Gunn Catrin Fossum Bøen

Feasibility and reproducibility of different echocardiographic indices of left ventricular function used for long-term follow-up to detect cardiotoxicity after breast cancer therapy

Master Thesis in Clinical Health Science

Applied Clinical Research

Trondheim, May 2018

Supervisors: MD, PhD, Professor Asbjørn Støylen

MD, PhD, Associate professor, Håvard Dalen

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

Abstract.

Background.

Cancer therapy may be cardiotoxic. There is a need for reproducible methods for longitudinal follow-up of cardiac function in these patients before, during and after cancer treatment.

Purpose.

The aim was to evaluate the feasibility and reproducibility of echocardiographic indices recommended for follow-up evaluation of left ventricular function in breast cancer patients.

Material and methods.

Echocardiographic recordings from 55 former breast cancer patients at their follow-up visit 7 years post treatment were analysed using 14 indices and investigated with respect to feasibility and reproducibility. The data obtained was analysed twice by the same operator for intra-rater reproducibility and by two separate operators for inter-rater reproducibility investigations.

Results.

The feasibility was excellent for Mitral Annular Plane Systolic Excursion (MAPSE) and all Doppler indices, and poor for tracking based Global Longitudinal Strain (AFI GLS). The semi-Automatic Left Ventricular Ejection Fraction (Auto EF), contrast enhanced LVEF ad. Simpson's rule and Early diastolic trans Mitral blood Velocity (MVE Vel) had significantly better reproducibility than 4 other indices.

Conclusions.

Intra-rater results are better than inter-rater results. Follow-up examinations by the same operator will improve the reproducibility compared to change of operator.

Relevance.

The number of breast cancer survivors is increasing due to better cancer therapy. For early detection of cardiotoxicity during or after cancer therapy there is a need for feasible and reliable measurements of cardiac function indices.

Keywords: Echocardiography; Feasibility; Reproducibility; Breast cancer; Cardiotoxicity.

Table of contents

..... ii
Abstract. ii
Relevance. ii
Abbreviations and acronyms. vi
1. Background. 1
1.1 Introduction. 1
1.2 Theoretical background. 3
2. Material and methods. 15
2.1 Literature. 15
2.2 Study population. 15
2.3 Study design. 16
2.4 Echocardiographic image acquisition. 17
2.5 The echocardiographic analyses. 19
2.6 The reproducibility analyses. 21
2.7 The statistical analyses. 22
2.8 Ethics. 23
2.9 Time planning and economy. 23
3. Results. 25
3.1 Study population. 25
3.2 Feasibility. 25
3.3 Reproducibility. 27
Intra-rater results. 28
Inter-rater results. 31
Significance of differences between methods according to inter-rater CoV. 33
Comparing intra-rater and inter-rater results. 34
4. Discussion. 37
4.1 Main findings. 37
4.2 Limitations 46
4.3 Clinical implication. 47
5. Conclusion. 49
References. 51
List of figures and tables. 59
Appendix 1-4. 60

Abbreviations and acronyms.

2D	Two-Dimensional data acquisition and presentation.
3D	Three-Dimensional data acquisition and presentation (4D includes time).
AFI	Automatic Functional Imaging, a tissue tracking based semi-automatic method for calculating GLS.
BA-plot	Bland-Altman plot, graphical presentation of differences.
BSA	Body Surface Area is found to be better correlated to echocardiographic indices than body weight or height.
CI	Confidence Interval, most often used 95 % CI = [mean ± (1.96 x SD)].
CMR	Cardiac Magnetic Resonance, see MRI.
CoV	Coefficient of Variation.
CT	Computed Tomography scan, computer-processed combinations of many X-ray measurements.
CTRCD	Cancer Therapeutics–Related Cardiac Dysfunction.
CVD	CardioVascular Disease.
EACVI	European Association of CardioVascular Imaging.
Echo	Echocardiography.
EF	Ejection Fraction. ((End-diastolic volume – end-systolic volume) / end diastolic volume) (%).
GLS	Global Longitudinal Strain. Short form of the more accurate GLPSS, global longitudinal Peak Systolic Strain.
GY	GraY, unit of absorbed dose of ionizing radiation.
HF	Heart Failure.
HUNT	Helse Undersøkelsen i Nord-Trøndelag, The Nord-Trøndelag Health Study.
ICC	Intraclass Correlation Coefficient, one of the reliability coefficients.
LV	Left Ventricle / Left Ventricular.
LVEF	Left Ventricular Ejection Fraction.
M-mode	Time Motion display of the ultrasound wave along a chosen ultrasound line.
MAPSE	Mitral Annular Plane Systolic Excursion, displacement of mitral annulus.

MRI	Magnetic Resonance Imaging (MRI) using strong magnetic fields, electric field gradients and radio waves to generate images of organs.
MUGA	Nuclear ventriculography using the radioactive material called technetium making a MUltiGated Acquisition scan of the heart. Also called nuclear heart scan, nuclear ventriculography, or radionuclide ventriculography (RNV).
PID	Patient IDentification number.
SD	Standard Deviation.
STE	Speckle-Tracking Echocardiography, tracking based.
%	Per hundred. Differences between two percentage values can be expressed as absolute or relative. The absolute differences are often called percentage points. The relative differences express change in true percentage.

1. Background.

This master thesis is written at the Norwegian University of Science and Technology (NTNU), Faculty of Medicine and Health Sciences (MH), Department of Circulation and Medical Imaging (ISB). The current investigation is part of a larger study called "Cardiovascular toxicity after treatment for breast cancer. The optimal diagnostic modality for early detection" and will evaluate the feasibility and reproducibility of the echocardiographic data. The reproducibility results will also work as quality control for all the echocardiographic data used in different co-projects in the main study.

Great thanks to my inspiring and knowledgeable supervisors:
MD, PhD, Professor Asbjørn Støylen and MD, PhD, Associate professor, Håvard Dalen,
both St. Olavs Hospital, Trondheim University Hospital and NTNU.

1.1 Introduction.

All biological measurements have a normal biological variation. In addition, every measurement has a method specific variability. Clinical studies of variability use different expressions dealing with these concepts. Thus, some of the main expressions used in this study are defined in the following paragraph.

Biological variation is the natural and true variation in biological values. **Variability** describes how much the values vary, often expressed as the mean and the Standard Deviation (SD) of the values. The measuring methods have additional variability. Measuring methods' variability is often expressed as their reproducibility. **Reproducibility** refers to the extent two repeated measures are likely to find the same result. Reproducibility is an umbrella term for the concepts of agreement and reliability (1). **Agreement** is the degree to which two repeated measurements are identical, often described by mean difference and Limits of Agreement (LoA). **Reliability** reflects both agreement and correlation between measurements and is often expressed as a reliability coefficient which is calculated considering patient and rater variability (2). Both agreement and reliability describe consistency of measures and will influence the validity of a measurement. **Validity** is an expression of how close the measured value is the true value. The true value must be defined by a reference method. With the best possible validity, one has perfect agreement between the score and the reference's score. In addition to studying the measurements' reproducibility, we need to know the feasibility of the methods. **Feasibility** means to which degree measures can easily or conveniently be done.

All clinical methods of measurements have qualities and limitations which must be considered. The true biological values vary and the method of measurement, the measuring equipment used, the population and finally the raters may also vary (1). Shrout et al. state that every measure involving people may be inaccurate (3). De Vet et al. underline the same by saying that measuring variations is the Achilles' heel of medical imaging (4). People make mistakes and their abilities to make analyses will vary. The reproducibility of analyses must be known to be able to make correct interpretations. To assess this, studies are needed to determine the reproducibility of the current methods in the current populations (3, 5) . Different populations' measures have different variability and will vary in many aspects, also in feasibility. Both feasibility and reproducibility of echocardiographic indices in our population will be considered in this study.

Reproducibility is not a fixed property of measurement tools, but rather a product of interaction between tools and subjects in a context. This is particularly important in echocardiography. To be able to trust investigation results, the reproducibility of measurements used in the current setting must be accounted for (6). Ideally, high feasibility, validity and reproducibility, full availability, a lack of side effects and low costs are desirable but is seldom an option. The choice of methods must consider all clinical demands and consequences. The acceptable size of measurement error is a clinical (not a statistical) decision (7). Method of measurement must be suitable for the actual setting.

When the reproducibility of the measures is crucial in the planning of lifesaving treatment, knowing the reproducibility of available methods is especially important.

Zamorano et al. sum up future perspectives and research directions in their position paper from 2016. Among other important issues, they emphasize a need of defining the most reliable cardiac monitoring approach for breast cancer patients (8).

The aim of the current study was to investigate the feasibility and reproducibility of different echocardiographic indices of left ventricular (LV) function in former breast cancer treated patients, all of them having been through potential cardiotoxic treatment programs.

Secondly, the study aims to compare feasibility and reproducibility of the different echocardiographic indices of LV function in former breast cancer treated patients.

1.2 Theoretical background.

Breast cancer.

Breast cancer is the most common cancer and the second most common cause of death in women in the western world (9). Radiotherapy, anthracyclines, hormones and immunotherapies (Herceptin) in different combinations and dosages are potential cardiotoxic therapies. Lancellotti et al. report a relative risk of radiation induced cardiac disease of 2 - 5.9 % proportional to radiation dose and number of exposures (10). In patients not developing clinical cardiac heart failure, a slow progressive and subclinical deterioration of cardiac function can be present in 20-30 % (11, 12). Cancer Therapeutics-Related Cardiac Dysfunction (CTRCD) may manifest acute or late. The acute manifestations are commonly subtle, while the late manifestations can occur several years after treatment (10, 13). CTRCD may manifest as left or right ventricular systolic and/or diastolic functional detoriation, valvular pathology, arrhythmias or pericardial diseases that may result in constrictive cardiomyopathy. These conditions may be caused by myocardial damage, microvascular dysfunction and coronary artery disease with potential myocardial ischaemia. Early detection of cardiac dysfunction allows for early start of treatment (10). Thus, as both cancer and heart failure are life threatening conditions, breast cancer patients are one group of patients that is highly dependent on reproducible methods for optimal therapy planning and follow-up.

Recommendations for echocardiographic follow-up in breast cancer patients.

As the number of breast cancer survivors is increasing due to better treatment, an increasing number of patients also need proper follow-up programs after their cancer treatments (9, 14, 15). Despite controversies around reproducibility, echocardiography is recommended as first line approach. Other options include MUltiGated Acquisition scan (MUGA) and Cardiac Magnetic Resonance (CMR). The alternatives have both advantages and disadvantages. MUGA has precise Left Ventricular Ejection Fraction (LVEF) calculations, but unfavourable radiation exposure. CMR provides excellent morphology, but has limited availability, difficulties in arrhythmic patients, risk of hazard problems with tissue expanders and metallic implants and higher cost (10, 16). Echocardiography is a quick and highly available modality with no side effects. It also gives a more complete investigation of the heart function beyond left ventricular function, including left ventricular diastolic function, right ventricular function, valvular and pericardial diseases, and findings suggestive of pulmonary hypertension (17, 18).

In the new position paper from the European society of cardiology, Zamorano et al. (8) summarize the main priorities for cancer follow-up as;

> *The same imaging modality and/or biomarker assay should be used for continued screening throughout the treatment pathway. Switching between modalities or assays is strongly discouraged. Modalities and tests with the best reproducibility are preferred.*
>
> *Imaging modalities that provide additional relevant clinical information are preferred (e.g. right ventricular function, pulmonary pressures, valvular function, pericardial evaluation).*
>
> *High quality radiation-free imaging is preferred, if available.*

For these reasons, echocardiography is recommended as method of choice for follow-up for detection of subclinical cardiac dysfunction from the European Association of Cardiovascular Imaging (EACVI) in their latest position paper from 2016 (8) . Similar recommendations are found in expert consensus papers from 2013 and 2014 (10, 14), and by the Norwegian health authorities' recommendations from 2017 for cancer treatment and follow-up (19) .

The EACVI recommendations list LV function investigated by echocardiographic ejection fraction (EF) as one of the recommended indices. Echocardiographic two-dimensional (2D) biplane LVEF by Simpson's method (from now on called LVEF) as a single measure has low reproducibility and thereby low sensibility for small and subclinical changes (1, 14, 20-24). Also, as a serial evaluation, the echocardiographic LVEF remains controversial, but has better reproducibility for contrast enhanced LVEF and highest for three-dimensional (3D) LVEF (14, 23). The echocardiographic definition of cardiotoxicity during or after cancer treatment is given in the EACVI expert consensus of 2014 as a decrease in the LVEF of >10 percentage points, to a value < 53% measured by modified biplane Simpson's technique in 2D echocardiography (14). In addition to LVEF, contrast enhanced LVEF and 3D LVEF, EACVI and others recommend additional echocardiographic methods such as Global Longitudinal Strain (GLS) as a useful and sensitive method, preferably compared to baseline values (14, 20, 25-27). For GLS used for follow-up during and after therapy in cancer patients, a relative percentage decrease of >15 % as compared to baseline is likely to represent significant cardiotoxicity (14).

The feasibility of echocardiography.

Echocardiography is doable in most patients, but image quality will vary between individuals. Reduced image quality may be caused by shadowing or noise, which can deteriorate the transmitted and the reflected ultrasound signal. Image quality can be affected by several factors such as obesity, lung diseases, smoking habits, scars, breast implants and more (14, 28, 29). It

is important to report the quality of cardiac imaging recorded, because the lack of some parameters can lead to inconclusive diagnosis when the quality of the images is suboptimal or poor. Under circumstances of suboptimal image quality, the use of advanced echo techniques such as strain and 3D LVEF, could be misleading (30). Different echocardiographic indices will have different feasibilities. Manual LVEF calculations are dependent on sufficient endocardial boarder definition in two planes for correct tracing. The feasibility of 2D biplane LVEF was reported by Malm et al. in 2004 as 86 %, where poor image quality of the 2-chamber view represented 92 % of the missed calculations (31). Szulik et al. reported in 2011 a feasibility of 93 % for 2D biplane LVEF and 90 % for the semi-automatic calculation, Auto EF (16). The feasibility of 3D LVEF was reported as 89 % from Chahal et al. in 2012 and as 73 % from Thavendiranathan et al. the same year (32, 33). Mitral Annular Plane Systolic Excursion (MAPSE) is an easily assessable method with feasibility reported as 95% by Støylen et al. in 2018 (34). Tissue velocities as peak Systolic velocity in mitral annulus (S') and Early diastolic velocity in mitral annulus (e') as well as trans Mitral blood flow Velocities in Early (MVE Vel) and late diastole (MVA Vel), all have good feasibility, often despite poor 2D image quality. Dalen et al. showed a feasibility for these indices of at least 96 % in 2010 (35). Strain is a method dependent on good image quality without considerable acoustic artefacts, often resulting in moderate feasibility scores (27). Barbier et al. investigated in 2015 GLS's feasibility both as excluded datasets and as percentage segments excluded. Exclusion of datasets were done because of incomplete LV visualization. They had 13.4 % excluded datasets, which gives an overall feasibility of 86.6 %. From the included datasets, their study showed highest feasibility in the basal anteroseptal wall and lowest for the basal posterior wall (36).

The population in the current study was expected to be challenging in respect of feasibility, both because of anatomical changes after surgery and because of shared risk factors with cardiac patients such as obesity and smoking habits (28, 37).

Contrast.

In patients with reduced image quality in every day practice, visibility may be improved by using ultrasound contrast to heighten the signal to noise ratio. Several studies indicate that contrast-enhanced echocardiography improves the evaluation of LVEF with respect to both feasibility and reproducibility (31, 38-41).

Ultrasound contrast is recommended when two contiguous LV segments are not well visualized (14). An example of clearly improved image quality with contrast is shown in Figure 1.

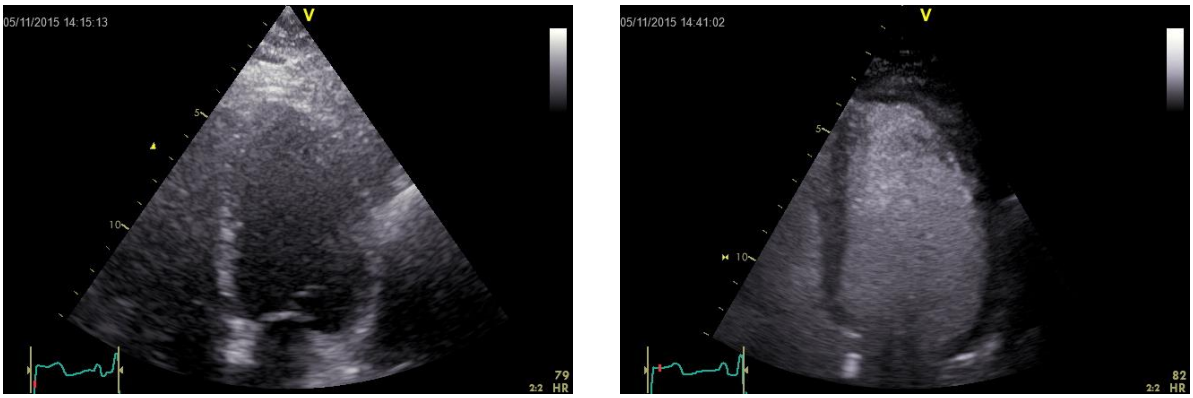


Figure 1. 4-chamber view of LV without contrast to the left and with contrast to the right, Patient IDentification number PID18355.

When reduced image quality is caused by shadowing, contrast may be of less help. Ultrasound contrast only helps when attenuation and thereby signal/ noise proportion is the problem. An example of less impressive improvements by ultrasound contrast is shown in Figure 2.

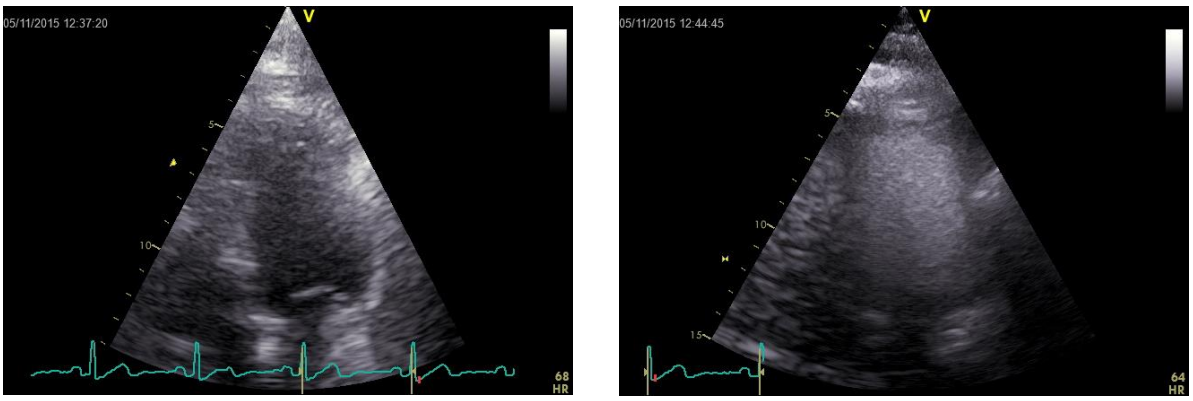


Figure 2. 4-chamber view of LV without contrast to the left and with contrast to the right, PID19183.

The reproducibility of echocardiography.

Feasibility is a prerequisite for performing echocardiography. Next, the required level of reproducibility must be specified. The reproducibility of echocardiography is studied by making two or more repeated analyses and comparing them in intra- and inter-rater computations (3). In the current study, the intra-rater results will give insight in how reproducible the same operator is from one analysis to the next, using the same set of recordings. The inter-rater results will show the reproducibility of different researchers doing repeated analyses using the same set of recordings (23). Many papers have discussed which statistical method to choose for reproducibility studies. Wood et al. and Erdei et al. published review articles in 2014 concerning reproducibility as agreement and reliability in echocardiography.

Agreement shown by Bland-Altman plot and the reliability coefficient called intraclass correlation coefficient (ICC) are widely recommended as statistical methods (6, 42-46). To calculate reproducibility, it is recommended to use more than one method (5). Bland-Altman plot makes a graphical visualisation of agreement showing all the differences, the mean of the differences (bias) and the LoA of the differences. The reliability coefficient tells how much of the variation in a sample is natural, and how much is due to errors. This can be illustrated by a general formula, as shown in Figure 3.

$$\text{reliability} = \frac{\text{variability between study objects}}{\text{variability between study objects} + \text{measurement error}}$$

Figure 3. The reliability coefficient. The expression variability represents the statistical expression variance, s^2 (SD^2).

