):330-7.
69. Andersen GN, Viset A, Mjolstad OC, Salvesen O, Dalen H, Haugen BO. Feasibility and accuracy of point-of-care pocket-size ultrasonography performed by medical students. BMC Med Educ. 2014;14:156.
70. Magelssen MI, Palmer CL, Hjorth-Hansen A, Nilsen HO, Kiss G, Torp H, 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(2):341-50.
71. Grue JF, Storve S, Dalen H, Mjølstad OC, Samstad SO, Eriksen-Volnes T, et al. Automatic quantification of left ventricular function by medical students using ultrasound. BMC Med Imaging. 2020;20(1):29.
72. Snare SR, Mjolstad OC, Orderud F, Haugen BO, Torp H. Fast automatic measurement of mitral annulus excursion using a pocket-sized ultrasound system. Ultrasound Med Biol. 2011;37(4):617-31.

73. Bhavnani SP, Sola S, Adams D, Venkateshvaran A, Dash PK, Sengupta PP. A Randomized Trial of Pocket-Echocardiography Integrated Mobile Health Device Assessments in Modern Structural Heart Disease Clinics. JACC Cardiovasc Imaging. 2018;11(4):546-57.
74. Singh S, Bansal M, Maheshwari P, Adams D, Sengupta SP, Price R, et al. American Society of Echocardiography: Remote Echocardiography with Web-Based Assessments for Referrals at a Distance (ASE-REWARD) Study. J Am Soc Echocardiogr. 2013;26(3):221-33.
75. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J. 2019;40(40):3297-317.
76. Roberts E, Ludman AJ, Dworzynski K, Al-Mohammad A, Cowie MR, McMurray JJ, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ. 2015;350:h910.
77. Zaphiriou A, Robb S, Murray-Thomas T, Mendez G, Fox K, McDonagh T, et al. The diagnostic accuracy of plasma BNP and NTproBNP in patients referred from primary care with suspected heart failure: results of the UK natriuretic peptide study. Eur J Heart Fail. 2005;7(4):537-41.
78. Dalen H, Thorstensen A, Vatten LJ, Aase SA, Stoylen A. Reference values and distribution of conventional echocardiographic Doppler measures and longitudinal tissue Doppler velocities in a population free from cardiovascular disease. Circ Cardiovasc Imaging. 2010;3(5):614-22.
79. Dalen H, Thorstensen A, Aase SA, Ingul CB, Torp H, Vatten LJ, et al. Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 healthy individuals: the HUNT study in Norway. Eur J Echocardiogr. 2010;11(2):176-83.
80. Kou S, Caballero L, Dulgheru R, Voilliot D, De Sousa C, Kacharava G, et al. Echocardiographic reference ranges for normal cardiac chamber size: results from the NORRE study. Eur Heart J Cardiovasc Imaging. 2014;15(6):680-90.
81. Bernard A, Addetia K, Dulgheru R, Caballero L, Sugimoto T, Akhaladze N, et al. 3D echocardiographic reference ranges for normal left ventricular volumes and strain: results from the EACVI NORRE study. Eur Heart J Cardiovasc Imaging. 2017;18(4):475-83.
82. Addetia K, Miyoshi T, Citro R, Daimon M, Gutierrez Fajardo P, Kasliwal RR, et al. Two-Dimensional Echocardiographic Right Ventricular Size and Systolic Function Measurements Stratified by Sex, Age, and Ethnicity: Results of the World Alliance of Societies of Echocardiography Study. J Am Soc Echocardiogr. 2021;34(11):1148-57.e1.
83. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, et al. Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. JAMA Cardiology. 2021;6(6):624-32.
84. Schneider M, Bartko P, Geller W, Dannenberg V, König A, Binder C, et al. A machine learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. The International Journal of Cardiovascular Imaging. 2021;37(2):577-86.

Paper I

Feasibility and Accuracy of Tele-Echocardiography, With Examinations by Nurses and Interpretation by an Expert via Telemedicine, in an Outpatient Heart Failure Clinic

Anna Katarina Hjorth-Hansen, MD, Garrett Newton Andersen, MD, PhD, Torbjørn Graven, MD, Guri Holmen Gundersen, RCN, Jens Olaf Kleinau, MD, Ole Christian Mjølstad, MD, PhD, Kyrre Skjetne, MD, Stian Stølen, MD, Hans Torp, PhD, Håvard Dalen, MD, PhD

Received January 10, 2020, from the Department of Internal Medicine, Levanger Hospital, Nord–Trøndelag Hospital Trust, Levanger, Norway (A.K.H.-H., G.N.A., T.G., G.H.G., J.O.K., K.S., S.S., H.D.); Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway (A.K.H.-H., O.C.M., H.T., H.D.); and Clinic of Cardiology, St Olav's University Hospital, Trondheim, Norway (O.C.M., H.D.). Manuscript accepted for publication April 26, 2020.

The Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, hosts a collaboration between the university, hospitals, and several vendors. GE Healthcare is one of these partners but had no role in the planning or performance of this study. Drs Mjølstad and Torp hold positions in this collaboration. Fimreite Software was a partner in a research-based innovation program hosted by the Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, and funded by the Norwegian Research Council.

Address correspondence to Anna Katarina Hjorth-Hansen, MD, Department of Internal Medicine, Levanger Hospital, Kirkegata 2, 7600 Levanger, Norway.

E-mail: anna.hjorth-hansen@hnt.no

Abbreviations

HF, heart failure; HFmrEF, heart failure with a midrange ejection fraction; HFpEF, heart failure with a preserved ejection fraction; HFrEF, heart failure with a reduced ejection fraction; LA, left atrial; LAVI, left atrial volume index; LV, left ventricular; LVEF, left ventricular ejection fraction

doi:10.1002/jum.15341

Objectives—To study the feasibility and accuracy of focused echocardiography by nurses supported by near–real-time interpretation via telemedicine by an experienced cardiologist.

Methods—Fifty consecutive patients were included from an outpatient heart failure (HF) clinic. Limited echocardiography was performed by 1 of 3 specialized nurses. The echocardiograms were transferred by a secure transfer model for near–real-time interpretation to 1 out-of-hospital cardiologist, assessing, among others, the left ventricular (LV) internal diameter, end-diastolic volume, ejection fraction, left atrial (LA) indexed end-systolic volume, mitral early inflow velocity (E), the ratio of E to mitral late inflow, and the ratio of E to the mitral annular early diastolic velocity. The reference method was echocardiography by 1 of 4 experienced cardiologists.

Results—The median age of the population (46% women) was 79 (range, 33–95) years. The assessment and quantification of LA and LV dimensions, volumes, and functional indices were feasible in 94% or more via the telemedical approach. The agreement with reference measurements was very high by the telemedical approach. The mean duration ± SD of the complete telemedical approach from the start of echocardiography until the cardiologist's report was received by the caregiving nurse was 1.32 ± 0.36 (range, 1.58) hours. The correlations with reference to the above-specified indices were $r = 0.75$ to 0.94.

Conclusions—Limited echocardiography by nurses in an outpatient heart failure clinic, supported by interpretation by an out-of-hospital cardiologist, was feasible and reliable. This may reduce geographic disparities and allow more patients to benefit from the advantages of implementing focused echocardiography by noncardiologists in diagnostics and follow-up.

Key Words—echocardiography; heart failure; nonexpert; nurse; telemedicine

Modern technology allows for fast and safe transfer of patient information, which has not previously been attainable. Telecommunication technology (telemedicine) provides delivery of health services where distance is a critical factor. Today, access to the Internet and, thus, the potential to use telemedicine is available in most places in the world through local area networks and wireless mobile telecommunication technology.

Heart failure (HF) is associated with a poor prognosis and a reduced quality of life, and the financial burden on the health care system is substantial.[1] Despite current treatment options, HF morbidity and mortality are still high, and 25% to 50% of all patients with HF are readmitted within 6 months of hospitalization after decompensated HF.[1–3] Guidelines advocate classification of HF by the left ventricular ejection fraction (LVEF), as this is decisive for both treatment and prognosis.[4] Patients with HF who have an EF of 50% or higher are classified as having heart failure with a preserved ejection fraction (HFpEF). Similarly, patients with HF who have an EF of 40% to 49% and an EF of 40% or lower are classified as having heart failure with a midrange ejection fraction (HFmrEF) and heart failure with a reduced ejection fraction (HFrEF), respectively. Furthermore, the size and function of the cardiac chambers are easily depictable by echocardiography.[5]

Fluid retention is the major consequence of decompensation, which usually happens over time with late-onset symptoms and an unpredictable course.[4] Patients with HF could benefit from frequent volume status assessments and more aggressive therapy.[2,5–7] The assessment of the volume status in patients with HF may be improved by evaluation of the presence of pleural effusion and the dimension and collapsibility of the inferior vena cava.[4,8] The classification of the HF category, estimation of filling pressures, and estimation of the volume status can be performed by echocardiography.[5,9]

Echocardiography is usually performed by specialized sonographers or experienced cardiologists. Interpretation of the recordings is usually performed at the same location as the examination. Implementation of telemedicine for interpretation of images at remote locations is common in the field of radiology to reduce the workload of local radiologists.[10] The research regarding tele-echocardiography is scarce. The first studies on tele-echocardiography were conducted by pediatric cardiologists as support for physicians in rural areas.[11] So far, telemedicine has not found its way into routine follow-up of patients with HF. However, tele-echocardiography allows for the performance of echocardiographic recordings at one location and interpretation by an expert at another; thus, patients can benefit from the positive impact of more precise diagnostics.

The aim of the study was to examine the feasibility and accuracy of tele-echocardiography in an outpatient HF clinic. We combined echocardiographic recordings by trained nurses with transfer of echocardiographic data by local area networks and wireless mobile telecommunication for interpretation by a cardiologist at a remote location. The purpose of the study was not to implement limited ultrasound in the routine follow-up of patients with HF, but the aim was to explore the benefit in HF follow-up where tele-echocardiography can overcome geographic challenges to improve HF diagnostics and care. Second, we aimed to evaluate the accuracy of tele-echocardiography for classification of HF and evaluation of filling pressures.

Materials and Methods

Study Population

Patients from an outpatient HF clinic were recruited at Levanger Hospital, Nord–Trøndelag Health Trust. All patients were followed for known HF and had previous echocardiographic examinations performed by a cardiologist. Only patients older than 18 years were eligible for inclusion between October 2016 and February 2017. All participants gave their informed written consent before inclusion. The study was conducted in conformity with the policy statement for the use of human subjects of the Declaration of Helsinki. The study was approved by the Regional Committee for Medical and Health Research Ethics (REK 2015/2312) and registered in the ClinicalTrials.gov database (NCT02936050).

Training and Education of Nurses

Three registered cardiac nurses with 6 to 12 years of clinical experience from a nurse-led outpatient HF clinic were trained in performing echocardiographic recordings by two cardiologists. The nurses had no previous experience in echocardiographic recordings or image analyses. However, they were familiar with the use of handheld ultrasound devices for evaluation of the size and respiratory variation of the inferior vena cava and assessment of pathologic fluid in the pericardium and pleural cavities to aid in their clinical work with patients with HF. Lung ultrasound was not included in the training. They underwent systematic training by cardiologists with experience in echocardiography and performed a mean of 67 (range, 47–97) lifetime examinations before patient inclusion. Initially, they performed approximately five echocardiographic examinations with hands-on training support.

Tele-Echocardiography With Recordings by Nurses and Interpretation by Telemedicine

A Vivid 7 scanner (GE Healthcare AS, Horten, Norway) was used by the nurses to obtain goal-directed echocardiographic recordings of the following standard views: parasternal long- and short-axis (with and without color Doppler), 3 standard apical views (4-chamber, 2-chamber, and long-axis) with and without color Doppler focusing on left ventricular (LV), left atrial (LA), and right ventricular subcostal views for assessment of the inferior vena cava, pulsed wave tissue Doppler imaging with a sample volume in the basal part of the septal and lateral walls (4-chamber), pulsed wave Doppler recordings of mitral inflow, and continuous wave Doppler imaging of the blood flow through the aortic valve and tricuspid regurgitations. All recordings contained at least 3 cardiac cycles. The recording of the inferior vena cava included both maximum and minimum dimensions by including a quick inspiration (sniff). Both pleural cavities were assessed in the midclavicular and midaxillary line in a sitting position with the transducer in the intercostal spaces, as described earlier,[12] and in cases of pleural effusion, longitudinal and transverse images were recorded. The amount of fluid was quantified as the distance from the diaphragm to the basal part of the lung, annotated as 0 in cases of no pleural effusion. No ultrasound examinations of the lungs were included. The nurses had access to patient histories and key words of previous echocardiographic examinations, but importantly, they did not have access to previous echocardiographic recordings.

Transfer of the recordings was done immediately after the examination. A commercial software- and hardware-based system (PaCentric; Fimreite Software, Stavanger, Norway) was used. PaCentric allows for secure transmission, interpretation, and reporting of medical Digital Imaging and Communications in Medicine images per the Internet. The data were stored securely and depersonalized. PaCentric is accredited (International Organization for Standardization 13485:2003) and certified (Conformitè Europëenne 0434), as well as approved by the US Food and Drug Administration (510 k100837). Transmission of the recordings was allowed by installing the software on the Vivid 7 scanner connected to the hospital's local area network. After the end of the examination, data were exported to the PaCentric server and stored on an ordinary, password-protected laptop computer by the cardiologist who performed all of the analyses. Both local area networks and mobile telecommunication network solutions were used, depending on the availability at the interpreter's actual location.

The interpretation of the recordings by the out-of-hospital cardiologist was performed in EchoPAC SWO (version BT12; GE Healthcare). The out-of-hospital cardiologist was blinded to all previous echocardiographic recordings and patient histories. All measurements reflect the average of at least 3 cardiac cycles. The LV endocardial borders were traced in end diastole and end systole in 4- and 2-chamber views. The LV internal length was measured from the traces and LV volumes (end-diastolic and end-systolic), and the EF was calculated by biplanar Simpson method. The LV internal diameter and wall thickness were measured at the level of the tip of the mitral leaflets in 2-dimensional grayscale recordings. The LA endocardial border was traced in end systole in 4- and 2-chamber views, and the volume was calculated by the area-length method and subsequently indexed per square meter body surface area (left atrial volume index [LAVI]). The pulmonary veins and the LA appendage were not included in the trace. Mitral inflow peak early (E) and late (A) velocities and the early filling mitral deceleration time were measured in pulsed wave Doppler recordings from the apical 4-chamber view. The peak velocity of the tricuspid regurgitation was measured by continuous wave Doppler imaging. Mitral annular peak systolic (S') and peak early diastolic (e') longitudinal velocities were assessed by pulsed-wave tissue Doppler imaging. The ratio of the early mitral inflow to the early diastolic mitral annular velocity (E/e') was calculated. The LV filling pressure was estimated according to recommendations by the European Association of Cardiovascular Imaging and the American Society of Echocardiography, based on the LAVI, e', E/e', and peak velocity of the tricuspid regurgitation for the HF subgroups as normal or elevated.[9]

Reference Echocardiography

Reference echocardiography was performed immediately after the nurses' recordings by 1 of 4 in-house physicians experienced in echocardiography (3 cardiologists and 1 experienced resident in cardiology). They were not blinded to medical histories or previous echocardiographic recordings from the patients. However, they were blinded to the echocardiographic examinations performed by the nurses and analyzed by the out-of-hospital cardiologist by telemedicine. The reference examinations were performed at the same department but in another room. High-end echocardiographic scanners (Vivid E9 and Vivid E95) were used. The reference imaging included the same recordings, and in addition, all other chambers and valves were assessed. The echocardiographic measurements were performed as indicated above.

Patient Flow

The patients were first examined by 1 of 3 registered cardiac nurses, who immediately transferred the echocardiographic data for further analyses by telemedicine. The recordings obtained by the nurses were interpreted in near real time via the tele-echocardiographic approach by an out-of-hospital cardiologist (Figure 1). No further follow-up or ultrasound examinations of the participants were performed during the study.

