Received: 18 February 2021 | Revised: 10 April 2021 | Accepted: 24 April 2021

DOI: 10.1111/echo.15073

Echocardiography WILEY

Variability of echocardiographic measures of left ventricular diastolic function. The HUNT study

Jon Magne Letnes MD[1,2] | Torfinn Eriksen-Volnes MD[1,2] | Bjarne Nes MSc, PhD[1,2] | Ulrik Wisløff MSc, PhD[1] | Øyvind Salvesen MSc, PhD[3] | Håvard Dalen MD, PhD[1,2,4]

[1]Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

[2]Clinic of Cardiology, St. Olavs University Hospital, Trondheim, Norway

[3]Unit of Applied Clinical Research, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway

[4]Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

Correspondence
Håvard Dalen, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, NTNU, Postbox 8905, 7491 Trondheim, Norway.
Email: havard.dalen@ntnu.no

Funding information
The project was funded by grants from The Liaison Committee for Education, Research and Innovation in Central Norway, St. Olavs University Hospital (Trondheim, Norway), Nord-Trøndelag Hospital Trust (Levanger, Norway), and Simon Fougner Hartmann's Family Fund (Copenhagen, Denmark).

Abstract

Objective: Investigate variability related to image acquisition and reading process for echocardiographic measures of left ventricular (LV) diastolic function, and its influence on classification of LV diastolic dysfunction (LVDD).

Methods: Forty participants (19 women) mean age 62 (28–88) years underwent echocardiographic examinations twice by different echocardiographers and blinded analyses by four readers in a cross-sectional design. Measurements included quantification of two- (2D) and three-dimensional (3D) recordings of the left atrium (LA) (maximal) volume (LAV_{max}) and spectral Doppler blood flow and tissue velocities for assessment of LV diastolic function. Variability and reproducibility measures were calculated using variance component analyses and Kappa statistics.

Results: Image acquisition influenced variability more than image reading (mean 24% and 4% of variance, respectively), but variability from image reading was especially important for 2D LAV_{max} (16% of variance) compared to 4% for 3D LAV_{max}, which was reflected in better agreement for 3D measures. The variability of measures used in classification of LVDD had clinical significance, and agreement across the four raters in classification using current recommendations was only fair (Kappa 0.42), but the agreement improved when using 3D LAV_{max} (Kappa 0.58). Agreement and reliability measures were reported for all measures.

Conclusion: Performing a new image acquisition influenced variability more than a introducing a new image reader, but there were differences across the different measures. LAV_{max} by 3D is superior to 2D with respect to lower variability. The variability of diastolic measures influences the reliability of LVDD classification, and this should be taken into account in the everyday clinic.

KEYWORDS

diastolic function, left atrium, repeatability, tissue Doppler

1 | INTRODUCTION

Left ventricular (LV) diastolic function offers important prognostic information in clinical practice.[1,2] Echocardiographic examination of these parameters is feasible and readily available, but the measurements are prone to variability from several sources such as image acquisition (examination) and image reading (interpretation). This is true both between and within echocardiographic examiners and readers. This variability adds to biological variability and temporal physiological changes. To precisely detect changes in patients' clinical status over time, measurement error must be low, and different echocardiographic methods are expected to have varying properties regarding variability, reliability, and agreement. For example, two-dimensional (2D) echocardiographic evaluation of left atrial (LA) volumes is susceptible to variation in identification of image planes during image acquisition[3,4] while three-dimensional (3D) techniques are not. On the other hand, 3D modalities have lower spatial and temporal resolution compared to 2D, and thus, comparisons of variability of different techniques and measures are of clinical interest.

Although several studies have reported results regarding reliability and agreement of, for example, LA and tissue Doppler measures as secondary objectives or supplementary analyses,[5–7] as well as some dedicated reproducibility studies,[8,9] there is a lack of studies quantifying and addressing the amount variability from image acquisition and image reading in the image acquisition process, which is of importance in planning and conducting patient follow-up. Therefore, the aim of the current study was (a) to quantify variability in the image acquisition and reading process for measures used in quantification of LV diastolic function and its influence in classification of LV diastolic dysfunction (LVDD), and (b) to investigate the reliability and agreement of echocardiographic measures of LVDD.