The reliability coefficient is reported as a value between 0 and 1, where a ratio of 1 indicates perfect reliability, whilst 0 indicates no reliability (47). A reliability coefficient of 0.8 means that 80 % of the variability in these scores are real and 20 % of the variability represent random variation. There is no set standard for acceptable values of ICC. If important decisions are made on the basis of reliability, the ratio should according to Polit and Nunnally be at least 0.9 or 0.95, as cited by Kottner (6). ICC is like other correlation coefficients dependent by the variability between the subjects (3). More heterogenic measurements will give higher reliability coefficient. Therefore, it is important to make these calculations in the population of each study. The measurement errors need to be smaller than the size of the detoriation we want to detect (1). The reliability coefficient will thus reveal the ability to differentiate among subjects, and thereby reveal true changes (1, 6) . Many study designs require reproducibility investigations (48).

Left ventricular ejection fraction (LVEF) without and with ultrasound contrast.

LVEF is the fraction of the diastolic volume ejected in systole as described in the equation:
LVEF = (End-diastolic Volume – End-systolic LV Volume) / End-diastolic LV Volume (%).
Traditionally LVEF has been assessed by 2D biplane manually tracing of endocardial boarders in end-diastole and end-systole, in 4-chamber and 2-chamber views, described by modified Simpson's rule (29). This method is limited by several factors such as image quality, foreshortened operator recordings, malrotated and/or imprecisely angled images and calculations assuming normal geometry giving wrong results, particularly for ventricular remodeling (32). Exact planes are vital and demands experience to produce. Endocardial

boarder definition may be enhanced by ultrasound contrast, but the problem of foreshortening, malrotation and angling remains. Figure 4 shows an example of foreshortening.

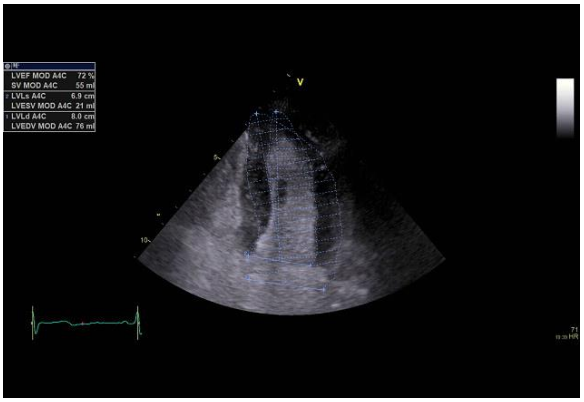


Figure 4. Tracing of LVEF with contrast enhanced endocardial boarders. The traces show the true apex which was clearly seen in diastole. The contrast filled area shows a shorter cavity, demonstrating foreshortening in PID19071.

Normal values for biplane LVEF in women according to Lang et al. in 2015 is mean (SD in percentage points) equal 64 (5) % giving a 2xSD range of [54, 74] (49). Chahal et al. found normal values for 2D LVEF as mean (SD) 63 (5) % for 161 European women (32).

Malm et al. reported LoA as mean ± 2xSD of inter-rater differences of biplane LVEF in percentage points as [–16.6, 14.2] without contrast which narrowed to [– 5.9, 6.9] with contrast. Their intra-rater differences had LoA [–11.1, 7.8] without contrast and LoA [–2.8, 2.4] with contrast. Mean inter- and intra-rater variability for biplane LVEF were reduced from 13.9% to 9.6% and from 5.4% to 2.5% without and with contrast, respectively. Their population had known or suspected heart disease (31).

Stanton et al. reports good reliability for LVEF. They report intra-rater intraclass correlation coefficients within and across observers as ICC of 0.67, and an inter-rater ICC of 0.80. Their population had known or suspected LV impairment as cardiovascular risk factors or illnesses, but no known cancer disease (50).

In 2016, Fei et al. investigated whether LVEF and GLS were able to identify recovery of CTRCD. Their work is of special interest because their population is similar to ours. The researchers found intra-rater variability calculated as Mean Error (SD in percentage points) as 5.1 (3.4) % and inter-rater Mean Error (SD) as 5.4 (3.2) % for LVEF (12).

LVEF has a marginal ability to discover a left ventricle function detoriation of 10 percent points. Strategies using newer echocardiographic technology, such as speckle-tracking (STE) derived strain imaging for the early detection of subclinical LV systolic dysfunction, have been actively investigated (14). Our study is continuing this project.

Semi-automatic assessment of left ventricular ejection fraction (Auto EF).

Auto EF (vendor specific method from GE, Vingmed Ultrasound, Horten, Norway) is a new semi-automatic, STE based method to assess 2D LVEF. Like traditional biplane 2D LVEF by modified Simpson's rule, this method also underestimates the LVEF value compared to CMR but proves to have less intra- and inter-rater variability than the traditional LVEF.

Szulik et al. found significantly lower intra- and inter-rater variability for the STE based LVEF (Auto EF) than for LVEF using modified biplane Simpson's rule. Analyzed by Bland-Altman plot their inter-rater differences (±1.96xSD in percentage points) were ±10.7 % and ±18.6 % and their intra-rater differences were ±6.1 % and ±15.5 % respectively. Their conclusion emphasizes the advantage of good reliability in follow-up situations (16).

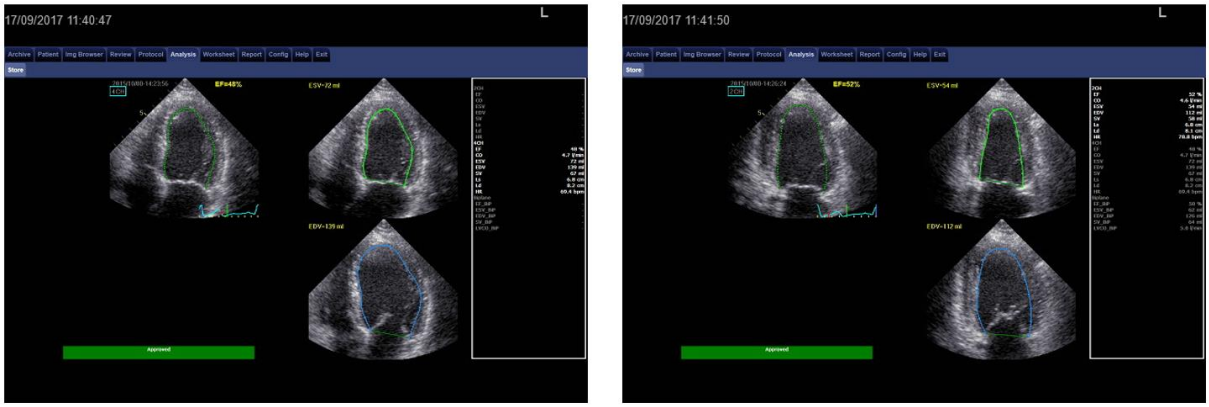


Figure 5. Biplane Auto EF in PID18388. 4-chamber view in the left and 2-chamber view in the right panel.

Three-dimensional ejection fraction (3D LVEF).

Echocardiographic 3D LVEF is recommended before 2D LVEF because there are less errors caused by the geometric assumptions which are used by 2D LVEF calculations. 3D LVEF also works around some of the well-known problems concerning foreshortening and inaccurate orientation of the views. The method is therefore less affected by acquisition differences in serial analyses and is thus a more automated process with less variation. On the negative side, 3D LVEF acquisition and analysis are dependent on good image quality and operator experience (14).

Thavendiranathan et al. showed in 2013 that 3D LVEF is far better than 2D LVEF concerning agreement and reliability. They expressed the variability by Standard Error of Measurement (SEM) and had chosen to express absolute changes as decimals. Their study found an intra-rater variability of 0.017 and an inter-rater variability of 0.027 for 3D LVEF without contrast.

Their population consisted of women with breast cancer undergoing chemotherapy (23). Chahal et al. report reference values of 3D LVEF for European women as mean (SD in percentage points) as 62 (5) % (32).

Systolic tissue velocity (S').

S' is the peak systolic tissue velocity in mitral annulus assessed by pulsed wave tissue Doppler imaging (pwTDI). In the current study, tissue velocities were measured in 4-chamber view from the septal and the lateral mitral annulus to be averaged and presented as cm/s. The measurements were done at the peak of the upper edge of the solid Doppler curve, with scale optimized and low gain settings as recommended by Dalen et al. in 2010. Normal values of S' were reported by Dalen et al. as mean (SD) 8.2 (1.3) cm/s in women, averaged for four walls (35).

Mitral annular plane systolic excursion (MAPSE).

MAPSE is a robust measure which is detectable in most patients. It is defined as the length of the displacement of the mitral annuls from diastole to peak systole measured by M-mode, reported in mm or cm. MAPSE should theoretically be measured in all six walls to be averaged, but is traditionally measured in only two walls. Støylen et al. recently reported age dependent and gender independent normal values of MAPSE. Mean (SD) for MAPSE averaged for two walls was reported as 1.55 (0.24) cm for women 40 – 60 years old and 1.39 (0.25) cm in women over 60 years. Four-wall average was suggested as optimal for global MAPSE with respect to reproducibility (34).

Left ventricular length (LV length).

LV length is measured to calculate generic longitudinal strain. The LV length was acquired by averaging the end diastolic lengths from the most distant point of epicardial apex to 1) septal mitral annulus and 2) lateral mitral annulus, reported in mm. Støylen et al. reports age and gender dependent normal values of LV length in data from The Nord-Trøndelag Health Study (HUNT). Averaged for two walls, the normal values of left ventricle wall length as mean (SD) was 9.1 (1.7) cm for women 40 - 60 years old and 8.9 (1.3) cm for women over 60 years old, with reproducibility reported as CoR equal 0.7 cm and Mean Error equal 2.8 % (51).

Global longitudinal strain (GLS).

GLS is a measure of myocardial deformation as contraction or lengthening, based on tissue-velocity or STE. STE is the favored method because of less angle dependency (14). The formula is simply: (L0 – L1) / L0) (%), where L0 is baseline length and L1 is the changed length, giving longitudinal strain to be presented as percent. Systolic myocardial shortening is presented as negative, diastolic lengthening as positive strain values. Lower strain may be described as less negative values, but to avoid misunderstandings we should use increase or decrease in the absolute value of strain. For example, a change in GLS from – 19 % to – 16.2 % or higher would be a clinical significant decrease (> 15% relative change) in global longitudinal strain (14), also termed reduction. Global longitudinal strain is listed as the optimal parameter for the early detection of subclinical LV dysfunction in many studies and in EACVI's expert consensus by Plana et al. (14) .

Normal values of GLS is reported by Yingchoncharoen et al. in their meta-analysis from 2013 as mean -19.7% with a 95% CI of [-20.4%, -18.9%] (52). Dalen et al. reported age and gender dependent normal values of GLS as mean (SD in percentage points) of -17.6 (2.1) % in women 40 – 60 years old and -15.9 (2.4) % in women over 60 years old. The Vivid 7 ultrasound machine from General Electric (GE) was used calculating GLS based on a combination of tissue Doppler and speckle-tracking (27). There are different normal values published, also exemplified by the large NORRE study publishing higher normal values of GLS than values published from the HUNT study (27, 53).

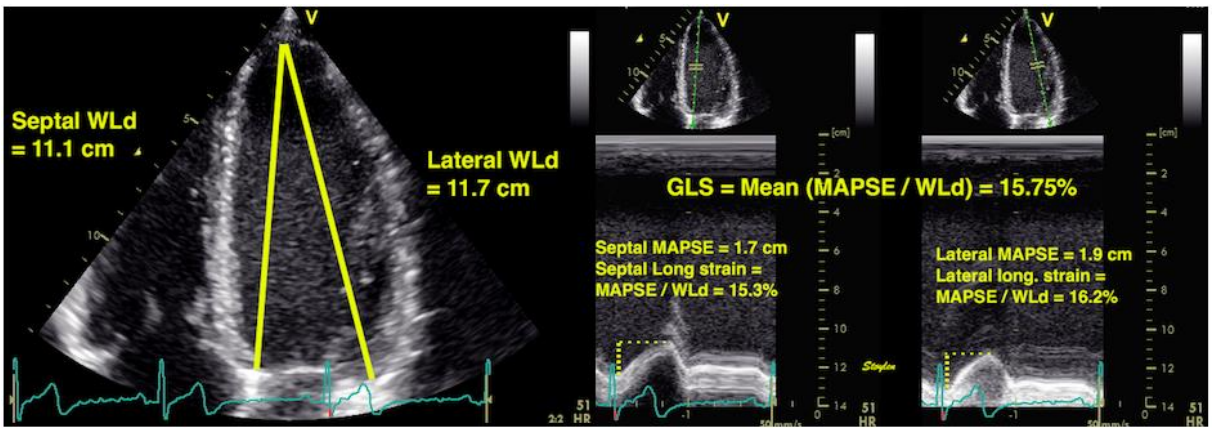
Stanton et al. suggest in 2009 GLS as the optimal method for assessment of left ventricular systolic function. With GE's Vivid 7, they reported reliability for manually traced GLS by STE method as good. Their study found intra-rater ICC within and across observers as 0.92, and inter-rater ICC as 0.92. Their population had known or suspected LV impairment as cardiovascular risk factors or illness, but no known cancer disease (50).

Cheng et al. studied a population from the Framingham study in 2013 and found an inter-rater ICC equal to or larger than 0.84 for all their GLS calculations, with an average CoV equal to or lower than 4 % and intra-rater values as 0.91 and 6% respectively. (54). Fei et al. investigated a population comparable to ours and found intra- and inter-rater variability of GLS as mean (SD in percentage points) 0.8 (0.6) % and 1.3 (0.8) %, respectively (12). Both studies used the speckle-tracking based 2D Cardiac Performance Analyzing (CPA) program from TomTec Imaging systems. Plana et al. lists several reproducibility studies for tracking based GLS in cancer patients using GE-machines, which thereby is comparable to our study (14). GLS is dependent on age, gender and vendor, and therefore meta-analyses might have limited interest.

As the different vendors have their own patented data algorithms, comparing measures is challenging (14). Takigiku et al. found normal values for females over 60 years as mean (SD) -20.9 (2.1) % from GE, -17.3 (2.3) % from Philips and -18.6 (2.3) % from Toshiba (55). Yingchoncharoen et al. found equipment vendor not to be significantly associated with mean GLS in normal subjects in their meta-analysis of 24 studies (52).

Generic longitudinal strain.

Generic longitudinal strain is a basic strain calculation as annular displacement normalized for LV length. This way to calculate global strain is the simplest way and includes only robust parameters (56). The formula is shown in Figure 6. The method is to our knowledge not previously investigated. We expected to find good feasibility and at least fair reproducibility for generic longitudinal strain in our study.

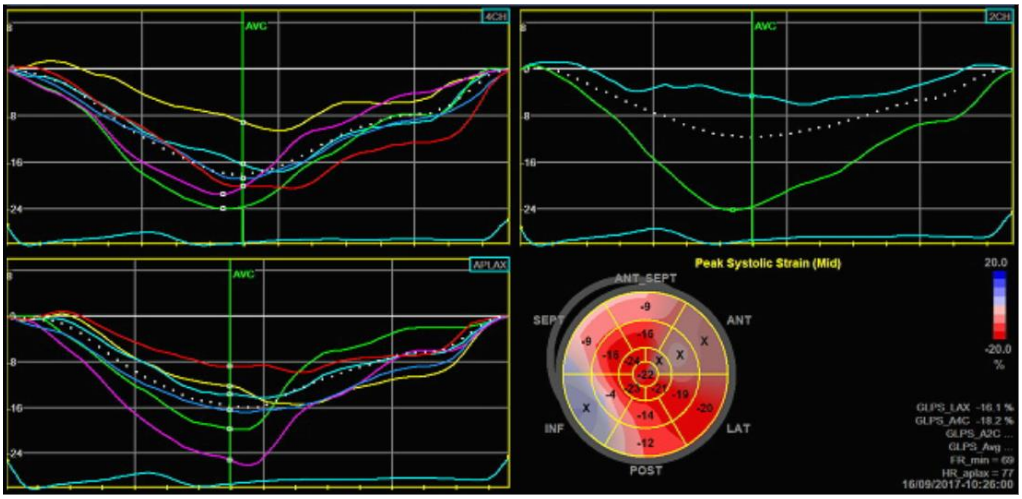


WLd: Wall length in diastole. GLS: Generic longitudinal strain. MAPSE: Mitral annular plane systolic excursion.

Figure 6. Generic longitudinal strain illustrated by Støylen (56).

Automated functional imaging (AFI).

AFI is a semi-automatic, vendor specific method from GE Vingmed Ultrasound AS, used for measuring STE based strain. Strain may be measured as longitudinal, circumferential and radial. In our study we investigated the STE based longitudinal peak systolic strain. This semi-automatic method can be manually adjusted by the operator whenever needed. AFI can be calculated in single planes, as our study did for 4-chamber view (AFI 4CH). The results of triplane AFI is called global strain (AFI GLS). AFI GLS may be presented in different ways, for example as strain curves and Bull's Eye shown in Figure 7.



This figure shows left ventricular regional strain as strain curves and as Bull's eye plot. The strain curves from the 4-chamber view is shown in the upper left panel. The strain curves from the 2-chamber view is shown in the upper right panel and strain curves from the apical long axis view is shown in the lower left panel. All the curves are color coded to address the specific segments of the left ventricle, corresponding the Bull's eye plot. Apical segments are shown in the center and basal segments shown in the outer ring of the Bull's eye. Mean strain is calculated for each segment in each view and for all views averaged as global strain (GLS). In the Bull's eye we see the strain values color coded in the range of max contraction as – 20.0 % colored dark red, via lighter red towards light blue to dark blue as max lengthening as + 20.0 %. Segments colored grey and marked with x show rejected segments due to poor image quality, here in all segments of the anterior wall.

Figure 7. AFI, strain curves from all segments in three planes and Bull's eye plot in PID19045.

Diastolic function.

Left ventricular diastolic function has shown to provide important prognostic information (57, 58). The use of the calculation E/e' remains questionable in oncologic settings because of the loading dependency of mitral flow. Fluctuating loading conditions due to nausea, vomiting and diarrhea must be considered in this patient group (14). The following diastolic parameters were investigated in our study:

MVE Vel: Trans mitral early diastolic blood-velocity. Normal values for MVE Vel is reported from the large HUNT database in 2009 as mean (SD) 75 (16) cm/s (27).

e' (cm/s): Early diastolic tissue-velocity measured in septal and lateral mitral annulus to be averaged and presented as cm/s.

Normal values reported as mean (SD) 11.8 (3.2) cm/s in women from the HUNT study. From the same study feasibility is reported superior to deformation imaging, but not feasible in regional dysfunction (35).

E/e': The proportion of MVE Vel over e`.

MV E/A: The proportion of early trans mitral diastolic blood velocity (MVE Vel) over late trans mitral diastolic blood velocity (A = MVA Vel). In patients with depressed LVEF, MV E/A < 0.8 indicates normal left atrial pressure and MV E/A > 2 indicates elevated left atrial pressure (57).

MV Dec T (ms): The time used for trans mitral early filling, deceleration time (from start to zero flow) measured in milliseconds. The variable is age and gender dependant. Dalen et al. reported in 2010 normal values from 673 healthy Norwegian women of mean age 47.8 years, mean (SD) MV Dec T = 218 (66) ms (27).

2. Material and methods.

2.1 Literature.

All referred articles are found during three years of working with these theories. The MeSH expression for our main topic is “Reproducibility of results”. To get into the echocardiographic field, search words such as echocardiography, LVEF, EF, ejection fraction, GLS, global strain, Cancer Therapeutics-Related Cardiac Dysfunction, CTRCD and cardiotoxicity were used searching online. For all articles found in unfamiliar journals, the impact factor was checked. The statistical articles are ranged from older to somewhat newer, while the echocardiographic articles are mostly newer, including newer methodology and equipment. The European Association of Cardiovascular Imaging (EACVI) has expert consensuses, recommendations and position papers used as background for this study (8, 10, 14, 30, 49, 57, 58).

2.2 Study population.

The population in this study is a sample from 250 patients in a former study named “Radiotherapy of breast cancer patients.” The researchers investigated the association between treatment, side effects and health-related quality of life (59, 60). Patients who had gone through treatment programs with high risk of treatment related cardiotoxicity, were asked to join the Echo and CMR group in this new and large study called “Cardiovascular toxicity after treatment for breast cancer. The optimal diagnostic modality for early detection.” The patients were included after giving informed and written consent during the years 2014 – 2015.

