

Reproducibility of Echocardiographic Techniques for Sequential Assessment of Left Ventricular Ejection Fraction and Volumes

Application to Patients Undergoing Cancer Chemotherapy

Paaladinesh Thavendiranathan, MD, MSc, Andrew D. Grant, MD, Tomoko Negishi, MD, Juan Carlos Plana, MD, Zoran B. Popović, MD, PhD, Thomas H. Marwick, MD, PhD, MPH
Cleveland, Ohio

- Objectives** The aim of this study was to identify the best echocardiographic method for sequential quantification of left ventricular (LV) ejection fraction (EF) and volumes in patients undergoing cancer chemotherapy.
- Background** Decisions regarding cancer therapy are based on temporal changes of EF. However the method for EF measurement with the lowest temporal variability is unknown.
- Methods** We selected patients in whom stable function in the face of chemotherapy for breast cancer was defined by stability of global longitudinal strain (GLS) at up to 5 time points (baseline, 3, 6, 9, and 12 months). In this way, changes in EF were considered to reflect temporal variability of measurements rather than cardiotoxicity. A comprehensive echocardiogram consisting of 2-dimensional (2D) and 3-dimensional (3D) acquisitions with and without contrast administration was performed at each time point. Stable LV function was defined as normal GLS ($\leq -16.0\%$) at each examination. The EF and volumes were measured with 2D-biplane Simpson's method, 2D-triplane, and 3-dimensional echocardiography (3DE) by 2 investigators blinded to any clinical data. Inter-, intra-, and test-retest variability were assessed in a subgroup. Variability was assessed by analysis of variance and compared with Levene's or *t* test.
- Results** Among 56 patients (all female, 54 ± 13 years of age), noncontrast 3D EF, end-diastolic volume, and end-systolic volume had significantly lower temporal variability than all other methods. Contrast only decreased the temporal variability of LV end-diastolic volume measurements by the 2D biplane method. Our data suggest that a temporal variability in EF of 0.06 might occur with noncontrast 3DE due to physiological differences and measurement variability, whereas this might be >0.10 with 2D methods. Overall, 3DE also had the best intra- and inter-observer as well as test-retest variability.
- Conclusions** Noncontrast 3DE was the most reproducible technique for LVEF and LV volume measurements over 1 year of follow-up. (J Am Coll Cardiol 2013;61:77-84) © 2013 by the American College of Cardiology Foundation

Sequential measurement of ejection fraction (EF) is used in a variety of conditions, perhaps most commonly in the assessment of potential cardiotoxicity from chemotherapy or immune therapy in patients with malignancies (1). Cardiotoxicity is most commonly defined as a reduction of the left ventricular (LV) EF of $\geq 5\%$ to $<55\%$ with symptoms of heart failure or an asymptomatic reduction of the LVEF of $\geq 10\%$ to $<55\%$ (2). Because decisions regarding cessation of lifesaving therapy (3) are based on

changes in EF values (4,5), it is important that EF measurement should not only be accurate but also have the lowest temporal variability such that a change in EF truly represents cardiotoxicity.

See page 85

Echocardiography remains the most common modality for EF measurements, but the method for EF measurement with the lowest temporal variability is unknown. Although 3-dimensional echocardiography (3DE) has been shown to be more accurate than 2-dimensional echocardiography (2DE) for both ventricular volume and EF measurements when compared with cardiac magnetic resonance imaging (6,7), no head-to-head comparison of 3DE with 2D meth-

From the Department of Cardiology, Heart and Vascular Institute, Cleveland Clinic, Cleveland, Ohio. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received June 20, 2012; revised manuscript received August 24, 2012, accepted September 4, 2012.

**Abbreviations
and Acronyms**

2DBi	= 2-dimensional biplane
2DE	= 2-dimensional echocardiography
2DTri	= 2-dimensional triplane
3DE	= 3-dimensional echocardiography
CI	= confidence interval
COV	= coefficient of variation
EDV	= end-diastolic volume
EF	= ejection fraction
ESV	= end-systolic volume
GLS	= average systolic global longitudinal strain
LV	= left ventricle/ventricular
SEM	= standard error of measurement

ods for temporal variability is available. Also, because 3DE acquisition and post-processing are still not routine, 2D methods including the biplane (2DBi) and triplane (2DTri) Simpson's methods are more commonly used in routine echocardiography. We sought to identify which of the commonly used echocardiographic techniques had the lowest temporal variability for EF and ventricular volumes, on the basis of multiple echocardiograms done over 1 year in women with breast cancer receiving chemotherapy who were clinically stable and had normal global longitudinal strain at each visit. The quantification methods compared include: 1) 2DBi; 2) 2DTri; and 3) 3D full volume acquisition. All methods were compared with and without LV opacification with contrast agents.

Methods

Patient selection. All patients referred to our echocardiography laboratory for LV function assessment before or during chemotherapy were prospectively enrolled in a database approved by the Institutional Review Board from between May 2010 and October 2011. We selected patients in whom stable function in the face of chemotherapy for breast cancer was defined by stability of global longitudinal strain (GLS) at up to 5 time points (baseline, 3, 6, 9, and 12 months). In this way, changes in EF were considered to reflect temporal variability of measurements rather than cardiotoxicity.