2 | MATERIALS AND METHODS

2.1 | Study population and setting

Participants from a sub-study of the fourth wave of the Trøndelag Health Study (HUNT4, 2017–2019) were included. The HUNT4 Echocardiography and Fitness studies were two co-organized sub-studies collecting data from October 2017 to June 2018. Of 2448 participants included, 40 participants examined in June 2018 were randomly selected to participate in the reproducibility study based on the participation dates. Inclusion criteria to the study were participation in the HUNT4 baseline examination, *and* participation in the HUNT3 (2006–2008) Fitness or Echocardiography studies *or* atrial fibrillation (validated from HUNT3 or self-report from HUNT4). Further information on inclusion and exclusion criteria (briefly, disease or disability prohibiting exercise testing) as well as information on clinical measurements is available in the Appendix (Appendix S1). All participants consented to participation. The current study was approved by the regional ethical committee (REC Central Norway 2018/2416).

2.2 | Echocardiography

Transthoracic echocardiography was performed after a brief interview and anthropometric measures, but before ergospirometry (as a part of the HUNT4 Fitness study). Two transthoracic echocardiographic examinations were performed by two experienced echocardiographers (one sonographer and one cardiologist [TE]) within 5–15 minutes of each other using the same imaging protocol. Participants were examined in the left lateral decubitus position using a Vivid E95 scanner (GE Ultrasound) with a 4Vc-D matrix transducer for both 2D and 3D imaging. The relevant methodology for image acquisition is included in the Appendix S1. The echocardiographers were blinded to each other's examinations, and data from the examinations were stored digitally for later offline reading by four different readers (two echocardiographers [HD, TE], and two sonographers) blinded to each other's readings summing up to eight readings per echocardiographic measure per participant. All readers were experienced with the basic methodology. One reader (HD) was the most experienced with 3D volume and strain analyses of the LA, and the others underwent 1 day of supervised training including 10 separate 3D volume and strain measurements. All readers were aware of taking part in a reproducibility study. Analyses were performed using EchoPAC SWO (version 203; GE Ultrasound). Measurements were made on one of at least three cardiac cycles where the readers were free to choose the best-suited cycle, meaning that measurements between readers not necessarily were made on identical cycles. To investigate the influence of restricting measurements to a prespecified cardiac cycle, new measurements were performed on a prespecified cardiac cycle (defined by one of the readers [HD]) for a selection of 20 examinations (10 participants).

All measurements were performed in accordance with recommendations where available,[10] and it was up to discretion of the individual reader to assess the appropriateness of image quality for analysis of the given measures. Comprehensive information regarding echocardiographic reading methodology is available in the Appendix (Appendix S1). Echocardiography included 2D maximal LA volumes (LAV_{max}) by the area–length (A-L) and summation of disks (MOD) method. 3D recordings were used for semi-automated calculations of LA volumes (LAV_{max}, minimal and preatrial contraction [LAV_{min} and LAV_{preA}]) and functional LA measures (ejection fraction [LAEF] and ejection volume [LAEV]). LA global longitudinal and circumferential reservoir (LS_r, CS_r), conduit (LS_{cd}, CS_{cd}), and contractile (LS_{ct}, CS_{ct}) strain were derived from the 3D recordings. 3D LA volumes and deformation measurements were performed using the LAQ package in EchoPAC. Peak early and late diastolic mitral annular (e' and a') and transmitral inflow velocities (E and A), mitral deceleration time, and pulmonary vein diastolic (D) and systolic (S) velocities were measured in pulsed-wave Doppler recordings. The E/A, E/e', and S/D ratios were calculated.

2.3 | Statistical analyses

We performed variance component analyses by linear mixed effect models to estimate the four different sources of variation in the acquisition of echocardiographic measurements, namely variation from (a) subject, (b) introducing a new echocardiographic image acquisition (two per subject), (c) a new reader performing image reading (four readings per echocardiographic image acquisition), and (d) unexplained sources (residual). Normality of error terms was tested by inspection of Q-Q plots. Logarithmic transformation was performed where necessary. In patients with atrial fibrillation during echocardiography (n = 4), measures dependent on atrial contraction were excluded (A, E/A, septal and lateral a', LS_{ct}, CS_{ct}) and included as a covariate if inclusion increased model performance for other measures. Based on the results from the variance component analysis, we calculated repeatability coefficients (RCs) as 1.96×√2 times the squared within-subject variance (standard error of measurement).[11–15] Using the relevant variance components, we estimated RCs for different clinical situations: (a) intrarater intra-acquisition RC; (b) interrater intra-acquisition RC; (c) interrater inter-acquisition RC; and (d) intrarater inter-acquisition RC. The intrarater intra-acquisition RC is presented scaled to the mean of the given variable (percentage), and in absolute terms. The other RCs are presented as a percentage of the mean, except for LA strain measures as they include both positive and negative values. Reliability, the ability to discriminate between subjects,[16] was assessed by the intraclass correlation coefficient (ICC) as the proportion of variance explained by subject to the total of subject plus residual variance ($ICC_{consistency}$).[17] In the subset of participants where readings were repeated on the same images on a prespecified cardiac cycle, we estimated the RC (a) by using the same linear mixed model approach. Pathological cutoffs for LAVmax (2D A-L), E/e', TRVmax, and e', and classification to LVDD and elevated filling pressures were based on recent recommendations.[18] LVDD was dichotomized to «diastolic dysfunction» or «normal» based on presence of >2/4 or ≥2/3 or 2/2 available variables meeting established cutoffs, or by an LVEF <0.40. Filling pressures were dichotomized to «elevated»/«normal», where the same algorithm was used both for normal or reduced LVEF. Indeterminate diastolic function and filling pressure were coded as normal. Light's Kappa was calculated as agreement between four raters within the same image acquisition, and 95% confidence intervals (CIs) were constructed by bootstrapping using 500 replications. A Kappa of 0.01–0.20 was defined as poor, 0.21–0.40 as slight, 0.41–0.60 as fair, 0.61–0.80 as substantial, and 0.81–1.0 as almost perfect.[19] Analyses were performed using R (www.r-project.org, packages *lme4* and *irr*).