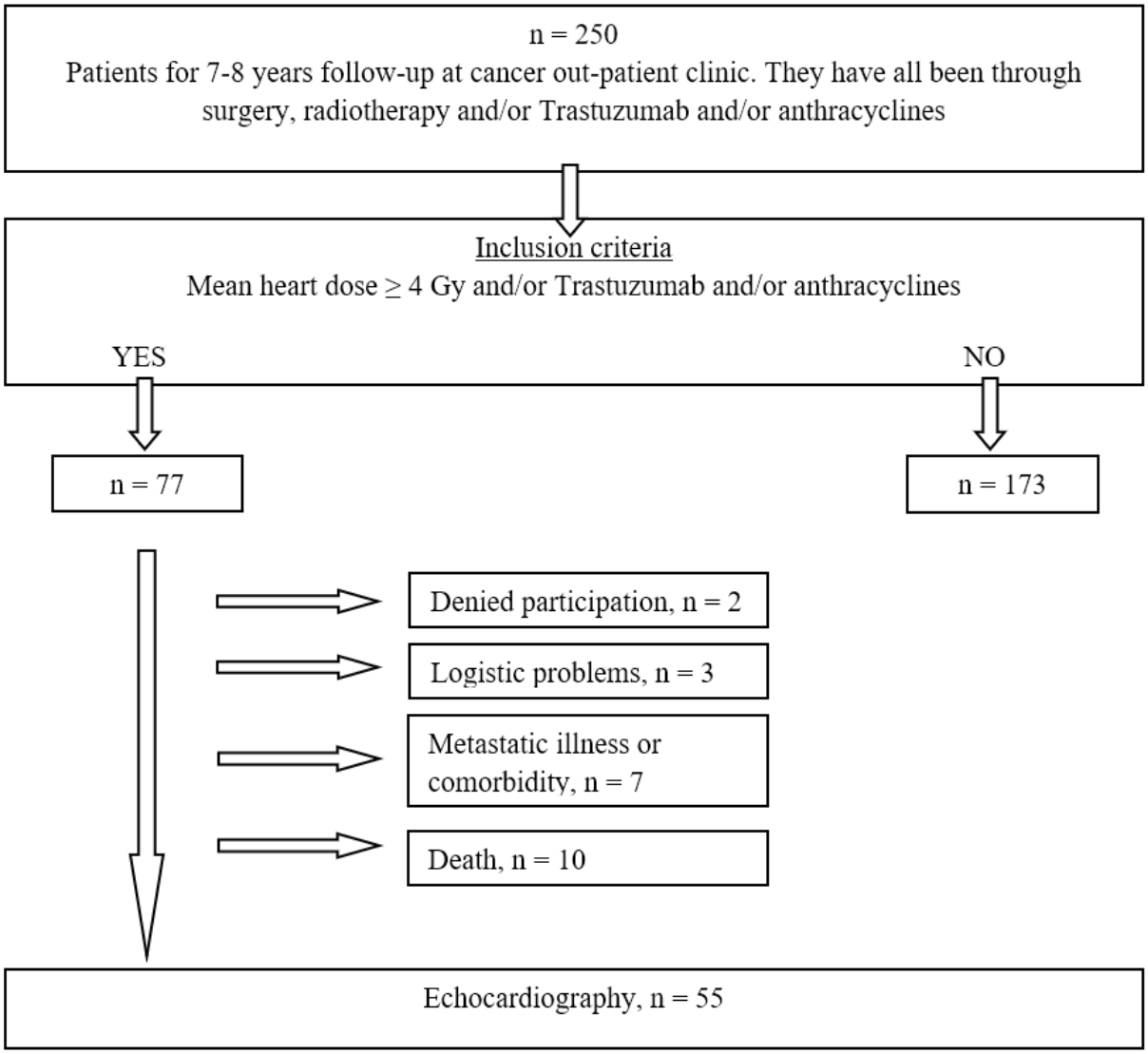


Figure 8. Flowchart of study population.

2.3 Study design.

Echocardiographic recordings were acquired from all 55 patients.

Echocardiographic analyses were done for the whole group once. Feasibility was investigated in the whole group of 55. The reproducibility analyses were performed at a random sample of 20 patients. The randomization was done by a random number generator. Reproducibility investigations for the selected echocardiographic indices were organised as an intra- and inter-rater investigation with two raters.

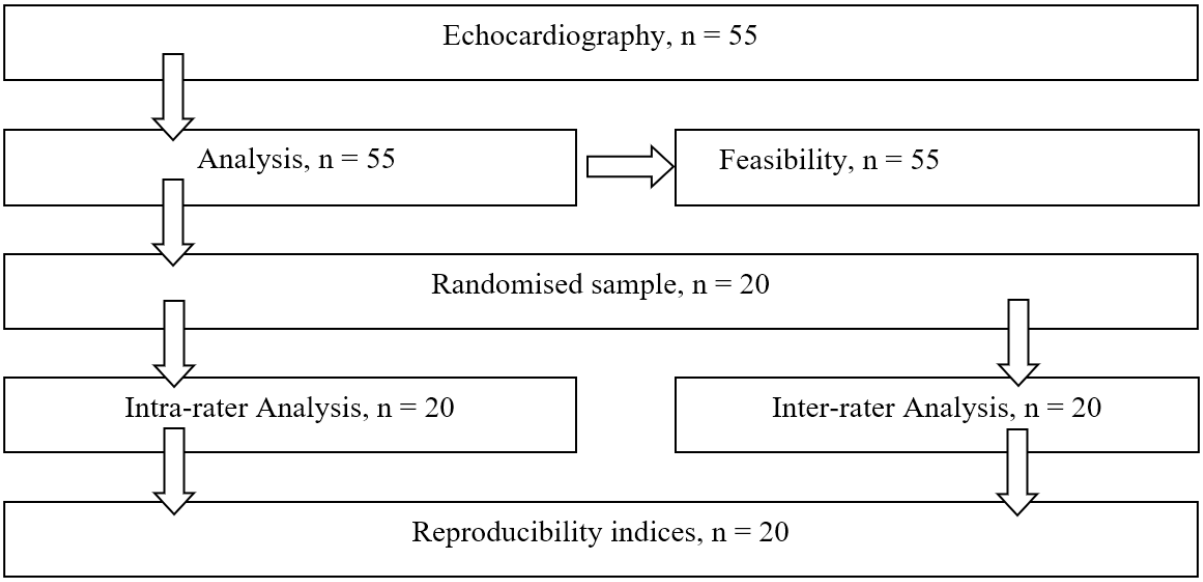


Figure 9. Flowchart of study design.

The intra-rater analyses were done by rater 1 (GCB). A full set of echocardiographic analyses was obtained during a few weeks, following a two weeks break. The second round of echocardiographic analyses was then completed for the intra-rater investigations, without any thought of the former analyses. A copy of all the 20 full recordings was given to rater 2 (AS) to be analysed for the inter-rater investigations. The analyses from rater 2 were completed in a different time and place, without communication between the two raters regarding the cases. The two raters are colleagues but met only occasionally and emphasized not to discuss any matter concerning the analyses.

2.4 Echocardiographic image acquisition.

In this study a set of the echocardiographic indices concerning left ventricular function were acquired specifically with regards to feasibility and reproducibility. The parameters are all present in recommendations from EACVI (8, 14).

Recordings were done in the left lateral supine position. For each view, at least one set of three consecutive cardiac cycles were recorded during quiet respiration.

All 55 complete recordings were acquired by the same operator (GCB). Our study procedure called" Echocardiographic Acquisition Procedure" is based on European guidelines (14, 49, 57). The procedure is shown in Appendix 1. Table 1 gives a listed overview of acquisitions. Recordings of the right ventricle and the valves are kept for future investigations.

All saved recordings were anonymously archived using Patient study IDentification number (PID). The patients also reported their height, weight and year of birth.

Table 1. A listed overview of the echocardiographic image acquisitions.

- Left ventricle in parasternal long-axis (PLAX) two-dimensional(2D) and M-mode.
- Left ventricle in parasternal short-axis (PSAX) two-dimensional(2D) and M-mode.
- Left ventricle outflow tract (LVOT) in Zoomed mode
- Left ventricle in the three standard apical views, 4-chamber, 2-chamber and long-axis as 2D, Tissue Velocity Imaging (TVI) and contrast (CO).
- Left ventricle pulsed wave (PW) tissue Doppler and M-mode both in septal and lateral mitral annulus.
- Left ventricle as a full 3D dataset.
- Right ventricle in apical right ventricle-optimized view 2D, TVI.
- Right ventricle tricuspidal annulus pulsed tissue Doppler, M-mode.
- Blood velocities through the Mitral valve, from the pulmonary veins, in LVOT and through the aortic valve and the tricuspid valve were obtained by pulsed and continuous Doppler.
- Valvular function was documented with grey scale recordings, spectral and colour Doppler.

Feasibility.

Feasibility was calculated for the whole group of 55 patients, presented as percent satisfactory analyses for each of our investigated echocardiographic index. Feasibility is of vital importance to express applicability of the measuring methods. Feasibility scores above 85 % was defined as good in our study.

Contrast.

Ultrasound contrast was used according to recommendations to enhance the left ventricular endocardial border (39, 40). In our study, ultrasound contrast was given to all the patients, independent of initial image quality, to investigate the impact of ultrasound contrast on feasibility and reproducibility issues in our population.

SonoVue (Bracco, Milan, Italy) was used for the contrast recordings. SonoVue is a microbubble based, non-radiation contrast solution with bubbles of mean size 2.5µm. The suspension was injected intravenously as a bolus of 1 – 2.5 ml followed by a 5ml saline injection.

Equipment.

The echocardiographic investigations were completed using a high-end ultrasound machine from General Electric (GE) Vingmed in Horten, Norway, equipped with their special heart probes for 2D and 3D (4D) data acquisitions.

Ultrasound machine:	GE Vivid E9 with XD clear, v.113, HW number 6975.
Probes:	M5Sc-D (1.7-3.3 MHz) and 4V (1.7-3.3 MHz).
Analysing program:	GE clinical workstation software, EchoPAC, application software version 113, rev. 1.1.

2.5 The echocardiographic analyses.

Investigated parameters.

In this study a set of the echocardiographic indices concerning left ventricular function were analysed with respect of feasibility and reproducibility. The indices are all present in recommendations from EACVI (8, 14, 30, 49, 57). All these indices can show changes in systolic or diastolic left ventricular function. Both traditional and newer measurements and methods are investigated. Table 2 gives an overview of all fourteen investigated parameters. The complete analysing process is described in the" Reading protocol" in Appendix 3.

Raters.

The two raters are experienced echocardiographers working at St. Olavs Hospital, Trondheim University Hospital. Rater 1 (GCB) works as a cardiovascular technologist and senior engineer and has approximately 12 years of experience in echocardiography. Rater 2 (AS) MD, PhD works as a cardiologist and has more than 20 years of experience in echocardiography. The two raters arranged a recommended training session to agree on reading techniques (23, 61-63). The training took approximately one hour, analysing all parameters involved. Measuring methods were agreed on to ensure the same basic understanding of what and how to measure and analyse. A reading protocol based on these agreements was written by rater 1, shown in Appendix 2. Rater 1 did three complete analyses, one for all 55 patients to investigate feasibility and to serve an early co-project. Then two complete analyses for the sample of 20 patients for intra- and

inter-rater analyses. Rater 2 did one complete analysis of the same 20 patients. The two raters did all the analysis blinded to each other's results.

Table 2. List of investigated parameters.

- Ejection fraction (LVEF) (%)
    - Biplane Simpson's method manually traced
    - Biplane Simpson's method manually traced with ultrasound contrast
    - Auto EF
    - 3D LVEF
- Mitral annular plane systolic velocity, S' (cm/s)
- Mitral annular plane systolic excursion, (MAPSE) (mm)
- Strain (%)
    - Generic method, MAPSE/ LV length
    - Automatic Functional Imaging, 4-chamber view (AFI 4CH)
    - Automatic Functional Imaging, triplane (AFI GLS)
- Diastolic function
    - MVE Vel (cm/s)
    - e' (m/s)
    - E/e'
    - MV E/A
    - MV Dec T (ms)

Ejection fraction: (Left ventricle (LV) diastolic volume – LV systolic volume) / LV diastolic volume. Auto EF: Biplane, semi-automatic STE based calculation of LVEF. Generic strain: Mean (MAPSE/ WLd). Automatic functional imaging (AFI): Semi-automatic, Speckle-tracking based strain calculation. MVE Vel: Trans mitral early diastole blood velocity. e': Early diastolic tissue velocity. MV E/A: Trans mitral E velocity/ A velocity. MV Dec T: Mitral Valve flow deceleration time.

2.6 The reproducibility analyses.

The measurements' reproducibility was investigated by intra- and inter-rater analyses. Intra-rater reproducibility in the current study was defined as the reproducibility calculated from rater 1's analysis and re-analysis of all indices from the same set of recordings. Inter-rater reproducibility was defined as the reproducibility calculated from the two raters' analyses of all indices from the same set of recordings. The raters might randomly select different pictures and loops for analysis and reanalysis.

Bland-Altman-plot (BA-plot).

Reproducibility with regards to degree of agreement of intra- and inter-rater differences, was calculated as mean differences and standard deviation (SD) of the differences of the repeated analyses, graphically presented as BA-plots of the differences (7, 64). The mean of the differences is called bias and was marked in the graph. The LoA of the differences was calculated as bias ±1.96xSD to cover the range of values which with 95 % certainty includes 95 % of all the differences. LoA was marked as lines in the graphs. The graphical presentation was made with the normal value of each index as the total size of the y-axis of the plot to visualize the proportions.

Coefficient of variation (CoV).

As a supplement, mean CoV was calculated as SD of the difference between two repeated measures, divided by the mean of the two measures multiplied by 100 to be presented as percent for all patients in all indices. Mean CoV is their average. CoV will covary with the size of the measures and is not a recommended coefficient (65). CoV was included to be able to compare with other similar studies.

Coefficient of repetition (CoR).

CoR is defined by Bland and Altman as 2xSD of the differences, based on a linear relationship between errors and measurements directly related to the 95% LoA and expressed in measurement units (7). The value of this coefficient must be exceeded to know there is a significant difference between measures. CoR is not included in the main results in Table 5, but numerical values of CoR is calculated and discussed under clinical implications.

Intraclass correlation coefficient (ICC).

Reproducibility as reliability with regards to consistency and correlations was calculated as ICC. This model takes the systematic differences into account and models both the effect of different raters and different measures (1). To calculate the intra-rater reliability, the ICC (3,1), a two-way mixed model with one rater was used. For the inter-rater investigation ICC (2,2), a two-way random model with two raters was used, both with the absolute agreement type of analysis and the coefficient based on mean ratings (2, 5, 6, 62). Categorisation of the ICC results is done by the general assumptions of Portney and Koo et al. as ICC > 0.9 indicating excellent reliability, 0.9 > ICC > 0.75 good reliability and 0.75 > ICC > 0.5 moderate and ICC ≤ 0.5 indicationg poor reliability (2, 62). The ICC highly depends on the heterogenicity of the subjects. The more heterogenic measures in the sample, the higher the ICC. The computation of ICC depends on complete pairs. Missing measures, degrees of freedom and Cronbach's alpha is accounted for in tables available in Appendix 4.

2.7 The statistical analyses.

The statistical analyses were performed using IBM SPSS for Windows (SPSS, Inc., Chicago, Illinois, USA) statistics version 24 and 25. The data was tested for normality by using a QQ-plot with broad pencil method.

Analysis of variance (ANOVA).

A one-way ANOVA with post hoc analysis with the least significant different (LSD) correction was performed to compare the mean inter-rater CoV of all investigated methods. $P < 0.05$ was considered statistically significant.

Bland-Altman-plot (BA-plot).

Both visual inspection of the graphics and simple linear regression analysis was used to check for potential proportional bias. The intra- and inter-rater differences were tested for significant difference from zero by a One Sample T-Test.

Power.

In line with comparable studies such as (12, 33, 66) , the current study aims to achieve adequate power to reveal clinically significant differences. This is done by having a sample size of 20 patients and by using 2 raters. Unfortunately, precise power calculations were not completed for the current study. Completing these calculations would have allowed us to find whether this study has sufficient power to allow statistical comparisons.

2.8 Ethics.

The study was approved by REC midt (2009/108 4.2006.2856). Most of the participating patients were in good health, cured from their cancer. The participation was voluntary and included minimal risk and for most people no discomfort. Study participants received an information letter and were asked to sign a written consent. The information letter is shown in Appendix 3.

2.9 Time planning and economy.

Table 3. Timetable.

Autumn 14	Feb 15 - Jan 16	Feb 16 - Jun 16	May 17	Sept - Oct 17	Oct 17 - May 18	May 18
Protocol Main project (2009/108/REK midt)						
	Inclusion Echocardiography					
		Basis analyses				
			Protocol Master			
				Inter/intrarater analyses		
					Statistics and writing	
						Hand in

The patients participated in this study voluntarily with no cost for the project. All the echocardiographic recordings were paid as 61 hours, NOK 15000 in wages. The ultrasound contrast had a price of NOK 700 x 25. This sums up to NOK 32500. All expenses were covered by the main project with a budget from Helse Midt-Norge.

3. Results.

3.1 Study population.

The 55 previously breast cancer treated patients were all women, who at the starting point of this study had a mean age of 61, ages ranging from 41 to 76 years old. Their mean body surface area (BSA) was 1.8 m², ranging from 1.4 to 2.2 m². The sample of 20 patients were of mean age 61, ranging from 45 to 73 years old, mean BSA was 1.8 m², ranging from 1.6 to 2.0 m². Basic echocardiographic characteristics for the whole group are listed in Table 4. Further population data can be found in previously mentioned studies by Reidunsdatter et al. (59, 60).

3.2 Feasibility.

The obtained echocardiographic data for the whole group was analysed by rater 1 during the summer of 2016. The feasibility scores were calculated for all 55 patients as number of satisfactory analyses, presented as number and percent in Table 4.

All data were found normally or near normally distributed. Thus, data are presented as mean and standard deviation (SD).

A summary of our findings:

- Mean values for systolic and diastolic echocardiographic parameters were normal or near normal compared with previously published material (27, 32, 34, 35, 49, 51, 52, 55, 67).
- Feasibility was excellent for MAPSE and all Doppler indices.
- Feasibility was good for LVEF without contrast, 3D LVEF and Generic strain.

Excellent feasibility was experienced measuring MAPSE with the score of 100 %. Further, all Doppler measurements had very good feasibility calculated as 98 %. The deviation from 100 % for the Doppler indices, was because of one missed recording. LVEF manually traced with Simpson's method and 3D LVEF showed very good feasibility proportions, both 96 %. Contrast enhanced LVEF and Auto EF showed feasibilities as 81 % and 78 % respectively. The lowest feasibility was found for global longitudinal strain acquired by automatic functional imaging (AFI GLS) and for strain acquired only from 4-chamber view by automatic functional imaging (AFI 4CH) as 51 % and 55 % respectively.

According to these feasibility results, the most feasible methods for calculating left ventricular systolic or diastolic function in this population are MAPSE and all the Doppler indices.

Table 4. Echocardiographic characteristics and feasibility for the whole study population of 55 patients.

	Value	Feasibility
	Mean (SD)	n (%)
LVEF (%)	57 (8.9)	53 (96)
LVEF with contrast (%)	60 (8.7)	45 (81)
Auto EF (%)	54 (5.9)	43 (78)
3D LVEF (%)	50 (10.1)	53 (96)
S' (cm/s)	7.0 (1.3)	54 (98)
MAPSE (mm)	14 (1.9)	55 (100)
Generic Strain (%)	-17.1 (2.0)	47 (85)
AFI 4CH Strain (%)	-17.1 (3.2)	30 (55)
AFI Global strain (%)	-17.9 (3.0)	28 (51)
MVE Vel (cm/s)	66.7 (25.7)	54 (98)
e' (cm/s)	7.6 (2.8)	54 (98)
E/e'	9.0 (4.5)	50 (91)
MV E/A	0.9 (0.3)	54 (98)
MV Dec T (ms)	226 (60)	54 (98)

Mean (SD): mean values of all 55 patients and their standard deviations. Feasibility: number (percentages) of patients with satisfactory analyses. LVEF: Left ventricular ejection fraction, manually traced using Simpson's 2D Biplane Method without and with ultrasound contrast. Auto EF: a semi-automatic biplane tracing of LVEF. 3D LVEF: a semi-automatic tracing and calculation of LVEF from a 3D dataset. MAPSE: Mitral Annulus Plane Systolic Excursion, averaged septal and lateral measures. Generic strain: Calculated from mean (MAPSE/ Left ventricular length). AFI 4CH Strain: Automatic Functional Imaging calculating strain by speckle-tracking from only the 4-chamber view. AFI Global strain (AFI GLS): Automatic Functional Imaging measuring strain with speckle-tracking from all three standard views. MVE Vel: Trans mitral early blood velocity. e': Mean early diastolic tissue velocity measured by tissue Doppler in the septal and lateral mitral annulus. E/e': The proportion of early trans mitral blood velocity over mitral annulus mean tissue velocity. MV E/A: The proportion of trans mitral early filling blood velocity above trans mitral late filling blood velocity. MV Dec T: The deceleration time of the trans mitral early filling.

3.3 Reproducibility.

Table 5. Intra- and inter-rater reproducibility for the different echocardiographic methods based on 20 patients.