Inclusion criteria consisted of having had: 1) all echocardiography studies performed with Vivid 7 or E9 (GE Healthcare, Milwaukee, Wisconsin) ultrasound systems; 2) images for EF quantification along with high frame rate 2D acquisitions for speckle strain analysis at every follow-up; and 3) normal 2D systolic average global systolic longitudinal strain (GLS) at every follow-up study (defined as $GLS \leq -16\%$). Exclusion criteria consisted of: 1) a GLS value of $> -16\%$ at any time point; 2) baseline or interval diagnosis of coronary artery disease, other nonchemotherapy-induced cardiomyopathy, or more than moderate mitral or aortic regurgitation; 3) clinical diagnosis of heart failure during follow-up; and 4) initiation of cardiac medications such as beta-blockers during the follow-up period.

LV function by 2DE and 3DE. As part of the chemotherapy protocol, all patients receive a complete echocardiogram with and without contrast administration and myo-

cardial strain assessment. Apical 2- and 4-chamber and triplane acquisitions were obtained in each patient with optimization of image quality during acquisition. A 3D full volume dataset of the ventricle was obtained with gated (4 beats) acquisition with sector size, depth, and the number of heart beats optimized to obtain the highest possible volume rates. All acquisitions were then repeated after boluses of contrast administration (Definity, Bristol-Myers Squibb Medical Imaging, New York, New York, or Optison, GE Healthcare) with the mechanical index adjusted to between 0.15 and 0.3 and image settings optimized (8). All 2D post-contrast acquisitions were performed before 3D post-contrast acquisitions.

The LV volumes and EF were measured offline with EchoPAC (GE Healthcare) for all acquisitions except 3D with contrast where Tomtec (4D LV-Function 2.2.14.7, Unterschleissheim, Germany) analysis package was used. The 2DBi measurements were performed by manual contouring of the 4- and 2-chamber views, whereas 2DTri measurements required manual contouring of 3 long-axis views of the LV. The 3D noncontrast EF was measured with a semi-automated technique. Basal and apical guide points were placed in 1 long-axis image in the end-diastolic and end-systolic frames. Endocardial contours were then automatically generated and displayed on multiple long- and short-axis cine images for verification and manual adjustment as necessary. For the 3D EF with contrast, 3 long-axis images were manually contoured in end-diastole and end-systole. The contouring was then verified on long- and short-axis cine images and modified as necessary to ensure optimal endocardial tracking. The EF and volume measurements for each time point were measured blinded to the prior measurement.

LV systolic average GLS. Apical 3-, 2-, and 4-chamber high frame rate grayscale acquisitions (40 to 80 frames/s) were obtained with commercially available equipment (Vivid 7 or E9, GE Healthcare). Measurement of GLS was performed as part of the routine clinical study on the ultrasound system or offline with EchoPAC (GE Healthcare) and verified by a physician. GLS was measured for 3-, 2-, and 4-chamber views separately, and all 18 myocardial segments were averaged to obtain the GLS. Segments that were inadequately tracked were excluded from the analysis. Our strain cutoff -16% was based on calculation of approximately 2SD from the mean on the basis of known normal strain values from a previous publication (9).

Reproducibility analysis. Interobserver and intraobserver reproducibility and interobserver test-retest variability of the EF and volume measurements for all 6 methods were tested by 2 observers (P.T., A.G.) for 10 patients at 2 different time points (baseline and 3 months). Before performing these analyses the 2 observers first agreed on the method of contouring for all 6 methods (contrast and noncontrast studies) and practiced on 5 separate cases together. Each observer measured each study twice with each measurement performed at 2 separate time periods (2 to 4 weeks apart) in

a random manner to avoid any memory of measurement between time points. For interobserver and intraobserver variability a total of 20 studies (10 from baseline, and 10 from follow-up) were used. For the interobserver test-retest variability, measurement 1 and 2 for each patient by observer 1 was compared with measurements 1 and 2 at time point 2 by the second observer. The interobserver test-retest variability provides an estimate of the expected variability in EF and volumes in an echocardiography lab where the measurements might be performed by 2 different sonographers/physicians at different time points during the follow-up period. All measurements were made by each observer blinded to previous measurements and to the measurements of the other observer.

Statistical analysis. Continuous data are expressed as mean \pm SD, whereas categorical data are expressed as frequency or percentage. All absolute changes in EF values are represented as decimals (e.g., 10% absolute EF change is represented as 0.10), whereas any relative change in EF is represented as a percentage. Echocardiograms were categorized into 1 of 5 time periods: 1) before initiation of therapy or the first available study; 2) 2 to \leq 3 months; 3) 3 to \leq 6 months; 4) 6 to \leq 9 months; and 5) 9 to \leq 12 months after initiation of therapy or the first echocardiogram. All EF, volume, and strain measurements were normally distributed as determined by the Kolmogorov-Smirnov test. To assess whether there was a significant temporal variability in GLS in our population, we used Linear Mixed Model analysis with unstructured covariance for random effects (10) with strain as a dependent variable, patients as random effect, and time (in days) as a covariate. One-way analysis of variance (ANOVA) was performed to obtain the mean squared error for each technique studied. The square root provided the standard error of the measurement (SEM) in the EF and volume measurements for each method over the measurement period. The 95% confidence interval (CI) for each SEM was calculated with the formula:

$$\frac{(n-1)S^2}{\chi_U^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_L^2}$$

where χ_U^2 is the upper-tail value of chi-square for $df = n - 1$ with area $\alpha/2$ to its right and χ_L^2 is the lower-tail value with area $\alpha/2$ to its left. In addition to the SEM the coefficient of variation (COV) and 95% CI (calculated as described in the preceding text) for the temporal variability for each method was also calculated. Levene's test was used to test for differences in the SEM and COV between the techniques. Because 5 comparisons of SEM and COV were made (3D noncontrast technique vs. others) Bonferroni correction was used with $p < 0.01$ considered to be statistically significant. The minimal change in EF or volumes over time beyond which 2 measurements could be considered to be different was calculated as $2 \times$ the SEM for any technique.

Intraobserver and interobserver variability were determined with 2-way ANOVA approach described by Eliasziw et al. (11) with observers treated as random factors. With this method the interobserver variability includes both the variability among measurements of observers and the variability within measurements of observers (11). In addition, the interobserver test-retest variation was also calculated with 2-way ANOVA. This measure consists of variability within observers, among observers, and over time. For these 3 measurements, with the formula in the following text, the 95% CI was calculated as a measure of the minimal difference between 2 measurements beyond which the measures can be considered to be truly different (11–13).

$$\Delta EF_{\text{intra}} = z_{\alpha} \times \sqrt{2} \times \text{MSE}_{\text{intra}}$$

$$\Delta EF_{\text{inter}} = z_{\alpha} \times \sqrt{2} \times \text{MSE}_{\text{inter}}$$

The differences in the calculated observer variability were compared with the t test described by Tong et al (13). Because 5 comparisons were made (3D noncontrast vs. others) Bonferroni correction was used with a $p < 0.01$ considered to be statistically significant. All statistical analysis was performed with SPSS (version 19.0.0, SPSS, Chicago, Illinois).

Results

Patient population. A total of 88 patients with at least 3 echocardiograms were screened; 32 were excluded due to at least 1 abnormal strain value during follow-up ($> -16\%$) or development of other reasons for LV dysfunction (e.g., coronary artery disease and viral myocarditis) or initiation of beta-blockers during the follow-up period. Demographic data of the 56 included patients and cancer therapy received are summarized in Table 1. Patient follow-up ranged from a total of 3 visits 3 months apart to as many as 5 visits (baseline, 3, 6, 9, and 12 months).

Systolic GLS. The GLS values at each follow-up period are summarized in Table 1. A total of 39 of 4,248 segments ($< 1.0\%$) were not used in the strain calculation due to poor myocardial tracking. The average GLS for all patients and all time points was $-19.6 \pm 2.0\%$ and ranged from -16.0% to -24.9% . There was no significant temporal variability in strain comparing all patients for all available time points ($p = 0.79$).

Temporal variability of EF. The EF measurements were performed on 236 echocardiographic studies. However, due to occasional refusal of contrast agent, missed acquisition of 1 or more methods of EF assessment in some studies, or poor image quality, we were able to perform the following number of measurements: 235 2DBi, 186 2DBi with contrast, 225 2DTr, 177 2DTr with contrast, 215 3D, and 174 3D with contrast. The median (95% CI) frame/volume rates for 2DBi, 2DBi with contrast, 2DTr, 2DTr with contrast, 3D, and 3D with contrast were 74 (71 to 74),

Table 1 Patient Demographic Data (N = 56)	
Age, yrs	54 ± 13
Diabetic	6 (11)
HTN	17 (30)
Smoker	19 (34)
Hyperlipidemia	15 (27)
Chemotherapy type	
TCH	28 (50)
ACT + trastuzumab	15 (27)
Other	13 (23)
Mastectomy	52 (93)
Chest radiation	32 (57)
Strain, %	
Baseline (n = 56)	-19.6 ± 1.8
3 months (n = 56)	-19.6 ± 2.0
6 months (n = 56)	-19.4 ± 2.1
9 months (n = 45)	-19.5 ± 2.0
12 months (n = 23)	-19.9 ± 2.3

Values are mean ± SD or n (%).

ACT = adriamycin, cyclophosphamide, and paclitaxel + trastuzumab; HTN = hypertension; Other = various combinations including trastuzumab; TCH = Taxotere, Carboplatin, Herceptin (Trastuzumab).

71(64 to 74), 47 (46 to 47), 46 (45 to 46), 38 (35 to 41), and 30 (26 to 33), respectively. The EF by each technique for the entire follow-up period for all patients is shown in Figure 1, whereas the mean ± SD values for each follow-up time point are illustrated in Figure 2.

The temporal variability in EF measurements of each technique is illustrated in Figure 3, with numerical SEM and 95% CI values shown in the lower panel in Figure 3. Among the techniques, noncontrast 3DE had the lowest temporal variability compared with all the other EF methods. Contrast increased the temporal variability for 2DBi (p = 0.02), 2DTri (p = 0.17), and 3D (p < 0.001). However, among the 3 contrast-enhanced methods, 3D had the lowest temporal variability although only significantly different in comparison with the contrast 2DTri technique (p = 0.001) and a trend toward improvement in comparison with the 2DBi technique (p = 0.09). The temporal vari-

ability of EF measurements for each method is provided as COV in Online Table A.