3 | RESULTS

Baseline clinical and echocardiographic characteristics are shown in Table 1. Mean age was 62 years (range 28–88), and 2.5% (n = 1) and 5% (n = 2) of the population had known heart failure or a previous myocardial infarction, and 43% had atrial fibrillation. Fourteen participants (35%) had diastolic LVDD, seven indeterminate (18%) diastolic function and 19 (48%) normal diastolic function, while 5% had elevated LV filling pressures, as graded by one of the readers (HD) on one of the image acquisitions. Echocardiographic indices of diastolic function are presented in Table S1.

3.1 | Variance components and reliability measures

Variability expressed by variance components for subject, image acquisition and reading, and residual variance for the different echocardiographic measures, are shown in Figure 1. Variability explained by subject dominated for most measures. For all measures, the variability from image acquisition (average 24% for all measures) was considerably larger compared to reading (average 4%), but especially for the 2D LAV_{max} measures, the variability from image reading was notable as well. For 2D LAV_{max} measures, image acquisition and reading contributed with average 23% and 16% of the total variance, compared with average 17% and 5% for the 3D measures. The total variance explained by acquisition and reading was low for tissue and transmitral inflow Doppler measures (15% and 1%, respectively), compared to 34% and 4% for LA strain measures, respectively.

TABLE 1 Clinical and echocardiographic characteristics of participants

	Men n = 21	Women n = 19
Age (years)	65 (12)	58 (13)
Weight (kg)	89 (18)	79 (12)
Height (cm)	178 (7)	166 (7)
Body mass index (kg/m^2)	28.0 (4.4)	28.6 (4.1)
Systolic BP (mm Hg)	132 (16)	130 (20)
Diastolic BP (mm Hg)	78 (11)	71 (10)
Current smoker	0 (0%)	0 (0%)
Former smoker	9 (43%)	10 (53%)
Myocardial infarction	1 (5.0%)	1 (5.9%)
Heart failure	0 (0%)	1 (5.9%)
Atrial fibrillation	10 (48%)	7 (37%)
Tricuspid regurgitation present	4 (19%)	5 (26%)
LV end-diastolic internal diameter (mm)	50 (8)	49 (5)
LV end-diastolic volume (mL)	129 (37)	109 (27)
LV EF (%)	58.3 (7.5)	60.6 (5.4)
TAPSE (cm)	2.29 (0.32)	2.37 (0.44)
S' (cm/s)	8.37 (1.91)	7.75 (1.57)

Note: Values are mean (SD) or n (%).

Abbreviations: BP = blood pressure; EF = ejection fraction; HbA1c = glycosylated hemoglobin; HDL = high-density lipoprotein; LV = left ventricle; S' = peak mitral annular systolic velocity; SD = standard deviation; TAPSE = tricuspid annular plane systolic excursion.

The large residual (unexplained) error proportions for the LA strain measures are reflected by the low ICCs ranging from 0.29 to 0.47 (Figure 1, Table S1). In contrast, the mitral inflow and tissue Doppler measures had ICCs from 0.88 to 0.97. The ICCs for the 3D LAV measures (0.84–0.85) were higher than for the 2D LAV measures (0.71–0.74).

In analyses exploring the individual components of 2D LA volume calculation, agreement was better for measurements of length compared to area. The RCs for LA area (mean of two- and four-chamber views) indexed to the mean for LA area were 17% for the same rater on the same acquisition (intrarater intra-acquisition), 21% for different rater on the same acquisition (interrater intra-acquisition), and 45% for different rater on different acquisition (interrater inter-acquisition). The corresponding RCs for LA length (mean of two- and four-chamber views) indexed to the mean LA length were 25%, 32%, and 70%, respectively. Figure 2 illustrates the difference in variability of 2D LAV_{max} introduced by acquisition and reading.