Methods	Mean inter-rater (n=20)	Intra-rater Reproducibility (n=20)					Inter-rater Reproducibility (n=20)				
	Value (SD)	Mean (SD) diff.	LoA diff.	CoV (%)	ICC (3,1)	95 % CI	Mean (SD) diff.	LoA diff.	CoV (%)	ICC (2,2)	95 % CI
LVEF (%)	57 (7.7)	-0.1 (6)	[-11.8, 11.6]	5.9	0.82	[0.52, 0.93]	6.1 (9.8)*	[-13.0, 25.2]	10.9	0.17	[-0.66, 0.65]
LVEF with contrast (%)	59 (7.4)	-0.4 (4.4)	[-9.1, 8.2]	4.1	0.87	[0.66, 0.95]	-0.3 (6)	[-12.1, 11.4]	5.1	0.82	[0.51, 0.93]
Auto EF (%)	54 (5.4)	0.8 (2.8)	[-4.8, 6.3]	3.3	0.93	[0.83, 0.98]	0.6 (4.5)	[-8.3, 9.5]	4.2	0.82	[0.49, 0.93]
3D LVEF (%)	50 (10.3)	-0.4 (6.1)	[-12.3, 11.5]	6.7	0.89	[0.69, 0.96]	-0.4 (7.7)	[-15.6, 14.8]	10.4	0.87	[0.67, 0.95]
S` (cm/s)	7.1 (1.2)	-0.2 (1.0)	[-2.1, 1.8]	8.6	0.81	[0.51, 0.93]	-1.0 (1.1)*	[-3.2, 1.3]	12.1	0.50	[-0.15, 0.8]
MAPSE (mm)	12.6 (1.8)	-0.2 (1.3)	[-2.8, 2.4]	5.7	0.88	[0.68, 0.95]	1.2 (0.8)*	[- 0.4, 2.7]	7.3	0.85	[-0.10, 0.96]
Generic Strain (%)	-14.6 (2.3)	0.4 (1.7)	[-2.9, 3.6]	5.8	0.87	[0.67, 0.95]	-1.7 (1.0)*	[-3.6, 0.3]	8.6	0.81	[-0.18, 0.95]
AFI 4CH Strain (%)	-17.7 (2.9)	-0.9 (1.9)	[-4.7, 2.9]	6.9	0.73	[0.29, 0.90]	-1.7 (1.9)*	[-5.5, 2.1]	8.1	0.64	[-0.06, 0.88]
AFI Global strain (%)	-17.9 (2.6)	-0.8 (0.9)*	[-2.6, 1.1]	3.9	0.93	[0.61, 0.98]	-1.5 (1.0)*	[-3.4, 0.4]	6.4	0.70	[-0.07, 0.91]
MVE Vel (cm/s)	67.2 (15.4)	-0.5 (4.2)	[-8.8, 7.8]	3.7	0.98	[0.96, 0.99]	-4.4 (5.4)*	[-15.0, 6.2]	6.2	0.95	[0.75, 0.99]
e` (cm/s)	8.1 (1.6)	-0.4 (0.9)	[-2.2, 1.4]	6.4	0.86	[0.64, 0.95]	- 1.3 (0.9)*	[- 2.9, 0.4]	11.4	0.76	[-0.20, 0.94]
E/e'	8.6 (1.6)	0.2 (1.4)	[-2.7, 2.9]	9.1	0.86	[0.62, 0.95]	0.7 (1.0)*	[- 1.3, 2.7]	7.5	0.84	[0.45, 0.95]
MV E/A	0.95 (0.2)	0.0 (0.1)	[-0.1, 0.1]	3.6	0.98	[0.94, 0.99]	-0.1 (0.1)	[- 0.3, 0.2]	6.3	0.92	[0.76, 0.97]
MV Dec T (ms)	219 (46.8)	19.1 (43.7)	[-66.7 - 104.8]	8.3	0.53	[-0.15, 0.82]	-19.3 (33.1)*	[-84.2, 45.7]	9.3	0.82	[0.48, 0.94]

Mean (SD) diff: mean and standard deviation of the differences. LoA: Limits of agreement as mean difference +/- (1.96 x SD). CoV: Coefficient of variation. ICC: Intraclass correlation coefficient. 95 % CI: 95 % Confidence interval as mean ICC +/- (1.96 x SD of ICC). LVEF: Left ventricular ejection fraction, manually traced using Simpson´s Biplane Method without and with ultrasound contrast. Auto EF: a semiautomatic biplane STE tracing of LVEF. 3D LVEF: a semiautomatic tracing and calculation of LVEF from a 3D dataset. MAPSE: Mitral Annulus Plane Systolic Excursion, averaged septal and lateral measures. Generic strain: Calculated from mean MAPSE/ mean LV Length. AFI 4CH Strain (AFI 4CH): Automatic functional imaging calculating strain by speckle-tracking from only the 4-chamber view. AFI Global strain (AFI GLS): Automatic functional imaging STE based calculation of global strain from all three standard views. MVE Vel: Transmitral early diastolic blood velocity. e`: Mean early diastolic tissue velocity measured by tissue Doppler in the septal and lateral mitral annulus. E/e`: The proportion of early transmitral blood velocity over mitral annulus mean tissue velocity. MV E/A: The proportion of transmitral early filling blood velocity above transmitral late filling blood velocity. MV Dec T: The decelleration time of the transmitral early filling. n: number of patients. SD: standard deviation. (* mean difference significantly different from zero, accounted for in the text.)

The main reproducibility results are listed in Table 5. Data from the random sampled group of 20 patients were analysed by rater 1 in August 2017 and rater 2 in March 2018. The first column shows the mean characteristics from the inter-rater investigations of all the echocardiographic indices. From column two and further, the intra- and inter-rater data are presented as mean differences, SD of the differences, LoA for the differences and mean CoV of the differences in addition to ICC and their confidence intervals (95 % CI).

Intra-rater results.

According to numerical values of ICC for intra-rater analyses and categorisation based on general assumptions, the reliability was:

- Excellent for MVE Vel, MV E/A, Auto EF and AFI GLS (29 % of the indices).
- Good for LVEF, LVEF with contrast, 3D LVEF, S', MAPSE, Generic strain, e' and E/e' (57 % of the indices).
- Moderate to poor for AFI 4CH and MV Dec T (14 % of the indices).

The intra-rater intraclass correlation coefficient (ICC (3,1)).

4 out of 14 indices (4/14) (29 %) had ICC > 0.9 and 10/14 (71 %) had ICC > 0.85. Categorized as good indices with ICC > 0.75 was found for 12/14 (86 %).

The reliability coefficients categorized as excellent with ICC > 0.9, corresponds well with CoV < 4. The ICC results indicating good reliability with ICC > 0.75, corresponds with a wider range of 4.1 < CoV < 9.1.

The intra-rater coefficient of variation (CoV).

All the intra-rater analyses (100 %) had CoV < 10 %, ranging from 1.4 % to 9.1 %. Ten out of fourteen indices (10/14) (71 %) had CoV < 6 %. The numerical highest CoV in the intra-rater analysis was found for S' as 8.6 % and for E/e` as 9.1 %. The numerical lowest CoV was found for Auto EF, MV E/A, MVE Vel, AFI GLS and LVEF with contrast 3.3, 3.6, 3.7, 3.9 and 4.1 % respectively.

Intra-rater agreement by Bland-Altman plot (BA-plot).

The Bland-Altman plots visualize the distribution of measurement differences. The mean values of the differences representing bias, are drawn as red lines. The upper and lower 95 % LoA are drawn as green lines. The scales of the Y-axes showing the differences are set to ± half the normal value of the present index. This is to illustrate the proportion of the differences against the normal values. Only one index showed mean intra-rater difference significantly different from zero. The index is marked with * in the result table and in the BA-plot, and is

accounted for in the text. In spite of the assumptions not being met, the BA-plot is shown to give a general impression of the differences, with reservation to the precision of further calculations for this index. The complete list of p-values is available in Appendix 4.

Evaluating intra-rater agreements of left ventricular ejection fraction (LVEF).

LVEF without ultrasound contrast had wider LoA compared with contrast enhanced LVEF, shown in Figure 10. The LV Auto EF had the narrowest LoA and 3D LVEF had the widest LoA of all the LVEF methods, shown in Figure 11.

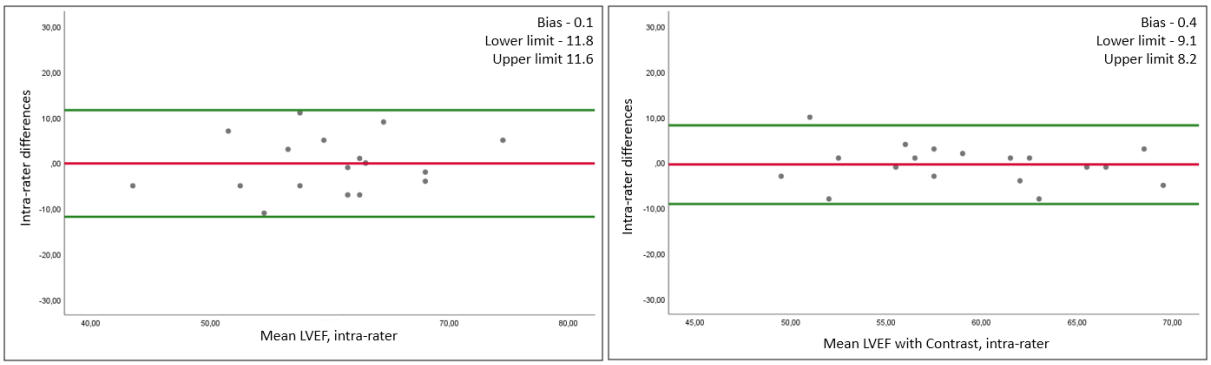


Figure 10. Bland-Altman plots of the intra-rater differences in manually traced LVEF ad Simpson's method to the left and the intra-rater differences for manually traced LVEF ad Simpson's method enhanced with ultrasound contrast to the right. Bias, upper and lower 95% LoA noted in the graphs.

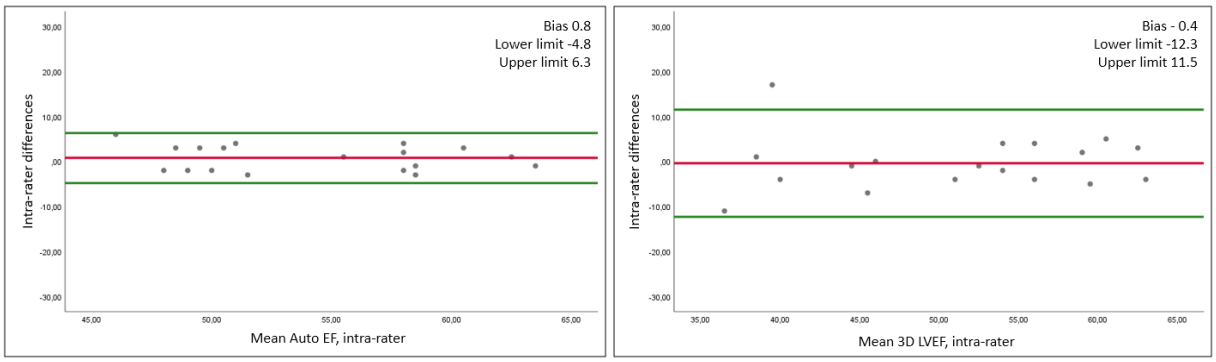


Figure 11. Bland-Altman plots of the intra-rater differences for Auto EF to the left and the intra-rater differences for 3D LVEF to the right. Bias, upper and lower 95% LoA noted in the graphs.

By summing up the four methods of measuring LVEF, the Bland-Altman plots clearly show how they all have large variability and moderate agreement. The method with numerically narrowest LoA in this investigation was Auto EF. As expected, the manually traced LVEF with contrast gives numerically narrower LoA than the traditional manual traced LVEF method without contrast (31, 38).

Evaluating intra-rater agreements of generic strain and tracking based global longitudinal strain (AFI GLS).

Generic strain is calculated from MAPSE and LV length, both known to be robust indices. Ultrasound contrast only gave a minimal and non-significant reduction of measurement errors in LV lengths. MAPSE was measured solely without contrast. MAPSE is a robust index, but still shows some intra-rater variability, with a bias of -0.2, 95 % LoA [2.8, 2.4], as shown in Figure 13.

The STE based AFI GLS analyses had several drop-out data, with a feasibility of 51 %. In most cases, this was because the 2-chamber analyses had been rejected. AFI GLS showed significantly different intra-rater differences from zero, p= 0.01, with a mean difference of -0.8. Some of the explanation of the significant bias might be the lowered degree of freedom, which in this case was only 12. A full list of degrees of freedom is shown in Appendix 4.

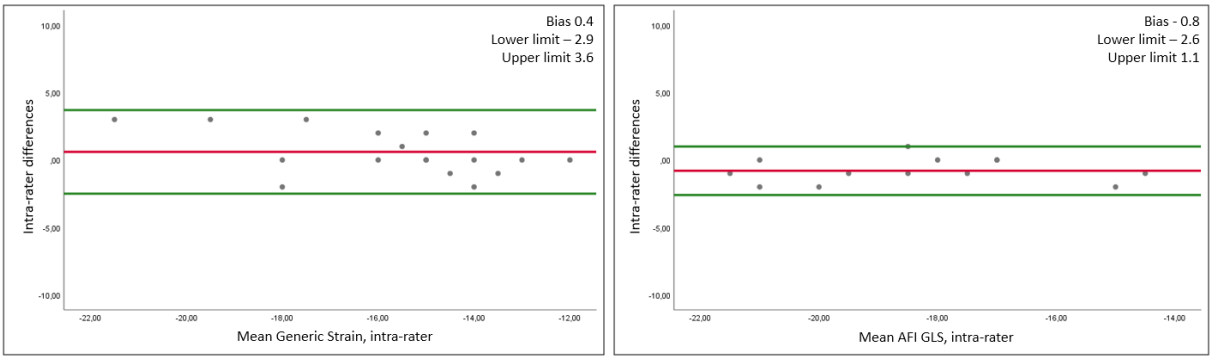


*Figure 12. Bland-Altman plots of the intra-rater differences of Generic strain to the left and of the intra-rater differences for tracking based global longitudinal strain (AFI GLS) to the right *. Bias, upper and lower 95% LoA noted in the graphs.*

Evaluating intra-rater agreements for mitral annular plane systolic velocity (S') and mitral annular plane systolic excursion (MAPSE).

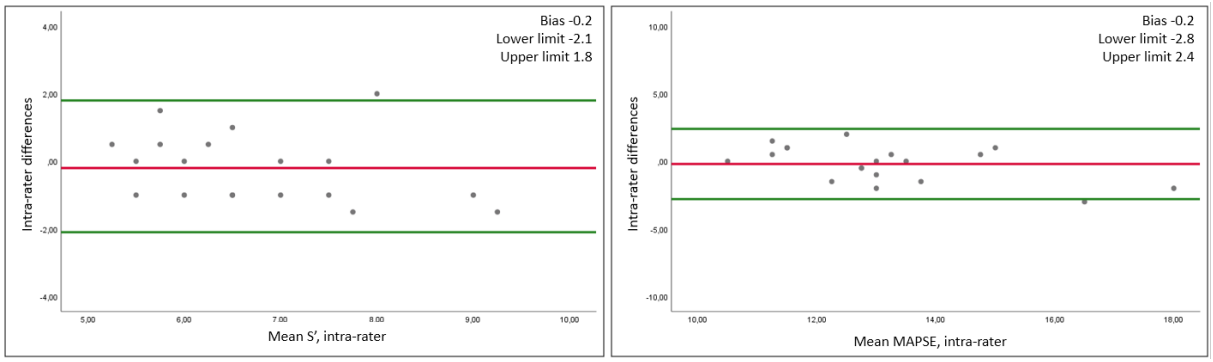


Figure 13. Bland-Altman plots of the intra-rater differences of the mitral annulus plane systolic velocity (S') to the left and of the intra-rater differences for mitral annular plane systolic excursion (MAPSE) to the right. Bias, upper and lower 95% LoA noted in the graphs.

In the BA-plot of the MAPSE differences, some proportional bias was observed having slightly higher counts of positive differences for the smaller means of MAPSE. A simple linear regression analysis was conducted, resulting in a probability of $p < 0.05$, which confirms a proportional bias in these differences. The narrowest LoA of the differences for the intra-rater analyses in proportion to the normal values was found for Auto EF, AFI GLS and MAPSE, [-4.6, 6.3] / 64, [-2.6, 1.1] / 17, [-2.8, 2.4] / 15, respectively.

Inter-rater results.

According to the numerical values of ICC for inter-rater analyses and categorisation based on general assumptions, the reliability was:

- Excellent for MVE Vel and MV E/A (14 % of the indices).
- Good for LVEF with contrast, AutoEF, 3D LVEF, MAPSE, Generic strain, e', E/e' and MV DecT (57 % of the indices).
- Moderate to poor for AFI GLS, AFI 4CH, S' and LVEF without contrast (29 % of the indices).

The inter-rater intraclass correlation coefficient (ICC (2,2)).

Two out of fourteen indices (2/14) (14 %) had ICC > 0.9 and 4/14 (29 %) had ICC > 0.85. Categorized as good; ICC > 0.75 was found for 10/14 (71 %).

As expected, the numerical values for ICC were lower for inter-rater than for intra-rater calculations. Very poor reliability was found as ICC of 0.17 for LVEF manually traced by Simpson's biplane method without ultrasound contrast. Using general assumptions, this is not at all acceptable as the ICC ≤ 0.5 is set as borderline for poor reliability. Using contrast enhanced recordings, the reliability of LVEF calculation became good with an inter-rater ICC of 0.82. The measures of S', AFI 4CH and AFI GLS had moderate to poor reliability with ICC of 0.5, 0.64 and 0.7 respectively. Very good reliability scores were found for MVE Vel with an ICC of 0.95 and MV E/A presenting an ICC of 0.92. The rest of the indices has ICC ranging from 0.76 to 0.87, by which we can conclude that most indices have good inter-rater reliability.

The inter-rater coefficient of variation (CoV).

All inter-rater analyses (100 %) had CoV < 12.2 %. Ten out of fourteen indices (10/14) (71 %) had CoV < 10 % ranging from 4.2 % to 9.3 % and 2/14 (14 %) had CoV < 6 %.

Inter-rater agreement by Bland-Altman plot (BA-plot).

Mean inter-rater differences were significantly different from zero for most indices. The indices are marked with * in the result table and in the BA-plots, and is accounted for in the text. All p-values are available in Appendix 4. Despite assumptions not being met, the plots are shown to give a general impression of the differences, with reservation to the precision of further calculations for these indices.

LVEF was evaluated for four methods. Ejection fraction manually traced by Simson's method without contrast (LVEF) showed non-acceptable agreement with LoA [- 13.0, 25.2]. A One Sample T-test showed the mean difference as bias of 6.1, which was significantly different from zero, p = 0.02. Thereby the assumptions for Bland-Altman plot were not met. LVEF with contrast, Auto EF and 3D LVEF had inter-rater differences which were not significantly different from zero.

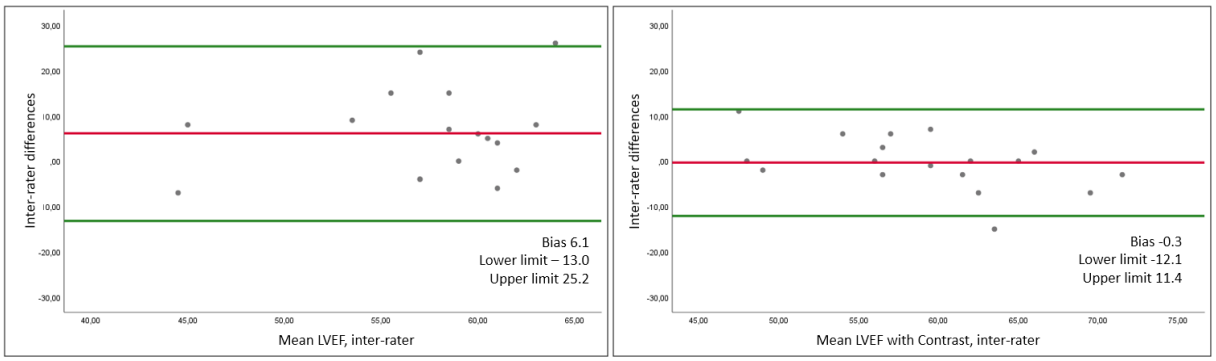


*Figure 14. Bland-Altman plots of the inter-rater differences of LVEF manually traced by Simpson's method witout ultrasound contrast to the left * and LVEF manually traced by Simpson's method with ultrasound contrast to the right.*

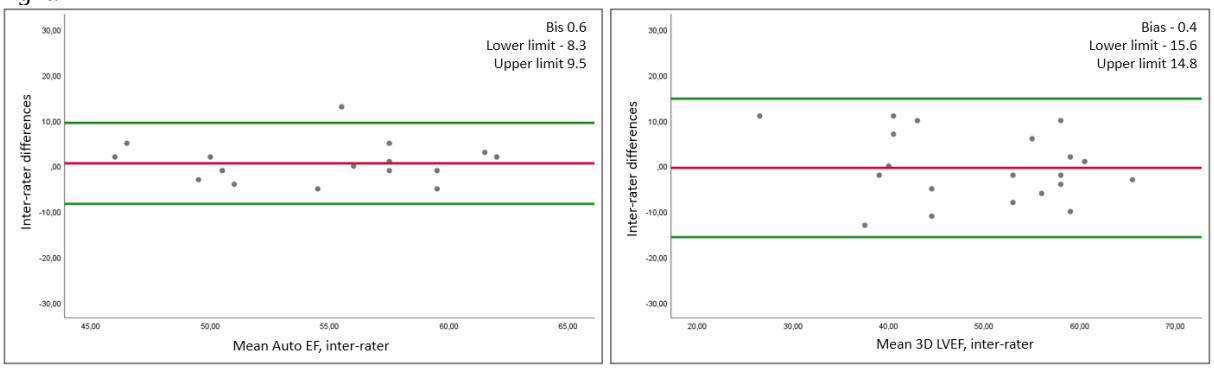


Figure 15. Bland-Altman plots of the inter-rater differences of Auto EF to the left and 3D LVEF to the right.