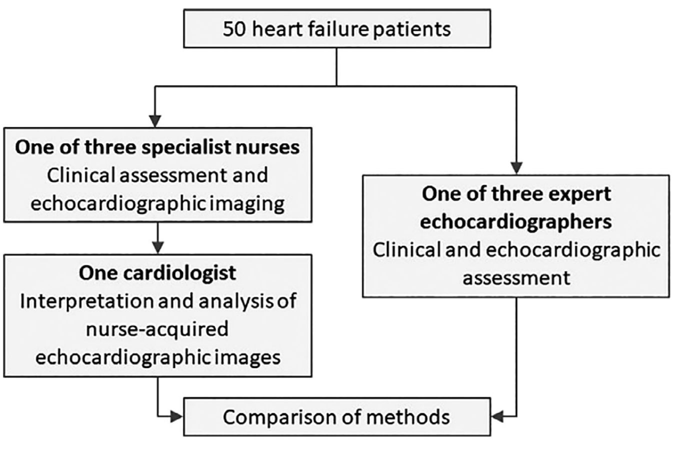
Before echocardiography, blood samples were drawn the same day and analyzed at the in-hospital accredited laboratory. Serum N-terminal pro-brain natriuretic peptide, serum creatinine, and estimated glomerular filtration rate (calculated by the Cockcroft-Gault equation) values were measured for characterization of

Table 1. Baseline Data of the 50 Study Participants

Parameter	Value
Age, years	77 ± 12 (62)
Women, n (%)	23 (46)
BMI, kg/m^2	25.5 ± 5.3 (28.3)
Systolic blood pressure, mm Hg	131 ± 22 (101)
Diastolic blood pressure, mm Hg	76 ± 11 (50)
NT-Pro-BNP, ng/L	4,320 ± 7,014 (44,867)
eGFR, mL/min/1.73m^2[a]	44 ± 26 (126)
NYHA functional class	1.9 ± 0.7 (3)
I (N)	12
II (N)	25
III (N)	7
IV (N)	1
Atrial fibrillation, n (%)	24 (48)
Diuretics, n (%)	42 (84)
Beta-blocker, n (%)	40 (80)
ACEI/ARB, n (%)	18 (36)

Data are presented as mean ± SD (range) unless otherwise specified. ACEI/ARB indicates angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker; BMI, body mass index; eGFR, estimated glomerular filtration rate; NT-Pro-BNP, N-terminal pro-B-type natriuretic peptide; and NYHA, New York Heart Association.
[a]Cockcroft-Gault equation.

Figure 1. Patient flow throughout the study. Patients with HF were examined by 1 of 3 specialist nurses and immediately after by 1 of 3 expert echocardiographers. The images acquired by the nurses were interpreted and analyzed by an out-of-hospital cardiologist via telemedicine.



the population. The New York Heart Association functional classification was scored by the nurses, and the body weight (kilograms), body height (centimeters), and blood pressure (millimeters of mercury) were measured. Anthropometric measurements were rounded up to the nearest multiple of 1.

Statistical Analyses

Descriptive statistics were used for describing the study population. Data are presented as mean ± SD, but data not following a normal distribution are presented as median (range). Categorical data are reported as numbers and percentages. The agreement of the measurements by the telemedical approach and reference was tested by Bland–Altman statistics, the coefficient of variation, and the Pearson or Spearman correlation coefficient. Proportions were analyzed by the χ^2 test. Agreement with respect to the correct HF classification was analyzed by the weighted κ statistic. Comparisons of means were tested by paired *t* tests or the related-sample Wilcoxon signed rank test. Two-sided $P < .05$ was considered statistically significant. Semiquantitative data were further assessed by calculations of sensitivity and specificity and negative and positive predictive values. The association of the type of HF with the correct classification of LV filling pressures was analyzed by logistic regression analyses.

Table 2. Time Consumption for the Components of the Tele-Echocardiographic Approach

Parameter	Value
Time used from start echocardiography to the finalized report, hours	1.32 ± 0.36 (1.58)
Time used for echocardiographic recordings by nurse, hours	0.48 ± 0.25 (0.93)
Time used for transfer of echocardiograms, hours	0.36 ± 0.26 (1.20)
Time used from echocardiograms uploaded to finalized report by cardiologist, hours	0.56 ± 0.16 (1.20)
Time used for analyses of echocardiograms by cardiologist, hours	0.20 ± 0.06 (0.27)

Data are presented as mean ± SD (range).

Results

Population

Baseline data for the 50 participants are shown in Table 1. The mean age was 77 ± 12 years; 46% were women; and the mean body mass index was 25.6 ± 5.9 kg/m^2. The mean estimated glomerular

Table 3. Echocardiographic Indices by the Telemedical Approach and Reference Cardiologist

Parameter	n_{tele}/n_{ref}	Telemedical Approach	Reference Echocardiography	*P* for Difference
LV end-diastolic volume, mL	48/49	113 ± 39	115 ± 49	.382
LVEF, %	48/50	42 ± 12	43 ± 11	.395
LV internal end-diastolic diameter, mm	49/50	49 ± 8	52 ± 10	<.001
LA end-systolic volume index, mL/m^2	49/48	62 ± 19	61 ± 22	.689
Mitral early diastolic velocity, cm/s	48/50	85 ± 41	84 ± 41	.544
Mitral annular early diastolic velocity, cm/s[a]	50/48	6.1 ± 2.3	5.7 ± 2.3	.052
Mitral annular systolic velocity, cm/s[a]	50/48	4.8 ± 1.3	4.8 ± 1.3	.621
Mitral E/A ratio	28/28	1.67 ± 1.22	1.43 ± 0.91	.177
E/e′ ratio[a]	48/48	16.9 ± 12.3	15.9 ± 11.4	.562
Tricuspid regurgitation peak velocity, m/s	32/41	2.65 ± 0.57	2.79 ± 0.46	.018
LV end-diastolic length, mm	48/50	84 ± 8	79 ± 9	<.001
LA end-systolic length, 4Ch, mm	49/49	6.2 ± 0.9	6.2 ± 0.9	.641
IVS end-diastolic thickness, mm	49/48	11.1 ± 3.0	10.3 ± 2.7	.164
LV posterior wall end-diastolic thickness, mm	49/49	11.1 ± 2.9	10.2 ± 2.9	.013
Pleural effusion, mm[b]	100/100	2.8 (0–60)[b]	2.9 (0–70)[b]	.906[b]

Data are presented as mean ± SD unless otherwise specified. LV and LA lengths are means of measurements from 4- and 2-chamber views. 4Ch indicates 4-chamber view; IVS, interventricular septum; n_{ref}, numbers analyzed by the reference cardiologist; and n_{tele}, numbers analyzed by the telemedical approach.
[a]Mean of peak septal and lateral tissue velocities.
[b]Left and right pleural cavities were treated separately. The numbers relate to the millimeter distance from the diaphragm to the basal part of the lung. In case of no effusion, the measurement is annotated as 0. Values are presented as mean (range).

Figure 2. Bland–Altman plots illustrating the agreement between the telemedical approach and reference echocardiography for the LVEF (**A**), LAVI (**B**), E/e′ ratio (**C**), and maximal tricuspid regurgitation pressure gradient (TR max P; **D**). The differences of the respective measurements are plotted against the means of the measurements.

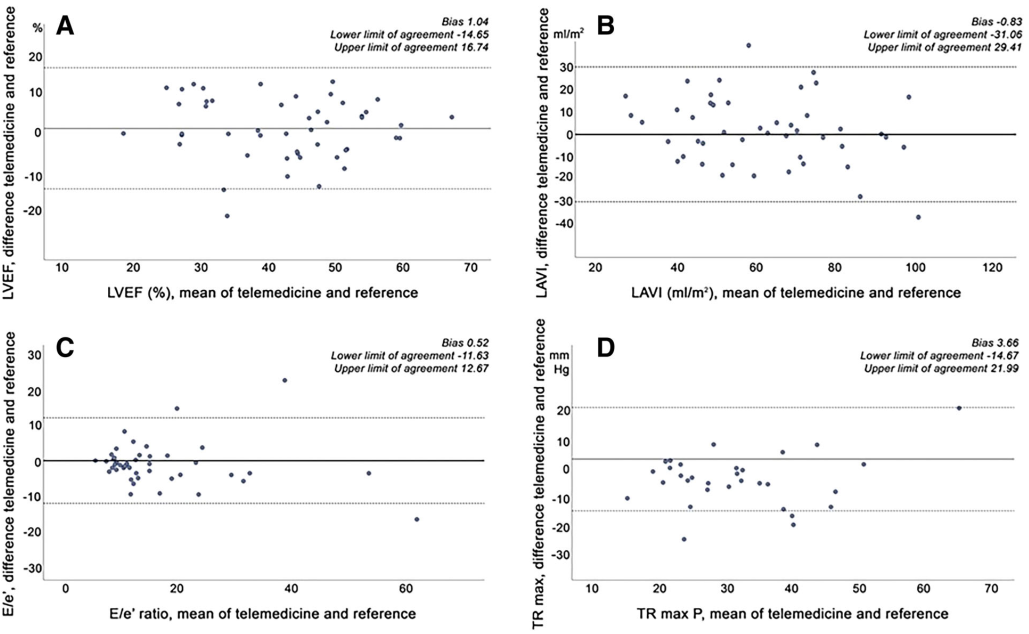


Table 4. Agreement Between Echocardiographic Indices by the Telemedical Approach and Reference Cardiologist

Parameter	n (Pairs)	Correlation (*P*)	CoV, %	Bias
LVEF	48	0.78 (.002)	11.7	1.0
LV end-diastolic volume	47	0.85 (<.001)	12.6	−3.3
LV internal end-diastolic diameter	49	0.80 (.01)	8.0	−3.6
LA end-systolic volume index	46	0.75 (.004)	14.8	0.7
Mitral early diastolic velocity	48	0.94 (<.001)	10.2	1.3
Mitral annular early diastolic velocity	48	0.82 (<.001)	12.8	0.3
Mitral annular systolic velocity	48	0.80 (.001)	10.9	−0.1
Mitral E/A ratio	21	0.88 (.001)	13.5	0.2
E/e′ ratio	46	0.88 (<.001)	16.7	−0.6
Tricuspid regurgitation peak velocity	31	0.71 (.007)	8.4	−0.2
LV end-diastolic length	48	0.74 (.004)	6.7	0.7
LA end-systolic length	49	0.72 (.006)	6.1	0.4
IVS end-diastolic thickness	48	0.62 (.02)	13.2	0.7
LV posterior wall end-diastolic thickness	48	0.60 (.03)	15.0	1.00
Pleural effusion[a]	100	0.88 (<.001)		0.1

Bias was measured as telemedical approach mean values minus mean reference values. CoV indicates coefficient of variation; and IVS, interventricular septum.
[a]As described in Table 3.

filtration rate was 44 ± 29 mL/min. Diuretics, beta blockers, and angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers were used by 84%, 80%, and 36% of the population, respectively. The low proportion of patients treated with angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers was related to the high prevalence of renal failure in the population and the fact that optimal HF therapy was not yet achieved.

Feasibility of Tele-Echocardiography

Table 2 shows the time used for the telemedical approach. The mean duration of the examination by the nurses from the start of echocardiography until the report was finalized and reported back electronically by the out-of-hospital cardiologist was 1.32 ± 0.36 (range, 1.58) hours. Image transfer was efficient in most cases, and delays were mainly due to technological difficulties and occurred in 20 of 50 cases. Delays were primarily caused by computer crashes and long downloads, leading to inaccessibility of images within reasonable time.

At the time of data transmission, the computer of the out-of-hospital cardiologist was connected to the Internet by fiber-optic cables in 30 (60%) examinations, an asymmetric digital subscriber line in 9 (18%), and wireless mobile networks (3G and 4G) in 11 (22%). The mean duration of transfer varied by the mode of telecommunication, being shortest for fiber-optic cables (0.32 ± 0.25 [range, 0.07–1.27] hours), followed by the asymmetric digital subscriber line (0.35 ± 0.20 [range, 0.10–0.67] hours) and wireless mobile networks (0.48 ± 0.28 [range, 0.22–1.22] hours), respectively.

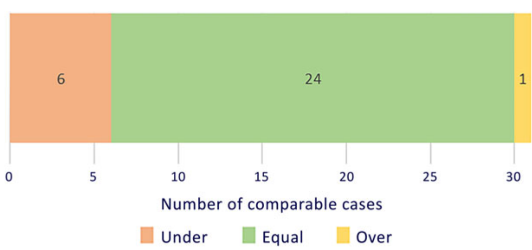
Comparison of the tele-echocardiographic data with the reference measurements is shown in Table 3. Feasibility was high for all indices, except for the mitral E/A ratio and peak velocity of the tricuspid regurgitation. Only the LV internal end-diastolic diameter, LV internal end-diastolic length, LV posterior wall end-diastolic thickness, and tricuspid regurgitation peak velocity were significantly different between the methods ($P < .02$). By tele-echocardiography, LV endocardial borders in end diastole were underestimated by 3 mm (49 versus 52 mm), and LV length was overestimated by 5 mm (84 versus 79 mm).

The agreements of measurements by telemedicine and the reference for the EF, LAVI, maximal tricuspid regurgitation gradient, and E/e′ ratio are illustrated in Figure 2 and Table 4. The biases for the different measurements were close to 0, and there was no significant relationship between the errors and the magnitudes of the measurements. Coefficients of variation for the above-mentioned central indices were all in the range of 6% to 15% (Table 4). The correlations were high for all echocardiographic indices ($r \geq 0.71$; $P \leq .007$), except for measurements of wall thickness, for which the correlations were moderate for both the interventricular septum and the posterior wall (both $r \geq 0.60$; $P \leq .03$). Pleural effusion was revealed in a total of 9 pleural cavities by either the reference or the telemedical approach. The latter detected pleural effusion in 7 of 8 cavities in which reference imaging results were positive.

Tele-echocardiography showed substantial agreement with the reference for classification of the category of HF, with a weighted κ of 0.73 ($P < .001$). Importantly, no participants were misclassified between rEF and pEF, but 17 cases were misclassified between mEF and pEF or rEF. The LV filling pressure was determined by the telemedical approach and reference in 39 and 41 of the 50 cases, respectively, and in 31 cases, the filling pressure was determined by both approaches (Figure 3). Among the misclassified cases, HFpEF and HFmrEF were numerically more prevalent, including all cases by the telemedical approach and 5 (71%) by reference.

Valvular disease was evaluated semiquantitatively. The sensitivity and specificity for tele-echocardiography to detect at least moderate mitral stenosis, mitral regurgitation, and tricuspid regurgitation were 100% and

Figure 3. Agreement of grading of LV filling pressure between the telemedical approach and the reference method in 31 available cases. Left ventricular filling pressures were assessed as high or normal. "Under," "equal," and "over" refer to the comparison of the LV filling pressure classification described by the telemedical approach compared to the reference echocardiography.



95% or higher, respectively. For detection of at least moderate aortic stenosis (n = 8, but only 7 cases available for the analyses) the sensitivity was lower (43%) but still with excellent specificity (97%).

Discussion

We are currently unaware of other studies evaluating tele-echocardiography with recordings by nonphysician personnel at a single geographic location combined with near–real-time interpretation by a cardiologist at another. The telemedical approach was feasible and reliable in this HF population. The most important finding of this study is that by using expert support by telemedicine, more patients with HF can gain the benefit of diagnostic ultrasound. Such an approach may improve diagnostics and care when distance and available resources matter.

The limited echocardiographic approach presented here may be used during the initial evaluation of a patient with suspected HF, making information available. Thus, the time delay to diagnosis can be reduced by improving the basis for decisions on the right workup. Importantly, the aim was not to replace comprehensive echocardiography by this approach but to evaluate whether telemedicine could support clinical decision making when the echocardiographic recordings were in the hands of inexperienced users.

The time spent on the echocardiographic recordings by the nurses was, on average, 0.5 hour and within range of what is acceptable and feasible in the everyday clinical practice for a nurse-lead outpatient clinic. Similarly, the time spent for transfer, analyses, and reporting was short and allows for implementation. Even though the time used for transfer of the recordings depended on the local area network available, the approach was feasible for near–real-time interpretation, also when a 3G/4G mobile network was used. This was in line with previous studies.[13] The out-of-hospital cardiologist's categorization of the type of HF via the telemedical approach was comparable to the in-hospital cardiologist's reference echocardiography. Thus, geographic challenges can be overcome with the use of telemedicine for support of dedicated health care personnel in remote areas where traveling is a burden.[14,15]

The telemedical software used is approved and complies with the regulations set by the Norwegian Data Protection Authority. With this software, sensitive data can be transferred and directly imported into EchoPAC (GE Healthcare) software for analyses, which presents a great advantage in simplifying the work flow. Today, several vendors provide similar software for transfer of imaging data, in accordance with the European Union general data protection regulations.

The three nurses performing the recordings had undergone dedicated, but limited training. Their training exceeded the recommendations for focused cardiac ultrasound examinations but did not reach the level recommended for comprehensive echocardiography.[16–18] In line with others, our results may add knowledge, and consequently, more patients can benefit from the diagnostic yield of echocardiography. As shown both in HF and other populations, diagnostic ultrasound may add important information, even when those performing the examinations have limited experience.[8,19,20] In this study, all patients were additionally examined by a cardiologist, and the results indicate that echocardiographic recordings by nurses combined with interpretation by a cardiologist add important information to the clinical decision making. To evaluate the clinical benefit of this approach, larger clinical studies are warranted.

As shown by the semiquantitative assessment of valvular disease, more training of the users may be needed to safely exclude valvular disease. Other studies evaluating handheld ultrasound devices by inexperienced users have also shown that evaluating valvular disease may be challenging,[15] and our group has previously shown that inexperienced users of diagnostic ultrasound perform better in assessments of global LV function than valvular assessments.[21] However, a valvular assessment was not the purpose of this study; thus, the results presented are in line with what was expected. The results highlight the need for dedicated training in any given task for operators of diagnostic ultrasound.