Kappa analysis showed fair agreement for classification of LVDD (Kappa 0.42, 95% CI 0.21–0.58) using 2D A-L method for LAV_{max} (raw data in Table S2). Using 3D for LAV_{max} in classification of LVDD Kappa was higher (0.58, 95% CI 0.12–0.82). Compared to the most experienced rater, the other three raters agreed with this rater in 71% of the cases when analyzing on the same echocardiographic dataset. The most experienced rater agreed with his own classification in 70% of cases based on reading of the second dataset. Classification to enlarged LAV_{max} by 2D LAV_{max} showed lower agreement (Kappa 0.43, 95% CI 0.28–0.59) compared to 3D (Kappa 0.72, 95% CI 0.45–0.93). Analyses for classification of LV filling pressures showed almost perfect agreement (Kappa 0.86), but due to very few having elevated filling pressures (prohibiting construction of bootstrapped CIs), these results should be interpreted with caution.

3.2 | Measures of agreement

Left atrium maximal measures (2D and 3D) showed relatively similar intrarater intra-acquisition RCs scaled to their respective means (32% and 27%), but when also considering variance from introducing a new reader in image reading (interrater intra-acquisition RC) the 3D LAV_{max} measure showed considerably better agreement than 2D measures (31% vs 44% for interrater intra-acquisition RC, Figure 3, Table S1). Both mitral inflow, pulmonary vein, and tissue Doppler indices consistently showed low RCs compared to volumetric measures with intrarater intra-acquisition RCs ranging from 14% to 26%, but the relative contribution of performing a new image acquisition was higher than for the volumetric measures. LA strain values showed large absolute RCs.

In sensitivity analyses restricted to one prespecified cardiac cycle (n = 10), mean intrarater intra-acquisition RCs was notably lower; for Doppler measures 13% compared to the 18% from the main analyses, and 26% and 32% for the volumetric measures, respectively.

4 | DISCUSSION

Our results demonstrate that variance from echocardiographic image acquisition is considerable and contributes more to the total variance compared to changing image readers. Reliability and agreement were good to excellent for most measures of diastolic function, but when combining several measures in classification of LVDD agreement was only fair. Image reading contributed considerably to lower agreement for 2D LAV_{max} compared to 3D LAV_{max} which showed better agreement and reliability compared to 2D.

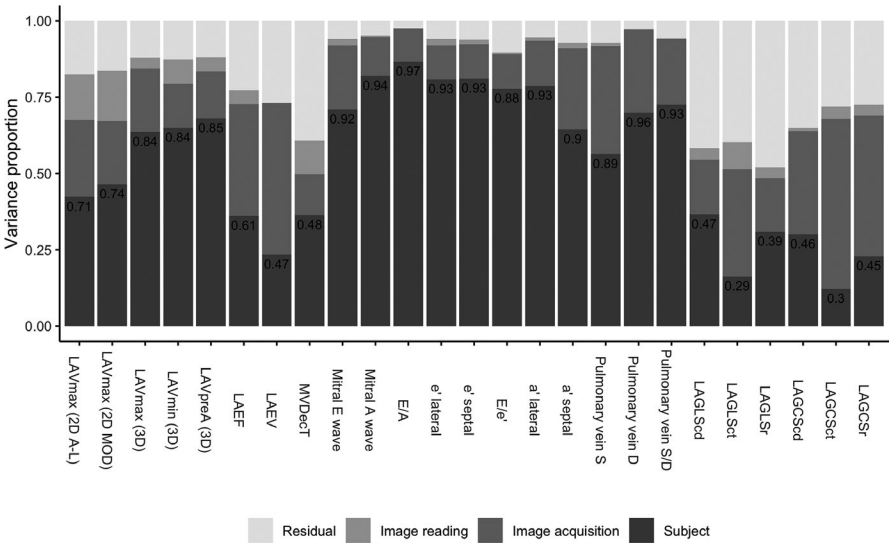


FIGURE 1 Variance proportions for different echocardiographic measures. Intraclass correlation coefficient ($ICC_{consistency}$) given for each variable at the respective columns