Temporal variability of LV volume. The end-diastolic volume (EDV) and end-systolic volume (ESV) measurements by each technique for the entire follow-up period are illustrated in Figure 1, and the mean values at each time point are illustrated in Figure 2. The temporal variability in EDV and ESV for each technique is illustrated in Figure 4, with numerical SEM and 95% CI values shown in the lower panel in Figure 4. Noncontrast 3DE had the lowest temporal variability for both EDV and ESV and was significantly lower than all the other methods compared. Administration of contrast agents decreased the temporal variability in EDV measurements for the 2DBi method (p = 0.02) but increased it for the 3D method (p < 0.001). Contrast did not improve the ESV measurements for any of the techniques. Among the contrast-enhanced techniques when the 3D method was compared with the other 2 methods, there was no difference in the temporal variability for both EDV (p = 0.12 vs. 2DBi, p = 0.19 vs. 2DTri) and ESV (p = 0.20 vs. 2DBi, p = 0.31 vs. 2DTri). The temporal variability of volume measurements for each method is provided as COV in Online Table A.

Interobserver and intraobserver variability. The interobserver and intraobserver variability for all 6 methods expressed as SEM and 95% CI for both EF and volumes are summarized in Tables 2 and 3. For EF measurements, noncontrast 3DE had the lowest intraobserver and interobserver variability with minimal detectable change in EF of 0.048 and 0.075, respectively. Among the contrast-enhanced techniques 3DE had the lowest interobserver and intraobserver variability. For EDV and ESV the intraobserver variability was the lowest for the noncontrast 3D technique, with minimal detectable change of 14.3 and 8.1 ml, respectively. For interobserver variability, the 2DBi with contrast method showed marginally smaller variability compared with noncontrast 3D. For both the 2DBi and 2DTri techniques, contrast decreased the interobserver and intrao-

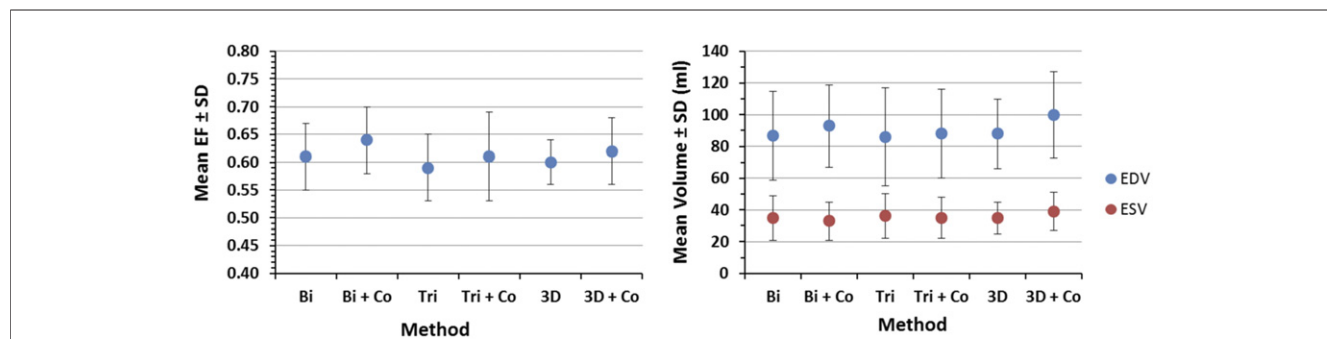


Figure 1. Mean EF and Volumes for Entire Follow-Up Period

For each method compared, the mean ejection fraction (EF), end-diastolic volume (EDV), and end-systolic volume (ESV) for all patients at all available time points are illustrated. Error bars illustrate 1 SD. 3D = 3-dimensional; Bi = biplane; Co = contrast; Tri = triplane.