The telemedical calculations of dimensions, volumes, and flow measurements were well in line with reference measurements. Most differences were nonsignificant. There were only small but significant differences for the LV internal end-diastolic diameter, end-diastolic length, and posterior wall end-diastolic thickness (mean differences of 3, 5, and 0.9 mm, respectively) and tricuspid regurgitation peak velocity (mean difference of 0.14 m/s). The data presented are quite similar to what has been presented from other studies evaluating test–retest variation of echocardiographic indices.[22–24] The agreement for correct classification of the HF category was substantial, with a κ of 0.73, and all of the inconsistencies between HF classification by telemedicine and the reference related to HFmrEF versus HFrEF or HFpEF. The recently introduced new HF class HFmrEF has caused intense debate, as it is based on the idea that the EF can correctly classify HF into different categories. Furthermore, on the basis of the published repeatability data from the Atherosclerosis Risk in Communities study, approximately 40% of patients will move from HFmrEF to HFrEF or HFpEF in repeated analyses.[25] The prevalence of HFpEF in the study population was similar to what has been shown other HF populations.[4] The agreement with respect to estimation of the LV filling pressure was good, with misclassification by the telemedical approach in only 17%. Studies performing similar tasks with repeated echocardiograms are scarce, but our data were similar to a study evaluating agreement between echocardiographers in 105 single echocardiograms.[26]

The views and parameters included in the limited echocardiographic examination were based on the need for a proper assessment of a patient with HF, allowing for classification of subtypes and an assessment of the LV filling pressure. These are less than what is recommended for a comprehensive echocardiogram but substantially more than what is included in focused cardiac ultrasound.[27,28]

Very little research has been done in which personnel not previously skilled in echocardiography performed echocardiographic recordings with interpretation by specialists, but the results are promising when compared to, for instance, the robotic-arm approach.[27,29] This approach differs from the training of echocardiography technicians and sonographers, as the nurses did not perform quantitative analyses but only dedicated recordings without interpretations of the recordings. The nurses had years of clinical experience evaluating the volume status with handheld ultrasound devices but had not performed echocardiography until training for this study.[8,30] This confirms previously shown results aiming to implement telemedicine to overcome geographic challenges. Thus, a limited echocardiographic examination by nonexperts can achieve quality that allows for a reliable assessment.[10,11] Considering the efforts and costs to transport patients to a hospital with an available specialist versus the efforts and costs of training nurses, the study results can justify implementation into everyday clinical practice.

Limitations

The aim of the study was to evaluate the LV function, volume status, and indices important for classification of subpopulations of HF. Thus, the results cannot be generalized for other tasks. The agreement between the methods for determination of an elevated filling pressure was not validated invasively; thus, only the agreement of the telemedical approach with the reference echocardiography could be assessed.

Most of the patients had previous echocardiographic examinations at the same hospital. The reference cardiologist had access to previous echocardiographic recordings. The out-of-hospital cardiologist who performed the interpretation of the nurses' recordings was blinded to all data from previous echocardiograms and medical histories, and it is not likely that the reference echocardiographic examination was influenced by previous echocardiography. Thus, we find it unlikely that the knowledge of previous echocardiograms biased the results, as the complete blinding of the telemedical cardiologist interpreting the echocardiograms would, if having any influence, tend to introduce a negative bias, which was not observed. The limited echocardiograms recorded by nonexperts described here should not be considered equal to comprehensive echocardiography by experts. However, the goal-directed examinations presented in this study add valuable quantitative information for clinical decision making, which may guide therapy beyond what is achievable by semiquantitative cardiac ultrasound from handheld devices.[17,28]

Conclusions

Tele-echocardiography, in the form of image acquisition by registered cardiac nurses supported by interpretation by a cardiologist, is feasible and provides reliable results of central indices for quantification of left-sided cardiac size and function, HF classification, and LV filling pressures. Implementing tele-echocardiography at remote locations where

echocardiography experts are not available may improve diagnostics and therapy.

References

1. Maggioni AP, Dahlström U, Filippatos G, et al. EURObservational Research Programme: the Heart Failure Pilot Survey (ESC-HF Pilot). *Eur J Heart Fail* 2010; 12:1076–1084.
2. Ambrosy AP, Pang PS, Khan S, et al. Clinical course and predictive value of congestion during hospitalization in patients admitted for worsening signs and symptoms of heart failure with reduced ejection fraction: findings from the EVEREST trial. *Eur Heart J* 2013; 34:835–843.
3. Khan H, Greene SJ, Fonarow GC, et al. Length of hospital stay and 30-day readmission following heart failure hospitalization: insights from the EVEREST trial. *Eur J Heart Fail* 2015; 17: 1022–1031.
4. 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: 2129–2200.
5. 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; 16:233–271.
6. Albert NM. Fluid management strategies in heart failure. *Crit Care Nurse* 2012; 32:20–32.
7. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? *Circulation* 2012; 126:501–506.
8. 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.
9. Nagueh SF, Smiseth OA, Appleton CP, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2016; 17:1321–1360.
10. Jarvis L, Stanberry B. Teleradiology: threat or opportunity? *Clin Radiol* 2005; 60:840–845.
11. Grant B, Morgan GJ, McCrossan BA, et al. Remote diagnosis of congenital heart disease: the impact of telemedicine. *Arch Dis Child* 2010; 95:276–280.
12. Graven T, Wahba A, Hammer AM, et al. Focused ultrasound of the pleural cavities and the pericardium by nurses after cardiac surgery. *Scand Cardiovasc J* 2015; 49:56–63.
13. Barbier P, Dalla Vecchia L, Mirra G, Di Marco S, Cavoretto D. Near real-time echocardiography teleconsultation using low bandwidth and MPEG-4 compression: feasibility, image adequacy and clinical implications. *J Telemed Telecare* 2012; 18:204–210.
14. Choi BG, Mukherjee M, Dala P, et al. Interpretation of remotely downloaded pocket-size cardiac ultrasound images on a Web-enabled smartphone: validation against workstation evaluation. *J Am Soc Echocardiogr* 2011; 24:1325–1330.
15. Evangelista A, Galuppo V, Mendez J, et al. Hand-held cardiac ultrasound screening performed by family doctors with remote expert support interpretation. *Heart* 2016; 102:376–382.
16. Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr* 2019; 32:1–64.
17. Neskovic AN, Edvardsen T, Galderisi M, et al. Focus cardiac ultrasound: the European Association of Cardiovascular Imaging viewpoint. *Eur Heart J Cardiovasc Imaging* 2014; 15:956–960.
18. Spencer KT, Kimura BJ, Korcarz CE, Pellikka PA, Rahko PS, Siegel RJ. Focused cardiac ultrasound: recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr* 2013; 26:567–581.
19. Kobal SL, Trento L, Baharami S, et al. Comparison of effectiveness of hand-carried ultrasound to bedside cardiovascular physical examination. *Am J Cardiol* 2005; 96:1002–1006.
20. Panoulas VF, Daigeler AL, Malaweera AS, et al. Pocket-size handheld cardiac ultrasound as an adjunct to clinical examination in the hands of medical students and junior doctors. *Eur Heart J Cardiovasc Imaging* 2013; 14:323–330.
21. 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: 1195–1202.
22. Thorstensen A, Dalen H, Amundsen BH, Aase SA, Stoylen A. Reproducibility in echocardiographic assessment of the left ventricular global and regional function: the HUNT study. *Eur J Echocardiogr* 2010; 11:149–156.
23. Baron T, Berglund L, Hedin EM, Flachskampf FA. Test-retest reliability of new and conventional echocardiographic parameters of left ventricular systolic function. *Clin Res Cardiol* 2019; 108: 355–365.
24. Karlsen S, Dahlslett T, Grenne B, et al. Global longitudinal strain is a more reproducible measure of left ventricular function than ejection fraction regardless of echocardiographic training. *Cardiovasc Ultrasound* 2019; 17:18.
25. Shah AM, Cheng S, Skali H, et al. Rationale and design of a multicenter echocardiographic study to assess the relationship between

cardiac structure and function and heart failure risk in a biracial cohort of community-dwelling elderly persons: the Atherosclerosis Risk in Communities study. *Circ Cardiovasc Imaging* 2014; 7: 173–181.

26. Chapman CB, Ewer SM, Kelly AF, Jacobson KM, Leal MA, Rahko PS. Classification of left ventricular diastolic function using American Society of Echocardiography guidelines: agreement among echocardiographers. *Echocardiography* 2013; 30:1022–1031.
27. Sengupta PP, Narula N, Modesto K, et al. Feasibility of intercity and trans-Atlantic telerobotic remote ultrasound: assessment facilitated by a nondedicated bandwidth connection. *JACC Cardiovasc Imaging* 2014; 7:804–809.
28. Cardim N, Dalen H, Voigt JU, 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; 20:245–252.
29. Boman K, Olofsson M, Berggren P, Sengupta PP, Narula J. Robot-assisted remote echocardiographic examination and teleconsultation: a randomized comparison of time to diagnosis with standard of care referral approach. *JACC Cardiovasc Imaging* 2014; 7:799–803.
30. Dalen H, Gundersen GH, Skjetne K, et al. Feasibility and reliability of pocket-size ultrasound examinations of the pleural cavities and vena cava inferior performed by nurses in an outpatient heart failure clinic. *Eur J Cardiovasc Nurs* 2015; 14:286–293.

Paper II

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

Check for updates

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.

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.

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.

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

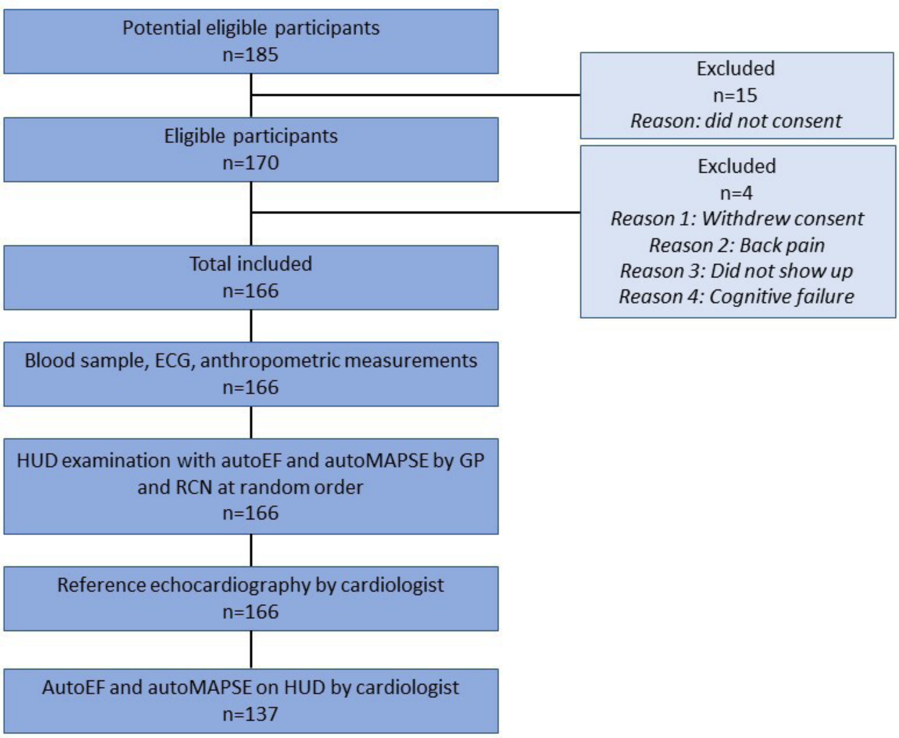


Figure 1 Study flow. AutoEF, automatic measurement of left ventricular ejection fraction; autoMAPSE, automatic measurement of mitral annular plane systolic excursion; GP, general practitioner; HUD, hand-held ultrasound device; RCN, registered cardiac nurse.

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 ClinicalTrial.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 themself 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 recommendations[18] 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[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)				
Mean of septal and lateral position				
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 21 s (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.

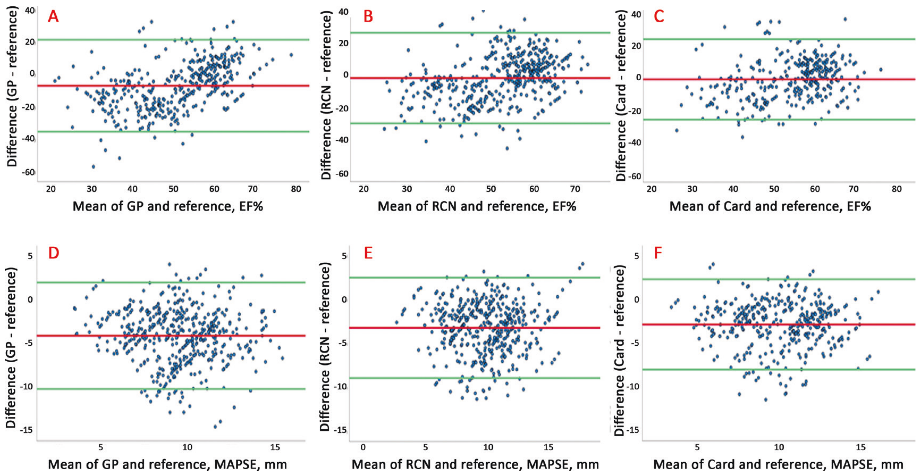


Figure 2 Bland-Altman plots illustrating the agreement between all autoEF and autoMAPSE recordings taken by GPs, RCNs and cardiologists compared to reference echocardiography for all recordings with automatic decision-support software irrespective of image score. Upper panel: autoEF by (A) GPs, (B) RCNs 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

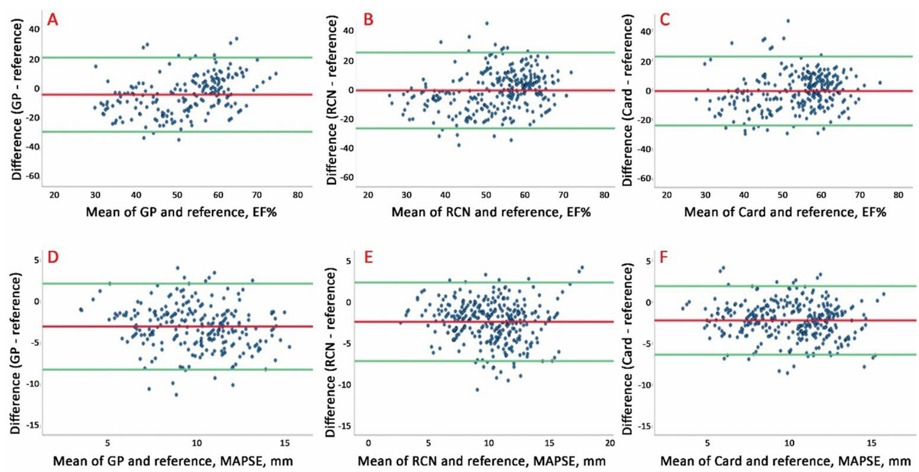


Figure 3 Bland-Altman plots illustrating agreement between the autoEF and autoMAPSE in recordings deemed acceptable for clinical use by evaluation of the blinded cardiologist (blinded image score ≥2). Upper panel: autoEF recorded by (A) GPs, (B) RCNs and (C) Card. Lower panel: autoMAPSE by (D) GPs, (E) 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

[1]Department of Internal medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway
[2]Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway
[3]Clinic of Cardiology, St. Olavs University Hospital, St. Olavs University Hospital, Trondheim, Norway
[4]Department of Research, Nord-Trøndelag Hospital Trust, Levanger, Norway
[5]Department of Health Sciences, Mid Sweden University, Östersund, Sweden
[6]Innherred 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.

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, Nicolaides 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.

Paper III

openheart

User experience and image quality influence on performance of automated real-time quantification of left ventricular function by handheld ultrasound devices: a diagnostic accuracy study with data from general practitioners, nurses and cardiologists

Anna Katarina Hjorth-Hansen, Malgorzata Izabela Magelssen, Garrett Newton Andersen, Torbjørn Graven, Jens Olaf Kleinau, Bodil Landstad, Lasse Løvstakken, Kyrre Skjetne, Ole Christian Mjølstad, Havard Dalen

To cite: Hjorth-Hansen AK, Magelssen MI, Andersen GN, *et al*. User experience and image quality influence on performance of automated real-time quantification of left ventricular function by handheld ultrasound devices: a diagnostic accuracy study with data from general practitioners, nurses and cardiologists. *Open Heart* 2022;**9**:e002083. doi:10.1136/openhrt-2022-002083

Received 6 July 2022
Accepted 26 September 2022

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

ABSTRACT

Background and objectives Echocardiography is the cornerstone of heart failure (HF) diagnosis, but expertise is limited. Non-experts using handheld ultrasound devices (HUDs) challenge the clinical yield. Left ventricular (LV) ejection fraction (EF) is used for assessment and grading of HF. Mitral annular plane systolic excursion (MAPSE) reflects LV long-axis shortening. Automatic tools for quantification of EF (autoEF) and MAPSE (autoMAPSE) are available on HUDs. We aimed to explore the importance of user experience and image quality for autoEF and autoMAPSE on HUDs, and how image quality influences the feasibility, agreement and reliability in patients with suspected HF.

Methods General practitioners, registered cardiac nurses and cardiologists represented the novice, intermediate and expert users, respectively, in this diagnostic accuracy study. 2543 images were evaluated by an external, blinded cardiologist by a five-parameter, prespecified score (four-chamber view, LV alignment, apical mispositioning, mitral annular assessment and number of visible endocardial segments) graded 0–6.