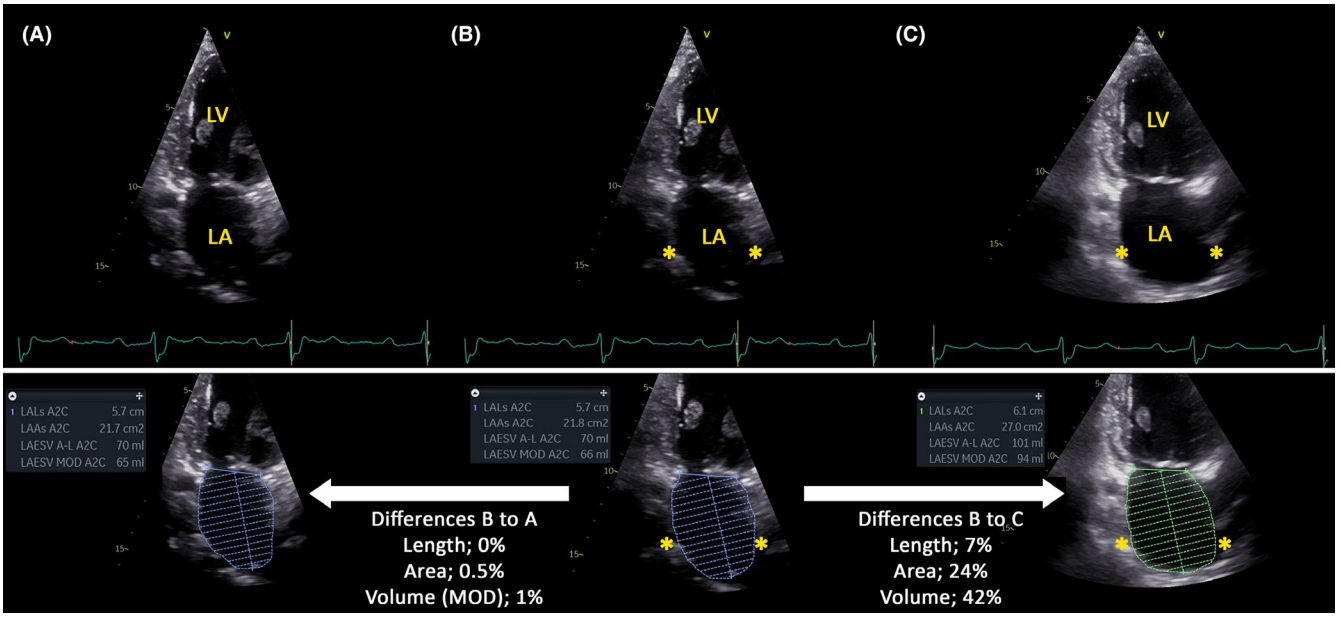
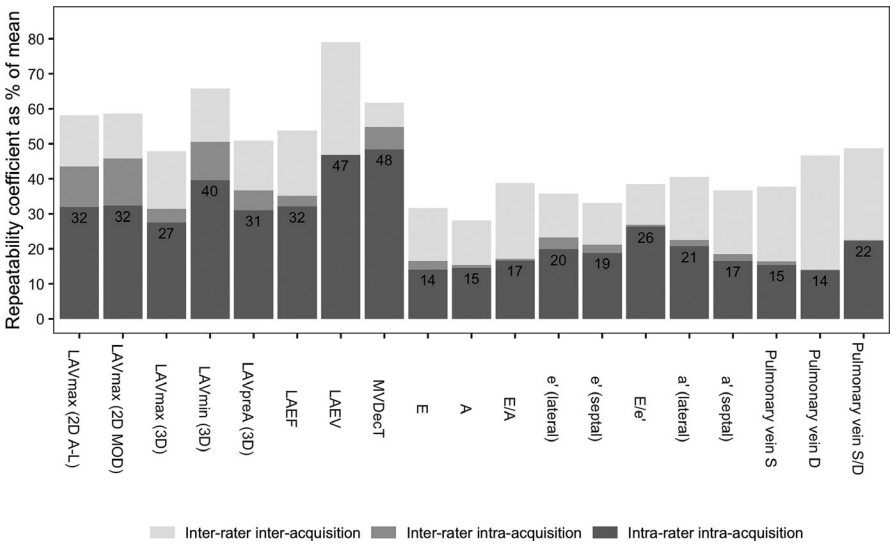


FIGURE 2 Illustration of the impact of acquisition and reading to the variability. In the upper panels, the frame used for actual measurement performed is shown. In the lower panels, the tracings provided measurements and the differences between measurements are shown. Panels A and B represent two separate readings of the same acquisition. This relate to both intrarater intra-acquisition and interrater intra-acquisition, but the illustration is an illustration of intrarater intra-acquisition with low variability. Panel C is from the second acquisition, and in comparison with panel B, this relates to interrater inter-acquisition with significant differences in the acquisition (see differences in LA endocardial borders (yellow stars)). Abbreviations: LA = left atrium; LV = left ventricle