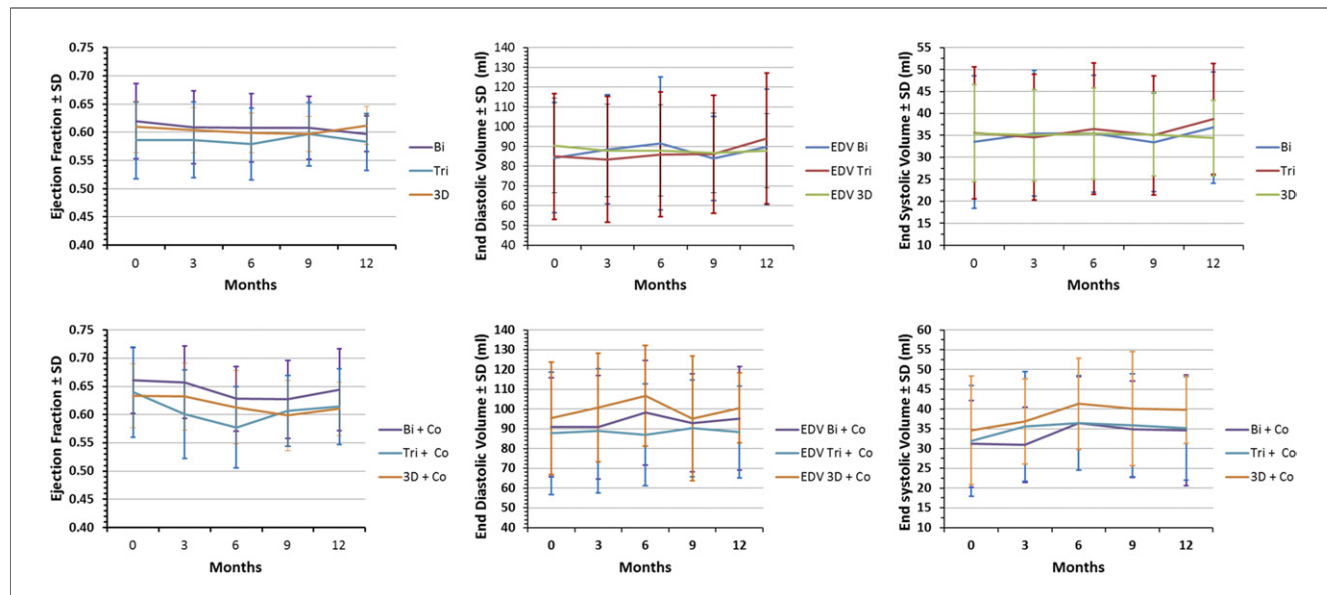


Figure 2 Mean EF and Volumes for Each Time Point

For each method compared, the EF, EDV, and ESV at each time point is the mean for all patients with available data at that time point. Error bars illustrate 1 SD. Abbreviations as in Figure 1.

observer variability, but this was not seen with the 3D technique.

Test-retest variability. The overall interobserver test-retest variability for EF and volumes for each method is also summarized in Tables 2 and 3, respectively. For EF and ESV measurements, noncontrast 3DE had the lowest variability, with minimal detectable EF and ESV differences of 0.060 and 14 ml, respectively. For EDV measurement, 3D with contrast had the lowest variability followed by 3D

without contrast. The minimal detectable changes for EDV with these methods were 30.1 and 34.8 ml, respectively.

Discussion

This study provides a comparison of temporal and observer variability in EF and volume measurements with commonly used echocardiographic methods in patients with stable ventricular systolic function undergoing chemotherapy for

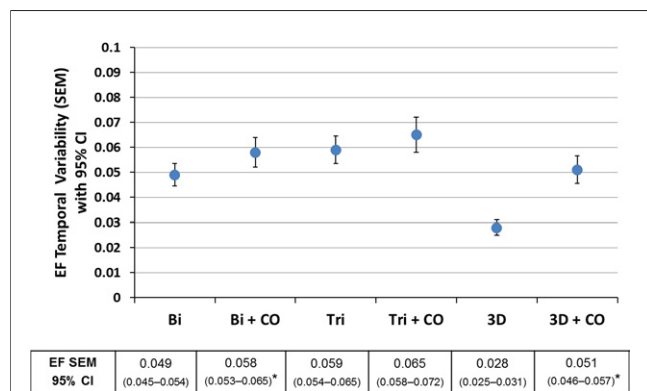


Figure 3 Temporal Variability in EF

The temporal variability is defined as the standard error of measurement (SEM) and 95% confidence intervals (CIs) for each technique for the entire follow-up period. Noncontrast 3D had the lowest temporal variability and 95% CI for EF measurements (lower panel) compared with all methods ($p < 0.01$ for all comparisons against noncontrast 3D). * $p < 0.05$ for comparison of contrast enhanced to noncontrast acquisition for the respective technique. Abbreviations as in Figure 1.

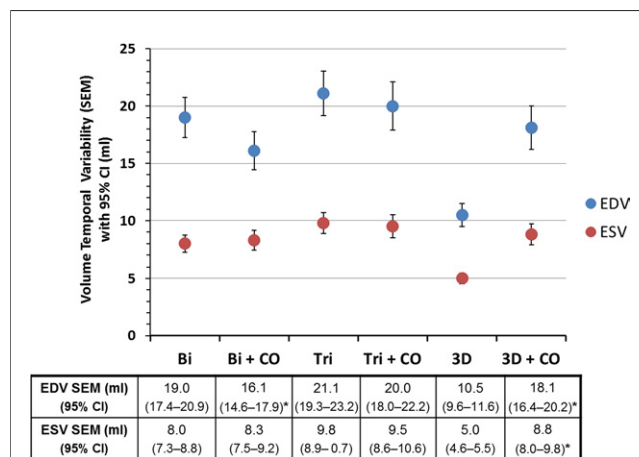


Figure 4 Temporal Variability in EDV and ESV

Temporal variability is defined as the SEM and 95% CIs for each technique for the entire follow-up period. Noncontrast 3D had the lowest temporal variability and 95% CI for EDV and ESV measurements (lower panel) compared with all methods ($p < 0.01$ for all comparisons against noncontrast 3D). * $p < 0.05$ for comparison of contrast enhanced to noncontrast acquisition for the respective technique. Abbreviations as in Figures 1 and 3.