Results Feasibility was higher with increasing image quality. In all recordings, irrespective of user, the average image quality score and the five prespecified scores were associated with the feasibility of autoEF and autoMAPSE (all p<0.001). Image quality was more important for the feasibility of autoMAPSE than autoEF. Image quality was not important for the agreement of autoEF (R^2 2%) and autoMAPSE (R^2 7%). Combining all user groups, the reliability was lower with larger within-patient variability in image quality of the repeated recordings (p≤0.005). Similar associations were not found in user group specific analyses (p≥0.16). Patients' characteristics were only weakly associated with image quality score (R^2≤4%).

Discussion Image quality was important for feasibility but does not explain the low agreement with reference or the modest within-patient reliability of automatic decision-support software on HUDs for all user groups in patients with suspected HF.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Handheld ultrasound devices (HUDs) have been used by non-experts for a decade, while decision support software to aid in evaluation of cardiac function is recently introduced.

⇒ In patients with suspected heart failure, we aimed to study how user experience and image quality influenced automatic quantification of left ventricular function with respect to feasibility, agreement and reliability.

WHAT THIS STUDY ADDS

⇒ Image quality was positively associated with feasibility across inexperienced general practitioners, intermediate experienced registered cardiac nurses and experienced cardiologists. However, image quality did not explain a modest agreement and reliability of the automatic decision support software for quantification of ejection fraction and mitral annular excursion.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Further refinement of the automatic decisionsupport software is needed before implementation into clinical practice.

INTRODUCTION

Echocardiography is the cornerstone for diagnosis and follow-up of heart failure (HF), but echocardiographic expertise is limited to a few selected occupational groups. Left ventricular (LV) ejection fraction (EF) is widely used for assessing HF and grading of the severity.[1] Furthermore, mitral annular plane systolic excursion (MAPSE) is a sensitive and robust measure reflecting LV long-axis shortening and is less dependent on echogenicity.[2] Image quality is a mandatory prerequisite for a correct diagnosis by echocardiography where patient specific factors (eg, obesity, hyperinflated lungs and arrhythmias) will impair the acoustic environment and complicate image acquisition irrespective of the user's experience.[2] In addition, the users' experience may interfere with image quality.[3] Non-experts commonly use hand-held ultrasound devices (HUDs), which challenges the clinical yield of ultrasound diagnostics.[4] Advances in user support as real-time automatic measurements of cardiac structure and function may improve the diagnostic yield for non-experts, but initial results are conflicting.[5–7] How user characteristics and the quality of the recorded images influence the feasibility and reliability of automatic decision support software is not well known. To our knowledge, no study has explored the importance of user experience and image quality for quantification of EF (autoEF) and mitral annular systolic plane excursion (autoMAPSE) by HUDs.

The aim of this study was to evaluate how the users' experience may influence image quality of HUD recordings. In addition to how various categories of image quality may influence the feasibility, agreement and reliability of real-time automatic decision support software for quantification of LV function by HUDs.

METHOD

Population and study design

Patients with suspected HF referred for cardiac evaluation at Levanger Hospital, Norway, were invited to take part in the study. Exclusion criteria were age <18 years, previous cardiac imaging within the last decade and known HF. The inclusion period was between June 2018 and May 2020, including a pause from March to May 2020 due to the COVID-19 pandemic. All participants gave their informed, verbal consent prior to inclusion and written consent the day of inclusion.

The general practitioners (GPs) were selected by the municipality administration based on their previously defined positions in municipality organised healthcare. The registered cardiac nurses (RCNs) were chosen according to their position at the outpatient HF clinic at Levanger Hospital, where all available RCNs participated. All available board-certified cardiologists holding a position at Levanger Hospital participated as reference examiners. The participants were chosen irrespective of personal motivation and previous experience with

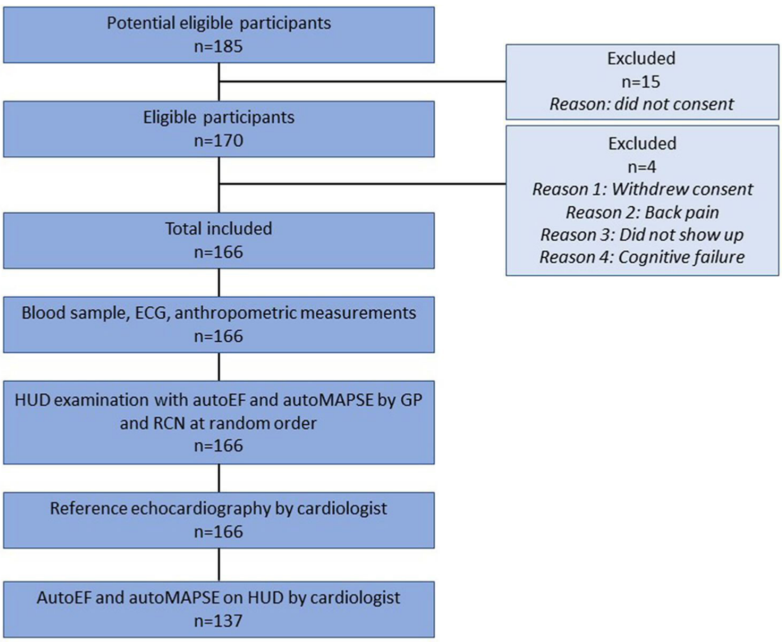


Figure 1 Patient flow through the study. Illustration of the patient inclusion, exclusion and order of examination throughout the study. In total, 160 patients were included. The first 29 patients were not examined with automatic HUD applications by the cardiologists due to logistic reasons. Abbreviations:AutoEF, automatic measurement of left ventricular ejection fraction by decision-support software; GP, general practitioner; HUD, handheld ultrasound device; autoMAPSE, automatic measurement of mitral annular plane systolic excursion by decision-support software; RCN, registered cardiac nurse.

Table 1 Image quality score

Score value	0	1	2	3	4	5	6
Four-chamber view		Two-chamber view, long axis view or others		Five-chamber view or posterior four-chamber view			Four-chamber view
LV alignment (misalignment)	≥45°		30°–44°		15°–29°		<15
Apical mispositioning			≥15 mm		<15 mm		None
Mitral annular assessment		Not judgeable	Poor	Fair	Good	Near excellent	Excellent
Number of visible endocardial segments		One	Two	Three	Four	Five	Six

This table represents five prespecified categories used to evaluate the image quality with the applied automatic decisionsupport software. Scores 0–6 were given for each category. The mean of each score per image stands for the overall quality of the image.
LV, left ventricular.

ultrasound. As shown in figure 1, the study participants were examined at random order by one of five GPs and one of three RCNs blinded to their respective results. One of five cardiologists performed comprehensive echocardiography serving as the reference method, as well as apical HUD recordings for comparison. All users used the decision support software for automatic measurements of LV function by the HUDs; however, the first 29 patients were not examined by HUD by the cardiologists due to logistic reasons. There were no other examinations organised by the study.

Training and education

Details of training and education are comprehensively described previously.[8] In short, the GPs, RCNs and cardiologists represented the novice, intermediate and expert operators, respectively. The novices (n=6) underwent six days of one-to-one training supervised by one of two experienced cardiology fellows in addition to two evening lectures. They had access to private HUDs in their day-to-day practice for the whole training and inclusion period. Only one of the GPs had previous experience with ultrasound diagnostics (only seven examinations). One GP changed occupation and did not take part in the study leaving five GPs for the analyses. The intermediate group (n=3) were experienced in evaluation of pleural effusion, the inferior vena cava and limited ultrasound examinations of the heart.[5] Thus, they did not undergo systematic training, but were instructed on how to use the HUDs and initialise the automatic algorithms approximately four weeks prior to inclusion. The expert group consisting of in-house cardiologist experienced in echocardiography (n=5) were instructed at the day of inclusion on how to use the automatic measurements but did not receive any further instructions or training.

Ultrasound examinations

The study specific protocol has been described previously.[8] All participants underwent three HUD examinations using a VScan Extend (GE Ultrasound AS, Horten, Norway) in addition to the reference echocardiography (Vivid E9 or E95 scanner, GE Ultrasound AS). All HUD

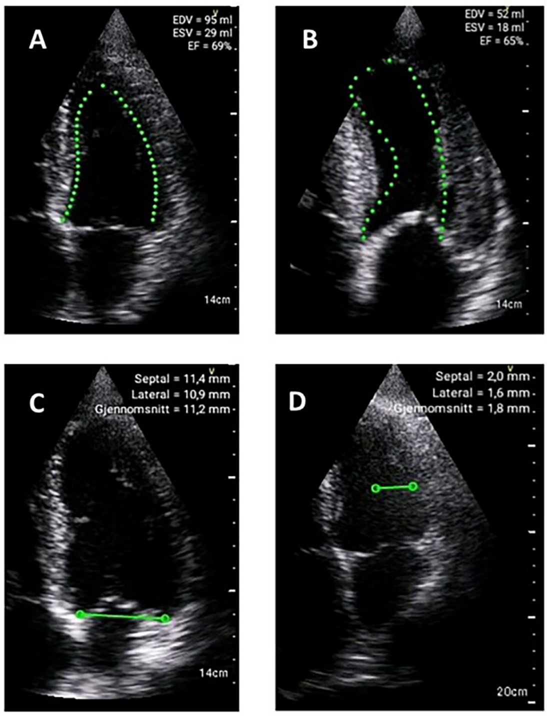


Figure 2 Examples of satisfactory and unsatisfactory decision support software results for ejection fraction and mitral annular plane systolic excursion. All images had an average score ≥5.6. (A) Satisfactory autoEF measurement with corresponding values. (B) Unsatisfactory autoEF measurement with values not reflecting true LV function. (C) Satisfactory autoMAPSE measurement with corresponding values. (D) Unsatisfactory autoMAPSE measurement with values not reflecting the true LV function. autoEF, automatic measurement of left ventricular ejection fraction by decision support software; autoMAPSE autoEF, automatic measurement of left ventricular mitral annular plane systolic excursion by decision-support software; LV, left ventricle.

Table 2 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 (67-89)
Haemoglobin (g/L)	144±15
N-terminal probrain natriuretic peptide (ng/L)	295 (66–864)
NYHA functional class, n (%)	
I	63 (37)
II	80 (47)
III	12 (7)
IV	1 (1)
Diuretics, n (%)	41 (25)
Beta blockers, n (%)	51 (31)
ACE inhibitor or angiotensin-receptor blocker, n (%)	32 (19)
Atrial fibrillation diagnosis, n (%)	49 (29)
Atrial fibrillation at day of inclusion, n (%)	40 (24)
COPD/asthma, n (%)	26 (16)
Diabetes mellitus type 2, n (%)	23 (14)
Coronary artery disease, n (%)	19 (11)

Normal distributed data are expressed as mean±SD, skewed data are presented as median (IQR) and proportions are n (%). Medications listed refer to the current use.
*Calculated by the Cockcroft-Gault equation.
COPD, chronic obstructive pulmonary disease; NYHA, New York Heart Association.

recordings by the three different user groups included apical four-chamber recordings with the addition of fully automated measurements of EF and MAPSE. All users ideally performed three separate recordings per automatic algorithm, a total of six four-chamber recordings per patient. The decision support software (autoEF or autoMAPSE) was then applied for fully automated measurements after each recording, and the analysed recordings were stored on the HUD. Subsequently, the recordings automatically transferred and stored on the Tricefy secure cloud-based server (Trice Imaging Inc, California, USA).

One of five cardiologists performed reference echocardiography on all patients in accordance with international recommendations for standardised echocardiographic examination.[9] LV endocardial borders were traced at end-diastole and end-systole in the four and two chamber views, and LV volumes and EF were calculated by the Simpson's biplane method. The mitral annular septal and lateral points in four-chamber views using motion mode measures MAPSE. All measurements were performed using EchoPAC SW Only, versions 202 and 203 (GE Ultrasound).

Automatic tools for quantification of LV function and image analysis

Details of the fully automated decision support software for quantification of LV function (autoEF and autoMAPSE) have been described elsewhere.[7 8 10] Shortly, the automatic measurements of LV end-diastolic volume (EDV), end-systolic volume and EF was done by the commercially available artificial intelligence aided LVivo EF software (DiA Imaging Analysis, Be'er Sheva, Israel). Fully automated tracing of the endocardial border in four-chamber recordings estimated LV volumes (figure 2). EF was calculated from the LV volume estimates based on the traces. The fully automated autoMAPSE software tracked the septal and lateral points of the mitral annulus in four-chamber recordings, and MAPSE was calculated as the average displacement of the septal and lateral points (figure 2).[7]

Image quality assessment

Image quality was evaluated by an external cardiologist experienced in echocardiography blinded to details of the operators and patients. All HUD recordings including either automatic desision-support software was reviewed. The image quality was scored by evaluating five prespecified categories, and the mean of the scores represents the averaged image quality score (table 1). Additionally, the external reviewer evaluated whether the automatic measurements were recommended for clinical use based on: (1) the image quality scores, (2) the quality of the tracking of the endocardial border for autoEF or the mitral annular points for autoMAPSE, respectively, and (3) the performance and numerical output of the autoEF and autoMAPSE algorithms. This was scored as following: (1) discard measurement (not for clinical use); (2) accept, but needs adjustment of the result due to suboptimal performance of the automatic software; (3) accept as it is.

During preliminary analyses by our group, detection of a system error in the autoEF software initiated a software revision by the vendor (LVivo EF, DiA Imaging Analysis, Be'er Sheva, Israel). The first 103 patients were examined with the first version of the autoEF software (version 1), and the following 63 patients were examined with the revised software (version 2).

Other data

Anthropometric measurements (body weight (kg), body height (cm) and blood pressure (mm Hg)) were measured, and New York Heart Association functional classification was scored by nurses the day of inclusion. Blood samples at the day of inclusion were analysed at the in-hospital accredited laboratory.

Statistics

Continuous variables were expressed as mean and SD or as median and IQR as appropriate. Normality was

Table 3 Pictures analysed per automatic function, per user and the average scores per parameter

	Novice	Intermediate	Expert
Automatic measurement of left ventricular ejection fraction			
Images, n	403	445	360
Total average score	3.7±0.9	4.2±0.9	4.8±0.8
Four-chamber view	3.9±1.7	4.7±1.7	5.1±1.5
LV alignment	4.7±1.3	5.5±1.0	5.6±0.8
Apical mispositioning	3.6±1.5	3.9±1.5	4.5±1.4
Mitral annular assessment	2.8±1.0	3.1±1.1	3.5±1.0
Number of visible LV endocardial segments	3.5±1.3	3.8±1.2	4.2±1.2
Automatic measurement of mitral annular plane systolic excursion			
Images, n	476	470	389
Total average score	3.2±1.9	3.7±0.9	4.4±0.8
Four chamber view	3.3±1.6	4.1±1.8	4.9±1.5
LV alignment	4.4±1.5	5.1±1.2	5.5±0.9
Apical mispositioning	3.0±1.3	3.3±1.4	4.0±1.4
Mitral annular assessment	2.5±0.9	2.9±1.0	3.5±1.0
Number of visible LV endocardial segments	2.8±1.3	3.2±1.3	3.9±1.3

The table illustrates the number and image quality score of the images by applications and users for the five prespecified parameters as well as the average score. All scores are given as mean±SD.
LV, left ventricular.

evaluated by histograms and Q–Q plots. Categorical variables were presented as frequencies and proportions. Student's t-test and Wilcoxon test were used for comparison of groups as appropriate. Proportions were compared using the χ^2 test and Fisher's exact test as appropriate. McNemar's test was used to compare paired nominal data. Repeated measure analysis of variance with post hoc Bonferroni correction was used to analyse variance in the groups. The influence of the image quality parameters with performance of the automatic applications, as well as patients' characteristics, was evaluated by logistic regression and general linear models as appropriate. The importance of image quality for feasibility and agreement with reference was first evaluated on the whole dataset of images from all three users and within the three user groups. The agreement with reference was assessed at the level of all available automatic measurements. The importance of the different image quality category for the within-patient reliability of the automatic applications was evaluated using the maximum difference in measurements of autoEF and autoMAPSE and the maximum difference in image quality scores. Analyses were performed in the whole dataset and within user groups. A p value <0.05 was considered statistically significant. All statistical analyses were performed using IBM SPSS Statistics, V.28 (SPSS Inc).

Initial calculations of sample size were 104 patients estimated by Sample Power (SPSS, Inc) based on diagnostic performance; however, in such a small population significant pathology would be scarce. Therefore, to account for likely low rate of pathological findings, the sample was expanded to 150 patients. Preliminary analyses revealed an error in the autoEF algorithm initiation a software upgrade, so recalculations of sample power led to an increase the population to 170 patients. No additional power analysis was performed in this study. The planned number of inclusions exceeds the number of participants needed for reliable evaluation of feasibility, reliability and agreement with paired analyses.