FIGURE 3 Scaled repeatability coefficients for echocardiographic measures. Stacks should be interpreted additively to each other, meaning the interrater intra-acquisition repeatability coefficient is the sum of the two bars in darkest shading and so forth. The value corresponding to the intrarater intra-acquisition repeatability coefficient is given for each variable on top of the respective column



4.1 | Comparison with other studies

Few studies have drawn attention to variability introduced in the image acquisition-image reading process. However, a recent study reported variability from acquisition and interpretation of images and reported larger variability from image acquisition than image reading.[20] Although they used a different methodology comparing acquisition-reading situations using direct Bland–Altman limits of agreement, the trend in their results was similar to ours. Doppler measures showed consistently good reliability and agreement in our study, similar to previous studies,[8,9,21,22] but it should be acknowledged that also for these measures variability from image acquisition was noteworthy, affecting agreement in inter-acquisition settings. The impact of image acquisition on variability for the Doppler measures may relate to both methodology and physiology. The alignment of the ultrasound beam with the direction of movement of blood and tissue may be one cause of variation, while physiological variation may be present even over a short time span. Even though the two separate echocardiograms were done immediately after each other, the participant had to walk 10 meters before the second echocardiogram, as well as being introduced to the second operator.

Our data show superior measures of agreement and reliability for the 3D measures of LAV_{max} compared to 2D. Especially variability from image reading was lower for 3D measures, while 2D measures showing considerable variation at both image acquisition and reading. Lower reproducibility of 2D LAV_{max} measurements is expected due to challenges in identifying the optimal plane of the LA, as well as tracing of the endocardium for determination of area and length. As the LA is more distant from the transducer compared to the LV, the resolution is poorer, and thus, endocardial border detection may be impaired. As measurements of length showed better agreement compared to area, we argue that the lower lateral resolution poses challenge for the reading process. Thus, the variability of 2D LA volumes is mostly related to optimizing the image plane during acquisition and tracing the LA contour during reading. Furthermore, 3D methods use algorithms based on artificial intelligence or geometrical models for the endocardial border delineation, and do also automatically set the timing of events. By the method used, it was up to the operator to manually adjust the endocardial border and the timing of the LA events. In general, more automation relates to less variability as the algorithm acts similar in repetitive cases. Figure 4 shows the potential impact on LA area vs length by slight changes in the endocardial tracing. As the 3D method is a more automated method the variability will be lower, but this study does not indicate which method is the most correct. Previous studies have highlighted the importance of dedicated atrial views for LAV_{max} measurement.[3,4] In this study the 3D full volumes were stitched, and thus, averaged over two to four cycles, while there may be some in and out of plane motion in 2D-recordings. Several studies have reported measures of reproducibility for LAV_{max} measures such as RCs[23] and CVs,[23–25] and ICCs[26] generally in line with our data. Rohner et al reported very high ICCs for real-time 3D assessment of LA volumes (0.99),[7] but as the ICC is sensitive to the population studied and the type of ICC presented[11,17] comparisons should be made with caution when limited information on methods is available.

When combining measures of LVDD for classification to normal or abnormal diastolic function, the agreement was only fair. This may be due to many of the participants having values close to thresholds for enlarged LAV_{max} and high e', meaning that small differences in measurements could lead to different conclusions. Classification to LVDD and enlargement of LAVI showed better agreement when using 3D compared to 2D as shown in Kappa analyses. This is in line with results from the variance component analysis. The World Alliance Societies of Echocardiography (WASE) Normal Values Study have recently published data indicating that the upper normal limit for LAV_{max} is higher than presented in the recommendations.[27] We have previously presented epidemiological data in line with the WASE Study.[28] Important for the presented results is that a lower number of participants would fulfill the criteria for abnormal diastolic function, but due to the variability of LA volume measurements, we would expect that some patients with LVDD were not recognized as abnormal. When interpreting the Kappa analyses, one should keep in mind the differences in means for LAV_{max} by 2D and 3D when classifying by a fixed cutoff (LAV_{max} 34 mL/m^2).