Table 2 Interobserver and Intraobserver Variability and Minimal Detectable Change for EF Measurements by All 6 Techniques

	Bi	Bi + Co	Tri	Tri + Co	3D	3D + Co
Intraobserver	0.033*	0.035*	0.038*	0.037*	0.017	0.026
Min Δ detectable	0.090	0.098	0.104	0.102	0.048	0.072
Interobserver	0.040	0.051*	0.049*	0.048*	0.027	0.038
Min Δ detectable	0.111	0.142	0.135	0.133	0.075	0.100
Interobserver test-retest	0.047*	0.055*	0.058*	0.069*	0.022	0.042*
Min Δ detectable	0.013	0.152	0.162	0.192	0.060	0.115

Noncontrast 3D had the lowest intraobserver and interobserver test-retest observer variability and the smallest minimal detectable change. An ejection fraction (EF) of 0.033 corresponds to 3.3%. *p < 0.01 t test compared with noncontrast 3D.

Bi = biplane Simpson's; BP + Co = biplane Simpson's with contrast; Tri = triplane; Tri + Co = triplane + contrast; 3D = 3-dimensional; 3D + Co = 3-dimensional with contrast.

breast cancer. In addition, this work also provides an estimate of the minimal change in EF and volumes above which 2 measurements should be considered to significantly differ in excess of temporal and observer variability. Non-contrast 3DE was most consistently superior for temporal and observer variability for both EF and volume measurements. Contrast administration did not improve variability in measurements of EF or volumes in this population. Our data indicate that a change in EF as high as 0.05 to 0.06 can be seen over time with noncontrast 3D EF, due physiological changes, acquisition differences, and observer variability. In comparison this variability can be close to 0.10 to 0.13 with 2D techniques.

Sequential follow-up by echocardiography. In patients with cancer, multiple studies are repeated during chemotherapy to monitor changes in EF as a marker of cardiotoxicity, and this and other indications for repeat testing are considered appropriate practice (14). However, the wide CIs with 2DE reported in previous studies raises concern about erroneously stopping chemotherapy due to changes in EF that occur only due to variability on repeat testing. Indeed, the utility of 2DE for sequential follow-up of patients for LV volumes and function has been disappointing, with 11% being the smallest change in EF that can be recognized with 95% confidence (15). A significant proportion of this variability has been attributed to test-retest variability (15).

The recommendation to perform sequential echocardiograms in these patients would be strengthened if measures of LV function were sufficiently robust to more faithfully

reflect variations in physiological differences between the repeated studies. Cross-sectional studies have shown 3DE to have less variation between studies than 2DE in comparison with magnetic resonance imaging (6,7). Only 1 study has illustrated more reproducible LV measurements with 3DE on temporal follow-up on the basis of 2 time points in patients with coronary artery disease (16). To our knowledge, this is the first study to compare 2D and 3D methods with and without contrast administration in a longitudinal study with multiple repeated echocardiograms to determine the method with the lowest temporal variability.

In this study, controlling for nonphysiological changes in EF and volumes with strain and clinical parameters, non-contrast 3DE was superior to 2DE methods and contrast 3DE with respect to temporal variability in EF and volumes. The superiority compared with 2DE methods is a reflection that 3DE does not make any geometric assumptions for EF and volume measurements and is less affected by acquisition differences from 1 scan to the next, as often seen with 2DE (7,17). In addition noncontrast 3DE volumes and EF were measured with a semi-automated method where automatically generated contours are only modified if necessary. This is in contrast to manual endocardial contouring in all other methods where there can be significant differences in the interpretation of endocardial borders. The improved reproducibility of semi-automated versus manual contouring has been previously illustrated by others both with 2DE and 3DE (18,19). Specifically, this latter reason likely accounts for the higher variability of

Table 3 Interobserver and Intraobserver Variability for EDV and ESV Measurements and Minimal Detectable Change by All 6 Techniques

	Bi		Bi + Co		Tri		Tri + Co		3D		3D + Co	
	EDV	ESV	EDV	ESV	EDV	ESV	EDV	ESV	EDV	ESV	EDV	ESV
Intraobserver	8.4	5.3*	6.5	3.7	7.8	3.8	5.7	3.8	5.1	2.9	6.1	4.7
Min Δ detectable	23.2	14.7	18.0	10.4	21.5	10.5	15.9	10.4	14.3	8.1	16.8	13.0
Interobserver	16.5†	7.4*	8.7	4.1	15.1†	9.2*	10.2	5.2	9.0	4.4	10.0	4.9
Min Δ detectable	45.7	22.9	24.2	11.4	41.9	25.4	28.3	14.5	25.0	12.2	27.6	13.5
Test re-test	21.4†	10.3	13.8	6.9	13.8	8.7	12.6	6.9	12.6	5.0	10.9	6.9
Min Δ detectable	59.3	28.8	37.6	19.2	38.3	24.1	35.0	19.0	34.8	13.9	30.1	19.1

Interobserver and intraobserver variability for end-diastolic volume (EDV) and end-systolic volume (ESV) measurements (ml) and minimal detectable change (ml) by all 6 techniques. *p < 0.01 t test compared with noncontrast 3D ESV; †p < 0.01 t test compared with noncontrast 3D EDV.