RESULTS

Study population

In total 185 patients with suspected heart failure were invited to take part, 15 did not consent, 1 withdrew consent, 1 could not complete the examinations due to back pain, 1 did not show up and 1 was excluded due to cognitive failure. In total, 166 participants were included in the analyses (figure 1). Population baseline characteristics are previously published but are shown in table 2.[11] Almost half the population was female, and mean age±SD was 70±13 years. Most patients were overweight with mean BMI±SD of 28.7±5.3 and a substantial proportion presented with atrial fibrillation (24%). Furthermore, chronic obstructive pulmonary disease (COPD) was present in 16% of patients.

Feasibility and image quality

In total, 2543 images were scored for assessment of image quality (table 3). Figure 3 shows that image quality score was consistently lower for novices versus intermediate experienced versus experts for both modalities. The image quality score was highest for the *LV alignment* and lowest for *mitral annular assessment*, with consistent findings across user groups and methods.

Feasibility was higher with higher image quality score. In univariate logistic regression analyses including all recordings irrespective of the user group, both the average image quality score and the five prespecified scores were associated with the feasibility of both automatic applications (all p<0.001). In multivariate analyses including all five image quality score categories, we found that all were significantly associated with the feasibility for autoEF (all p<0.001, except *four-chamber view* (p=0.02)). For autoMAPSE, *apical misposition* (p=0.94) and *number of visible LV endocardial segments* (p=0.06) were not significantly associated with the feasibility, while the other categories were (*four-chamber view*, p=0.046, others p<0.001).

Table 4 shows that image quality was more important for the feasibility of autoMAPSE than autoEF. Additionally, there was a gradient in adjusted R^2 ranging from 41% within novices to 22% within experts with respect to

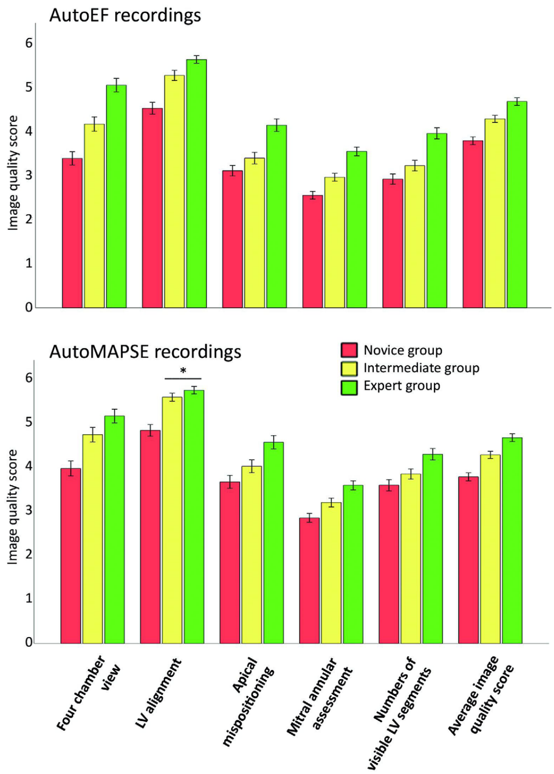


Figure 3 Image quality score parameters in handheld ultrasound recordings for automatic assessment of ejection fraction and mitral annular plane systolic excursion by user groups. figure 3 shows the image scores by user groups in recordings for automatic assessment of ejection fraction (autoEF) and mitral annular plane systolic excursion (autoMAPSE). All image scores were significantly different across user groups (p≤0.01) except for *LV alignment* between intermediate and expert users in autoEF recordings (*). autoEF, automatic measurement of left ventricular ejection fraction by decision support software; LV, left ventricle.

the feasibility of autoMAPSE, while no gradient was seen across user groups for autoEF. Among the image score categories, the *numbers of visible LV endocardial segments* were the most important predictor for the feasibility of autoEF for the two most experienced groups but not for novices. Correspondingly, *mitral annular assessment* explained most of the variability related to image quality for autoMAPSE across user groups.

The averaged image quality score was weakly associated with body mass index and systolic blood pressure for both autoEF and autoMAPSE (R^2≤4%, p≤0.04) when analysed in the whole dataset. Systolic blood pressure was not associated with image quality in experts (p≥0.42), while BMI showed stronger associations with image quality in experts (R^2 12% and 9% for autoEF and autoMAPSE, both p<0.001). In novices and the intermediate group, the associations with systolic blood pressure were very weak (R^2≤2%, p≤0.06).

Table 4 The importance of image quality score components for feasibility of automatic assessment of ejection fraction and mitral annular plane systolic excursion

	Novice	Intermediate	Expert
All (five) image quality parameters	EF 20%/MM 41%	EF 18%/MM 37%	EF 24%/MM 22%
Four-chamber view	MM 13%		
LV alignment	EF 13%/MM 17%	EF 8%/MM 10%	
Apical mispositioning			EF 2%
Mitral annular assessment	EF 10%/MM 33%	MM 32%	MM 20%
Number of visible LV endocardial segments		EF 13%	EF 14%

The table shows the proportion of explained variance (adjusted R^2) for feasibility of automatic assessment by autoEF and autoMAPSE according to user groups, respectively. Each of the prespecified image score parameters were evaluated in univariate log-linear regression analyses, and parameters significantly associated with feasibility are shown. Subsequently, the five prespecified image quality scores were included in a multivariate regression analysis, where all shown data were still significantly associated with feasibility with p<0.001, except for the four-chamber view in novices (p=0.02).
AutoEF, Automatic measurement of ejection fraction; AutoMAPSE, Automatic measurement of mitral annular plane systolic excursion; EF, ejection fraction; LV, left ventricular; MAPSE, mitral annular plane systolic excursion; MM, motion mode.

Image quality was not significantly associated with known hypertension or COPD (R^2<1%, p>0.09).

In analyses comparing the importance of image quality for the feasibility of the different autoEF software versions, minor differences were revealed. The adjusted R^2 for version 1 was 23% for novices, 19% for the intermediate group and 27% for experts, with corresponding R^2 of 32%, 19% and 30% for version 2.

Agreement of HUD recordings with reference and image quality

Figure 4 shows that the image quality of the HUD recordings was not important for the agreement of the automatic decision support software measurements compared with reference. Image quality of the HUD recordings explained only 2% of the variability (R^2=2%) between the autoEF and reference measurements in the whole dataset. In analyses within user groups, the findings were similar (R^2=1% for all three user groups). Furthermore, the associations of less underestimation by the decision support software on HUDs with better image quality were only significant for the novices and the intermediate group. Similarly, image quality of the HUD recordings explained only 7% of the variability between the autoMAPSE and reference measurements in the whole dataset. In analyses within user groups, we found a gradient in the explained variance ranging from 7% for novices to only 1% for experts, however still significant across user groups (p<0.05 for all except autoEF in expert group p=0.056).

Reliability of decision support software measurements on HUDs and image quality

Figure 5 illustrates the within-patient differences for repeated measurements of EF, EDV and MAPSE by the decision support software according to within-patient differences in image quality in the whole dataset. In analyses combining all user groups, there were significant

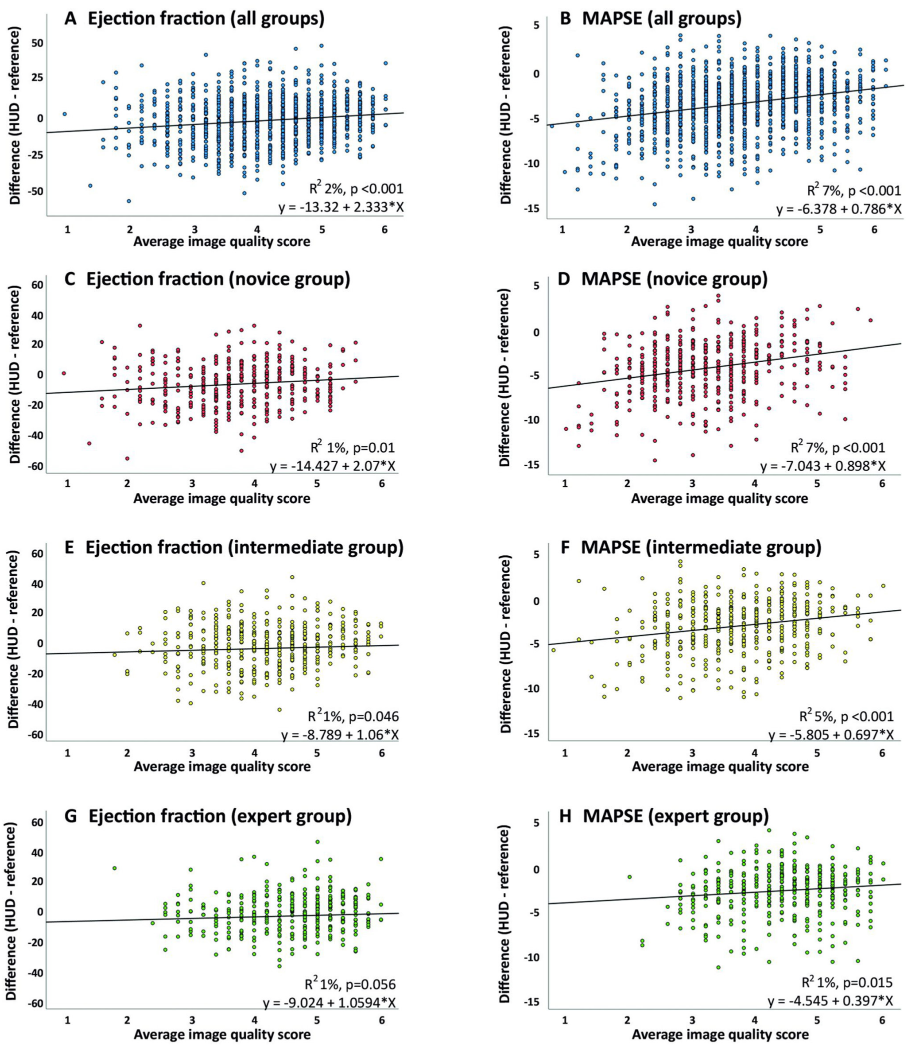


Figure 4 Agreement between measurements of automatic ejection fraction and MAPSE by different operators using handheld ultrasound and reference echocardiography according to image quality. Scatter plots with applied 'line of best fit' of the difference in automatic ejection fraction and MAPSE measurements by HUDs and reference according to user groups by average image quality score. Panels A–H show measurement and user group. Explained variance of the difference between measurements are shown by the R^2, and the equation for the best fitted line (y-axis) is showed by image quality score (x-axis). MAPSE, mitral annular plane systolic excursion; HUD, handheld ultrasound device.

associations of lower reliability with larger within-patient variability in image quality of the repeated recordings (all p≤0.005). In user group specific analyses, the reliability was not significantly associated with image quality for neither of the three specified measurements in experts (all p≥0.16). For the novices and intermediate group, the reliability was significantly associated with within-patient differences in image quality for decision support software measurements of EF, but not for EDV (p≥0.051) or MAPSE (p≥0.12).

Discussion

This study evaluated the influence of operators' experience and image quality for fully automatic decision-support software measurements of LV EF, EDV and MAPSE by HUDs in three user groups with varying experience. Blinded evaluation of 2543 four chamber HUD recordings by novices, intermediate experienced users and experts showed that image quality was significantly associated with the feasibility of the decision support software measurements. Image quality was more closely

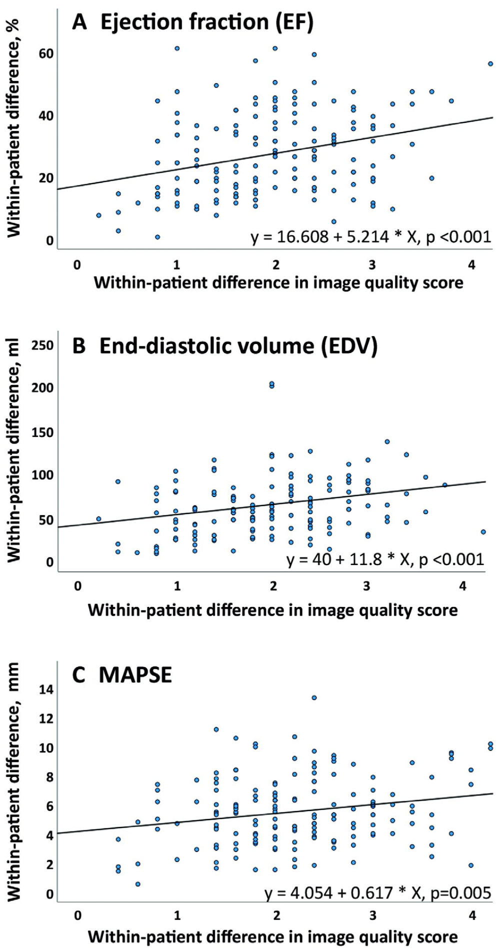


Figure 5 Within-patient differences in automatic measurements of ejection fraction, end-diastolic volume and MAPSE by handheld ultrasound devices plotted against the within-patient differences in image quality. The maximum within-patient difference for automatic HUD measurements of EF (A), EDV (B) and MAPSE (C) is plotted against the maximum within-patient difference in image quality score. Lines of best fit have been applied by linear regression calculation. MAPSE, mitral annular plane systolic excursion.

related to feasibility in the less experienced user groups and explained 18%–24% of the variability in feasibility of autoEF and volumes, and 22%–41% of the variability in the feasibility of autoMAPSE, respectively. Of five prespecified image quality categories, the *number of visible LV endocardial segments* was most important for autoEF and volumes, while *mitral annular assessment* was most important for autoMAPSE. Contradictory, the agreement of the automatic decision support software measurements was less dependent on image quality (adjusted $R^2 \leq 7\%$). Furthermore, image quality did not explain the low test–retest reliability of the decision support software measurements. In user-specific analyses, the reliability of the decision support software measurements was significantly associated with image quality only for the less experienced user groups for autoEF measurements.

Population

The finding of elevated blood pressure and body mass index in significant proportions are expected as they represent relevant risk factors for HF, and the population studied was referred to hospital for evaluation of suspected HF. Similarly, since the population includes both healthy and diseased individuals, the distribution of relevant patient characteristics are wider compared with more strictly selected samples.[5 7 12] The presented associations of image quality with body mass index and systolic blood pressure do not seem to be of clinical importance with respect to the study aims.

Decision support software and image quality. Until recently, evaluation of LV function on HUDs has been done by visual assessment ('eyeballing') only, which has several limitations.[13] Easy to perform focused cardiac ultrasound performed by inexperienced users on HUDs is feasible and has showed promising results.[4 12 14] Image quality is essential and a major challenge within all ultrasound diagnostics. As overweight, atrial fibrillation and COPD were common in the studied population this challenges the image quality of the ultrasound recordings. As shown in this study the image quality was closely related to the experience of the users, even though the body mass index and systolic blood pressure were of importance as well.

The feasibility of both automatic decision support software was higher when image quality score was higher. As shown, image quality influenced the feasibility of the automatic measurements more for the novices, and intermediate group, compared with the experts. This is related to less variation in image quality score for the experts and that the performance of the decision support software was not solely dependent on relevant image quality. The corresponding explained variance for the feasibility of autoMAPSE was nearly twice the explained variance for autoEF, indicating that the feasibility of autoMAPSE was closer associated with image quality. The *number of visible LV endocardial segments* category explained the majority of variance in feasibility for autoEF, and similarly the *mitral annular assessment* was most important for the feasibility of autoMAPSE. This finding is in line with clinical experience on echocardiographic requirements for EF and MAPSE. The finding of less influence of image quality for autoEF compared to autoMAPSE may be due to technological characteristics of the software. The autoEF software is assisted by artificial intelligence,[10] and it may be hypothesised that the training of the algorithm was not optimal for the HUD recordings used in this study. Second, the autoMAPSE software used grayscale images only, while the robustness of MAPSE is commonly shown for methods using tissue Doppler.[15]

Agreement and reliability

In a recent publication, we showed that the coefficient of repeatability for the presented automatic decision support software ranged 19%–24% (*reference article is not yet published but is currently in for review*). This is significantly higher than shown by experts using high-end ultrasound equipment[16], and a recent publication showing limits of agreement 14.5% using artificial intelligence assisted decision support software for assessment of EF by a novel HUD.[6] In the latter study, image quality evaluated by the number of LV walls where the endocardial border was not clearly identifiable in end-diastole did not significantly influence the agreement with reference.[6] By exploring the unacceptably high variability of the automatic desicion-support software presented in this study, only a minor part of the low agreement was explained by image quality. As shown by figure 4, both automatic decision-support software underestimated EF and MAPSE more compared to reference when image quality was low. Importantly, even for the experts where image quality overall was good, the agreement with reference was below recommendation for clinical use, and only 1% of the variation compared with reference was explained by image quality. Thus, this adds in the disfavour of clinical implementation of the presented automatic decision support software for LV evaluation on HUDs.

Similarly, for autoMAPSE, there was a linear relation of more underestimation compared with reference when image quality was low for all user groups. Still, the variability was too high even when image quality was good, and overall, only 7% of the variability compared with reference was due to image quality.