The measurements of Generic longitudinal strain and AFI GLS had their mean differences tested by One Sample T-tests as significantly different from zero, p = 0.001 and 0.006 respectively. Seen from this, the Bland-Altman plot's assumptions are not met.

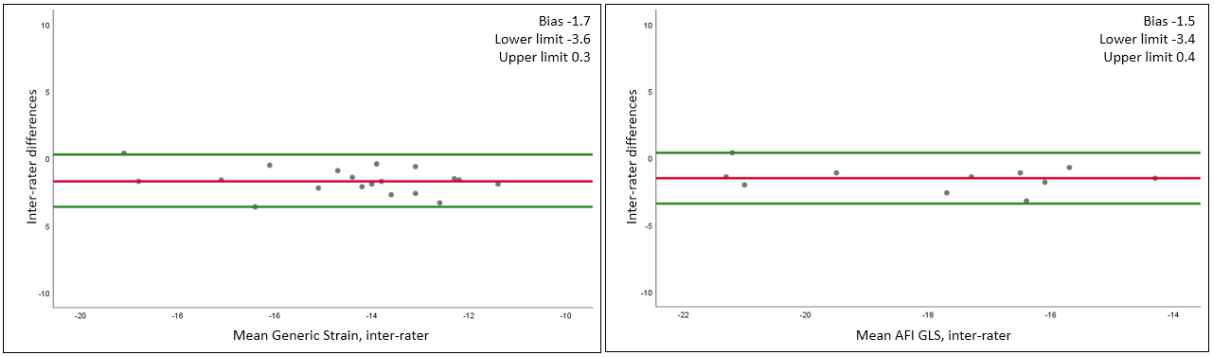


*Figure 16. Bland-Altman plots of the inter-rater differences of Generic longitudinal strain to the left * and tracking based global longitudinal strain (AFI GLS) to the right *.*

The measurements of S' and MAPSE had mean differences tested by One Sample T-test as significantly different from zero, $p = 0.002$ and $p \leq 0.001$ respectively. Seen from this, the Bland-Altman plot's assumptions are not met.

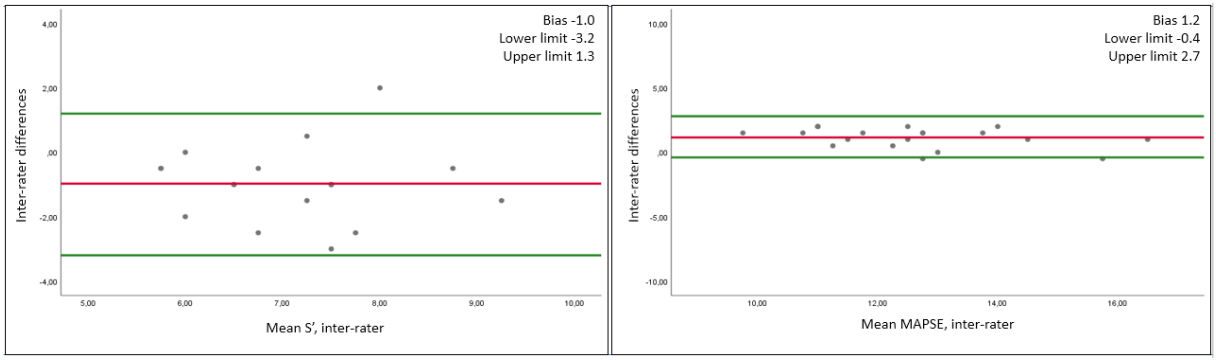


*Figure 17. Bland-Altman plots of the inter-rater differences of Mitral annular plane systolic velocity (S') to the left * and Mitral annular plane excursion (MAPSE) to the right *.*

Significance of differences between methods according to inter-rater CoV.

Comparing all the CoV results from the inter-rater analyses for all 14 indices against each other in a one-way ANOVA analysis, there were 91 comparisons in total. Only 22 comparisons showed significant differences, which represents 24 % of the comparisons, leaving 76% with no significant differences of CoV.

The method with the best numerical inter-rater CoV was Auto EF with a CoV of 4.2 %. Comparisons showed significant differences to 6 methods: S', MV Dec T, LVEF without contrast, 3D LVEF, Generic strain and e'. Auto EF was not significantly better than the remaining 7 methods: MAPSE, AFI GLS, E/e', MV E/A, LVEF with contrast, AFI 4CH and MVE Vel.

The method with the second lowest numerical inter-rater CoV was LVEF with contrast with a CoV of 5.1 %. Comparisons showed significant difference to 5 methods: S', MV DecT, EF witout contrast, 3D LVEF and e'. LVEF with contrast was not significantly better than the remaining 8 methods: MAPSE, AFI GLS, E/e', MV E/A, Auto EF, Generic strain, AFI 4CH and MVE Vel.

The method with the third lowest numerical inter-rater CoV was MVE Vel with a CoV of 6.2 %. MVE Vel was significantly different from 4 methods: S', LVEF without contrast, 3D LVEF and e'. MVE Vel was not significantly better than the remaining 9 methods: MAPSE, AFI GLS, E/e', MV E/A, MV Dec T, LVEF with contrast, Auto EF, Generic strain and AFI 4CH.

The method with the fourth lowest CoV was AFI GLS with a CoV of 6.4 %. AFI GLS showed significant difference to 2 methods: S' and e'. AFI GLS was not significantly different from any of the remaining 11 methods.

This indicates that the numerical value of CoV correlates to number of methods that is significantly different. Lower numerical value of CoV meaning good reproducibility, correlates with significant difference to increasing number of methods.

The methods with the four lowest numerical CoV, Auto EF, LVEF with contrast, MVE Vel and AFI GLS were all significantly better than S' and e'.

If AFI GLS is excluded, the three remaining methods: Auto EF, LVEF with contrast and MVE Vel, were additionally significantly better than LVEF without contrast and 3D LVEF.

According to inter-rater CoV, the reproducibility was significantly better for:

Auto EF, LVEF with contrast and MVE Vel than for S', e', LVEF without contrast and 3D LVEF.

Comparing intra-rater and inter-rater results.

CoV < 10 % was found for 100 % of the intra-rater analyses and 71 % of the inter-rater analyses.

CoV < 6 % was found for 71 % of the intra-rater analyses and 14 % for the inter-rater analyses.

Our results, similarly to many other studies' results, showed a generally lower numerical reproducibility for inter-rater analyses than for intra-rater analyses. There was a significant difference between the intra- and the inter-rater CoV. We found;

Mean (SD) intra-rater CoV of 5.8 (1.9) %

Mean (SD) inter-rater CoV of 8.1 (2.4) %

$p \leq 0.001$

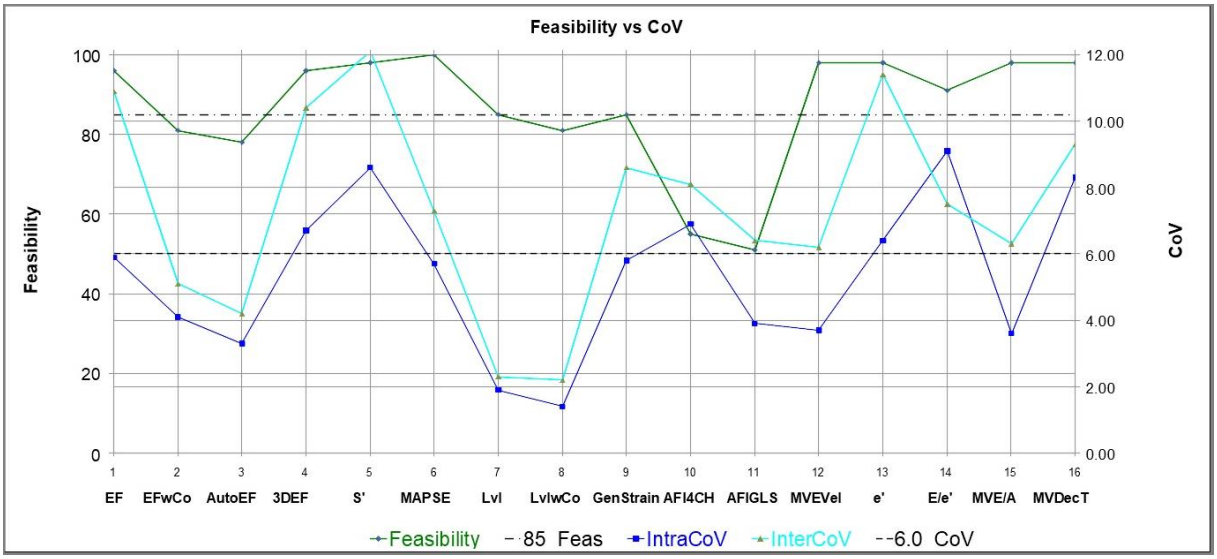
ICC > 0.85 was found for 71 % of the intra-rater analyses and 29 % of the inter-rater analyses.

ICC > 0.75 was found for 86 % of the intra-rater and 71 % of the inter-rater analyses.

Comparison of variations: feasibility and reproducibility.

The variations of feasibility, intra- and inter-rater agreement and reliability was plotted into two diagrams to get an overview, as shown in Figures 18 and 19.

Figure 18 shows the intra- and inter-rater CoV variation compared with feasibility variation. The inter-rater CoV has slightly higher numerical values than intra-rater CoV for most indices, covarying with largest difference for LVEF without contrast, 3D LVEF, S' and e'. Ideally, the feasibility should be high and the CoV low to offer good feasibility and good accuracy. Feasibility >85% and CoV < 6 from the intra-rater analyses was found for MAPSE, MVE Vel and MV E/A. LVEF with contrast and Auto EF had good reproducibility, but slightly lower feasibility of approximately 80 %. The least successful combination of feasibility and CoV was seen for AFI 4CH and AFI GLS.

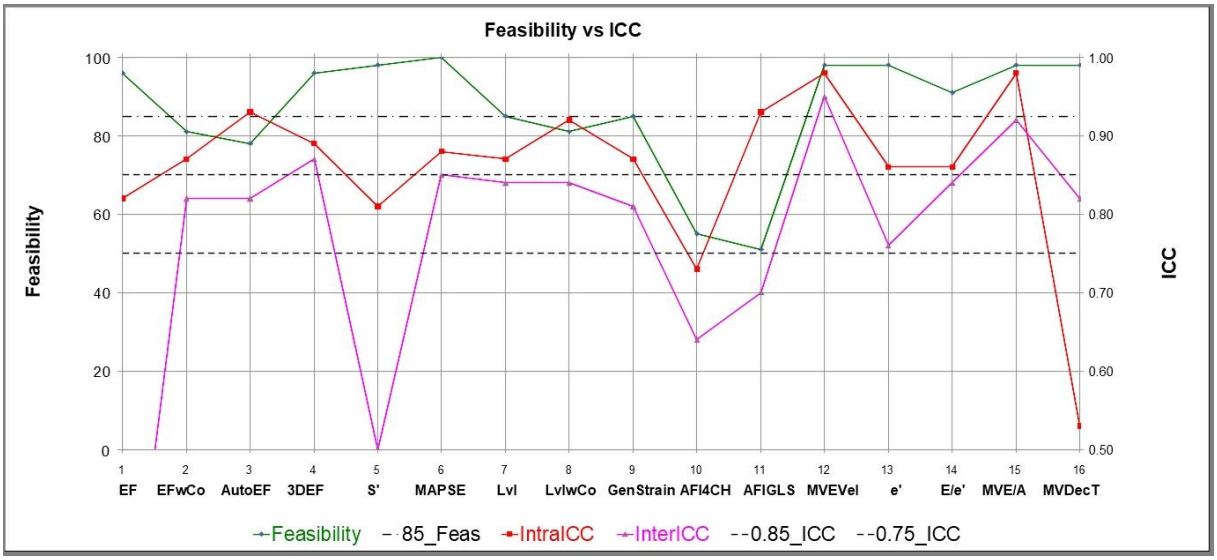


Feasibility (%) is presented in green lines, scale to the left. CoV (%): Coefficient of Variation. Intra-rater CoV (%) is presented in dark blue lines, scale to the right. Inter-rater CoV (%) is presented in turquoise lines, scale to the right. Limit of good feasibility marked as a dashed line at 85 %. Limit of good CoV marked as a dashed line at CoV = 6.

EF: Left ventricular ejection fraction, manually traced using Simpson's Biplane Method without and with ultrasound contrast. Auto EF: a semi-automatic biplane tracing of LVEF. 3D LVEF: a semi-automatic tracing and calculation of LVEF from a 3D dataset. MAPSE: Mitral Annulus Plane Systolic Excursion, averaged septal and lateral measures. LV Length measured as apex to septal Mitral annulus and apex to lateral Mitral annulus to calculate the mean, without and with contrast. Generic strain: Calculated as mean(MAPSE/ LV length). AFI 4CH: Automatic functional imaging calculating strain with speckle-tracking from only the 4-chamber view. AFI Global strain (GLS): Automatic Functional Imaging measuring strain with speckle-tracking from all three standard views. e': The early diastolic tissue-velocity measured by tissue Doppler in the septal and lateral mitral annulus, to be averaged. E/e': The proportion of early Mitral blood-velocity over mitral annulus mean tissue-velocity. MV E/A: The proportion of Mitral valve early filling blood-velocity above late filling blood-velocity. MV Dec T: The deceleration time of the Mitral early filling.

Figure 18. Comparison of variation of feasibility and CoV with lines for limits of good feasibility and CoV as discussed.

Figure 19 shows intra-rater ICC and inter-rater ICC variations, with slightly higher numerical values for intra-rater than for inter-rater measures. The largest differences were found for LVEF without contrast, S' and MV DecT. Ideally, the feasibility and ICC scores should both be high to offer good feasibility and good reliability. Feasibility > 85 % and intra-rater ICC > 0.75 was found for many indices. Feasibility > 85 % and intra-rater ICC > 0.85 was seen for 3D LVEF, MAPSE, generic strain, MVE Vel, e', E/e' and MV E/A. The best combination was seen for 3D LVEF, MAPSE and the diastolic Doppler indices. Good reproducibility was also seen for LVEF with contrast and 3D LVEF, but with slightly lower feasibility of approximately 80 %. The least successful combination of ICC and feasibility was seen for AFI 4CH.



Feasibility (%) is presented in green lines, scale to the left. ICC: Intraclass Correlation Coefficient. Intra-rater ICC as ICC (3,1) is presented in red lines, scale to the right. Inter-rater ICC as ICC (2,2) is presented in pink lines, scale to the right. Limit of good feasibility marked as a dashed line at 85 %. Limit of excellent and good ICC is marked as a dashed line at ICC equal 0.95 and 0.75, respectivly. Otherwise as listed in Figure 18.

Figure 19. Comparison of variation of feasibility and ICC with lines for limits of good feasibility and ICC as discussed.

Larger discrepancies between intra- and inter-rater indices were seen for ICC than for CoV. These two reproducibility indices are calculated slightly differently, as ICC also considers variability of the measurements and raters. For the indices LVEF witout contrast, S' and MV Dec T there were large ICC discrepancies which will be discussed.

4. Discussion.

This study provides an evaluation and comparison of the feasibility and reproducibility of several commonly used echocardiographic indices recommended for follow-up examination after potential cardiotoxicity (10, 14).
Prior studies report different calculation indices of reproducibility and are difficult to compare. Recalculations might therefore be necessary, if sufficient data is available.

4.1 Main findings.

There were six main findings in our analyses of feasibility and reproducibility:

- The feasibility scores of MAPSE, S', all Doppler indices, LVEF without contrast, 3D LVEF and Generic strain were all 85 % or more. All other indices had lower scores, AFI GLS the lowest.
- Reproducibility was significantly better for intra-rater analyses than for inter-rater analyses, according to CoV.
- Intra-rater reproducibility was found best for Auto EF, MV E/A, MV E Vel, AFI GLS and LVEF with contrast, here listed with increasing values of CoV.
- Inter-rater reproducibility was found best for Auto EF, LVEF with contrast, MVE Vel, MV E/A and AFI GLS, here listed with increasing values of CoV.

Feasibility.

MAPSE and Doppler indices are well-known, robust parameters, which do not dependent on image quality. Støylen et al. and Dalen et al. found feasibility > 95 % for all these indices (34, 35). Further, better feasibility for the analysis of LVEF without contrast than for the contrast enhanced analysis of LVEF was found. These findings seem quite contrary to immediate logic. An explanation might be our built-in eyeballing skills. The manually traced LVEF is done with an automatic proportion of eyeballing. When the ultrasound contrast is present, the rater clearly realize all the uncertainties, and may more often reject the analyses. Our non-contrast LVEF feasibility result of 96 % is better than that reported by Malm et al. (86 %) in 2004 (31). This might support the theory of an extended use of eyeballing in our study. Our feasibility result for 3D LVEF of 96 % is higher than 3D LVEF feasibility of 89 % reported by Chahal et al. in 2012 (32). This might be due to rater rejection practise, or it can be a result of improved software quality available in newer equipment. The most promising and recommended index, AFI GLS, shows low numerical values of feasibility as only 51 % in our study. Barbier et al. report

feasibility of segmental AFI GLS as averaged to 93 %. Unfortunately, their feasibility results cannot be generalized to everyday practice, as they left out 45 out of 290 participants' data because of suboptimal image quality (36). In our data, AFI GLS had a feasibility score of 51 %, AFI 4CH of 55 % and generic strain had a feasibility score of 85%. Among the methods mentioned specifically in the recommendations for evaluating CTRCD (10, 14), our study demonstrated good feasibility scores for MAPSE, S', LVEF with contrast and 3D LVEF, but not for Global longitudinal strain (AFI GLS).

Image quality.

Image quality is difficult to investigate. In our study, image quality was considered according to reading protocol for the sample of 20 patients. In 2D mode, the 2-chamber view often has the lowest image quality (31), which affects biplane LVEF and triplane AFI GLS. Good image quality was acquired from 2/20 patients, suboptimal images from 12/20 patients and poor image quality from 6/20 patients. Images with good quality enabled safe analyses. Suboptimal image quality required corrective adjustments and eyeballing to enable further analyses. It is important to report the quality of the cardiac images recorded. Under suboptimal image quality, the use of advanced echocardiographic techniques such as strain and 3D may present misleading results (30). To our experience, image quality varied by several factors, one beeing surgical outcome. The investigations of image quality were inaccurate and are not integrated in the main results.

Contrast.

Ultrasound contrast has limited effect when shadowing cause suboptimal image quality. When using echocardiographic contrast, patients with serious allergies should be excluded and security precautions available for the specific drug should be followed (41, 68). For the population in the current study, difficult venous approaches were also experienced. Echocardiographic volumes calculated with contrast are larger than volumes calculated witout contrast, as compared to CMR. Tissue harmonic imaging does not track the true endocardial surface as well without contrast and tracking noise in the LV cavity may be perceived as the endocardial border (40). This will influence volume calculations more so than LVEF calculations, as LVEF is based on proportions. Different LV abnormalities give different filling patterns and therefore a varying contrast saturation throughout the ventricle. Because of this, ventricles with abnormalities will have the lowest reproducibility when analysing contrast enhanced LVEF. In one of our patients, pericardial effusion caused some confusion as the liquid could be misinterpreted as myokard. Opposite colours may confuse the analyser and previous

experience in echocardiography is of vital importance in such examinations. In contrast mode, the mitral valve closure can be difficult to see, and thereby the raters' timing of end diastole may be inaccurate.

Study population.

Echogenicity is a personal characteristic influencing feasibility. The population in our study is prone to developing CTRCD. Data suggest that pre-existing Cardio Vascular Disease (CVD) or cardiovascular risk factors substantially increases the risk of developing CTRCD (10). Obesity and lung disease are known challenges for this patient group (14, 28). Our feasibility investigation showed varying individual feasibility for the different methods.

Surgical implications.

Anatomic changes after surgery was expected to correlate with image quality. In our population, 12 out of 20 patients had undergone left sided surgery with possible impact on echocardiographic image quality and feasibility. Different treatments give different anatomical conditions. Our patients had their surgical and reconstruction methods distributed as shown in Table 6.