Abbreviations as in Table 2.

contrast 3DE compared with noncontrast 3DE in this study. Additional reasons for lack of improvement with contrast 3DE include lower volume rates of acquisitions (i.e., 30 vs. 38 VPS) as well as limitations of the software used for contrast 3DE analysis with respect to freedom in endocardial contour adjustment.

Contrast administration has been shown to improve accuracy and observer reproducibility with 2DE in patients with sub-optimal echocardiographic windows (20,21). However, other studies have shown the lack of additional benefit over harmonic imaging with only limited improvement or even worsening in reproducibility of EF measurements in some patient populations (22–24). The latter might be attributed to blooming and attenuation artifacts that hinder delineation of structures such as the mitral valve, resulting in variability in contouring (23). With 3DE, previous work suggested improved accuracy with contrast administration, even with good quality images (7). However, the lack of improvement in temporal variability in EF or volume measurements with contrast administration except for 2DBi EDV likely reflects the population studied. With harmonic imaging most of our patients had adequate acoustic windows and would not have met traditional criteria for contrast administration. This is also supported by the fact that we were able to analyze 235/236 2D images for 2DBi EF calculations. Unfortunately we were unable to identify a subgroup of patients in our study cohort who had at least 3 consecutive studies that would meet criteria for contrast administration to perform a subgroup analysis.

In our study, the 2DTri method did not perform better than the 2DBi method, for several reasons. The triplane method was more susceptible to off-axis views of the ventricle due to the difficulty in optimizing 3 separate views in a single acquisition. Second, the endocardial border delineation of the triplane images was inferior to the 2D images used for biplane calculations. Also, the frame rates for the triplane acquisition (47/s) were significantly lower than those of 2D images used for biplane measurements (74/s).

Interobserver and intraobserver variability. Previous studies have also shown 3DE to have lower interobserver and intraobserver variability probability, due to some of the factors discussed in the preceding text (6,16). Our findings are congruent with these studies. Another important variability is that seen between 2 studies at different time points with measurements performed by different individuals (interobserver test-retest variability). This is of direct clinical importance, because patients are likely imaged by different sonographers at different time points and the post-processing is performed by different individuals. In this study 3DE had the lowest interobserver test-retest variability with minimal detectable difference in EF of 0.060. This was significantly higher for the 2D techniques.

Study limitations. The average systolic GLS and clinical criteria were used as a marker of stability of ventricular function in these patients. Although, 2D speckle tracking-

based strain measurement has limitations, the usefulness of strain measurements as a marker of subclinical ventricular dysfunction has been illustrated in several studies (25), including in patients receiving chemotherapy (26,27). Alternative study designs, including using sequential reference measurements with cardiac magnetic resonance, were considered less feasible.

We were unable to obtain contrast enhanced images and all 6 EF measurement methods of all patients at all time points. This was due to the refusal of some patients to have IV access or the inability to obtain IV access during echocardiography or patient discomfort necessitating selective image acquisition. Likewise, although our data are based on a small patient population, the critical variable is the total number of echocardiographic studies.

All our EF and volume acquisitions and post-processing were performed with specific vendor equipment and software. Therefore our findings cannot be directly extrapolated to other 3D acquisition and post-processing methods. However, we do not feel that the longitudinal reproducibility data are reflective of a single vendor but rather the semi-automated 3D measurements. Finally, this study seeks to define reliability and reproducibility rather than to validate the agreement or the diagnostic accuracy of EF and volume measurements against an external reference standard. The accuracy of 2D and 3D methods has been well-described (7).

Conclusions

The results of this study confirm that the 95% CI for 2D EF (approximately 0.10 variability) is analogous to the >0.10 change in EF that defines cardiotoxicity in asymptomatic patients (2). That noncontrast 3D EF measurements have lower temporal variability suggesting that this approach should be considered in patients with good acoustic windows. To achieve this degree of temporal variability with 3DE, an automated or semi-automated endocardial contouring method should be used. Although these data are based on patients receiving chemotherapy, our findings have broader implications for patients in whom clinical decisions or prognostic markers are based on changes in EF and ventricular volumes, including patients with valvular diseases, heart failure, and those undergoing surgical or percutaneous revascularization.

Reprint requests and correspondence: Dr. Thomas H. Marwick, Menzies Research Institute of Tasmania, 17 Liverpool Street, Hobart, TAS 7000, Australia. E-mail: tom.marwick@utas.edu.au.

REFERENCES

1. Yeh ET, Bickford CL. Cardiovascular complications of cancer therapy: incidence, pathogenesis, diagnosis, and management. *J Am Coll Cardiol* 2009;53:2231–47.
2. Martin M, Esteva FJ, Alba E, et al. Minimizing cardiotoxicity while optimizing treatment efficacy with trastuzumab: review and expert recommendations. *Oncologist* 2009;14:1–11.