Adding to the low agreement of the automatic measurements by HUDs, we found that only a minor part of test–retest reliability was caused by differences in image quality. For the experts, we have recently published moderate to good intrarater intraclass correlation (0.72 for autoEF and 0.83 for autoMAPSE), and within-patient differences in image quality did not explain the modest reliability (p=0.16 for EF, p=0.45 for EDV and p=0.99 for MAPSE) (*reference article is not yet published but is currently in for review*). To our knowledge, image quality of repeated recordings and its relevance for the reliability of automatic decision-support software measurements of LV function has not been evaluated on HUDs previously. Figure 5 shows the importance of image quality for the reliability within patients, but these associations were not always present when performing the analyses per user groups. Importantly, in the experts' recordings, we found no signs that higher image quality improved the reliability of the automated decision-support software. This shows the inconsistency of the automatic measurements, indicating image quality alone not to be sufficient for reliable performance of the automatic decision-support software. Two other studies have evaluated the agreement of automatic evaluation of LV EF by HUDs.[6 10] However, direct comparison is difficult as the published data on image quality characteristics were scarce in these studies and both included only one experienced operator each.

Until recently no decision support software for evaluation of LV function evaluation has been available on HUDs. Automatic decision-support software for estimation of EF performed by experts has showed promising results in recent publications.[6 10] Furthermore, in a previous publication from our group, we showed a slight underestimation of autoMAPSE compared with reference.[7] However, experts do not usually seek or require decision support. Differences in the studied populations in LV function, arrhythmias and body composition may partly explain the differences between the studies. Further, we evaluated the two versions of the automatic decision-support software for EF calculations but to be consistent with the planned study aims, we did not reanalyse the patients analysed by the first software version. In the future, additional refinement of decision support software based on better training of the algorithms and artificial intelligence may improve the software. More advanced decision support software including deformation analyses will also be available for HUDs.

Strengths and limitations

The main strength of this study is the comprehensive blinded analyses of five distinct categories of image quality and the performance of the decision support software. Another strength is that the recruitment of inexperienced operators was based on positions in the community healthcare system and not based on motivation for participation. Furthermore, the three groups of operators (in total 13 different users) had different experience ranging from no previous experience to level III experience according to the American Society of Echocardiography.[17] However, with respect to reduce potential bias related to the user specific results, even larger groups of operators would have been preferred. The most important limitation is that we only reviewed images being able to run the decision support software. Thus, cases where the cardiac cycle or image quality did not allow for the applications to run were consistently excluded from the image quality analyses. This may influence the results between the user groups, as less recordings were able to run the decision support software among the less experienced user groups. Even though the cardiologists did not perform HUD examinations on the first 29 participants, the findings across user groups were consistent also in analyses of subgroups (data not shown).

CONCLUSIONS

Image quality was important for the feasibility of decision support software for automatic analyses of left ventricular ejection fraction, volumes and mitral annular plane systolic excursion by novices, intermediate experienced and expert groups using HUDs in a population with suspected heart failure. However, neither the low agreement with reference nor the modest within-patient reliability are explained solely by image quality. Further refinement of the decision support software is warranted before implementing these into everyday practice for non-expert users of HUDs.

Contributors AKH-H is the main author and has been involved in study design, data collection, data analyses, drafting and revision of the manuscript. MIM contributed in data collection, drafting and revision of the manuscript. GNA, TG, JOK and KS collected data and revised the manuscript. BL was involved in funding of the study, supervision of the first author and revision of the manuscript. LL and OCM contributed to the study design and revised the manuscript. HD provided funding and designed the study, participated in data collection and revision of the manuscript, and acts as the guarantor of this manuscript. All authors provided final approval of the manuscript version to be published. All authors have agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding The study was funded by grants from the European Interreg A initiatives (Norwegian-Swedish initiative), Research Council of Norway and Nord-Trøndelag Hospital Trust, Norway.

Competing interests This work was supported by GE Ultrasound lending the HUD devices through a research contract with the project leader (HD). GE Ultrasound had no role in performance of the study, including data correction, data interpretation or drafting and revision of the manuscript. MIM, OCM, LL and HD hold positions in Centre for Innovative Ultrasound Solutions where GE Ultrasound is one of the industrial partners. LL acts as part-time consultant for GE Ultrasound.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Cristin-prosjekt-ID: 569755. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Deidentified participant data are available from HD (ORCID 0000-0003-1192-3663) upon reasonable request.

ORCID iDs
Anna Katarina Hjorth-Hansen http://orcid.org/0000-0003-4626-2113
Havard Dalen http://orcid.org/0000-0003-1192-3663

REFERENCES

1. Ponikowski P, Voors AA, Anker SD, *et al*. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;2016:2129–200.
2. Hu K, Liu D, Niemann M, *et al*. Methods for assessment of left ventricular systolic function in technically difficult patients with poor imaging quality. *J Am Soc Echocardiogr* 2013;26:105–13.
3. Popescu BA, Stefanidis A, Fox KF, *et al*. Training, competence, and quality improvement in echocardiography: the European association of cardiovascular imaging recommendations: update 2020. *Eur Heart J Cardiovasc Imaging* 2020;21:1305–19.
4. 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.
5. 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.
6. 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. *EHJDH* 2022;3:29–37.
7. 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.
8. Magelssen MI, Hjorth-Hansen AK, Andersen GN, *et al*. The clinical influence of hand-held ultrasound examinations by general practitioners in patients with suspected heart failure supported by tools for automatic quantification and telemedicine. *Eur Heart J Cardiovasc Imaging* 2022;23:2022.
9. 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:e14:1–39.
10. 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.
11. 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.
12. 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.
13. Marwick TH. Ejection fraction pros and cons: JACC state-of-the-art review. *J Am Coll Cardiol* 2018;72:2360–79.
14. 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.
15. 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.
16. Baron T, Berglund L, Hedin E-M, *et al*. Test-retest reliability of new and conventional echocardiographic parameters of left ventricular systolic function. *Clin Res Cardiol* 2019;108:355–65.
17. Wiegers SE, Ryan T, Arrighi JA, *et al*. 2019 ACC/AHA/ASE advanced training statement on echocardiography (revision of the 2003 ACC/AHA clinical competence statement on echocardiography): a report of the ACC competency management Committee. *Circ Cardiovasc Imaging* 2019;2019:12.

Paper IV

Ultrasound in Medicine & Biology 000 (2022) 1−8

Contents lists available at ScienceDirect

Ultrasound in Medicine & Biology

journal homepage:

Original Contribution

Clinical Influence of Handheld Ultrasound, Supported by Automatic Quantification and Telemedicine, in Suspected Heart Failure

Malgorzata Izabela Magelssen [a,b,*], Anna Katarina Hjorth-Hansen [a,c], Garrett Newton Andersen [c], Torbjørn Graven [c], Jens Olaf Kleinau [c], Kyrre Skjetne [c], Lasse Løvstakken [a], Håvard Dalen [a,b,c], Ole Christian Mjølstad [a,b]

[a] Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway
[b] Clinic of Cardiology, St. Olavs University Hospital, Trondheim, Norway
[c] Department of Internal Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

ARTICLE INFO

Keywords:
Echocardiography
Heart failure
General practitioner
Diagnosis
Ejection fraction
Mitral annular plane systolic excursion
Telemedicine

Early and correct heart failure (HF) diagnosis is essential to improvement of patient care. We aimed to evaluate the clinical influence of handheld ultrasound device (HUD) examinations by general practitioners (GPs) in patients with suspected HF with or without the use of automatic measurement of left ventricular (LV) ejection fraction (autoEF), mitral annular plane systolic excursion (autoMAPSE) and telemedical support. Five GPs with limited ultrasound experience examined 166 patients with suspected HF (median interquartile range = 70 (63−78) y; mean ± SD EF = 53 ± 10%). They first performed a clinical examination. Second, they added an examination with HUD, automatic quantification tools and, finally, telemedical support by an external cardiologist. At all stages, the GPs considered whether the patients had HF. The final diagnosis was made by one of five cardiologists using medical history and clinical evaluation including a standard echocardiography. Compared with the cardiologists' decision, the GPs correctly classified 54% by clinical evaluation. The proportion increased to 71% after adding HUDs, and to 74 % after telemedical evaluation. Net reclassification improvement was highest for HUD with telemedicine. There was no significant benefit of the automatic tools ($p \geq 0.58$). Addition of HUD and telemedicine improved the GPs' diagnostic precision in suspected HF. Automatic LV quantification added no benefit. Refined algorithms and more training may be needed before inexperienced users benefit from automatic quantification of cardiac function by HUDs.

Introduction

Symptoms of heart failure (HF) are non-specific and present a challenge to the diagnostic workflow. One of six patients older than 65 years presenting to their general practitioner (GP) with dyspnea on exertion will have undiagnosed HF [1,2]. Early and correct diagnostics are essential to improve patient care and to reduce the burden on the health care system. Echocardiography is the cornerstone of HF diagnostics and the method of choice when evaluating function of the left and right ventricles [2,3]. Handheld ultrasound devices (HUDs) are established diagnostic tools that enable on-site imaging [4]. After a period of training, less experienced users can evaluate cardiac morphology and function by HUDs [5−7]. The European Association of Cardiovascular Imaging (EACVI) supports the use of HUDs to screen for cardiac pathology under the condition that proper training has been performed [4].

We believe that inexperienced HUD users would benefit from automatic quantification of left ventricular (LV) function and telemedical support by an expert when assessing patients with possible HF. Evaluation of LV ejection fraction (EF) and mitral annular plane systolic excursion (MAPSE) are two methods for quantification of LV function [8,9]. Automatic measurement of EF is commercially available and implemented on HUDs. An algorithm for automatic quantification of MAPSE has been implemented on HUDs for research purposes [10]. Telemedicine is an established method in cardiology and other specialties [11−13] and feasible when evaluating cardiac function [7,12]. Automatic LV quantification tools, implemented on HUDs have until now only been tested in a few single-center studies with one operator each [14,15].

The primary aim was to evaluate the clinical influence of HUD examinations by GPs in patients with suspected HF without or with the use of supportive tools such as automatic quantification of EF/MAPSE and telemedical support compared with experienced cardiologists' decisions based on comprehensive assessment of medical history, clinical evaluation and echocardiography as reference.

Methods

The study was conducted at the outpatient cardiology clinic at Levanger Hospital, Levanger, Norway, from January 2018 until June

* Corresponding author. Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
E-mail address: malgorzata.izabela.magelssen@ntnu.no (M.I. Magelssen).

https://doi.org/10.1016/j.ultrasmedbio.2022.12.015
Received 19 April 2022; Revised 18 November 2022; Accepted 22 December 2022

2020. Inclusion was paused during part of the COVID-19 pandemic (March−May 2020).

Study population

Patients with suspected HF referred to the outpatient clinic at Levanger Hospital were invited to participate in the study. Eligible patients were contacted by study researchers and gave their oral and written consent. The inclusion criteria were suspicion of HF, N-terminal pro-brain natriuretic peptide (NT-pro-BNP) ≥125 ng/L, age >18 y and ability to give consent. The exclusion criteria were known HF or known results from cardiac imaging examination within the last 10 y. The study was performed in conformity with the policy statement for the use of human subjects of the Declaration of Helsinki. It was approved by the regional committee for medical and health research ethics (REK 2017/2054) and registered in the ClinicalTrial.gov database (identifier: NCT03547076).

Study design

On arrival at the outpatient clinic, blood samples (NT-pro-BNP, creatinine, sodium, potassium and hemoglobin), blood pressure and an electrocardiogram (ECG) were taken. Five GPs with limited experience in echocardiography were randomly selected by the study administration to participate in the study. They had access to the initial patient referrals and the ECGs, but not in-hospital medical records. The GPs examined patients in chronological order. A standard clinical evaluation including patient history and a physical examination was performed, followed by a focused cardiac ultrasound examination by HUD, addition of automatic quantification tools and, finally, telemedical supportive image analyses by out-of-hospital cardiologists. At each stage, the GPs considered whether the patients had HF and whether they would refer them for a cardiac examination. Because of laboratory delay, NT-pro-BNP was not always available for the GPs during their examination. After the GP evaluation, all patients were examined by one of five in-hospital cardiologists who performed a complete reference echocardiography (Fig. 1). Five to eight patients were included per 30 inclusion days.

Education and training

The GPs received six days (6 h/d) of practical training and two theoretical lectures. The training focused on visualizing parasternal long- and short-axis, apical four-chamber and subcostal views by HUD, as well as the inferior vena cava (IVC) and pleural cavities with respect to pleural effusion. On average, seven HUD examinations were performed per day. In addition, The GPs performed on average 13 unsupervised focused ultrasound examinations by HUD in their daily practice.

Handheld ultrasound

The focused ultrasound was performed using Vscan Extend (GE Ultrasound, Horten, Norway) with the capability of storing cine loops of one cardiac cycle without the need for ECG [16]. The commercially available LVivo application (DiA Imaging Analysis Ltd, Be'er Sheva, Israel) for automatic EF quantification (autoEF) was implemented on the HUDs. It detects the LV endomyocardial wall in the apical four-chamber view (Fig. 2) [14,17]. A customized research version of the automatic MAPSE (autoMAPSE) application was implemented on the HUD. MAPSE was calculated by tracking the basal LV points' movement using a customized method implemented on the HUDs—real-time contour tracking library (RCTL) (Fig. 2). The method was originally developed by our group a decade ago [10]. The specific details were recently described [18]. The RCTL provides segmentation of the left ventricle using a model composed of 12 control points which are updated by detecting the LV border in 75 equally spaced edge profiles. When tracking is enabled, the RCTL returns the septal and lateral points of the mitral annulus. The operators are unable to see, control or adjust the segmentation process. However, the tracking of the basal LV points and a line between these two points during the whole cardiac cycle is presented to the operator of the HUD (Fig. 2). There was no automatic feedback of the quality of the recordings or the robustness of the measurements for the autoEF or the autoMAPSE software. The tracking of the regions of interest throughout a cardiac cycle was available to the users. Examinations were performed with patients in the left lateral supine position and included the main cardiac views (parasternal long and short axis, apical four chamber and subcostal) and identification of IVC and pleural cavities. LV function was categorized by the GPs as normal, moderately reduced or severely reduced, while the cardiologists categorized LV function by EF (≤40%, 41%−49% or ≥50%) [9]. The IVC was described as dilated or not. The presence of pericardial and pleural effusion was evaluated. GPs were instructed to measure autoMAPSE and autoEF three times each per patient.

Telemedicine

Pseudonymized images stored on the HUDs, were transferred in near real time to a cloud-based server (Trice Imaging, Inc., Del Mar, CA, USA), an integrated software that allows for secure and anonymous sharing of medical images [19]. One of two external cardiologists (both localized at St. Olav's Hospital, Trondheim University Hospital,

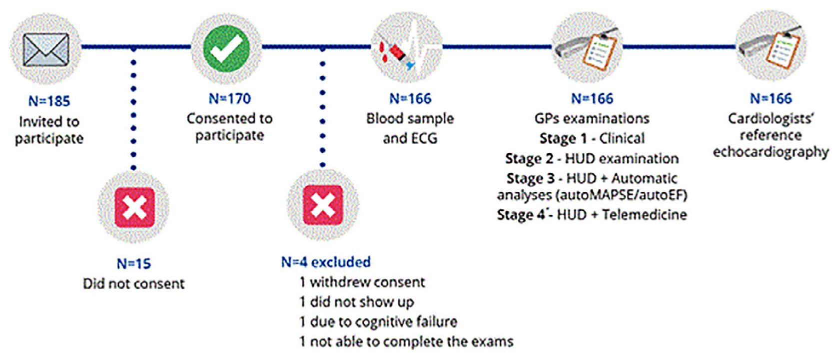


Figure 1. Flowchart of study participants. Real time response from cardiologist used in GPs' decision making. GPs decided whether the patients had heart failure at each stage. ECG, electrocardiogram; autoEF, automatic analyses of ejection fraction; autoMAPSE, automatic analyses of mitral annular plane systolic excursion; GP, general practitioner; HUD, handheld ultrasound device.

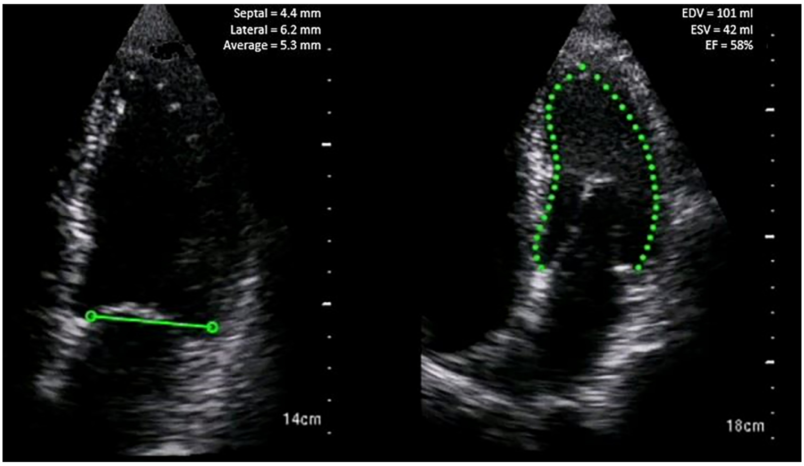


Figure 2. Tools for automatic quantification of left ventricular function. Automatic quantification of left ventricular function by mitral annular plane systolic excursion (left panel) and ejection fraction (right). Values of the systolic excursion of the mitral annulus are annotated at top of the left panel, and values for the automatic quantification of ejection fraction and volumes are at top of the right panel. EDV, end-diastolic volume; EF, ejection fraction; ESV, end-systolic volume.