Table 6. Distribution of surgery and reconstruction procedures.

RECONSTRUCTION / SURGERY	Radical	Conservative	Number
Conservative		xxxxxxx	7
None	xx		2
Diep	x		1
NaCl implant	x		1
Silicone implant	x		1
Number	5	7	12

Conservative surgery is also called minimal surgery. NaCl implant is also called water implant. Radical surgery is also called mastectomy. DIEP is short form of Deep Inferior Epigastric Artery Perforator, meaning reconstruction from one's own tissue.

Conservative surgery gave us acceptable image quality. Silicone implants represent a well-known risk of reduced image quality and low feasibility. The relationship between the surgery and the image quality was not clarifying in our sample. As expected, the two patients with good image quality had no implants, one had right sided surgery and one had conservative surgery.

The six patients with poor image quality consisted of one with silicone implants, one with a NaCl implant, one with Diep, but also two with conservative surgery and one with right sided surgery. This shows that there are many factors influencing image quality and feasibility. No conclusions can be made with regards to image quality based on this small number of patients.

Reproducibility.

Like in many other studies, reproducibility for intra-rater analyses was found better than for inter-rater analyses. Barbier et al. reports up to double variability for inter-rater than for intra-rater (36). Intra- and inter-rater analyses can be reported analysing the same recordings or different recordings. Considered theoretically, analysing different recordings is expected to present the lowest reproducibility of all test - retest methods in echocardiography. Reproducibility levels can be investigated analysing the same cine-loops or analysing different cine-loops from the same recordings. Analysing different loops will give lower reproducibility than analysing the same loop. Inter-rater analyses of the same cine-loops underestimate the clinically relevant inter-rater reproducibility analysing different cine-loops by 40 % for most measurements of LV function (66).

Reanalysing priorly saved recordings for comparison with new recordings will give better reproducibility, equivalent to intra-rater reproducibility. Reanalysing also facilitates comparison of angle dependencies with the last recording. Suboptimal angles will often be similar in repeated recordings, and comparisons can safely be made. It is of great importance to save and keep standardized recordings from all previous examinations. Barbier et al. state that feasibility and reproducibility data from clinical practices are few (36). We contribute to broadening this field with our results.

The discrepancies in the inter-rater analyses of LVEF witout contrast, S' and MV Dec T were considerable in our study, as shown in Table 5 and Figure 19. Some of the differences might be due to different measuring techniques despite having completed a training session. A more extensive training and agreement session might have reduced the differences. Another explanation is the possibility of having chosen different loops and images for the analyses. This study design (as saving more than one set of cine-loops), is suitable because it provides generalizable results, similar to results likely found in everyday practice.

Table 7. Mean and standard deviation of inter-rater differences in three sets of measures of the discussed indices LVEF witout contrast, S' and MV Dec T from rater 1 and rater 2.

Method	Rater 1	Rater 2
LVEF without contrast, mean (SD) (%)	60 (8)	54 (6)
S', mean (SD) (cm/s)	6.7 (1.1)	7.6 (1.1)
MV DecT, mean (SD) (ms)	211 (45)	227 (48)

LVEF: Ejection fraction without enhancement of echocardiographic contrast. S': Peak systolic tissue-velocity. MV DecT: Deceleration time of the trans mitral early diastolic velocity.

The semi-automatic Auto EF had the best numerical values in reproducibility scores in our study. Some of the explanation might be that semi-automatic procedures include less personal variability, and thus better reproducibility is achieved. Szulik et al. found significantly lower variability for Auto EF than for LVEF ad Simpson's rule (16). Furthermore, 3D LVEF is one of the semi-automatic methods which presents high numerical reproducibility scores. It was however tested, showing significantly lower reproducibility scores than Auto EF and LVEF with contrast in our study. 3D LVEF is the preferred method for investigating LVEF for detection of CTRCD in the recommendations (14). 3D LVEF is often referred to as promising in the case of reproducibility (23). In our BA-plots, 3D LVEF presents equally inaccurate as the criticized method of manually traced LVEF ad Simpson's method without ultrasound contrast. Our study suggests that Auto EF and LVEF with contrast has better reproducibility than 3D LVEF.

Barbier et al. concludes in their large study from 2014 that AFI GLS has excellent reproducibility and is suitable for diagnosis and follow-up examinations of LV global and regional systolic function (36). The population in their study had various left ventricular function. They report intra- and inter-rater reproducibility for AFI GLS as CoV 2.6 % and 4.8 % respectively. Our results were CoV 3.9 % and 6.4 % respectively, which is numerically slightly worse. They reported an inter-observer ICC of 0.99 for a population with coronary heart disease. Our inter-rater ICC of 0.7 was very low in comparison. The large difference might be explained by Barbier et al. measuring on the same loop, while we might have measured on different loops.

When it comes to semi-automatic methods, our study suggests the superiority in terms of reproducibility cannot be generalized. Auto EF is the semi-automatic method showing best reproducibility levels in our study.

Left ventricular ejection fraction (LVEF).

Looking into the data for LVEF without contrast, large standard deviations of the inter-rater mean differences were seen for several patients, worst for patient PID18279 and 19451. The data sets from these two patients were both categorized as suboptimal in terms of image quality having SD equal to 17 and 18.4 percentage points, respectively. Figure 14 shows the BA-plot of LVEF without contrast, with both a huge bias and wide spread differences. The CoV of 10.9 % was the third highest in our inter-rater results, however the reliability coefficient ICC of 0.17 was found extremely low. Wood et al. reports inter-rater reliability of LVEF by Simpson's method as ICC = 0.79 – 0.94. To be able to compare one should read the articles reviewed (45). Due to the use of different LVEF cutoffs and methods of measurement, the comparability of results between different studies are limited. EACVI have found limited amounts of evidence-based data, not sufficient to formulate evidence-based screening and follow-up recommendations for CTRCD (14).

Some explanation of the intra- and inter-rater differences in tracing LVEF, might be related to apex contouring. Our study had four different LVEF analyses. Manually traced LVEF ad Simpson's method without contrast, with contrast, Auto EF and 3D LVEF. Epicardial tracing of apex was selected to improve the expected poor reproducibility of manually traced, non-contrast LVEF. Endocard was traced in basal- and mid- segments and when approaching the apex, the tracing aimed at the epicardial boarder at the tip of apex. As the apex is thin, the difference between the traditional and the epicardial tracing method for LVEF will be small. The result will be more valid as this tracing approach to some extent works around the problem of reduced apical image quality and foreshortening.

The contrast enhanced LVEF was traced along the enlightened endocardial boarders.

The Auto EF automatically detects endocard. Therefore, the analysed values may differ slightly. Because we are studying reproducibility within indices, slightly different values between the indices is not emphasized.

Semi-automatic tracings are good when the image quality is good, when the image quality is poor, there is less help from automation. Semi-automatic procedures are good for standardization and reproducibility, as results have shown (16). When analyses are complex however, the automatic steps can cause confusion. When semi-automatic analyses demand several savings, the analyser might get off track and miss some savings. Thus, the screen might show a mix of old and new measurements. If the merge of unfinished earlier measurements and new measurements is overlooked, the averaged measurement score might present wrong results. Seen from this, automation could either enhance reproducibility or confuse the operators.

Different methods provide different measurement values. This applies to 2D biplane LVEF witout contrast, 2D biplane LVEF with contrast and 3D LVEF (45). In reproducibility studies this is not emphasized. In a routine situation, LVEF measurements are often repeated when the results are deemed incompatible with the visual impression of the reader (16). This visual evaluation highly depends on the skills and experiences of the raters.

Experience.

The raters' level of experience is crucial when it comes to reproducibility. Unexperienced raters have shown larger variations than experienced raters in measurements and analyses (69). Measurements from two experienced raters provide realistic figures, like everyday practice in echocardiographic laboratories. Our study showed reproducibility results with experienced raters in normal surroundings with an unselected population.

Diastolic indices.

Diastolic function is an independent predictor of outcome in different cohorts (70). The clinical value of diastolic parameters is still not proven as suitable for detecting CTRCD and is highly sensitive to changes in loading conditions. Diastolic indices are however recommended and have value in follow-up examinations (10, 14). All diastolic indices in our study showed numerically excellent feasibilities. The reproducibility of MVE Velocity was found to be significantly better than the other researched indices. Unfortunately, the reproducibility for early diastolic tissue velocity (e') was found significantly lower than MAPSE, AFI GLS, LVEF with contrast, Auto EF and MVE Vel. e' was averaged between only two walls, which might be some of the explanation for the low score. It is shown that tissue velocities should be averaged over four instead of two walls to improve reproducibility levels (66). MV Dec T had large standard deviation, most likely due to measuring different images. This index is not central in evaluating cardiotoxicity.

Echocardiographic image acquisition.

All loops were recorded as three consecutive cardiac cycles to ensure good quality. Echocardiographic recordings were often recorded and saved as more than one set of three cine-loops and more than one image. This was meant to ensure saving the best possible images and loops for optimal analyses. The raters could then later choose the best image or set of cine-loops. While doing intra- and inter-rater analyses, the risk was present that the raters re-analysed different cine-loops, according to different selections. During the training session, it was discussed whether to clean up the recordings before analysing and leaving only the one

chosen set of three cine-loops and one image for each index. The raters agreed on leaving all the recordings as they were, so that the rater would have to choose individually for each analysis which cine-loop from which set of cine-loops or which image to analyse. The raters might therefore have analysed different images and loops. This is found to be close to everyday practice, where examiners often must choose the best images and loops out of several recordings for analysis. The reproducibility results would likely have been better if the raters had analysed the same cine-loops.

Equipment.

During this project there has been software (SW) updates available to our analysis software. However, all our analyses were completed by the same SW version, to maintain consistency. Improvements due to updates might have been missed. New projects will show if image quality and algorithms have improved. In the current study, 3D LVEF and Auto EF analyses were not available in contrast mode. Therefore, some patients' data could not be analysed by 3D, because the image quality without contrast was poor, while image quality with contrast was good. The implementation of contrast enhanced 3D and Auto EF analysations will likely improve the feasibility scores.

Reproducibility- and Statistic indices.

There are several reproducibility and reliability indices and several statistical methods used in echocardiographic studies. As reproducibility issues are more commonly reported, there is an increased need for standardisation. Kottner et al. give an excellent guideline for reporting reliability and agreement, suggesting standardisation. They underline that reliability is a product of interaction between tools, subjects, objects and contexts. Reliability is affected by the variability of these factors and by statistical approaches. Therefore, study results are only interpretable when the measurement settings are sufficiently described, and the method of calculation or graphical presentation is fully explained (6).

Intra- and inter-rater agreement by Bland-Altman plot (BA-plot).

Generally, there were only small levels of bias in most of our intra-rater BA-plots. Some proportional bias was found in the distribution of differences when investigating MAPSE. A possible explanation of the discrepancies is that there is a numerical uncertainty in small numbers. Millimetres are shown very small on the screen and in the analysed pictures. In the current study the raters did not use zoom mode measuring MAPSE. To enhance accuracy, zoom mode and a strict gain setting could have been used.

We found significant bias for most of the inter-rater differences, with a considerable bias in the inter-rater LVEF witout contrast.

Intra- and inter-rater reliability by intraclass correlation coefficient (ICC).

The reliability of echocardiographic LVEF in this patient group was not expected to be any better than in other investigated groups. The results confirmed it is not. The reliability coefficients depend on the range of the measurements (6). Large variations will give high ICC. Making ICC calculations, only complete pairs are approved. In LVEF without contrast calculations, rater 1 and 2 had rejected different patients, PID19045 and PID17943,19183 respectively, giving the total number of three rejected pairs. Performing calculations on LVEF with contrast, the two raters rejected the same two patients, PID17943 and 19183, giving a total of two rejected pairs. Calculating Auto EF, the two raters both rejected PID18427, and in addition rater 1 rejected PID19045 and rater 2 rejected PID19183, giving a total of 3 rejected pairs. S' had only one common missing value, due to a missing recording. For the AFI GLS, a total of seven pairs were rejected, of these only one pair was rejected from both raters, PID18427. In addition, PID18787, 18947 and 19506 was rejected by rater 1 and PID18960 and 19183 was rejected by rater 2. This might suggest the raters had chosen different loops for analysing. e' had only one missing analysis from PID18279 who was failed to be recorded for e'. MV E/e' missed three complete pairs due to missed recordings in PID 17943,18279 from both raters and in addition PID19506 from rater 1 where mitral flow was only recorded inaccurately as combined measuring IVRT. Valid pairs and degrees of freedom is accounted for in Appendix 4. SPSS gives us Cronbach's alpha coefficient and good correlation was seen between inter-rater analyses as expected. Correlation is not weighted in our study, but available in Appendix 4.

Comparing ICCs in different studies requires fully explained models, types and definitions of selections in the study design (2), which was not always found.

Figure 18 and 19 visualize more variation between intra- and inter-rater ICC than between intra- and inter-rater CoV. This might be due to ICC taking consistency into account, while CoV is a more strict agreement index.

Power.

In many clinical trials a desirable value of ICC is set, and the investigation is designed to ensure this (71). Rankin and Stokes suggest having at least 50 participants to ensure good reliability estimates (47). Koo et al. suggests as a rule of thumb to involve at least 3 raters and at least 30 heterogeneous samples when conducting reliability studies (2). As an example, Jenkins et al.

had 60 patients allowing a 20% difference between End Diastolic Volume measurements and a 34% difference in End Systolic Volume measurements to be identified at 80% power (72). The most important issue for a valid result is probably to have a decent description of the population and the raters. It is also important that the raters are well trained (73). In the case of limited power, the detection of small differences in reproducibility between different measurements can be difficult. However, these differences can be assumed to be of less clinical importance. The aim of this study was to reveal true reproducibility indices in true settings. The findings of this study may be used for power calculations in future studies.

4.2 Limitations

Literature.

European heart association's (EACVI) different recommendations were found interesting to clarify as they indicate different levels of recommendations. One expert consensus, one recommendation and one position paper from EACVI were presented in only four years, 2013, 2014 and 2016 (8, 10, 14). In 2014, the European Cystic Fibrosis Society (ECFS) published their policy on the matter (74). They defined the consensus papers as state of the art, evidence based, mostly commissioned developed via a specific process, commissioned and organized by their organization. Their guidelines were meant as proposed practical approaches, not binding, multidisciplinary developed, grading of evidence of recommendations endorsed by their board prior to publication. ECFS's definition of a position paper was as an opinion about an issue that was written by a working group and reviewed by their board. According to this explanation of ranging papers, we should have confidence in our expert consensus from Plana et al. and Lancellotti et al. (10, 14). However, Plana et al. claim in their final notice and disclaimer, that their paper is primarily based on opinions of experts rather than on scientifically verified data. This makes our expert consensus of 2014 questionable used as one of the main sources in our study and in our everyday practice.

Reproducibility of results is the standardized Mesh term of our study. Kottner et al.'s suggestion to include reliability as a Mesh term for easier identification of current material is supported (6).

Investigated parameters.

Our selection of parameters investigated is based on formal recommendations, however it lacks the WMS method. The WMS method is time-consuming, is seldom used in everyday practise,

has shown weak correlation with LVEF (50) and is therefore left out. Image quality indices could have been more precisely planned to increase accuracy. All parameters chosen are recommended for cardiotoxicity follow-up examinations. These parameters are investigated with regards to feasibility and reproducibility. The current study has not evaluated these methods' ability to reveal cardiotoxicity. This needs to be clarified in future work.

Reproducibility.

Ideally both raters should perform individual intra-rater reliability analyses before the inter-rater analyses (62). Because of time-consuming procedures only rater 1 completed intra-rater analyses.

Experience and training.

The raters' level of prior experience was good in the current study, however the training and preparation for the analyses could have been improved. The raters had a small, recommended training session to standardize procedures between the two raters (62). The manual tracing method of LVEF was slightly adjusted after the training session, which lead to slightly different analyses of the sample of 20 patients, compared to the first basic analyses of all 55 patients. This adjustment did not influence the reproducibility investigations.

4.3 Clinical implication.

Feasibility.

Our results show large variability in the feasibility scores for the different methods. Categorisation of breast cancer treated patients based on image quality, might be useful to simplify the choice of follow-up method. The need of echocardiography and echogenicity should be considered against advantages and disadvantages of CMR and MUGA. Well-known parameters such as body weight and lung challenges, including smoking habits, influence the echocardiographic image quality. Soreness is a potential problem, as some patients experience discomfort from pressure against the breast. The Echocardiographic procedure requires different amounts of pressure from the probe. The pressure feels differently for all the patients. This aspect was not included in our study, but we did briefly look into surgical methods as discussed.

In cancer patients' fast track system, it would be preferable with methods both feasible and reproducible for all patients. MAPSE, 3D LVEF, MVE Vel and MV E/A showed best combination of feasibility and reproducibility out of all measures included in our study.

Reproducibility.

To optimize the comparison of analyses, the same machine with the same software and preferably the same investigator should be used for all follow-up examinations. This might be difficult to organize in practise. Another alternative is that the investigator reanalyses earlier recordings.

The ICC, the mean and SD of the differences and the CoV defines reproducibility. The limits of significant reduction evaluating cardiotoxicity is given by EACVI as an absolute change of 10 % in LVEF or a relative change of 15 % in GLS (14). These limits are challenging to reveal for the methods tested in the current study. Coefficient of repetition (CoR) represents by definition the smallest significant difference between repeated measurements (64). CoR thereby gives a direct indication of the methods' ability to detect small changes.

Our results have shown a preferation of intra-rater investigations. From the intra-rater investigation, we calculate the CoR of LVEF with contrast and Auto EF as 8.8 % and 5.6 % respectively, which means the difference between two analyses of LVEF with contrast must exceed 8.8 % and Auto EF must exceed 5.6 % in absolute differences to reveal a true change. Seen from this, Auto EF is the safest method for evaluating a detoriation of 10 % (percentage points). Generic strain and AFI GLS has CoR equal to 3.4 % and 1.8 % respectively. A relative decrease of 15 % in GLS from a normal value of -20 % will be + 3 % (percentage points). AFI GLS can reveal this change, however generic strain cannot. To my knowledge there is no recommended limits given for significant reduction in MAPSE, MVE Vel and MV E/A Vel indices to indicate CTRCD.

Good reproducibility is not enough to choose a method of measurement. Sensitivity, specificity and likelihood ratios is needed to determine the clinical meaningfulness of a test.

5. Conclusion.

The numerically most feasible echocardiographic methods were MAPSE, all Doppler indices, LVEF without contrast and 3D LVEF. The most reproducible methods in intra-rater analyses were Auto EF, LVEF with contrast and MVE Vel. We found better reproducibility when echocardiographic recordings were reanalysed by the same operator, compared to analyses made by two operators. Thus, follow-up examinations by the same operator may improve reproducibility in everyday clinic.

References.

1. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033-9.
2. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.
3. Shrout PE, Fleiss JL. Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull. 1979;86:420 - 8.
4. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.
5. Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? Physiotherapy. 2000;86(2):94-9.
6. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet Glob Health. 1986.
8. Zamorano JL, Lancellotti P, Rodriguez Munoz D, Aboyans V, Asteggiano R, Galderisi M, et al. 2016 ESC Position Paper on cancer treatments and cardiovascular toxicity developed under the auspices of the ESC Committee for Practice Guidelines: The Task Force for cancer treatments and cardiovascular toxicity of the European Society of Cardiology (ESC). Eur Heart J. 2016;37(36):2768-801.
9. Brystkreft Oslo: Kreftregisteret; 2016 [cited 2016 01.10]. Available from: https://www.kreftregisteret.no/Generelt/Fakta-om-kreft/Brystkreft-Alt2/.
10. Lancellotti P, Nkomo VT, Badano LP, Bergler-Klein J, Bogaert J, Davin L, et al. Expert consensus for multi-modality imaging evaluation of cardiovascular complications of radiotherapy in adults: a report from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. J Am Soc Echocardiogr. 2013;26(9):1013-32.
11. Jensen BV. Cardiotoxic consequences of anthracycline-containing therapy in patients with breast cancer. Semin Oncol. 2006;33(3 Suppl 8):S15-21.
12. Fei HW, Ali MT, Tan TC, Cheng KH, Salama L, Hua L, et al. Left Ventricular Global Longitudinal Strain in HER-2 + Breast Cancer Patients Treated with Anthracyclines and

Trastuzumab Who Develop Cardiotoxicity Is Associated with Subsequent Recovery of Left Ventricular Ejection Fraction. Echocardiography. 2016;33(4):519-26.