4.2 | Strengths and limitations

The inclusion was preset by dates, and no selection of individuals to be included was performed. Most of the preset dates included participants included in the main project due to atrial fibrillation, and thus, the proportion with atrial fibrillation was high. Contrary to some previous studies, analyses were performed in all images without selection of image quality.[3] Also, some studies have prespecified a cardiac cycle to perform measurements on, and our sensitivity analysis showed the considerable effect this may have on the measures of reproducibility. Thus, although the presented agreement and reliability may be lower than in optimized settings where time for imaging and quality of echocardiographic images is not a limitation, this study effectively mimic clinical practice further increasing generalizability to real-world clinical situations. As breath hold was performed only for the stitched 3D acquisitions, this may have favored these measures. It would have been preferable to perform repeated echocardiographic image acquisitions within each examiner, and repeated intrareader readings for each image acquisition. However, this was not feasible. The intrarater intra-acquisition RC from the main models should be

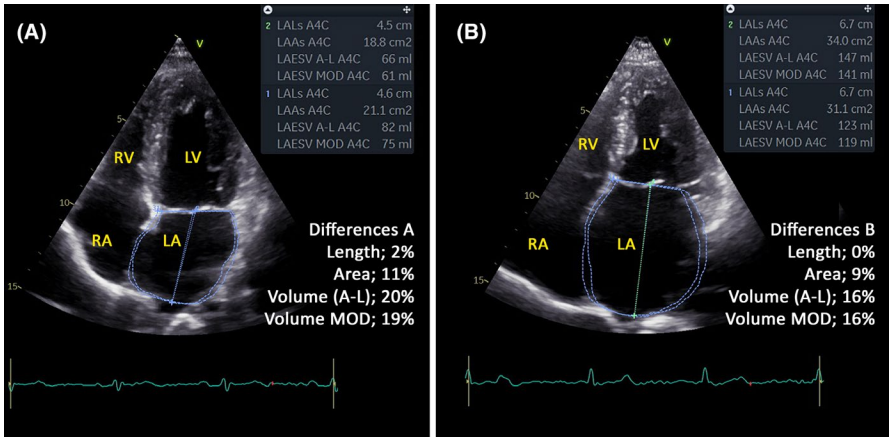


FIGURE 4 Two tracings of the left atrium (LA) for two different participants, with a patient with indexed LA volume of normal size A and enlarged in B. The differences shown represent the effect on calculated LA length, area, and volume by tracing the endocardial trace on the luminal border as opposed to tracing slightly into the signals. Abbreviations: LA = left atrium; LV = left ventricle; RA = right atrium; RV = right ventricle

interpreted as a conservative estimate since it was estimated based on excluding the other variance components.

4.3 | Clinical implications

We provide data on variability from different sources for echocardiographic acquisition and reading which is of importance when planning and conducting patient follow-up. Firstly, Doppler measures in general seem to be well-suited in longitudinal follow-up for detecting clinically meaningful change and to discriminate normal from pathological measures. In line with previous findings,[29] the 3D LAV_{max} measure seems better suited for longitudinal follow-up than the 2D measures due to superior agreement. By 3D less variability is introduced by a new image reader compared to 2D. However, the accuracy compared to gold-standard reference is important for clinical use, but we did not have such data. The presented data also offer realistic data of reproducibility for several echocardiographic measures which should be transferable to clinical practice. These data can aid clinicians in interpreting if a repeated measure is likely to reflect a true biological change. The agreement of the clinically important classification of LVDD was only fair, which is important when interpreting echocardiographic reports in the everyday clinic.

5 | CONCLUSIONS

In echocardiographic assessment of LV diastolic function variability from a new image, acquisition contributes more to the total variability compared to changing the image reader. Measures of agreement and reliability were good for most measures used in LV diastolic function evaluation. However, 3D measures of LAV_{max} are better suited for longitudinal follow-up than 2D measures, as more variability is introduced when changing image readers in 2D recordings. Accumulated, variability of LVDD measures is of clinical significance when classifying individuals by presence of LVDD. The presented measures of agreement may help clinicians plan and correctly interpret change in echocardiographic measures in serial follow-up.

ACKNOWLEDGMENTS

The HUNT (Trøndelag Health Study) is a collaboration between the HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology), Nord-Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. We recognize the invaluable effort by employees and students from our research group throughout the data collection process at the HUNT4 Fitness and Echocardiography Studies, and especially the effort by sonographers Kristofer Soler and Even Olav Jakobsen.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The data from used in this study are available on application to the HUNT Data Access Committee in accordance with the policy on data availability (further information and contact information: https://www.ntnu.edu/hunt/data).