Trondheim, Norway) downloaded the recordings to EchoPAC SWO (Version 203, GE Ultrasound, Horten, Norway) for interpretation. The cardiologists had access to the initial referrals, but not the results from the clinical evaluation, ECG, blood samples or hospital records. They provided feedback to the GPs electronically.

Reference examinations by cardiologists

A comprehensive echocardiographic evaluation was performed within 1 h after the GP`s examination. The echocardiograms were recorded using high-end equipment (Vivid E9 or E95, GE Ultrasound) and included the main cardiac views [9]. EF was measured using Simpson's biplane in apical four- and two-chamber views. Systolic LV function was defined as normal if the EF was >50%, mild to moderately reduced and significantly reduced if the EF was 40%−49% and <40%, respectively. MAPSE was measured in the septal and lateral mitral annular points by M-mode or reconstructed M-mode. All measurements represent the average of three consecutive cardiac cycles [9]. The echocardiographic analyses were performed using EchoPAC SWO (Version 202, GE Vingmed Ultrasound).

Statistics

Continuous data are expressed as the mean ± standard deviation (SD) or median (interquartile range [IQR]) as appropriate. Categorical data are expressed as frequencies and percentages.

A difference of <15% of correctly identified patients with and without HF between the GPs and reference cardiologists was considered of little importance. Use of Sample Power (SPSS, Inc., Chicago, IL, USA) resulted in a sample size of 104 and power of 0.80 ($p < 0.05$). The number of patients with significant pathology was expected to be small; furthermore, we expected failure of both GPs' recordings and some of the automatic measurements. Thus, a sample size of 150 was chosen. Throughout the study period, the autoEF software was upgraded and the sample size was adjusted to 170 to account for the new version.

McNemar's test was used to compare paired nominal data. An indication of the prognostic gain was calculated using net reclassification improvement (NRI), which is aimed at quantifying whether a new method or marker improves the prediction of a disease [20]. For this test, the clinical examination was used as a reference. NRI combines NRI event and NRI non-event. NRI event represents the net reclassification proportion of HF diagnosis among those with HF after each stage (clinical and HUD, clinical, HUD and addition of automatic tools, clinical, HUD, automatic tools and addition of telemedicine) compared with clinical examination alone. Similarly, an NRI non-event represents the net reclassification proportion in non-events (non-HF). The overall NRI is the sum of the net proportions of correctly reclassified exams. The positive (PPV) and negative (NPV) predictive values were calculated as the probability of correct classification.

A p value <0.05 was considered to indicate statistical significance for all analyses. SPSS (version 26, IBM Corp, Armonk, NY, USA) and Excel were used for the analyses.

Results

Study population

Of 185 patients invited, 170 agreed to participate in the study. After exclusion, 166 (78 women) completed the exams (Fig. 1). Table 1 outlines the population characteristics. Median (IQR) age was 70 (63−78) y, mean BMI 29 ± 5 kg/m^2 and median NT-pro-BNP 298 (65−870) ng/L. Patients with HF had a median NT-pro-BNP of 1302 (866−2626) ng/L compared with 148 (53−525) ng/L in patients without HF. Sinus rhythm was present in 126 (76%) patients. Systolic function was preserved in 131 (79%) patients; 21 (13%) had a mild reduction in LV function and 14 (8%) had moderately or severely reduced function. Mean EF was 53 ± 10%. Of the 166 patients, 118 (71%) presented with dyspnea, 24 (14 %) had reduced physical capacity, 40 (24 %) had peripheral edema, 22 (13 %) had fatigue and 16 (10 %) experienced palpitations.

Heart failure

Reference cardiologists diagnosed 28 patients with HF, excluded HF in 130 patients and were uncertain about the diagnosis in the remaining 8. Of the 28 patients, 13 (46%) had HF with preserved EF (HFpEF). After the clinical examination, the GPs correctly classified 92 (55%) patients (15 with HF, 77 without). The corresponding numbers increased to 118 (71%) after HUD examinations (19 with HF, 99 without) and 123 (74%) after telemedical evaluation (20 with HF, 103 without) (Fig. 3). The difference between the clinical examination and HUD or telemedicine was highly significant (p <0.001), but that between HUD and telemedicine was not (p = 0.44). There was no improvement in diagnostic precision after

Table 1
Baseline characteristics of the study population

Variable[a]	Entire population	Heart failure[b]
Number	166	28
Age, y	73 (63−78)	77 (71−80)
Female sex, n (%)	78 (47)	11 (39)
Height, cm	172 ± 10	174 ± 9
Weight, kg	85 ± 19	90 ± 23
Body mass index, kg/m^2	29 ± 5	29 ± 6
NT-pro-BNP,[c] ng/L	298 (65−870)	1302 (866−2626)
Creatinine,[c] μmol/L	84 (73−97)	89 (83−116)
Heart rate, bpm	77 ± 16	86 ± 24
Sinus rhythm, n (%)	126 (76)	13 (46)
Ongoing atrial fibrillation, n (%)	40 (24)	15 (54)
Bundle branch block, n (%)	15 (9)	5 (18)
Systolic blood pressure, mm Hg	150 ± 22	142 ± 25
Diastolic blood pressure, mm Hg	83 ± 11	83 ± 9
Hypertension, n (%)	60 (35)	10 (36)
Diabetes mellitus type 2, n (%)	23 (14)	3 (11)
Coronary artery disease, n (%)	19 (11)	4 (14)
Valvular heart disease, n (%)	2 (1)	1 (4)
COPD/asthma, n (%)	26 (16)	6 (21)
Diuretics, n (%)	41 (25)	10 (36)
Beta blockers, n (%)	51 (31)	14 (50)
ACEI/ARB, n (%)	32 (19)	6 (21)
LV EF biplane, %	53 ± 10	44 ± 13
LV EDV, mL	106 ± 44	109 ± 38
LV ESV, mL	50 ± 27	64 ± 33
Left atrial ESV index, mL/m^2	42 ± 16	54 ± 13
MAPSE,[d] mm	12 ± 3	8 ± 2

ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor inhibitor; COPD, chronic obstructive pulmonary disease; EDV, end-diastolic volume; EF, ejection fraction; eGFR, estimated glomerular filtration rate; ESV, end-systolic volume; LV, left ventricle; MAPSE, mitral annular plane excursion rate; NT-pro-BNP, N-terminal pro B-type natriuretic peptide.

[a] All values are expressed as the mean ± standard deviation unless otherwise specified.

[b] Heart failure diagnosis according to the reference examination.

[c] Median (interquartile range) or specified elsewhere.

[d] Mean of lateral and septal MAPSE. Echocardiographic data are from the reference examination.

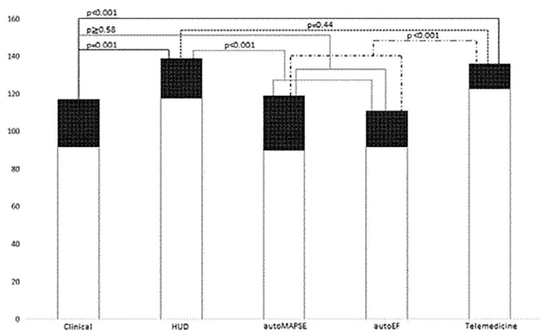


Figure 3. Diagnostic precision of heart failure diagnosis by general practitioners. Total number of patients with correct and incorrect classification after each stage of the examination. The columns include patients with and without heart failure. *Black* reflects incorrect classification; *white* reflects correctly classified individuals. Agreement between the stages was calculated with McNemar's test. The group with uncertain diagnoses is not included and reflects the proportion between 166 and the presented sum of correct and incorrect classifications. autoEF, automatic analyses of ejection fraction; autoMAPSE, automatic analyses of mitral annular plane systolic excursion; HUD, handheld ultrasound device.

adding automatic quantification (57% correctly classified after autoMAPSE and 55% after autoEF). AutoMAPSE and autoEF were not performed in 16 and 34 cases, respectively. In these cases, the GPs failed to run the applications because of suboptimal ultrasound images. On the basis of symptoms and physical examination, the GPs suspected 7 of the 13 patients with HFpEF of having HF. The NRI for HUD and telemedicine was 0.10 and 0.19, respectively. These data were based on 3.6% and 12.0% of correctly reclassified HF patients after HUD and telemedicine, respectively (NRI for events) and 6.5% (both HUD and telemedicine) correct reclassification of non-events (Fig. 4). The GPs were uncertain of the diagnosis in 43 patients after the initial assessment. There was a statistically significant decrease ($p < 0.05$) in uncertain cases after adding HUD and telemedicine (20 and 24, respectively), with no significant difference between the two ($p = 0.44$). There was non-significant reduction after autoMAPSE and autoEF (36 and 40 uncertain cases, respectively). The NPV was high at all stages (>0.91), while the PPV was lowest for autoMAPSE (0.40) and highest for telemedicine (0.71) (Table 2). When we evaluated the importance of the autoEF upgrade, we found no improvement in NPV and PPV (0.96 vs. 0.86 and 0.58 vs. 0.38, respectively). However, similar differences were found for autoMAPSE (data not shown). Coefficients of variation (COVs) for the GP recordings according to the reference were recently reported [21]. The COVs for autoEF and autoMAPSE were 15.4% and 24.3%, respectively.

Referral of patients

Even though all patients underwent a reference echocardiogram, the GPs still had to state whether they would refer the patients for a cardiac examination. Because of logistics, they were not presented with this possibility on the first day of inclusion. They intended to refer 113 (68%) patients after the clinical examination. They suspected HF in 35 of these patients, while in the remaining 78 (69%) there were other reasons for the referral. There was a significant decrease in the total number of referrals after addition of HUD, automatic quantification and telemedicine (all p values <0.02) (Fig. 5).

There was a non-significant decrease in the proportion of referred patients with suspected HF after adding HUDs, autoMAPSE and autoEF (31, 35 and 26 patients, respectively). After telemedical support, 23 of 98 (60%) referred patients were suspected of having HF (difference vs. clinical examination $p = 0.02$). Of the 23 patients, 21 (91%) were diagnosed with HF by the reference cardiologist. Of the 8 patients with a missed HF diagnosis, all but one was referred for cardiac evaluation (Table 3).

Discussion

In patients with suspected HF, the proportion of patients correctly diagnosed by GPs improved by ≥25% after HUD and telemedical support. The number of uncertain cases was reduced by approximately 50%. Adding automatic quantification tools for MAPSE or EF did not improve diagnostic precision. After telemedical support, the GPs would refer 35% fewer patients with suspected HF, and among those selected for referral, >90% were finally diagnosed with HF.

Population

In the present population, 28 (17%) were diagnosed with HF. The results agree with previously published studies in which one in six patients older than 65 years presenting to their primary care physician with dyspnea had undiagnosed HF [1]. In our study, 71% of the patients experienced dyspnea, either at rest or on exertion. The proportion of HF highlights that HF symptoms are unspecific and overlap with other disorders [2]. Improved selection of patients would allow for better use of restricted health care resources. NT-Pro-BNP has a high NPV and improves the ability to rule out HF. In our study, NT-pro-BNP was not always present during the GP assessment because of laboratory delay but the number of false-positive cases in patients with NT-pro-BNP below the usual threshold of 125 ng/L varied from 1 to 6 depending on the diagnostic stages. As

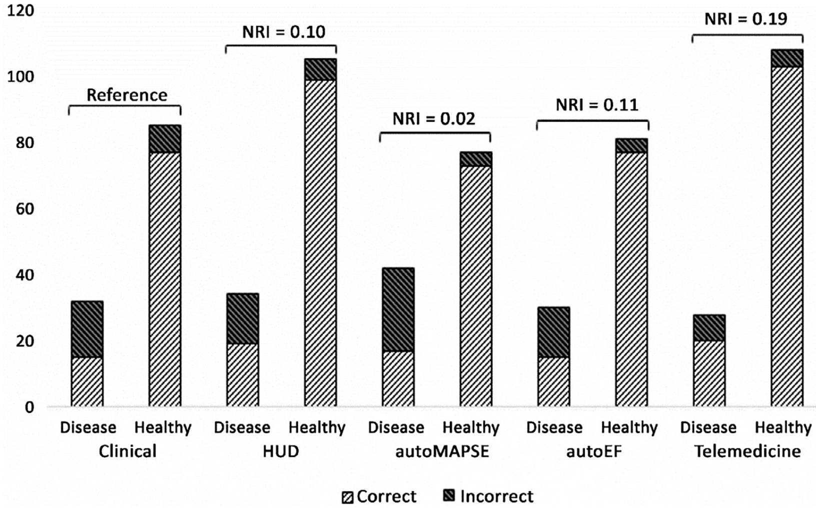


Figure 4. Reclassification of heart failure diagnosis. Number for correct/incorrect classification of patients with heart failure diagnosis dichotomized to "yes" or "no." Net reclassification improvement is shown at each stage (clinical + HUD, clinical + HUD + autoMAPSE/autoEF, clinical + HUD + autoMAPSE/autoEF + telemedicine) compared with clinical examination alone. autoEF, automatic analyses of ejection fraction; autoMAPSE, automatic analyses of mitral annular plane systolic excursion; HUD, handheld ultrasound device; NRI, net reclassification improvement.

NT-pro-BNP has a lower PPV, its availability would not necessarily reduce the number of false-positive diagnoses [22]. Moreover, it has been reported that only 25% of patients with NT-pro-BNP above the threshold of 125 ng/L and/or pathological ECG had HF [1]. The sensitivity and specificity of any test are influenced by the distribution of the studied population, and this also relates to the diagnostic performance of NT-pro-BNP, as well as the diagnostic decision-support software used in the present study. The results indicated that inclusion of HUD examinations improved the GPs' diagnostic precision, whereas the GPs were not able to improve their practice by adding the automatic tools providing measurements of LV function. There was no significant difference between diagnostics after the clinical examination and after adding automatic quantification tools. 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 in 24%) may have interfered with image acquisition and the accuracy of the automatic measurements. As the present population had characteristics similar to those of other studies evaluating HUDs and HF [7,23,24], we believe that the results are applicable to others.

Training

The skills and competence of users are important for operator-dependent diagnostics [4,25]. The training program was based on previous studies and recommendations [5−7,26]. After completing the program, the GPs were able to perform and interpret cardiac ultrasounds with improvement in diagnostic precision. However, they were not able to interpret and adjust to the false-positive (and false-negative) results provided by the algorithms. Future work is needed to determine if more training would improve image quality, the ability to correctly diagnose HF and the ability to use the advantages of HUD applications.

Clinical influence of HUD examinations and diagnostic supportive tools

The precision of HF diagnosis based on standard examination alone is suboptimal, with low accuracy and high false-positive rates [27,28]. In our study, clinical examination alone had a low PPV (0.47). Of the patients suspected of having HF, only 43% were diagnosed with HF by the reference.

The proportion of correct classification increased after HUD examinations with fewer false positive and uncertain cases. This confirms the diagnostic value of adding HUDs to a clinical examination. Mjølstad et al. [5] found that GPs were able to assess LV function with HUDs, and the benefits of HUDs have been reported by several groups across different scenarios [6,7,18].

AutoMAPSE and autoEF have been reliable when used on high-end equipment and by experts [18,29,30]. Automatic LV quantification tools

Table 2
General practitioner diagnostics in comparison with reference echocardiography

Diagnostic stage	Heart failure positive	Heart failure negative	False positive	False negative	PPV	NPV
Clinical	35 (21%)	88 (53%)	17 (10%)	8 (5%)	0.47	0.91
HUD	36 (22%)	110 (66%)	15 (9%)	6 (4%)	0.56	0.94
AutoMAPSE	45 (27%)	81 (49%)	25 (15%)	4 (2%)	0.40	0.95
AutoEF	31 (19%)	85 (51%)	15 (9%)	4 (2%)	0.50	0.95
Telemedicine	30 (18%)	112 (67%)	8 (5%)	6 (4%)	0.71	0.94

Data are expressed as the number (%). Uncertain cases are not included.
AutoEF, automatic ejection fraction; autoMAPSE, automatic mitral annular plane systolic excursion; HUD, handheld ultrasound device; PPV, positive predictive value; NPV, negative predictive value.