- Moja L, Tagliabue L, Balduzzi S, et al. Trastuzumab containing regimens for early breast cancer. *Cochrane Database Syst Rev* 2012; 4:CD006243.
- Ewer MS, Ewer SM. Cardiotoxicity of anticancer treatments: what the cardiologist needs to know. *Nat Rev Cardiol* 2010;7:564–75.
- Verma S, Ewer MS. Is cardiotoxicity being adequately assessed in current trials of cytotoxic and targeted agents in breast cancer? *Ann Oncol* 2011;22:1011–8.
- Jenkins C, Chan J, Hanekom L, Marwick TH. Accuracy and feasibility of online 3-dimensional echocardiography for measurement of left ventricular parameters. *J Am Soc Echocardiogr* 2006;19:1119–28.
- 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;30:98–106.
- Mulvagh SL, Rakowski H, Vannan MA, et al. American Society of Echocardiography Consensus Statement on the Clinical Applications of Ultrasonic Contrast Agents in Echocardiography. *J Am Soc Echocardiogr* 2008;21:1179–201; quiz 1281.
- Marwick TH, Leano RL, Brown J, et al. Myocardial strain measurement with 2-dimensional speckle-tracking echocardiography: definition of normal range. *J Am Coll Cardiol Img* 2009;2:80–4.
- Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley and Sons, 2004.
- Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther* 1994;74:777–88.
- Yuan XP, Bach D, Skanes A, Drangova M. Assessment of intra- and interobserver variability of pulmonary vein measurements from CT angiography. *Acad Radiol* 2004;11:1211–8.
- Tong S, Cardinal HN, McLoughlin RF, Downey DB, Fenster A. Intra- and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging. *Ultrasound Med Biol* 1998;24:673–81.
- Douglas PS, Garcia MJ, Haines DE, et al. ACCF/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 appropriate use criteria for echocardiography. A report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance Endorsed by American College of Chest Physicians. *J Am Coll Cardiol* 2011;57:1126–66.
- Otterstad JE. Measuring left ventricular volume and ejection fraction with the biplane Simpson's method. *Heart* 2002;88:559–60.
- Jenkins C, Bricknell K, Chan J, Hanekom L, Marwick TH. Comparison of two- and three-dimensional echocardiography with sequential magnetic resonance imaging for evaluating left ventricular volume and ejection fraction over time in patients with healed myocardial infarction. *Am J Cardiol* 2007;99:300–6.
- King DL, Harrison MR, King DL Jr., Gopal AS, Kwan OL, DeMaria AN. Ultrasound beam orientation during standard two-dimensional imaging: assessment by three-dimensional echocardiography. *J Am Soc Echocardiogr* 1992;5:569–76.
- Cannesson M, Tanabe M, Suffoletto MS, et al. A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. *J Am Coll Cardiol* 2007;49:217–26.
- Muraru D, Badano LP, Piccoli G, et al. Validation of a novel automated border-detection algorithm for rapid and accurate quantitation of left ventricular volumes based on three-dimensional echocardiography. *Eur J Echocardiogr* 2010;11:359–68.
- Yu EH, Sloggett CE, Iwanochko RM, Rakowski H, Siu SC. Feasibility and accuracy of left ventricular volumes and ejection fraction determination by fundamental, tissue harmonic, and intravenous contrast imaging in difficult-to-image patients. *J Am Soc Echocardiogr* 2000;13:216–24.
- Thomson HL, Basmadjian AJ, Rainbird AJ, et al. Contrast echocardiography improves the accuracy and reproducibility of left ventricular remodeling measurements: a prospective, randomly assigned, blinded study. *J Am Coll Cardiol* 2001;38:867–75.
- Whalley GA, Gamble GD, Walsh HJ, et al. Effect of tissue harmonic imaging and contrast upon between observer and test-retest reproducibility of left ventricular ejection fraction measurement in patients with heart failure. *Eur J Heart Fail* 2004;6:85–93.
- Kasprzak JD, Paelinck B, Ten Cate FJ, et al. Comparison of native and contrast-enhanced harmonic echocardiography for visualization of left ventricular endocardial border. *Am J Cardiol* 1999;83:211–7.
- Main ML, Asher CR, Rubin DN, et al. Comparison of tissue harmonic imaging with contrast (sonicated albumin echocardiography) and Doppler myocardial imaging for enhancing endocardial border resolution. *Am J Cardiol* 1999;83:218–22.
- Dedobbeleer C, Rai M, Donal E, Pandolfo M, Unger P. Normal left ventricular ejection fraction and mass but subclinical myocardial dysfunction in patients with Friedreich's ataxia. *Eur Heart J Cardiovasc Imaging* 2012;13:346–52.
- Hare JL, Brown JK, Leano R, Jenkins C, Woodward N, Marwick TH. Use of myocardial deformation imaging to detect preclinical myocardial dysfunction before conventional measures in patients undergoing breast cancer treatment with trastuzumab. *Am Heart J* 2009;158:294–301.
- Sawaya H, Sebag IA, Plana JC, et al. Early detection and prediction of cardiotoxicity in chemotherapy-treated patients. *Am J Cardiol* 2011; 107:1375–80.

Key Words: 3D echocardiography ■ chemotherapy ■ interobserver test re-test variability ■ interobserver variability ■ longitudinal variability.

 APPENDIX

For supplementary tables and text, please see the online version of this article.