ORCID

Håvard Dalen https://orcid.org/0000-0003-1192-3663

REFERENCES

1. Gupta S, Matulevicius SA, Ayers CR, et al. Left atrial structure and function and clinical outcomes in the general population. *Eur Heart J*. 2013;34:278–285.
2. Kuznetsova T, Thijs L, Knez J, Herbots L, Zhang Z, Staessen JA. Prognostic value of left ventricular diastolic dysfunction in a general population. *J Am Heart Assoc*. 2014;3:1–11.
3. Badano LP, Miglioranza MH, Mihăilă S, et al. Left atrial volumes and function by three-dimensional echocardiography. *Circ Cardiovasc Imaging*. 2016;9:1–12.
4. Kebed K, Kruse E, Addetia K, et al. Atrial-focused views improve the accuracy of two-dimensional echocardiographic measurements of the left and right atrial volumes: a contribution to the increase in normal values in the guidelines update. *Int J Cardiovasc Imaging*. 2017;33:209–218.
5. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R. Normative reference values for the tissue Doppler imaging parameters of left ventricular function: a population-based study. *Eur J Echocardiogr*. 2010;11:51–56.
6. Sugimoto T, Robinet S, Dulgheru R, et al. Echocardiographic reference ranges for normal left atrial function parameters: Results from the EACVI NORRE study. *Eur Heart J Cardiovasc Imaging*. 2018;19(6):630–638.
7. Rohner A, Brinkert M, Kawel N, et al. Functional assessment of the left atrium by real-time three-dimensional echocardiography using a novel dedicated analysis tool: Initial validation studies in comparison with computed tomography. *Eur J Echocardiogr*. 2011;12:497–505.
8. 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.
9. Frikha Z, Girerd N, Huttin O, et al. Reproducibility in echocardiographic assessment of diastolic function in a population based study (The STANISLAS Cohort Study). *PLoS One*. 2015;10:1–12.
10. 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. *Eur Hear J - Cardiovasc Imaging*. 2015;16:233–271.
11. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033–1039.
12. Gerke O, Vilstrup MH, Segtnan EA, Halekoh U, Høilund-Carlsen PF. How to assess intra- and inter-observer agreement with quantitative PET using variance component analysis: a proposal for standardisation. *BMC Med Imaging*. 2016;16:54.
13. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307–310.
14. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One*. 2013;8:e73990.

15. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001;10:571–578.
16. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011;48:661–671.
17. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1:30–46.
18. 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 Hear J - Cardiovasc Imaging*. 2016;17:1321–1360.
19. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257–268.
20. Morbach C, Gelbrich G, Breunig M, et al. Impact of acquisition and interpretation on total inter-observer variability in echocardiography: results from the quality assurance program of the STAAB cohort study. *Int J Cardiovasc Imaging*. 2018;34:1057–1065.
21. Yao G-H, Zhang M, Yin L-X, et al. Doppler echocardiographic measurements in normal chinese adults (EMINCA): a prospective, nationwide, and multicentre study. *Eur Hear J - Cardiovasc Imaging*. 2016;17:512–522.
22. Okura H, Takada Y, Yamabe A, et al. Age- and gender-specific changes in the left ventricular relaxation. *Circ Cardiovasc Imaging*. 2009;2:41–46.
23. Rodevan O, Bjornerheim R, Ljosland M, Maehle J, Smith HJ, Ihlen H. Left atrial volumes assessed by three- and two-dimensional echocardiography compared to MRI estimates. *Int J Card Imaging*. 1999;15:397–410.
24. Artang R, Migrino RQ, Harmann L, Bowers M, Woods TD. Left atrial volume measurement with automated border detection by 3-dimensional echocardiography: comparison with magnetic resonance imaging. *Cardiovasc Ultrasound*. 2009;7:16.
25. Tsang W, Salgo IS, Medvedofsky D, et al. Transthoracic 3D echocardiographic left heart chamber quantification using an automated adaptive analytics algorithm. *JACC Cardiovasc Imaging*. 2016;9:769–782.
26. Heo R, Hong GR, Kim YJ, et al. Automated quantification of left atrial size using three-beat averaging real-time three dimensional Echocardiography in patients with atrial fibrillation. *Cardiovasc Ultrasound*. 2015;13:1–9.
27. Miyoshi T, Addetia K, Citro R, et al. Left ventricular diastolic function in healthy adult individuals: results of the world alliance societies of echocardiography normal values study. *J Am Soc Echocardiogr*. 2020;33:1223–1233.
28. Letnes JM, Nes B, Vaardal-Lunde K, et al. Left atrial volume, cardiorespiratory fitness, and diastolic function in healthy individuals: the HUNT study, Norway. *J Am Heart Assoc*. 2020;9:e014682.
29. Jenkins C, Bricknell K, Marwick TH. Use of real-time three-dimensional echocardiography to measure left atrial volume: comparison with other echocardiographic techniques. *J Am Soc Echocardiogr*. 2005;18:991–997.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Appendix S1: Supplementary methods and results

How to cite this article: Letnes JM, Eriksen-Volnes T, Nes B, Wisløff U, Salvesen Ø, Dalen H. Variability of echocardiographic measures of left ventricular diastolic function. The HUNT study. *Echocardiography*. 2021;38:901–908. https://doi.org/10.1111/echo.15073