13. Stewart FA, Seemann I, Hoving S, Russell NS. Understanding radiation-induced cardiovascular damage and strategies for intervention. Clin Oncol (R Coll Radiol). 2013;25(10):617-24.
14. Plana JC, Galderisi M, Barac A, Ewer MS, Ky B, Scherrer-Crosbie M, et al. Expert Consensus for Multimodality Imaging Evaluation of Adult Patients during and after Cancer Therapy: A Report from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr. 2014;27(9):911-39.
15. Jacob S, Pathak A, Franck D, Latorzeff I, Jimenez G, Fondard O, et al. Early detection and prediction of cardiotoxicity after radiation therapy for breast cancer: the BACCARAT prospective cohort study. Radiat Oncol. 2016;11:54.
16. Szulik M, Pappas CJ, Jurcut R, Magro M, Peeters E, Goetschalckx K, et al. Clinical validation of a novel speckle-tracking-based ejection fraction assessment method. J Am Soc Echocardiogr. 2011;24(10):1092-100.
17. Hering D, Faber L, Horstkotte D. Echocardiographic features of Radiation-Associated valvular disease. The American Journal of Cardiology. 2003;92(2):226-30.
18. Hull M, Morris C, Pepine C, Medenhall N. Valvular dysfunction and carotid, subclavian, and coronary artery disease in survivors of Hodgkin lymphoma treated with radiation therapy. JAMA. 2003;290:2831-7.
19. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft (IS-2669) Oslo: Helsedirektoratet,; 2017 [cited 2018 02.10]. Available from: https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/1398/IS-2669_nasjonalt_handlingsprog_brystkreft.pdf.
20. Tsai HR, Gjesdal O, Wethal T, Haugaa KH, Fossa A, Fossa SD, et al. Left ventricular function assessed by two-dimensional speckle tracking echocardiography in long-term survivors of Hodgkin's lymphoma treated by mediastinal radiotherapy with or without anthracycline therapy. Am J Cardiol. 2011;107(3):472-7.
21. Sawaya H, Sebag IA, Plana JC, Januzzi JL, Ky B, Cohen V, et al. Early detection and prediction of cardiotoxicity in chemotherapy-treated patients. Am J Cardiol. 2011;107(9):1375-80.

22. Malm S, Sagberg E, Larsson H, Skjærpe T. Choosing Apical Long-axis Instead of Two-chamber View Gives More Accurate Biplane Echocardiographic Measurements of Left Ventricular Ejection Fraction: A Comparison with Magnetic Resonance Imaging. J Am Soc Echocardiogr. 2005;18(10):1044-50.
23. Thavendiranathan P, Grant AD, Negishi T, Plana JC, Popovic ZB, Marwick TH. Reproducibility of echocardiographic techniques for sequential assessment of left ventricular ejection fraction and volumes: application to patients undergoing cancer chemotherapy. J Am Coll Cardiol. 2013;61(1):77-84.
24. Stoodley PW, Richards DA, Boyd A, Hui R, Harnett PR, Meikle SR, et al. Left ventricular systolic function in HER2/neu negative breast cancer patients treated with anthracycline chemotherapy: a comparative analysis of left ventricular ejection fraction and myocardial strain imaging over 12 months. Eur J Cancer. 2013;49(16):3396-403.
25. Edvardsen T, Helle-Valle T, Smiseth OA. Systolic dysfunction in heart failure with normal ejection fraction: speckle-tracking echocardiography. Prog Cardiovasc Dis. 2006;49(3):207-14.
26. Erven K, Jurcut R, Weltens C, Giusca S, Ector J, Wildiers H, et al. Acute radiation effects on cardiac function detected by strain rate imaging in breast cancer patients. Int J Radiat Oncol Biol Phys. 2011;79(5):1444-51.
27. 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.
28. Pignatti M, Mantovani F, Bertelli L, Barbieri A, Pacchioni L, Loschi P, et al. Effects of silicone expanders and implants on echocardiographic image quality after breast reconstruction. Plast Reconstr Surg. 2013;132(2):271-8.
29. Otto CM. Textbook of Clincal Echocardiography. Philadephia, USA: Elsevier Health Sciences; 2018 [cited 2018 02.18]. Available from: https://books.google.no/books?id=qLXIDQAAQBAJ&source=gbs_navlinks_s.
30. Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, Di Salvo G, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging. 2017;18(12):1301-10.
31. Malm S, Frigstad S, Sagberg E, Larsson H, Skjaerpe T. Accurate and reproducible measurement of left ventricular volume and ejection fraction by contrast

echocardiography: a comparison with magnetic resonance imaging. J Am Coll Cardiol. 2004;44(5):1030-5.

32. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R. Population-based reference values for 3D echocardiographic LV volumes and ejection fraction. JACC Cardiovasc Imaging. 2012;5(12):1191-7.
33. Thavendiranathan P, Liu S, Verhaert D, Calleja A, Nitinunu A, Van Houten T, et al. Feasibility, accuracy, and reproducibility of real-time full-volume 3D transthoracic echocardiography to measure LV volumes and systolic function: a fully automated endocardial contouring algorithm in sinus rhythm and atrial fibrillation. JACC Cardiovasc Imaging. 2012;5(3):239-51.
34. Stoylen A, Molmen HE, Dalen H. Relation between Mitral Annular Plane Systolic Excursion and Global longitudinal strain in normal subjects: The HUNT study. Echocardiography. 2018.
35. 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.
36. Barbier P, Mirea O, Cefalu C, Maltagliati A, Savioli G, Guglielmo M. Reliability and feasibility of longitudinal AFI global and segmental strain compared with 2D left ventricular volumes and ejection fraction: intra- and inter-operator, test-retest, and inter-cycle reproducibility. Eur Heart J Cardiovasc Imaging. 2015;16(6):642-52.
37. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. Circulation. 2016;133(11):1104-14.
38. Hoffmann R, von Bardeleben S, ten Cate F, Borges AC, Kasprzak J, Firschke C, et al. Assessment of systolic left ventricular function: a multi-centre comparison of cineventriculography, cardiac magnetic resonance imaging, unenhanced and contrast-enhanced echocardiography. Eur Heart J. 2005;26(6):607-16.
39. Senior R, Becher H, Monaghan M, Agati L, Zamorano J, Vanoverschelde JL, et al. Clinical practice of contrast echocardiography: recommendation by the European Association of Cardiovascular Imaging (EACVI) 2017. Eur Heart J Cardiovasc Imaging. 2017;18(11):1205-af.
40. Senior R, Becher H, Monaghan M, Agati L, Zamorano J, Vanoverschelde JL, et al. Contrast echocardiography: evidence-based recommendations by European Association of Echocardiography. Eur J Echocardiogr. 2009;10(2):194-212.

41. Claudon M, Dietrich CF, Choi BI, Cosgrove DO, Kudo M, Nolsoe CP, et al. Guidelines and good clinical practice recommendations for contrast enhanced ultrasound (CEUS) in the liver--update 2012: a WFUMB-EFSUMB initiative in cooperation with representatives of AFSUMB, AIUM, ASUM, FLAUS and ICUS. Ultraschall Med. 2013;34(1):11-29.
42. Armstrong AC, Ricketts EP, Cox C, Adler P, Arynchyn A, Liu K, et al. Quality Control and Reproducibility in M-Mode, Two-Dimensional, and Speckle Tracking Echocardiography Acquisition and Analysis: The CARDIA Study, Year 25 Examination Experience. Echocardiography. 2015;32(8):1233-40.
43. Kottner J, Streiner DL. The difference between reliability and agreement. J Clin Epidemiol. 2011;64(6):701-2; author reply 2.
44. Spurney CF, McCaffrey FM, Cnaan A, Morgenroth LP, Ghelani SJ, Gordish-Dressman H, et al. Feasibility and Reproducibility of Echocardiographic Measures in Children with Muscular Dystrophies. J Am Soc Echocardiogr. 2015;28(8):999-1008.
45. Wood PW, Choy JB, Nanda NC, Becher H. Left ventricular ejection fraction and volumes: it depends on the imaging method. Echocardiography. 2014;31(1):87-100.
46. Erdei T, Smiseth OA, Marino P, Fraser AG. A systematic review of diastolic stress tests in heart failure with preserved ejection fraction, with proposals from the EU-FP7 MEDIA study group. Eur J Heart Fail. 2014;16(12):1345-61.
47. Rankin G. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil. 1998;12:187-99.
48. Hallgren KA. Computing Inter-rater Reliability for Observational Data: An Overview and Tutorial. Psychol Methods. 2012;8(1):23-34.
49. 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.
50. Stanton T, Leano R, Marwick TH. Prediction of all-cause mortality from global longitudinal speckle strain: comparison with ejection fraction and wall motion scoring. Circ Cardiovasc Imaging. 2009;2(5):356-64.
51. Stoylen A, Molmen HE, Dalen H. Importance of length and external diameter in left ventricular geometry. Normal values from the HUNT Study. Open Heart. 2016;3(2):e000465.

52. Yingchoncharoen T, Agarwal S, Popovic ZB, Marwick TH. Normal ranges of left ventricular strain: a meta-analysis. J Am Soc Echocardiogr. 2013;26(2):185-91.
53. Sugimoto T, Dulgheru R, Bernard A, Ilardi F, Contu L, Addetia K, et al. Echocardiographic reference ranges for normal left ventricular 2D strain: results from the EACVI NORRE study. Eur Heart J Cardiovasc Imaging. 2017;18(8):833-40.
54. Cheng S, Larson MG, McCabe EL, Osypiuk E, Lehman BT, Stanchev P, et al. Reproducibility of speckle-tracking-based strain measures of left ventricular function in a community-based study. J Am Soc Echocardiogr. 2013;26(11):1258-66 e2.
55. Takigiku K, Takeuchi M, Izumi C, Yuda S, Sakata K, Ohte N, et al. Normal Range of Left Ventricular 2-Dimensional Strain. Circ J. 2012;76(11):2623-32.
56. Stoylen A. Cardiac ultrasound. Strain rate imaging. Trondheim. Norway [cited 2018. Available from: http://folk.ntnu.no/stoylen/.
57. 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. J Am Soc Echocardiogr. 2016;29(4):277-314.
58. Nagueh SF, Appleton CP, Gillebert TC, Marino PN, Oh JK, Smiseth OA, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. J Am Soc Echocardiogr. 2009;22(2):107-33.
59. Reidunsdatter RJ, Albrektsen G, Hjermstad MJ, Rannestad T, Oldervoll LM, Lundgren S. One-year course of fatigue after post-operative radiotherapy in Norwegian breast cancer patients--comparison to general population. Acta Oncol. 2013;52(2):239-48.
60. Reidunsdatter RJ, Rannestad T, Frengen J, Frykholm G, Lundgren S. Early effects of contemporary breast radiation on health-related quality of life - predictors of radiotherapy-related fatigue. Acta Oncol. 2011;50(8):1175-82.
61. Palmieri V, Russo C, Buonomo A, Palmieri EA, Celentano A. Novel wall motion score-based method for estimating global left ventricular ejection fraction: validation by real-time 3D echocardiography and global longitudinal strain. Eur J Echocardiogr. 2010;11(2):125-30.
62. Portney LG, Watkins MP. Foundations of Clinical Research. Essex: Pearson Education Limited; 2014.

63. Johri AM, Picard MH, Newell J, Marshall JE, King ME, Hung J. Can a teaching intervention reduce interobserver variability in LVEF assessment: a quality control exercise in the echocardiography lab. JACC Cardiovasc Imaging. 2011;4(8):821-9.
64. Bland M. An introduction to medical statistics. 2015. 411 p.
65. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217-38.
66. 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.
67. Caballero L, Kou S, Dulgheru R, Gonjilashvili N, Athanassopoulos GD, Barone D, et al. Echocardiographic reference ranges for normal cardiac Doppler data: results from the NORRE Study. Eur Heart J Cardiovasc Imaging. 2015;16(9):1031-41.
68. Mulvagh SL, Rakowski H, Vannan MA, Abdelmoneim SS, Becher H, Bierig SM, et al. American Society of Echocardiography Consensus Statement on the Clinical Applications of Ultrasonic Contrast Agents in Echocardiography. J Am Soc Echocardiogr. 2008;21(11):1179-201; quiz 281.
69. Medvedofsky D, Kebed K, Laffin L, Stone J, Addetia K, Lang RM, et al. Reproducibility and experience dependence of echocardiographic indices of left ventricular function: Side-by-side comparison of global longitudinal strain and ejection fraction. Echocardiography. 2017;34(3):365-70.
70. AlJaroudi WA, Thomas JD, Rodriguez LL, Jaber WA. Prognostic value of diastolic dysfunction: state of the art review. Cardiol Rev. 2014;22(2):79-90.
71. Shoukri MM. Sample size requirements for the design of reliability study: review and new results. Stat Methods Med Res. 2004;13:1-21.
72. Jenkins C, Moir S, Chan J, Rakhit D, Haluska B, Marwick TH. Left ventricular volume measurement with echocardiography: a comparison of left ventricular opacification, three-dimensional echocardiography, or both with magnetic resonance imaging. Eur Heart J. 2009;30(1):98-106.
73. Cicchetti DV. Sample Size Requirements for Increasing the Precision of Reliability Estimates: Problems and Proposed Solutions. J Clin Exp Neuropsychol. 1999;21:567-70.
74. De Boeck K, Castellani C, Elborn JS, Board E. Medical consensus, guidelines, and position papers: a policy for the ECFS. J Cyst Fibros. 2014;13(5):495-8.

List of figures and tables.

Figure 1. 4-chamber view of LV without contrast to the left and with contrast to the right, Patient IDentification number PID18355. 6
Figure 2. 4-chamber view of LV without contrast to the left and with contrast to the right, PID19183. 6
Figure 3. The reliability coefficient. The expression variability represents the statistical expression variance, s^2 (SD^2). 7
Figure 4. Tracing of LVEF with contrast enhanced endocardial boarders. The traces show the true apex which was clearly seen in diastole. The contrast filled area shows a shorter cavity, demonstrating foreshortening in PID19071. 8
Figure 5. Biplane Auto EF in PID18388. 4-chamber view in the left and 2-chamber view in the right panel. 9
Figure 6. Generic longitudinal strain illustrated by Støylen (56). 12
Figure 7. AFI, strain curves from all segments in three planes and Bull's eye plot in PID19045. 13
Figure 8. Flowchart of study population. 16
Figure 9. Flowchart of study design. 17
Figure 10. Bland-Altman plots of the intra-rater differences in manually traced LVEF ad Simpson's method to the left and the intra-rater differences for manually traced LVEF ad Simpson's method enhanced with ultrasound contrast to the right. Bias, upper and lower 95% LoA noted in the graphs. 29
Figure 11. Bland-Altman plots of the intra-rater differences for Auto EF to the left and the intra-rater differences for 3D LVEF to the right. Bias, upper and lower 95% LoA noted in the graphs. 29
Figure 12. Bland-Altman plots of the intra-rater differences of Generic strain to the left and of the intra-rater differences for tracking based global longitudinal strain (AFI GLS) to the right *. Bias, upper and lower 95% LoA noted in the graphs. 30
Figure 13. Bland-Altman plots of the intra-rater differences of the mitral annulus plane systolic velocity (S') to the left and of the intra-rater differences for mitral annular plane systolic excursion (MAPSE) to the right. Bias, upper and lower 95% LoA noted in the graphs. 30
Figure 14. Bland-Altman plots of the inter-rater differences of LVEF manually traced by Simpson's method witout ultrasound contrast to the left * and LVEF manually traced by Simpson's method with ultrasound contrast to the right. 32
Figure 15. Bland-Altman plots of the inter-rater differences of Auto EF to the left and 3D LVEF to the right. 32
Figure 16. Bland-Altman plots of the inter-rater differences of Generic longitudinal strain to the left * and tracking based global longitudinal strain (AFI GLS) to the right *. 33
Figure 17. Bland-Altman plots of the inter-rater differences of Mitral annular plane systolic velocity (S') to the left * and Mitral annular plane excursion (MAPSE) to the right *. 33
Figure 18. Comparison of variation of feasibility and CoV with lines for limits of good feasibility and CoV as discussed. 35
Figure 19. Comparison of variation of feasibility and ICC with lines for limits of good feasibility and ICC as discussed. 36

Table 1. A listed overview of the echocardiographic image acquisitions. 18
Table 2. List of investigated parameters. 20
Table 3. Timetable. 23

Table 4. Echocardiographic characteristics and feasibility for the whole study population of 55 patients. 26
Table 5. Intra- and inter-rater reproducibility for the different echocardiographic methods based on 20 patients. 27
Table 6. Distribution of surgery and reconstruction procedures. 39
Table 7. Mean and standard deviation of inter-rater differences in three sets of measures of the discussed indices LVEF witout contrast, S' and MV Dec T from rater 1 and rater 2..... 41

Appendix 1-4.

Appendix 1. Echocardiographic Acquisition Procedure.
Appendix 2. Reading protocol.
Appendix 3. Information letter.
Appendix 4. Statistical supplement.

Echocardiographic Acquisition Procedure. Appendix 1

"Cardiovascular toxicity after treatment for breast cancer. The optimal diagnostic modality for early detection."

Before starting Establish venous port for the patient.
Choose the right settings including tree cine-loops.
Ensure to register all data with the correct patient code.
Note the year of birth, the height and weight of the patient.

Parasternal long-axis

1. 3 cine-loops 2D
2. M-mode
3. Zoom LVOT

Parasternal short-axis

1. 3 cine-loops 2D
2. M-mode

Apical 4-chamber-view

1. 3 cine-loops 2D
2. 3 cine-loops TVI
3. Pulsed tissue Doppler septal
4. Pulsed tissue Doppler lateral
5. M-mode MAPSE septal
6. M-mode MAPSE lateral

Apical 2-chamber-view

1. 3 cine-loops 2D
2. 3 cine-loops TVI

Apical long-axis

1. 3 cine-loops 2D
2. 3 cine-loops TVI

Apical right ventricle

1. 3 cine-loops 2D
2. 3 cine-loops TVI
3. TAPSE

Spectral-Doppler from apical four-/ five-chamber view

1. PW mitralflow
2. PW pulmonal veins flow
3. PW LVOT flow
4. CW aortic flow
5. IVRT left (minimal filter)

Colour-Doppler

1. Mitralflow/ regurgitation
2. Aortic flow/ regurgitation
3. Tricuspid flow/ regurgitation

3D/ 4D

Apical 4-chamber-view, multi-slice, 4 heartbeats, large, breath hold, adjust top bottom

Contrast, Sono Vue

3D/4D

1. ALLERGY? CONTRASTPROGRAM! MI = 0,14, Tissue visualisation off
2. Apical four chamber view, multi-slice, 4 heartbeats, large, breath hold

2D

1. ALLERGY? CONTRASTPROGRAM! Manually chose three cycles, Tissue visualisation off
2. Apical 4chamber-view
3. Apical 2-chamber-view
4. Apical long-axis-view

Later analyses:

Ejection fraction manually and with Auto EF with and without contrast in 2D and 3D

Standard systolic function parameters from M-mode og tissue Doppler

Standard diastolic function parameters from pulmonary veins flow, Doppler flow and tissue Doppler. Age corrections.

Valve failure mild/moderate/severe, stenosis/regurgitation.

2015-05-23, Gunn Catrin Fossum Bøen

Reading Protocol Appendix 2

"Cardiovascular toxicity after treatment for breast cancer. The optimal diagnostic modality for early detection."

Image Quality categorisation.

During the analyses, we visually consider image quality of all images in respect of ability to be measured and analysed correctly. We use three categories; Good (2), suboptimal (1) and poor (0). Good image quality is the category for recordings which has straight forward analyses. Suboptimal image quality is the category for recordings needing adjustments and eyeballing to make their analyses. Poor image quality is the category for recordings impossible to analyse. Individual comments are written under each index.

Manually traced EF bi-plan ad Simpsons without contrast.

In the non-contrast, left ventricle focused, apical 4-chamber and apical two-chamber views, EF is measured with Biplane Simpsons method. For each view, the recording with the best quality of three consecutive loops was chosen. The most convenient method was found to save this single best loop on its own to avoid confusion during the tracing procedure. The endocardium boarder is manually traced at the blood-tissue interface, in end diastole and end systole. End diastole is defined as the first frame with Mitral valve closed, which corresponds with S in the ECG. End systole is defined as when the ventricle is smallest. Trabecles and the papillary muscle is not traced out and thereby they are taken into the volume, according to European recommendations (Lang et al, 2015). At the apex, where the image quality most often is poor, we choose to trace up to the usually good visible epicardium. This method will also help avoiding foreshortened ventricles. When we see foreshortening we try to adapt the systolic position of apex close to the diastolic position.

We use the menu; Measure, Volume, Biplane, LVEDV A4K, LVESV A4K, LVEDV A2K, LVESV A2K.

Image quality; Visually considered as described above (Good/suboptimal/poor) in 4 CH with an add on of + (0.5) when there is great uncertainty about the tracing, but still can be done in an uncertain way.

Manually traced EF bi-plane ad Simpsons with contrast.

The analysis in contrast mode differs slightly from the non-contrast analysis. In contrast recordings, we more confidently can trace true boarders between lumen and endocardium. Therefore, we chose to trace true boarders along all segments, and not up to the epicardium as in non-contrast mode. This will probably give us slightly smaller volumes, in both diastole and systole, than in our modified non-contrast tracing. The contrast function helps us to better endocardial distinction and more correctly recognition of the papillary muscles and the trabecula. Contrast enhanced images usually shows a more traceable anterior wall. Unfortunately, the contrast does not completely help us out of the foreshortening problem. An example of foreshortening when tracing apex is shown in figure 1. When we see foreshortening we try to adapt the systolic position of apex close to the diastolic position.

> We use the menu; First; contrast settings. Then; Measure, Volume, Biplane, LVEDV A4K, LVESV A4K, LVEDV A2K, LVESV A2K.
>
> Image quality; Visually considered as described above (2/1/0) in 4 CH with an add on of 0.5 when there is great uncertainty about the tracing, but still can be done in an uncertain way.

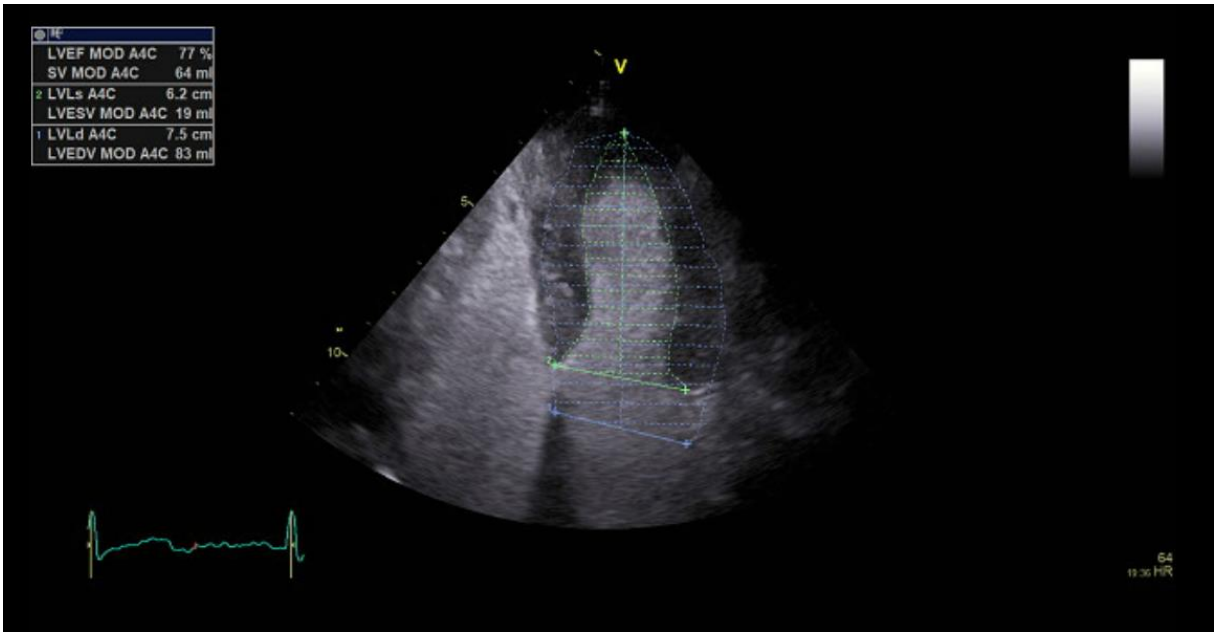


Figure 1. Example of tracing foreshortened, contrast enhanced EF on patient PID19071.

LV length without contrast.

After the non-contrast EF calculations, we use the same loop for recording the length of the left ventricle in end diastole. We use the caliper function to measure the length of the ventricle, both from the apex to the septal AV-plane and from apex to the lateral AV-plane.

The apical measuring site is chosen as the most distant point of the apex. At apex in end diastole the problems with foreshortening are insignificant and the myocardium is thin which make it easy to put the correct apical measuring point. The AV-plan site is at the top of the signal using moderate gain. Then we calculate the mean LV length as; (apex to septal AV + apex to lateral AV) / 2.

We use the menu; Caliper

Image quality; Visually considered as described above (2/1/0).

LV length with contrast.

After the contrast EF calculations, we use the same loop for recording the length of the left ventricle in end diastole. We use the caliper function to measure the length of the ventricle, both from the apex to the septal AV-plane and from apex to the lateral AV-plane. The apical measuring site is chosen as the most distant point of the apex. At apex in end diastole the problems with foreshortening are insignificant and the myocardium is thin which make it easy to put the correct apical measuring point. The apical measuring site is chosen as the boarder of contrast-no contrast. Then we calculate the mean LV Length as (apex to septal AV + apex to lateral AV) / 2.

We use the menu; Caliper

Image quality; Visually considered as described above (2/1/0).

Semi-automatic 3-Dimensional EF.

We choose the 3D cine-loop recording that seems of best quality. We first do a simple alignment of axis. To measure the EDV, end diastolic volume, we are asked to mark apex and AV-plan to assist the semi-automatic tracing. After the automatic tracing, we can adjust the lines if necessary. The frozen 3D images are not very easy to read, and adjustments often feel uncertain, which often ends up doing few corrections. Then we repeat the same procedure in systole.

We use the menu; Measure, Volume, 4D Auto LVQ.

The image quality; Good (2) when there are no problems. Poor (1) when we must adjust the lines and end up not fully satisfied. Partly impossible (0.5) when we cannot verify any of the machines tracings, but the machine gives an answer anyway. Impossible (0) when the analysis cannot be done.

Semi-automatic biplane AUTO EF.

With the best 4-chamber loop, we recognise end diastole and choose 4-chamber in the menu. The automatic tracing detects endocard, and the measurements will therefore be slightly different form our measures with the manually traced EF without contrast which we chose to draw all the way up to pericardia. Most often adjustments of the automatic traced line are needed. We get our result and continues to repeat the procedure in 2-chamber view. The 2-chamber and the 4-chamber results are then calculated to give us Biplane AUTO EF.

We use the menu; Measure, Volume, Auto EF.

Image quality; Categorized as good, poor or impossible, 2, 1 or 0.

Good (2) when everything is ok.

Poor (1) when the reduced image quality make the analysing result uncertain according to visual assessment or the program warn about problems, such as “Image quality not optimal to trace” or “Verify tracking” needing manual considerations and/ or adjustments, considered as quite right after visual quality control. Impossible when the analyse cannot be done, nor with or without manual adjustments, not passing a visual quality control.

Automatic functional imaging for global longitudinal strain, AFI.

AFI is a semiautomatic measure of global longitudinal strain, GLS. We need the apical 2D images, in all three standard views, 4-chamber-, two-chamber- and long-axis view. The automatic tracing was verified manually and readjusted to achieve

optimal tracking of the endocardial boarder correctly. The region of interest should include both endocard and myokard, but not the epicardium. In this study, we choose to decrease the width of the region of interest (ROI) with one step consequently in all the analyses. We start in long-axis view to be able to verify the automatic aortic closure. The image quality is most often too poor, which results in using the automatic setting /timing of aortic closure. When

tracking seems ok we proceed with the menu; Process and Approve. The program shows which segments who fails to be calculated. In the final Bull's eye, this will show as a grey cloud. This is to be repeated for all three standard views. When two segments fail to be traced, we cannot have the average reading. If we are lucky, 4-chamber view is not affected by poor image quality and we may be able to get this single plane reading of strain, although the global reading is not available.

We use the menu; Measure, AFI, proceeding with Process and Approve.

Image quality; We expand the quality scale for this measure only, giving 2,1, (0.5),0) Good (2) when everything is ok, and 0-1 segments fail to be analysed. Quality of the images will be set to poor (1) when 2 or more segments fail to be calculated or visually have bad tracking. Quality is set to partly impossible (0.5) when either average strain or 4CH strain cannot be analysed. Impossible (0) when we are unable to get neither average strain nor 4 CH strain.

MAPSE

In the M-mode images we measured the height of the AV-plane movement, both at septum and at the lateral wall, to calculate the mean MAPSE of the two.

$(MAPSE_{septal} + MAPSE_{lateral}) / 2.$

We use the menu; Caliper, moderate gain and horizontal sweep showing 4-6 heartbeats.

Image quality; Visually considered as described above (2/1/0).

Generic global strain.

The value is calculated as mean MAPSE/ mean LV Length. Mean LV Length is the average of the end diastolic lengths from the most distant point of epicardial apex to septal mitral annulus and from the most distant point of epicardial apex to lateral mitral annulus, reported in mm. We choose to use the length measured without contrast to be able to do sensible comparisons with non-contrast AFI GLS Strain.

Early Diastolic Tissue velocity, e'.

We measure the early diastolic tissue velocity in both septal and lateral mitral ring with pulsed Tissue-Doppler and analyse with the lowest gain in which spectrum is clearly visible as recommended by Dalen et al. in 2016. This means often gain sat down to lowest or near lowest value. Our measure is represented by the mean of the two measured; $e' = (e'_{septal} + e'_{lateral}) / 2$.

We use the menu; Measure, Mitral valve, e'.

Image quality; The quality of the image is categorised poor if there is great variability and noise in the recordings. We measure the most appropriate beat.

Peak Systolic Tissue velocity, S'.

We measure systolic tissue velocity S' with pulsed Tissue-Doppler and analyse with the lowest gain in which spectrum is clearly visible as recommended by Dalen et al. 2016.

We use the menu; Measure, Mitral valve, S'.

Image quality; The quality of the image is categorised poor if there is great variability over time in the recordings. We measure the most representative/ appropriate beat.

Mitral Inflow E-velocity, A-velocity and Deceleration time.

Blood velocity in early and late diastole through the mitral valve is measured at the tip of the leaflets and analysed with low gain. To get the most accurate measures we adjust horizontal sweep to approximately 66 mm/s, usually visualizing two heartbeats. We choose the beat with the best image quality for the analysis. By marking the E and A velocities together with the deceleration angle, the proportion E/A and the MV Dec T is calculated.

We use the menu; Measure, Mitral valve, MV E/A Vel.

Image quality; Visually considered as described above (2/1/0).

Mitral Inflow E-velocity/ Tissue-Doppler e`-velocity, E/e`.

Blood velocity through the mitral valve in early diastole, is measured at the tip of the leaflets and analysed with low gain. (ref) We measure the early diastolic Tissue velocity in both septal

and lateral mitral ring with pulsed Tissue-Doppler to give the mean value to be included in the proportion. The proportion E/e` is then calculated.

We use the menu; This value is automatically calculated from earlier measures.

Image quality; Visually considered as described above (2/1/0).

Isovolumetric relaxation time.

The isovolumetric relaxation time in diastole (IVRT) is measured with the Doppler beam pointing between the aortic valve and the mitral valve in a modified five-chamber-view. Both the aortic valve closure signal and the starting of mitral flow must be visible. We choose the best image quality beat available in CW mode and measure once. The signal of aortic valve closure is often wide, and we choose to measure at the end of the closure signal. The mitral valve does not have a clear opening signal. We therefore choose to measure at the start of mitral flow. Again, we adjust Doppler gain down to the level when noise is barely visible.

We use the menu; Measure, Aortic, IVRT.

Image quality; When there is uncertainty, poor quality is categorized.

The menu references.

The menu references relate to the equipment used in this study which is the GE, Vivid E9, Horten, Norway.

Information letter. **Appendix 3**

NTNU

Forespørsel om deltakelse i forskningsprosjektet

Reliabilitet for ekkokardiografi ved kontroll av hjertefunksjon etter brystkreftbehandling

Dette er et spørsmål til deg om å delta i et forskningsprosjekt for å undersøke om ekkokardiografi, dvs. undersøkelse av hjertet med ultralyd, er en god metode for å kontrollere din hjertefunksjon etter at du har blitt behandlet for brystkreft. Du er valgt ut siden du har fått den kombinasjonen av kreftbehandlinger som vi er mest nysgjerrige på å få vite mer om. Dette prosjektet har sitt utspring fra NTNU og St. Olavs Hospital.

Hva innebærer prosjektet?

Deltakelse vil innebære at du får en ultralydundersøkelse av hjertet ditt i tillegg til din ordinære 8-års kontroll på kreft poliklinikken. Undersøkelsen er ufarlig og smertefri og tar ca. 40 min. Vi utfører undersøkelsen på et rom innerst i gangen ved hjertepoliklinikken på St. Olavs hospital. Hjertepoliklinikken ligger i Akutten og hjerte-lunge-senteret i første etg. i sørfløya. Følg skilting fra hovedinngangen på senteret til hjertepoliklinikken.

I prosjektet vil vi innhente og registrere opplysninger om deg. Anonymiserte ultralydbilder vil bli lagret på en dataserver og tatt godt vare på så lenge prosjektet varer. Dette prosjektet skal se på bildene og gjøre målinger og analyser på dem for å vurdere nøyaktigheten på målingene. Et annet prosjekt vil bruke de samme måleverdiene for å sammenligne med MR- og MUGA-mål. Våre vurderingen og konklusjoner skal siden publiseres for at ny viten skal komme flere til nytte.

Mulige fordeler og ulemper

Fordelen ved å være med vil for deg vil være at du får en grundig hjerteundersøkelse. Skulle det være noe å bemerke, vil dette tas tak i og du vil få den behandlingen og oppfølgingen du trenger. Ulempen er ikke annet enn at dette tar litt av din tid. Under prosedyren vil du få ultralydkontrast, som er en ikke radioaktiv kontrast basert på bitte små gassbobler. Det er en minimal risiko for allergisk reaksjon på blandestoffet i denne kontrasten. Hvis du ikke har noen alvorlige allergier, er denne faren svært liten. Dine ordinære kontroller og avtaler går uforstyrret parallelt med denne undersøkelsen.

Frivillig deltakelse og mulighet for å trekke sitt samtykke

Det er frivillig å delta i prosjektet. Dersom du ønsker å delta, undertegner du samtykkeerklæringen på siste side. Du kan når som helst og uten å oppgi noen grunn trekke ditt samtykke. Dette vil ikke få konsekvenser for din videre behandling. Dersom du trekker deg fra prosjektet, kan du kreve å få slettet innsamlede prøver og opplysninger, med mindre opplysningene allerede er inngått i analyser eller brukt i vitenskapelige publikasjoner. Dersom du senere ønsker å trekke deg eller har spørsmål til prosjektet, kan du kontakte prosjektleder Randi Reidunsdatter, randi.reidunsdatter@ntnu.no eller masterstudent Kari Nordmann, kari.nordmann@ntnu.no, mobil 1232016.

Hva skjer med informasjonen om deg?

Informasjonen som registreres om deg skal kun brukes slik som beskrevet i hensikten med studien. Du har rett til innsyn i hvilke opplysninger som er registrert om deg og rett til å få korrigert eventuelle feil i de opplysningene som er registrert.

Alle opplysningene vil bli behandlet uten navn og fødselsnummer eller andre direkte gjenkjennende opplysninger. En kode knytter deg til dine opplysninger gjennom en navneliste som oppbevares separat og trygt.

Prosjektleder har ansvar for den daglige driften av forskningsprosjektet og at opplysninger om deg blir behandlet på en sikker måte. Informasjon om deg vil bli anonymisert eller slettet senest fem år etter prosjektslutt.

Forsikring

Undersøkelsen regnes som en utvidelse av standard oppfølging og dekkes derved av Norsk pasientskade erstatning (NPE).

Oppfølgingsprosjekt

Dette er andre gangen du blir med og støtter forskningen. Vi er takknemlige for dette og vil gjerne ha muligheten til å få spørre deg igjen, hvis nye prosjekter bygger videre på dette. Du kan selvsagt når som helst velge å avstå.

Økonomi

Du vil få refundert reiseutgifter etter statens satser.

Godkjenning

Prosjektet er godkjent av Regional komité for medisinsk og helsefaglig forskningsetikk, saksnr. hos REK midt (2009/108 4.2006.2856).

Samtykke til deltakelse i PROSJEKTET

Jeg er villig til å delta i prosjektet

_____	_____
Sted og dato	Deltakers signatur
	_____
	Deltakers navn med trykte bokstaver

Statistical supplement. Appendix 4

Tables showing intraclass correlation coefficients ICC (3,1) and ICC (2,2) with their respective degrees og freedom (df) and Cronbach's alpha correlation coefficient.

Intra-rater	Cronback's alpha	ICC (3,1)	95 % CI	Valid pairs	df
EF (%)	0.81	0.82	[0.52, 0.93]	19	18
EF with contrast (%)	0.87	0.87	[0.66, 0.95]	18	17
Auto EF (%)	0.93	0.93	[0.83, 0.98]	18	17
3D EF (%)	0.88	0.89	[0.69, 0.96]	18	17
S' (cm/s)	0.81	0.81	[0.51, 0.93]	19	18
MAPSE (mm)	0.87	0.88	[0.68, 0.95]	19	18
Generic Strain (%)	0.89	0.88	[0.69, 0.96]	18	17
AFI 4CH Strain (%)	0.75	0.73	[0.29, 0.90]	18	17
AFI Global Strain (%)	0.96	0.93	[0.61, 0.98]	13	12
e' (cm/s)	0.87	0.86	[0.64, 0.95]	19	18
MV E Vel (cm/s)	0.98	0.98	[0.96, 0.99	18	17
E/e'	0.86	0.86	[0.62, 0.95]	17	16
MV E/A	0.98	0.98	[0.94, 0.99]	18	17
MV Dec T (ms)	0.56	0.53	[-0.15, 0.82]	18	17

Inter-rater	Cronback's alpha	ICC (2,2)	95 % CI	Valid pairs	df
EF (%)	0.21	0.17	[-0.66, 0.65]	17	16
EF with contrast (%)	0.81	0.82	[0.51, 0.93]	18	17
Auto EF (%)	0.81	0.82	[0.49, 0.93]	17	16
3D EF (%)	0.87	0.87	[0.67, 0.95]	20	19
S' (cm/s)	0.64	0.5	[-0.15, 0.8]	19	18
MAPSE (mm)	0.95	0.85	[-0.10, 0.96]	20	19
Generic Strain (%)	0.9	0.81	[-0.18, 0.95]	20	19
AFI 4CH Strain (%)	0.74	0.64	[-0.06, 0.88]	16	15
AFI Global Strain (%)	0.81	0.7	[-0.07, 0.91]	13	12
e' (cm/s)	0.9	0.76	[-0.20, 0.94]	19	18
MV E Vel (cm/s)	0.97	0.95	[0.75, 0.99]	18	17
E/e'	0.88	0.84	[0.45, 0.95]	17	16
MV E/A	0.93	0.92	[0.76, 0.97]	18	17
MV Dec T (ms)	0.86	0.82	[0.48, 0.94]	18	17

Table showing p-values of One Sample T-test of the intra- and inter-rater differences.

Methods	One Sample T-test for the inter-rater differences	One Sample T-test for the intra-rater differences
EF (%)	p= 0.02	p= 0.940
EF with contrast (%)	p= 0.816	p= 0.673
Auto EF (%)	p= 0.519	p= 0.258
3D EF (%)	p= 0.815	p= 0.788
S' (cm/s)	p= 0.002	p= 0.379
MAPSE (mm)	p= 0.000	p= 0.501
Generic Strain (%)	p= 0.000	p= 0.163
AFI 4CH Strain (%)	p= 0.004	p= 0.068
AFI Global strain (GLS) (%)	p= 0.006	p= 0.011
MVE Vel (cm/s)	p= 0.003	p= 0.623
e' (cm/s)	p= 0.000	p= 0.095
E/e'	p= 0.013	p= 0.731
MV E/A	p= 0.048	p= 0.440
MV Dec T (ms)	p= 0.025	p= 0.082

Table showing results from ANOVA significance test of inter-rater differences of CoV.

ANOVA	S'	MAPSE	AFI_GLS	E/e'	E/A	DecT	EF_uten	EF_med_kontrast	AutoEF	EF_3D	Strain_Gen	AFI_4K	MV_E	e'
S'		0.016	0.014	0.026	0.004	0.163	0.558	0.001	0.000	0.391	0.081	0.062	0.004	0.729
MAPSE			0.675	0.919	0.613	0.325	0.078	0.273	0.136	0.119	0.510	0.699	0.597	0.038
AFI_GLS				0.621	0.986	0.212	0.056	0.603	0.380	0.083	0.331	0.466	0.971	0.030
E/e'					0.558	0.397	0.110	0.249	0.125	0.162	0.594	0.781	0.544	0.059
E/A						0.147	0.028	0.564	0.327	0.044	0.254	0.397	0.983	0.012
DecT							0.436	0.043	0.017	0.583	0.740	0.591	0.141	0.292
EF_uten								0.006	0.002	0.804	0.265	0.203	0.028	0.803
EF_med_kontrast									0.674	0.010	0.085	0.163	0.578	0.002
AutoEF										0.003	0.037	0.079	0.337	0.001
EF_3D											0.372	0.286	0.042	0.609
Strain_Gen												0.819	0.245	0.161
AFI_4K													0.385	0.122
MV_E														0.011
e'														