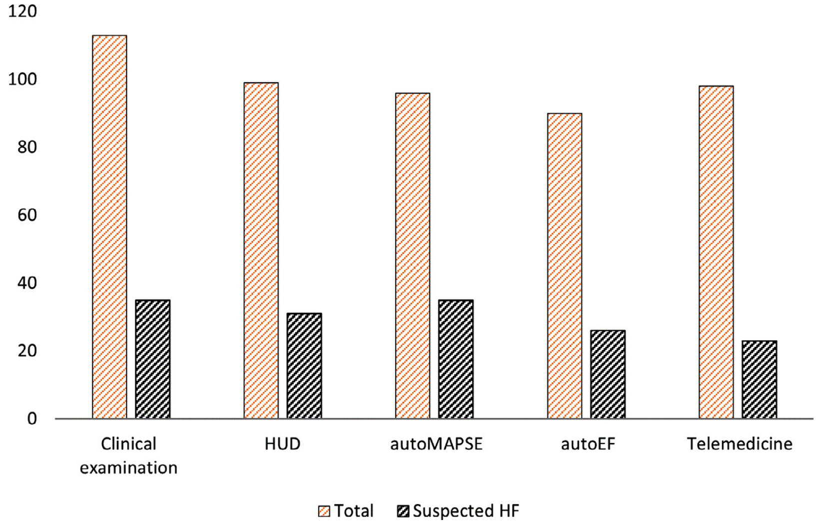


Figure 5. General practitioners' referrals of patients for cardiac examinations. The *orange columns* represent the total number of patients the GPs decided to refer for a cardiac examination no matter the diagnosis. The *black columns* represent how many of the total number of referred patients had suspected heart failure. autoEF, automatic analyses of ejection fraction; autoMAPSE, automatic analyses of mitral annular plane systolic excursion; HF, heart failure; HUD, handheld ultrasound device; NRI, net reclassification improvement.

Table 3
Characteristics of patients with heart failure incorrectly classified at one or more of the diagnostic stages

Characteristic	Symptoms	Reference echocardiography	Comment telemedicine	AutoEF/autoMAPSE
Male,[a] 78 y, HT, AF	Reduced physical capacity	HFmrEF EF 40% MAPSE 8.5 mm	Moderate dilated LV EF 30% Probable HF	AutoEF: 26%, 25%, 36% AutoMAPSE: 4.9, 7.0, 6.9 mm
Female, 72 y, HT, AF	Dyspnea on exertion	HFpEF EF 55% MAPSE 10.5 mm	Preserved LV HF unlikely	AutoEF: 64%, 71%, 72% AutoMAPSE: 9.0, 7.6, 9.9 mm
Male, 78 y, AF	Dyspnea on exertion.	HFpEF Visual EF 40% MAPSE not measured AF	AF. Slightly reduced LV. Aortic sclerosis?	AutoEF: 28%, 43%, 56% AutoMAPSE: 6.8, 7.5, 6.8 mm
Male, 70 y	Dyspnea on exertion	HFrEF Frequent VES EF 33% MAPSE 12 mm	Difficult diagnosis because of arrythmia	AutoEF not performed AutoMAPSE: 5.5, 8.1, 7.5 mm
Male, 59 y, HT	Dyspnea on exertion Systolic murmur Near syncope	HFpEF Severe aortic stenosis EF and MAPSE not stated	Preserved LV EF >50% Calcified aortic valve	AutoEF: 66%, 59%, 62% AutoMAPSE: 9.4, 6.5, 8.4 mm
Female, 56 y, DM2, asthma, OSAS	LV hypertrophy on ECG Reduced physical capacity Family history of cardiomyopathy	HFpEF Hypertrophic cardiomyopathy EF 74% MAPSE 11 mm	LV hypertrophy Preserved LV	AutoEF: 51%, 47%, 39% AutoMAPSE: 9.7, 10.4, 10.1 mm
Male, 76 y	Dyspnea on exertion	HFpEF Probable hypertrophic cardiomyopathy EF 58% MAPSE 8 mm	Normal EF LV hypertrophy	AutoEF: 66%, 33%, 53% AutoMAPSE: 6.8, 6.8, 6.8 mm
Female, 81 y, HT, RA, cerebellar stroke	Dyspnea on exertion Chest pain on exertion	HFpEF EF 59% MAPSE 9mm Severe mitral regurgitation AF	Preserved LV	AutoEF: 72%, 68%,71% AutoMAPSE: 6.6, 8.7, 9.5 mm

AF, atrial fibrillation; autoEF, automatic ejection fraction; autoMAPSE; automatic mitral annular plane systolic excursion; DM2, diabetes mellitus type 2; EF, ejection fraction; HF, heart failure; HFmrEF, heart failure with midrange ejection fraction; HFpEF, heart failure with preserved heart failure; HFrEF, heart failure with reduced ejection fraction; HT, hypertension; HUD, handheld ultrasound device; LV, left ventricle; OSAS, obstructive sleep apnea syndrome; RA, rheumatoid arthritis; VES, ventricular extrasystole.

[a] Patient misdiagnosed after the HUD examination, but HF suspected after automatic quantification and telemedicine.

implemented on HUDs have previously scarcely been evaluated by inexperienced users. There was no improvement in correctly diagnosed patients after automatic quantification compared with clinical examination alone. False-positive results increased after autoMAPSE, and the uncertainty was high for both applications. The difference between autoMAPSE and autoEF was not significant. After upgrading the autoEF software, there was no improvement in false-positive or false-negative cases. The COVs revealed a modest variation of autoEF between GPs and reference cardiologists (COV = 15.4%) and quite a large variation for autoMAPSE (COV = 24.3 %) [21]. Automatic quantification of MAPSE in B-mode images underestimates the mitral annular excursion [10,18], and can be compensated for by integrating tissue Doppler mode [31], which was not available in this study. In our study, both automatic quantification tools underestimated the LV measurements (mean MAPSE: 8 mm vs. 12 mm, and mean EF: 48% vs. 53%, respectively). The underestimation by autoMAPSE was larger than previously reported [10,18], and the potential importance of reduced image quality, patient characteristics and users must be addressed in future studies. The underestimation may have contributed to the GPs' uncertainty and overdiagnosis. Both automatic application tools were fully automatic, and it was not possible for the operators to adjust the endocardial tracking or mitral points. Further, no automatic feedback to optimize the recordings was provided. It is not known if novel tools providing automatic feedback to optimize the recordings could have improved the results [32]. The decrease in correctly diagnosed patients after adding the automatic quantification tools to the HUD examination may indicate that the GPs were unable to distinguish between correct and incorrect measurements. It may be hypothesized that feedback regarding the robustness of tracking of regions of interest and the measurements could potentially improve the GPs' interpretation. Further, atrial fibrillation (AF) poses a challenge to HF diagnostics. AF was overrepresented by 50% among the uncertain cases after autoEF (13 [33%]). AF was present in 54% of all HF patients and in 62% of patients with HFpEF. The combination of HFpEF and AF also represents a clinical challenge. Diagnostics are difficult because of overlapping symptoms [33]. AF might also have contributed to the failure of the automatic decision-support software.

Use of telemedicine may reduce time to diagnosis and treatment [12,34]. After telemedical support, the numbers of false-positive and uncertain cases decreased. There was a significant difference in the proportion of correct reclassification (NRI 0.19) compared with clinical evaluation and automatic quantification. Telemedical support of the HUD recordings had the numerically highest PPV (0.71), and the total number of correctly diagnosed patients improved in comparison with HUD alone (123 vs. 118). Improved image quality could further facilitate the evaluation [7,35]. The time between inclusion days for the participating GPs varied and may have influenced the image quality and subsequently reduced the quality of the telemedical support. Further, the GPs needed to fit the feedback into a clinical context, which may have been challenging in some cases. Access to the clinical information from the present exam could have improved the telemedical interpretation and feedback to the GPs.

There were false-negative cases at each stage of the examinations. In all but one instance, either HF was identified after telemedical support and/or the patients were referred for an echocardiography for other reasons such as AF, suspicion of aortic stenosis, coronary artery disease or hypertrophic cardiomyopathy. Of the patients with HF, 46% had HFpEF, which is challenging to diagnose [36]. As outlined in Table 3, the combined use of HUD, autoEF, autoMAPSE and telemedicine was not sufficient for precise diagnosis of HFpEF.

Despite a low proportion of patients with suspected HF, the GPs would still refer a large proportion to a cardiologist. In most cases, there was a valid reason for the referral, such as AF, suspicion of valvular disorders, pericardial effusion, suspicion of other cardiac pathology or poor image quality.

Limitations

This was a single-center study with a limited number of observers. The participating GPs were randomly selected by the health care administration in two municipalities and did not join because of a special interest. Further, the five GPs originated from five different GP offices. Evaluating an even broader set of operators would improve the ability to make stronger conclusions regarding the generalizability of the results. Diagnostic ultrasound is user dependent, and enthusiasm and interest would likely motivate participants to improve their recordings. The modest sample size may limit the power for secondary analyses. For example, including a larger population would have improved the power to detect subtle inter-operator differences. Additionally, we do not know whether differences between the two external experts influenced the results. Because of laboratory delay, the GPs did not have access to all the blood sample results, and the level of NT-pro-BNP was not available in all initial referrals to the outpatient clinic. NT-Pro-BNP is an important factor in the evaluation of suspected HF [2] which can be ruled out using a threshold of <125 ng/L [22,37]. However, in this study, most patients had elevated NT-pro-BNP, and we do not expect a substantially changed outcome if all results were available at the time of examination. Lastly, even though the GPs were not able to provide automatic measurements of the left ventricle in a substantial proportion of the examinations, the finding of ≥43% misclassified decisions after automatic measurements indicates that we had adequate power to conclude on the pre-set 15% limit for accepted misclassifications.

Conclusions

Addition of handheld ultrasound to examinations by general practitioners improved diagnostic precision in patients with suspected heart failure. The highest NRI was found after the HUD recordings were supported by telemedical interpretation. In the future, this may allow for better selection of patients with suspected HF in need for cardiac follow-up. The applications for automatic quantification of LV function added no significant benefit. Further refinement of the methods and more specific training of personnel may be needed before these methods add benefit to the diagnostic process.

Conflict of interest

L.L. is a part-time consultant for GE Vingmed Ultrasound, unrelated to this work. M.I.M., O.C.M., L.L. and H.D. hold positions at CIUS, a Centre of Research-based Innovation that is funded by the Research Council of Norway, Oslo, Norway, and industry. One of the industry partners is GE Vingmed Ultrasound, Horten, Norway. The company (GE) provided hand-held ultrasound devices with dedicated software for free use in the study but had no role in the planning of the study, data acquisition or drafting or revision of the manuscript.

Acknowledgments

We thank the clinicians and other employees at Nord-Trøndelag Hospital Trust for their support and for contributing to data collection in this research project. The study is funded by the Research Council of Norway and the Norwegian University of Science and Technology.

Data availability statement

Data will be made available on reasonable request.

References

[1] van Riet EE, Hoes AW, Limburg A, Landman MA, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. Eur J Heart Fail 2014;16:772–7.

[2] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, 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 J Heart Fail 2016;18:891–975.
[3] McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J 2012;33:1787–847.
[4] Cardim N, Dalen H, Voigt JU, Ionescu A, Price S, Neskovic AN, 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.
[5] Mjolstad OC, Snare SR, Folkvord L, Helland F, Grimsmo A, Torp H, et al. Assessment of left ventricular function by GPs using pocket-sized ultrasound. Fam Pract 2012;29:534–40.
[6] Andersen GN, Viset A, Mjolstad OC, Salvesen O, Dalen H, Haugen BO. Feasibility and accuracy of point-of-care pocket-size ultrasonography performed by medical students. BMC Med Educ 2014;14:156.
[7] Evangelista A, Galuppo V, Mendez J, Evangelista L, Arpal L, Rubio C, et al. Hand-held cardiac ultrasound screening performed by family doctors with remote expert support interpretation. Heart 2016;102:376–82.
[8] Matos J, Kronzon I, Panagopoulos G, Perk G. Mitral annular plane systolic excursion as a surrogate for left ventricular ejection fraction. J Am Soc Echocardiogr 2012;25:969–74.
[9] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, 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. Eur Heart J Cardiovasc Imaging 2015;16:233–70.
[10] Snare SR, Mjolstad OC, Orderud F, Haugen BO, Torp H. Fast automatic measurement of mitral annulus excursion using a pocket-sized ultrasound system. Ultrasound Med Biol 2011;37:617–31.
[11] Franken Jr EA, Berbaum KS. Subspecialty radiology consultation by interactive telemedicine. J Telemed Telecare 1996;2:35–41.
[12] Singh S, Bansal M, Maheshwari P, Adams D, Sengupta SP, Price R, et al. American Society of Echocardiography: remote echocardiography with web-based assessments for referrals at a distance (ASE-REWARD) Study. J Am Soc Echocardiogr 2013;26:221–33.
[13] Volterrani M, Sposato B. Remote monitoring and telemedicine. Eur Heart J Suppl 2019;21:M54–6.
[14] Filipiak-Strzecka D, Kasprzak JD, Wejner-Mik P, Szymczyk E, Wdowiak-Okrojek K, Lipiec P. Artificial intelligence-powered measurement of left ventricular ejection fraction using a handheld ultrasound device. Ultrasound Med Biol 2021;47:1120–5.
[15] Papadopoulou SL, Sachpekidis V, Kantartzi V, Styliadis I, Nihoyannopoulos P. 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. Eur Heart J Digital Health 2022;3:29–37.
[16] Aase SA, Snare SR, Dalen H, Stoylen A, Orderud F, Torp H. Echocardiography without electrocardiogram. Eur J Echocardiogr 2011;12:3–10.
[17] Aldaas OM, Igata S, Raisinghani A, Kraushaar M, DeMaria AN. Accuracy of left ventricular ejection fraction determined by automated analysis of handheld echocardiograms: a comparison of experienced and novice examiners. Echocardiography 2019;36:2145–51.
[18] Magelssen MI, Palmer CL, Hjorth-Hansen A, Nilsen HO, Kiss G, Torp H, 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.
[19] Bryczkowski CJ, Byrne MW. Workflow and middleware. In: Tayal VS, Blaivas M, Foster TR, editors. Ultrasound program management: a comprehensive resource for administrating point-of-care, emergency, and clinical ultrasound. Cham: Springer; 2018. p. 281–99.
[20] Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology 2014;25:114–21.
[21] Hjorth-Hansen AK, Magelssen MI, Anderssen GN, Graven T, Kleinau JO, Landstad B, 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.
[22] Zaphiriou A, Robb S, Murray-Thomas T, Mendez G, Fox K, McDonagh T, et al. The diagnostic accuracy of plasma BNP and NTproBNP in patients referred from primary care with suspected heart failure: results of the UK natriuretic peptide study. Eur J Heart Fail 2005;7:537–41.
[23] Kelder JC, Cramer MJ, van Wijngaarden J, van Tooren R, Mosterd A, Moons KG, Lammers JW, et al. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. Circulation 2011;124:2865–73.
[24] Zheng A, Cowan E, Mach L, Adam RD, Guha K, Cowburn PJ, et al. Characteristics and outcomes of patients with suspected heart failure referred in line with National Institute for Health and Care Excellence guidance. Heart 2020;106:1579–85.
[25] Chamsi-Pasha MA, Sengupta PP, Zoghbi WA. Handheld echocardiography: current state and future perspectives. Circulation 2017;136:2178–88.
[26] Mjolstad OC, Andersen GN, Dalen H, Graven T, Skjetne K, Kleinau JO, 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.
[27] Wheeldon NM, MacDonald TM, Flucker CJ, McKendrick AD, McDevitt DG, Struthers AD. Echocardiography in chronic heart failure in the community. Q J Med 1993;86:17–23.
[28] Sparrow N, Adlam D, Cowley A, Hampton JR. The diagnosis of heart failure in general practice: implications for the UK National Service Framework. Eur J Heart Fail 2003;5:349–54.
[29] Nagueh SF, Smiseth OA, Appleton CP, Byrd III BF, Dokainish H, Edvardsen T, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2016;29:277–314.
[30] Grue JF, Storve S, Dalen H, Mjølstad OC, Samstad SO, Eriksen-Volnes T, et al. Automatic quantification of left ventricular function by medical students using ultrasound. BMC Med Imaging 2020;20:29.
[31] Storve S, Grue JF, Samstad S, Dalen H, Haugen BO, Torp H. Realtime Automatic assessment of cardiac function in echocardiography. IEEE Trans Ultrason Ferroelectr Freq Control 2016;63:358–68.
[32] Pasdeloup D, Olaisen SH, Østvik A, Sabo S, Pettersen HN, Holte E, et al. Real-time echocardiography guidance for optimized apical standard views. Ultrasound Med Biol 2022;49:333–46.
[33] Kotecha D, Lam CS, Van Veldhuisen DJ, Van Gelder IC, Voors AA, Rienstra M. Heart failure with preserved ejection fraction and atrial fibrillation: vicious twins. J Am Coll Cardiol 2016;68:2217–28.
[34] Bhavnani SP, Sola S, Adams D, Venkateshvaran A, Dash PK, Sengupta PP. A Randomized trial of pocket-echocardiography integrated mobile health device assessments in modern structural heart disease clinics. JACC Cardiovasc Imaging 2018;11:546–57.
[35] Choi BG, Mukherjee M, Dala P, Young HA, Tracy CM, Katz RJ, et al. Interpretation of remotely downloaded pocket-size cardiac ultrasound images on a web-enabled smartphone: validation against workstation evaluation. J Am Soc Echocardiogr 2011;24:1325–30.
[36] Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J 2019;40:3297–317.
[37] Roberts E, Ludman AJ, Dworzynski K, Al-Mohammad A, Cowie MR, McMurray JJ, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ 2015;350:h910.

ISBN 978-82-326-7276-9 (printed ver.)
ISBN 978-82-326-7275-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology