

# Echocardiography Core Laboratory Reproducibility of Cardiac Safety Assessments in Cardio-Oncology



Michel G. Khouri, MD, Bonnie Ky, MD, MSCE, Gary Dunn, RDCS, Ted Plappert, CVT,  
Virginia Englefield, RDCS, Dawn Rabineau, RDCS, Eric Yow, MS, Huiman X. Barnhart, PhD,  
Martin St. John Sutton, MBBS, and Pamela S. Douglas, MD, *Durham, North Carolina;  
and Philadelphia, Pennsylvania*

**Background:** As the potential for cancer therapy–related cardiac dysfunction is increasingly recognized, there is a need for the standardization of echocardiographic measurements and cut points to guide treatment. The aim of this study was to determine the reproducibility of cardiac safety assessments across two academic echocardiography core laboratories (ECLs) at the University of Pennsylvania and the Duke Clinical Research Institute.

**Methods:** To harmonize the application of guideline-recommended measurement conventions, the ECLs conducted multiple training sessions to align measurement practices for traditional and emerging assessments of left ventricular (LV) function. Subsequently, 25 echocardiograms taken from patients with breast cancer treated with doxorubicin with or without trastuzumab were independently analyzed by each laboratory. Agreement was determined by the proportion (coverage probability [CP]) of all pairwise comparisons between readers that were within a prespecified minimum acceptable difference. Persistent differences in measurement techniques between laboratories triggered retraining and reassessment of reproducibility.

**Results:** There was robust reproducibility within each ECL but differences between ECLs on calculated LV ejection fraction and mitral inflow velocities (all CPs < 0.80); four-chamber global longitudinal strain bordered acceptable reproducibility (CP = 0.805). Calculated LV ejection fraction and four-chamber global longitudinal strain were sensitive to small but systematic interlaboratory differences in endocardial border definition that influenced measured LV volumes and the speckle-tracking region of interest, respectively. On repeat analyses, reproducibility for mitral velocities (CP = 0.940–0.990) was improved after incorporating multiple-beat measurements and homogeneous image selection. Reproducibility for four-chamber global longitudinal strain was unchanged after efforts to develop consensus between ECLs on endocardial border determinations were limited primarily by a lack of established reference standards.

**Conclusions:** High-quality quantitative echocardiographic research is feasible but requires a commitment to reproducibility, adherence to guideline recommendations, and the time, care, and attention to detail to establish agreement on measurement conventions. These findings have important implications for research design and clinical care. (J Am Soc Echocardiogr 2018;31:361–71.)

**Keywords:** Cancer, Cardiotoxicity, Echocardiography, Reliability, Reproducibility

As cardiac morbidity in patients with cancer is increasingly recognized, accurate diagnostic tools are critical to identify patients at risk for cancer therapy–related cardiac dysfunction (CTRCD).<sup>1</sup> Echocardiography provides essential structural, functional, and hemodynamic insights into cardiac

pathophysiology and, as a low-cost, widely available, and safe test is frequently used to assess the cardiac consequences of cancer and cancer therapy.<sup>2</sup> However, variability related to imaging quality, biologic variation, and interpretive differences can limit the reliability of

From the Division of Cardiology, Department of Medicine, Duke University Medical Center (M.G.K., P.S.D.), and the Duke Clinical Research Institute, Durham, North Carolina (M.G.K., G.D., D.R., E.Y., H.X.B., P.S.D.); and the Division of Cardiology, Department of Medicine (B.K., T.P., V.E., M.S.S.), Penn Cardiovascular Institute (B.K., M.S.S.), and Center for Clinical Epidemiology and Biostatistics (B.K.), Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania.

Conflicts of Interest: None.

The Penn and Duke Clinical Research Institute echocardiography core laboratories received funding from the National Cancer Institute, Division of Cancer Prevention, for work related to this study (grants U10CA45809 and U10CA81920).

Reprint requests: Michel G. Khouri, MD, Box 3254 DUMC, Durham, NC 27710 (E-mail: [michel.khouri@dm.duke.edu](mailto:michel.khouri@dm.duke.edu)).

0894-7317/\$36.00

Copyright 2017 by the American Society of Echocardiography.

<https://doi.org/10.1016/j.echo.2017.11.018>

**Abbreviations**

<b>2D</b> = Two-dimensional
<b>CP</b> = Coverage probability
<b>CTRCD</b> = Cancer therapy–related cardiac dysfunction
<b>DCRI</b> = Duke Clinical Research Institute
<b>ECL</b> = Echocardiography core laboratory
<b>GLS</b> = Global longitudinal strain
<b>GLS<sub>4CH</sub></b> = Apical four-chamber global longitudinal strain
<b>LV</b> = Left ventricular
<b>LVEF</b> = Left ventricular ejection fraction
<b>ROI</b> = Region of interest

echocardiographic results.<sup>3</sup> In clinical practice, variability of conventional echocardiographic parameters of left ventricular (LV) function (i.e., LV ejection fraction [LVEF]) can extend across treatment eligibility thresholds and affect critical decisions regarding cancer therapy.<sup>4</sup> In cancer trials, clinical LVEF data from site echocardiography laboratories are often used to determine study eligibility and evaluate the cardiac consequences of novel cancer therapies,<sup>5,6</sup> although reproducibility of results across clinical sites is rarely reported. Echocardiographic parameters that assess cardiac mechanics, including myocardial deformation and ventricular-arterial coupling, may improve sensitivity for early CTRCD beyond LVEF.<sup>2</sup> However, data in patients with cancer are predom-

inantly derived from single-center studies<sup>7</sup> that may not account for factors recognized to potentially diminish measurement reproducibility.<sup>8</sup>

As cardio-oncology progresses toward larger clinical trials, standard echocardiographic measurements and thresholds for treatment, with validated reproducibility of such measures, are vital. Multicenter cardiovascular clinical trials generally use echocardiography core laboratories (ECLs) to provide expertise and consistency for image acquisition and measurements as well as for assessments of imaging eligibility criteria and safety end points. In this regard, ECLs can reduce variability of imaging data and ensure the validity of study results.<sup>9–12</sup> Cardio-oncology studies have used ECLs,<sup>13</sup> but the practice is not widespread.

Against this background, the National Cancer Institute Division of Cancer Prevention awarded substudies of the PREDICT MDA 2007 0914 (ClinicalTrials.gov identifier NCT01032278) and SCUSF 0806 (ClinicalTrials.gov identifier NCT01009918) trials for the central review of echocardiograms to ECLs at the University of Pennsylvania (Penn) and the Duke Clinical Research Institute (DCRI), respectively. As a condition of the awards, the ECLs were instructed to collaborate with the potential goal of pooling echocardiographic data from the trials. To determine the feasibility of pooling the data, as well as the impact of central echocardiography review in cardio-oncology clinical trials, the ECLs at Penn and DCRI aimed to (1) determine the reproducibility of echocardiographic assessments in cardio-oncology within and across two academic ECLs, (2) identify sources of variability and corrective solutions, and (3) propose recommendations for echocardiographic research in the detection and monitoring of CTRCD, with potential implications for clinical care.

**METHODS****Penn and DCRI ECL Group Reads**

To align data collection elements, the Penn and DCRI ECLs reviewed two-dimensional (2D) and Doppler echocardiographic parameters of cardiac size and systolic and diastolic function relevant to clinical cardio-

vascular outcomes in patients with cancer.<sup>13,14</sup> A harmonization process between ECLs ensued. Sonographers and principal investigators at both ECLs conducted serial calls and multiple Web-conference group reads from October 2013 to March 2014 to share standard operating procedures, review sample echocardiograms, and align ECL perspectives on image quality, border selection, and tracing conventions. Group readings illustrated differences between ECLs on tracing conventions for certain parameters. These included nonconsensus on the angle of the minor axis for LV internal dimensions (e.g., parallel to mitral valve plane vs perpendicular to the LV long axis) and LV endocardial border definitions (e.g., depth of exclusion of trabeculae) during measurements of LV internal dimensions (Figure 1) and volumes (Figure 2), respectively. Harmonization efforts were aimed at achieving consensus on the application of measurement conventions outlined in national guidelines recommendations<sup>2,15,16</sup> and culminated in the development of consensus reading instructions (please see the Online Appendix, available at [www.onlinejase.com](http://www.onlinejase.com)) as well as a common comprehensive case report form across ECLs.<sup>17–20</sup>

**Echocardiographic Acquisition and Creation of Analysis Repository**

After developing consensus reading instructions, each ECL contributed echocardiograms for reproducibility analyses. A total of 25 patient echocardiograms were selected from transthoracic echocardiograms previously acquired at both institutions from patients who had completed treatment with potentially cardiotoxic anticancer agents (i.e., doxorubicin with or without trastuzumab) for breast cancer. More detailed clinical data were not made available, as each ECL was blinded to patient characteristics. Selected echocardiograms were required to have visible LV endocardium unobscured by undergaining or artifact and no significant apical foreshortening in 2D acquisitions. Echocardiograms were obtained by dedicated sonographer teams in the Intersocietal Accreditation Commission clinical laboratories at both institutions. All images were acquired using Vivid 7 or E9 machines (GE Healthcare, Milwaukee, WI) at 60 to 90 frames/sec and digitally archived at the acquisition frame rate. Digital echocardiographic images were deidentified and transferred in standard Digital Imaging and Communications in Medicine format to TomTec (TomTec Imaging Systems, Unterschleissheim, Germany) and DigiView (Digisonics, Houston, TX) analysis workstations at the Penn and DCRI ECLs, respectively. Both ECLs used TomTec 2D Cardiac Performance Analysis version 1.1 for strain analysis.

**Measurement of Echocardiography Parameters**

After image transfer, measurements of echocardiography parameters were assigned to two readers at each ECL ( $n = 4$  total readers). Readers included three highly experienced research sonographers and a cardiologist with level III certification in echocardiography. Each reader independently analyzed two uniquely identified copies of the 25 patient echocardiograms and recorded 50 measurement results per analyzed parameter. Each result was treated independently (i.e., no averaging within or between readers).

At laboratory A, LV volumes and strain were measured by a single reader; Doppler parameters (i.e., velocities and timing intervals) were measured by a separate reader, according to reader expertise and existent laboratory practices. At laboratory B, each reader measured every parameter. The measurement results generated per echocardiogram by the readers in each ECL are depicted in more detail in Figure 3A.

## HIGHLIGHTS

- Standardizing echocardiography measurements that guide treatment decisions in cancer is needed.
- The authors assessed the reproducibility of echocardiographic measurements in cancer across two echocardiography core laboratories.
- There was robust reproducibility within each laboratory but differences between laboratories.
- Interlaboratory reproducibility for LVEF and GLS was less concordant.
- Comparing small changes in LVEF and GLS across laboratories should be done cautiously.

### Intra- and Interlaboratory Reproducibility Testing

Intra- and interlaboratory reproducibility was evaluated by pairwise comparisons of all possible measurement differences (interpretative variability) among all readers for selected echocardiographic parameters, as described below.<sup>21</sup> For each of the 25 patient echocardiograms, there were four pairwise comparisons of measurement results between the two readers at laboratory B (Figure 3B) and eight pairwise comparisons of measurement results between the reader at laboratory A and two readers at laboratory B (Figure 3C) on any selected parameter.

The “acceptable difference,” or limit of measurement variability, for each parameter served as a benchmark for reproducibility. The intra- and interreader acceptable differences were defined prospectively on the basis of literature review of studies with reproducibility data for LV volumes and LVEF,<sup>22</sup> mitral inflow velocities,<sup>23,24</sup> global longitudinal strain (GLS),<sup>25</sup> and R-wave flow onset and flow end times.<sup>26</sup> All acceptable differences are expressed in absolute terms (Table 1).

After readers completed the echocardiographic interpretations, result reports were generated, and intra- and interreader pairwise comparisons were displayed in tables and illustrated in dot-plot graphs. Paired measurements with differences exceeding the prespecified acceptable differences (outliers) were reviewed in a group read. If reproducibility between readers was not acceptable for any echocardiographic parameter(s), a process of revisiting image analysis instructions, retraining and retesting ensued.

### Retraining and Retesting

Retraining involved Web conference–based sessions among all readers. Review of individual statistical results guided the retraining process by illustrating the extent and direction of outlier pairs. Open discussion among readers helped identify the source(s) of interpretation and/or measurement error(s). Illustrative case examples and the consensus reading instructions were discussed and revised to promote uniformity of interpretation and help eliminate individual idiosyncrasies. After retraining, all readers reinterpreted the 25 echocardiograms twice for selected parameters with unacceptable reproducibility on initial testing, and reproducibility was reevaluated.

### Statistical Analysis

Specific standards or a universally preferred index does not exist for assessing intrareader and interreader reproducibility in ECLs; therefore, multiple statistical approaches for assessing reproducibility

were considered.<sup>22</sup> The coverage probability (CP) method was selected given its desirable characteristics as an agreement index in the ECL setting, including its computational simplicity, rapid identification of group and individual reader variability and specific disagreements for review and retraining, and broad applicability to patient populations and continuous and categorical variables.<sup>21</sup> CP is the probability that the difference between any two measurements of a parameter on the same echocardiogram are within a prespecified acceptable difference.<sup>21</sup> Specifically, all possible interreader comparisons were examined to determine whether the measurement difference between paired readers was within the prespecified acceptable difference. The nonparametric approach was used to obtain the estimate of CP. The estimate of CP is the proportion of the number of pairwise interreader comparisons within the acceptable difference divided by the number of all possible pairwise comparisons. The higher the CP, the better the reproducibility. Perfect reproducibility corresponded to 100% CP (i.e.,  $CP = 1.00$ ), indicating that all measurements were within the prespecified acceptable difference for that parameter.

The prespecified acceptable differences and the cut point for the CP determine the standard for acceptable reproducibility. In this study, the cut point for the estimated CP was set at 0.80 on the basis of previous studies assessing reproducibility in an ECL setting.<sup>22,27</sup> Reproducibility was considered acceptable if  $\geq 80\%$  of all possible pairwise comparisons were within the prespecified acceptable difference for each parameter. Additionally, all pairwise comparisons (100% or  $CP = 1.00$ ) were required to be within twice the acceptable difference. To easily visualize the reproducibility data and rapidly identify outlier pairs, the results of the CP analysis were displayed graphically for continuous parameters. Statistical analyses were performed using SAS version 9.2 (SAS Institute, Cary, NC).

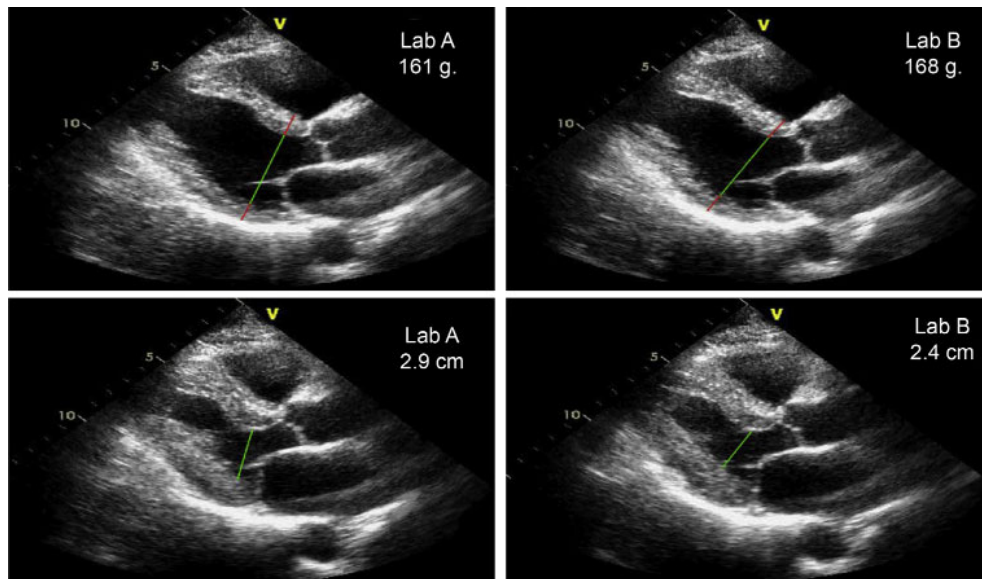
## RESULTS

On the initial testing, there was robust reproducibility within each ECL. Intrareader reproducibility was acceptable ( $CP \geq 0.80$ ) for the a priori defined acceptable difference for all parameters on the basis of analysis of 225 pairwise comparisons (25 for each of the nine parameters) for each reader. All readers achieved a CP of 1.00 for twice the acceptable difference for all parameters except one reader for mitral valve A peak velocity ( $CP = 0.960$ ). Interreader reproducibility between two readers at laboratory B demonstrated acceptable reproducibility for all parameters except mitral E velocity ( $CP = 0.770$ ) and mitral A velocity ( $CP = 0.770$ ) on the basis of analysis of 900 pairwise comparisons (100 for each of nine parameters). Reproducibility performance results within each ECL (i.e., intrareader CPs for all readers at both ECLs and interreader CPs at laboratory B) are shown in Table 2.

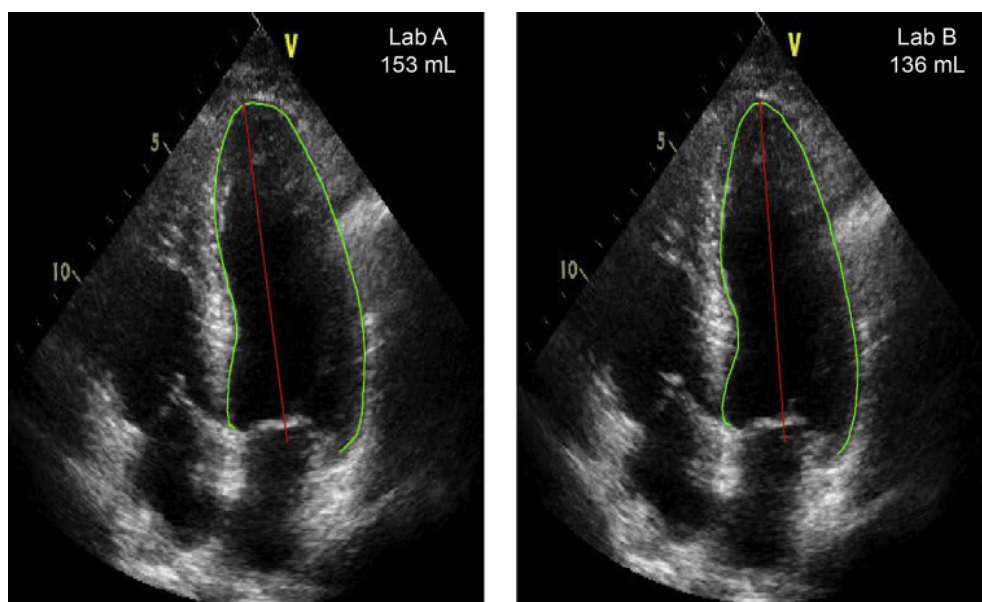
Reproducibility performance results between ECLs are shown in Table 3. Analysis of 1,800 pairwise comparisons (200 for each of nine parameters) among three readers (for any given parameter) across both ECLs demonstrated acceptable interlaboratory reproducibility for all parameters except calculated LVEF ( $CP = 0.675$ ), mitral E velocity ( $CP = 0.715$ ), and mitral A velocity ( $CP = 0.760$ ). Measurements of GLS in the apical four-chamber view ( $GLS_{4CH}$ ) bordered acceptable reproducibility ( $CP = 0.805$ ).

Intra- and interreader comparisons for parameters of LV structure and systolic function are illustrated in dot plots, which indicate the direction and absolute magnitude of the measurement differences within a single reader or between pairs of readers for each echocardiogram. Within-ECL comparisons are illustrated in





**Figure 1** LV internal dimension tracing conventions by ECL.



**Figure 2** LV endocardial border tracing conventions by ECL.

Figures 4 and 5. The differences for calculated LVEF measurements by a single reader at laboratory A (Figure 4A) and between two readers at laboratory B (Figure 4B) were typically within 5% and all were within 10%. The differences for GLS<sub>4CH</sub> measurements exceeded 4% (the prespecified minimum acceptable threshold) on only two echocardiograms at each ECL (Figures 5A and 5B). Between-ECL comparisons for LV volumes were acceptably reproducible; however, end-diastolic volumes (Figure 6A) were similar to slightly higher and end-systolic volumes (Figure 6B) were higher for laboratory A compared with laboratory B. Dot-plot comparisons for LVEF and GLS<sub>4CH</sub> measurements between ECLs are illustrated in the Online Appendix (Supplemental Figures 1 and 2, available at [www.onlinejase.com](http://www.onlinejase.com)).

After reviewing individual results, readers identified three potential sources of interlaboratory variability, as follows: (1) single-beat measure-

ments, (2) nonuniformity in image selection, and (3) endocardial border tracing conventions. Mitral inflow velocity differences were attributed to heterogeneity in the images selected for measurements, which was at the discretion of individual readers, and single-beat measurements. Heterogeneous image selection also contributed to variability in calculated LVEF and GLS<sub>4CH</sub> measurements; however, laboratory-specific differences in endocardial border definition, particularly at end-systole, more consistently affected determination of LV volumes and the manually defined region of interest (ROI) for TomTec longitudinal strain analyses.

Readers reconciled sources of variability and repeated a limited set of measurements on all 25 echocardiograms. To address the impact on variability related to (1) single-beat measurements, readers remeasured mitral valve velocities and averaged measurements over three beats; (2) image selection heterogeneity, one laboratory B reader acted as an image selection control, aligning image selection for mitral valve

A.

	Lab A				Lab B			
Echocardiography Parameter	Measurement Results Per Echocardiogram							
	Reader 1		Reader 2		Reader 3		Reader 4	
	1.1	1.2	2.1	2.2	3.1	3.2	4.1	4.2
Single plane LVEDV	X	X			X	X	X	X
Single plane LVESV	X	X			X	X	X	X
Single plane LVEF (calculated)	X	X			X	X	X	X
Single plane LVEF (visual)	X	X			X	X	X	X
MV E velocity			X	X	X	X	X	X
MV A velocity			X	X	X	X	X	X
Global longitudinal strain <sub>4CH</sub>	X	X			X	X	X	X
R wave to LVOT flow onset			X	X	X	X	X	X
R wave to LVOT flow end			X	X	X	X	X	X

B.

Reader 3	Reader 4		
	Measurement Result	4.1	4.2
	3.1	X	X
	3.2	X	X

C.

Lab A	Lab B				
	Measurement Result	3.1	3.2	4.1	4.2
	1.1 or 2.1	X	X	X	X
	1.2 or 2.2	X	X	X	X

**Figure 3** ECL assignment of echocardiography measurements and comparisons of results. **(A)** For each echocardiogram, there were two results per reader for every measured parameter. At laboratory A, LV volumes and strain were measured by a single reader; Doppler parameters were measured by a separate reader. At laboratory B, each reader measured every parameter. **(B)** Pairwise comparisons of measurement results between the two readers at laboratory B. For each measured parameter per echocardiogram, there were four pairwise comparisons of two measurement results per reader between the two readers at laboratory B. **(C)** Pairwise comparisons of measurement results between the readers at laboratories A and B. For each measured parameter per echocardiogram, there were eight pairwise comparisons of two measurement results per reader between one reader at laboratory A and two readers at laboratory B. *4CH*, Apical four-chamber; *LVEDV*, LV end-diastolic volume; *LVESV*, LV end-systolic volume; *LVOT*, LV outflow tract; *MV*, mitral valve.

**Table 1** Prespecified reproducibility standards for echocardiographic parameters

Echocardiographic parameter	Minimum acceptable difference (for 80% of pairwise comparisons)	Maximum acceptable differences (for 100% of pairwise comparisons)
LVEDV (mL)	30	60
LVESV (mL)	30	60
LVEF (calculated) (%)	10	20
LVEF (visual) (%)	10	20
MV E velocity (cm/sec)	5	10
MV A velocity (cm/sec)	5	10
GLS (%)	4	8
Time from R wave to LVOT flow onset (msec)	20	40
Time from R wave to LVOT flow end (msec)	40	85

*LVEDV*, LV end-diastolic volume; *LVESV*, LV end-systolic volume; *LVOT*, LV outflow tract; *MV*, mitral valve.

velocity measurements with the laboratory A reader (the second laboratory B reader's image selection remained blinded and independent); and (3) endocardial border definition, readers remeasured GLS<sub>4CH</sub> by selecting the speckle-tracking ROI at end-diastole, on the basis of previous results with LV volumes suggesting better interlaboratory agreement for end-diastolic endocardial border determinations.

On the second reproducibility testing, analysis of 400 pairwise comparisons for mitral valve E and A velocities (200 for each) among

three readers across both ECLs demonstrated acceptable reproducibility for the minimum acceptable differences. Coverage probabilities improved for E (from 0.715 to 0.925) and A (from 0.760 to 0.810) velocities when results were the average of measurements taken over three beats; independent analysis of the 100 pairwise comparisons for the laboratory B reader aligned with the reader at laboratory A's image selection demonstrated further improvement in CP's for E (to 0.990) and A (to 0.940) velocities for the minimum acceptable differences. There was no improvement in reproducibility for GLS<sub>4CH</sub> measurements when the strain ROI was defined using the endocardial border at end-diastole (CP = 0.775) versus end-systole (CP = 0.805). Comparisons of reproducibility from the initial and second assessments are shown in Table 4.

## DISCUSSION

### Main Findings

Our main findings after directly comparing echocardiographic measurements obtained in patients with cancer between two academic ECLs are as follows: (1) reproducibility was strong within each ECL, confirming the utility and robustness of standard practices among ECL readers at Penn and DCRI,<sup>26,27</sup> and (2) there was variability between ECLs for certain measurements, related to different ECL interpretations of guideline-recommended measurement conventions, suggesting a role for standardized application of measurement conventions across all ECLs.

### Exploring the Variability of Measurements

Despite strong intralaboratory reproducibility, interlaboratory reproducibility results for mitral inflow velocities, quantitative LVEF, and GLS<sub>4CH</sub> were less concordant. During subsequent reconciliation and

**Table 2** Coverage probabilities for echocardiographic reproducibility within each ECL

Echocardiographic parameter	Intrareader* (laboratories A and B)		Interreader (laboratory B)	
	CP minimum acceptable difference (target $\geq 0.8$ )	CP 2 $\times$ acceptable difference (target 1.00)	CP minimum acceptable difference (target $\geq 0.8$ )	CP 2 $\times$ acceptable difference (target 1.00)
Single-plane LVEDV (mL)	0.960–1.000	1.000–1.000	1.000	1.000
Single-plane LVESV (mL)	1.000–1.000	1.000–1.000	1.000	1.000
Single-plane LVEF (calculated) (%)	1.000–1.000	1.000–1.000	1.000	1.000
Single-plane LVEF (visual) (%)	1.000–1.000	1.000–1.000	1.000	1.000
MV E velocity (cm/sec)	0.840–0.960	1.000–1.000	0.770	0.990
MV A velocity (cm/sec)	0.840–0.920	0.960–1.000	0.770	0.990
GLS <sub>4CH</sub> (%)	0.920–1.000	1.000–1.000	0.940	1.000
R wave to LVOT flow onset (msec)	0.960–1.000	1.000–1.000	0.900	1.000
R wave to LVOT flow end (msec)	0.960–1.000	1.000–1.000	0.970	1.000

LVEDV, LV end-diastolic volume; LVESV, LV end-systolic volume; LVOT, LV outflow tract; MV, mitral valve.

\*Intrareader CP reported as a range across all readers at both laboratories.

**Table 3** Initial assessment of echocardiographic reproducibility between ECLs

Echocardiographic parameter	CP minimum acceptable difference (target $\geq 0.8$ )	CP 2 $\times$ acceptable difference (target 1.00)
Single-plane LVEDV (mL)	0.985	1.000
Single-plane LVESV (mL)	0.975	1.000
Single-plane LVEF (calculated) (%)	0.675	0.980
Single-plane LVEF (visual) (%)	0.900	1.000
MV E velocity (cm/sec)	0.715	0.910
MV A velocity (cm/sec)	0.760	0.915
GLS <sub>4CH</sub> (%)	0.805	1.000
R wave to LVOT flow onset (msec)	0.810	0.995
R wave to LVOT flow end (msec)	0.955	1.000

LVEDV, LV end-diastolic volume; LVESV, LV end-systolic volume; LVOT, LV outflow tract; MV, mitral valve.

retraining efforts, potential explanations for differences were considered by parameter, as follows: (1) mitral inflow velocities: single-beat measurements and interlaboratory heterogeneity in image and cardiac cycle selection likely accentuated differences due to inherent physiologic variability (e.g., respiratory changes); (2) LVEF: ventricular volumes were acceptably reproducible by pairwise comparisons, but divergent trends in volume measurements between ECLs likely affected LVEF estimations; and (3) GLS<sub>4CH</sub>: laboratory-specific differences in endocardial border tracing conventions likely extended to how each ECL manually defined placement and thickness of the speckle-tracking ROI at end-systole for myocardial strain analyses.

On reanalysis, reproducibility for mitral inflow velocities improved, demonstrating the importance of averaging multiple measurements to mitigate differences related to intrinsic physiologic variability and well-defined, prospectively determined image selection to generate

homogeneous data points for measurement. Despite retraining, variability for GLS<sub>4CH</sub> persisted, however, reflecting fundamental differences in how each ECL defined the endocardial borders that also affected quantitative LVEF measurements.

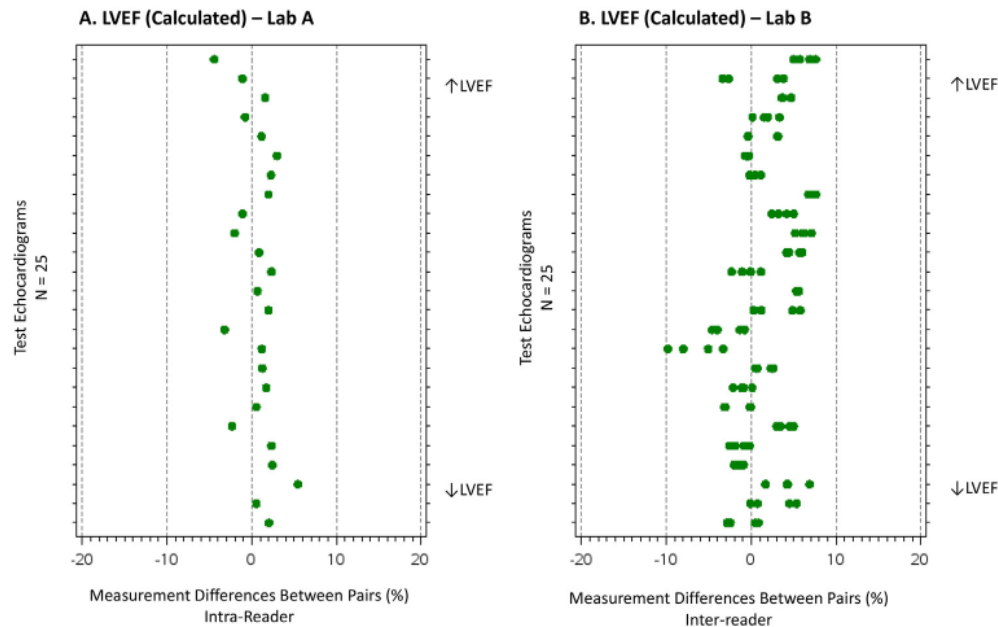
### Importance of Precision in the Echocardiographic Assessment of Patients with Cancer

In cancer, at least a baseline assessment of LVEF is recommended for potentially cardiotoxic cancer therapies,<sup>2</sup> and serial assessments are mandated by the US Food and Drug Administration for certain agents,<sup>28</sup> with the primary goal to minimize risk for CTRCD. Echocardiography is commonly used to monitor cardiac function in clinical oncology practice; an LVEF that is low or has declined may prompt withdrawal or postponement of potentially curative cancer therapy.<sup>29</sup> Despite inherent variability of LVEF measurements related to dynamic biologic conditions and inconsistent image quality, echocardiography laboratories should endeavor to report consistently reproducible results<sup>30</sup> given the implications for cardiac and oncologic outcomes.

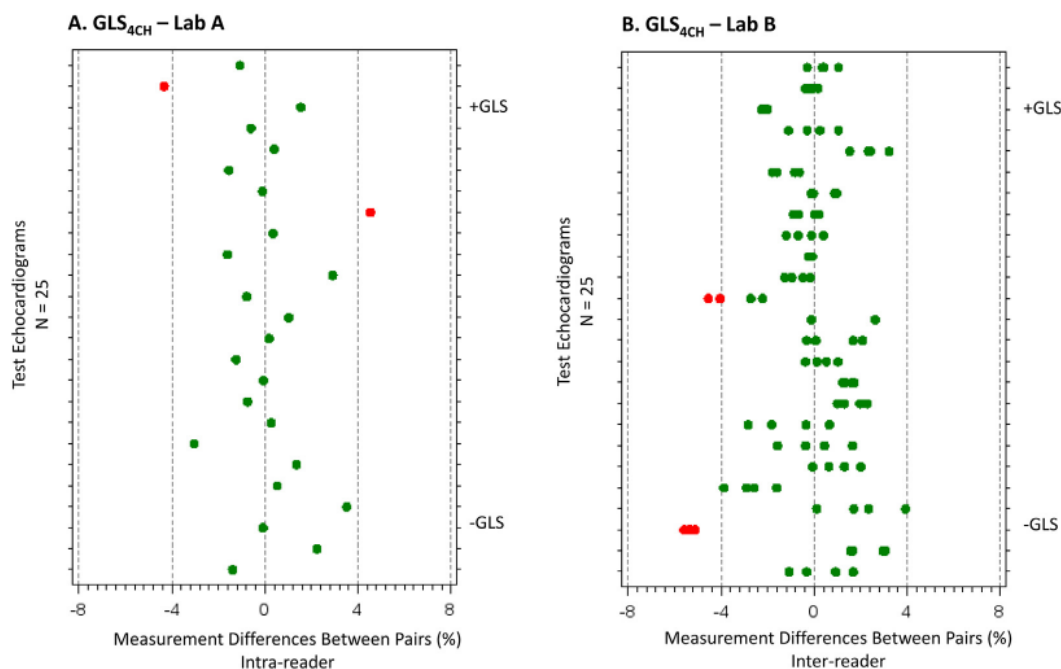
Alternative echocardiographic parameters of cardiac performance have recently emerged for the detection of early myocardial injury during cancer therapy. Echocardiography-derived measures of myocardial mechanics, including an impairment or relative decline in myocardial deformation by speckle-tracking longitudinal strain<sup>13,31</sup> and changes in ventricular-arterial coupling,<sup>26</sup> have demonstrated predictive utility for LVEF decline and/or symptomatic heart failure. Novel modalities such as strain echocardiography may ultimately improve cardiovascular risk stratification and treatment of patients with cancer, although dependence on high-quality images, multiple evolving vendor acquisition and analysis platforms,<sup>32</sup> and uncertain optimal cutoff values for positive tests in cancer<sup>1,2</sup> also raise concerns.

### Current State of Echocardiography for Identifying CTRCD

Current clinical investigations routinely use cardiac imaging to assess the potential of CTRCD. Oncology clinical trials assessing novel therapies often rely on LVEF assessment performed and interpreted at multiple clinical sites and often assessed visually. In general, cardio-oncology investigations have not accounted for reproducibility concerns related to different image acquisition methods, variable image quality, and multiple observers that previous multicenter,



**Figure 4** LVEF reproducibility results within ECLs. Pairwise comparisons on testing of calculated LVEF. **(A)** Intrareader at laboratory A (CP = 1.000). **(B)** Interreader at laboratory B (CP = 1.000). LVEF was determined in the apical four-chamber view. *Red dots* indicate reader pairs exceeding the minimum acceptable difference of 10%; *green dots* indicate those pairs within the acceptable difference. ↑ LVEF, Higher LVEF values within each ECL; ↓ LVEF, lower LVEF values within each ECL.

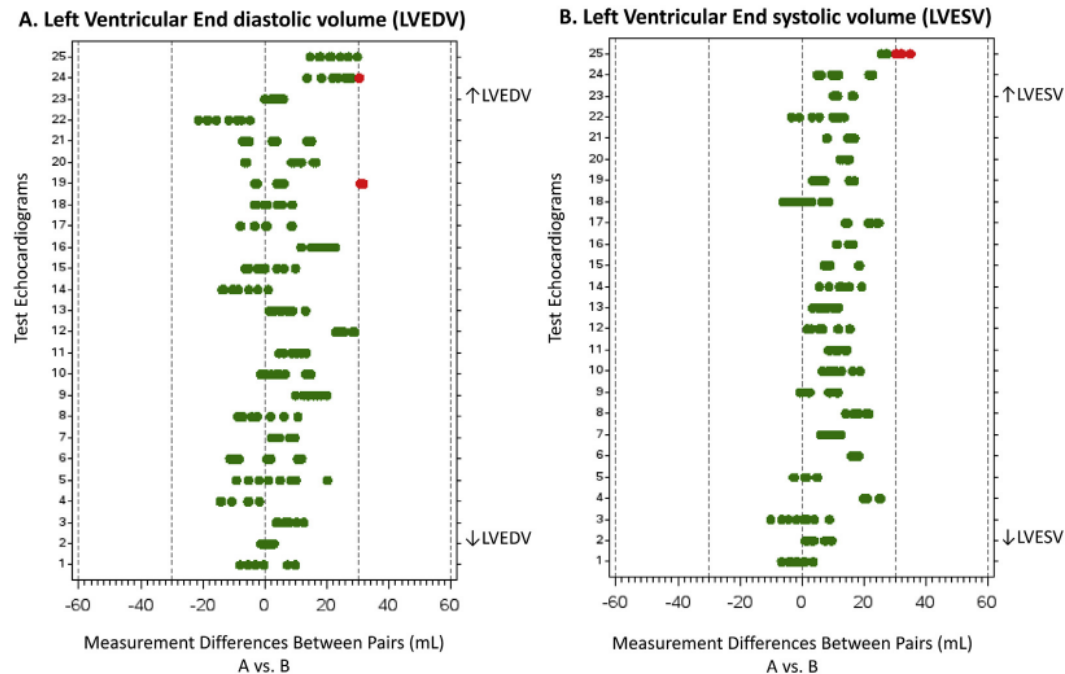


**Figure 5** LV GLS<sub>4CH</sub> reproducibility results within ECLs. Pairwise comparisons on testing of GLS<sub>4CH</sub>. **(A)** Intrareader at laboratory A (CP = 0.920). **(B)** Interreader at laboratory B (CP = 0.940). GLS<sub>4CH</sub> was determined in the apical four-chamber view. *Red dots* indicate reader pairs exceeding the minimum acceptable difference of 4%; *green dots* indicate those pairs within the acceptable difference. +GLS, More positive GLS<sub>4CH</sub> values within each ECL; –GLS, more negative GLS<sub>4CH</sub> values within each ECL.

multimodality cardiac imaging studies have reported.<sup>8,10</sup> Failure to account for these issues has led to the inclusion of apparent “outliers” or ineligible subjects in clinical trials, potentially undermining the validity of results.<sup>10,11</sup> In cardiology, the importance of a thoughtful, proactive approach to ensuring the validity of echocardiography results has been previously

demonstrated by the Predictors of Response to Cardiac Resynchronization Therapy study, in which marked interlaboratory variability contributed to inconclusive results.<sup>33</sup> Studies that use echocardiography measurements for inclusion and/or end points now routinely use ECLs.<sup>3</sup> In this setting, large multicenter cardiovascular clinical trials have demonstrated that the standardized approach





**Figure 6** LV volume reproducibility results between ECLs. **(A)** Interlaboratory pairwise comparisons on testing of LV end-diastolic volume (LVEDV; CP = 0.985). **(B)** Interlab pairwise comparisons on testing of LV end-systolic volume (LVESV; CP = 0.975). LVEDV and LVESV were determined in the apical four-chamber view. *Red dots* indicate reader pairs exceeding the minimum acceptable difference of 30 mL; *green dots* indicate those pairs within the acceptable difference. A, Laboratory A; B, laboratory B; ↑LVEDV, higher LVEDV values; ↓LVEDV, lower LVEDV values (range of LVEDV values depicted is referenced to Lab A results); ↑LVESV, higher LVESV values; ↓LVESV, lower LVESV values (range of LVESV values depicted is referenced to laboratory A results).

**Table 4** Echocardiographic reproducibility between ECLs before and after retraining

Assessment and intervention	CP*		
	Mitral valve E velocity	Mitral valve A velocity	GLS <sub>4CH</sub>
Initial assessment	0.715	0.760	0.805
Postretraining intervention			
Multiple averaged beats	0.925	0.810	
Multiple averaged beats and image selection control	0.990	0.940	
ROI defined at end-diastole			0.775

\*CP ≥ 0.80 represents acceptable reproducibility.

and consistency of an ECL minimizes measurement variability and improves the precision of study results,<sup>9</sup> with the potential also to enhance study outcomes.<sup>12</sup>

### Implications of Findings for Echocardiography in Cardio-Oncology Research

To our knowledge, our study is the first to compare academic ECLs in the assessment of CTRCD parameters being studied in cardio-oncology research. Strongly reproducible findings within the Penn and DCRI ECLs in our study confirm the value of a commitment in both laboratories to robust reproducibility and remediation processes to ensure temporal and reader validity<sup>9,26,27</sup> that is required

for high-quality quantitative echocardiographic research.<sup>3</sup> Our study also highlights, however, that a combination of small differences between ECLs in standard operating procedures and in the application of measurement conventions yielded slightly different “assays” for quantitative echocardiography data in each ECL, making interlaboratory reproducibility of certain echocardiography parameters potentially challenging. Going forward, the findings have reaffirmed the commitment of both ECLs to achieve consensus in the setting of multi-ECL studies and prompted consideration for creating a common set of images to serve as a potential reference standard across ECLs, similar to the international normalized ratio for prothrombin time ratio measurements.

Currently, our findings suggest the following: (1) use of a single ECL may be optimally suited to meet the needs of an individual clinical trial in minimizing variability and should similarly be considered for serial studies assessing a single drug or device; (2) in the absence of a single ECL, a strong commitment with adequate funding is needed to ensure quality and harmonization, including the possibility of several rounds of measurement testing and retesting; (3) results obtained from more than one ECL should not be combined unless proof of “poolability” is developed prospectively for parameters of interest, likely requiring extensive cross-training and process harmonization across ECLs; (4) comparability of results may be a more realistic goal, provided there is acceptable intralaboratory reproducibility and other aspects of studies are aligned; and (5) parameter cut points used in CTRCD criteria should be interpreted and determined cautiously, until these can be validated across multiple laboratories with additional attention to other potential sources of variability, such as vendor analysis and acquisition platforms.<sup>7,8,25</sup>



## Core Lab Comparisons Using Other Cardiac Imaging Modalities

The importance of standardizing measurement conventions among different core labs has been similarly noted for cardiac imaging modalities outside echocardiography.<sup>34,35</sup> In coronary angiography, clinical interpretations commonly rely on visual estimation of percentage stenosis, despite well-described intra- and interobserver variability.<sup>36</sup> Compared with visual estimations, quantitative coronary angiography improves the precision and accuracy of stenosis severity determinations.<sup>37,38</sup> However, core labs had marked interlaboratory variability in quantitative coronary angiography of bifurcation stenosis due to software and measurement methodologic differences; subsequent analyses after aligning the core labs' software and methodology demonstrated improved interlaboratory variability and altered trial results.<sup>34</sup> In cardiac magnetic resonance, a comparison of LV volumes and mass measurements across seven core labs demonstrated small intrareader variability but considerable interreader and interlaboratory variability due to different contour drawing practices at each laboratory, highlighting the need to achieve consensus before multicenter analyses.<sup>35</sup> Although standardization can improve reproducibility among core labs, it is also an inherently challenging process that demands time, care, and attention to detail. Nevertheless, as accuracy is unattainable without precision, standardizing measurement conventions has the potential to positively affect the primary intent for all diagnostic testing (i.e., to guide patient management and reduce future clinical events).<sup>39</sup>

## Implications of Findings for Echocardiography in Clinical Cardio-Oncology

Reproducibility findings between ECLs at Penn and DCRI may extend to clinical laboratories that quantify echocardiographic parameters during cardiac monitoring of patients with cancer. Guidelines<sup>2</sup> and prior studies<sup>13,31</sup> have emphasized the importance of changes in LVEF and GLS for identifying CTRCD. Our study suggests that comparing quantitative echocardiographic measurements across laboratories should be considered in a similar context to comparing blood biomarker (e.g., troponins, natriuretic peptides) results across institutions with different measurement assays and reference ranges. Accordingly, comparing small changes across readers and laboratories should be performed cautiously. The ideal clinical environment for cardio-oncology patients likely includes serial studies with a consistent vendor acquisition and analysis platform and sonographer or reader in a single laboratory that uses continuous quality monitoring and improvement processes to optimize image acquisition quality as well as robust reproducibility testing and remediation processes to ensure temporal and interreader validity of echocardiographic data.

## Limitations

Quantitative echocardiography studies are generally limited by a lack of established standards that define acceptable reproducibility for individual parameters. For instance, the absence of an external reference "gold" standard technique such as cardiac magnetic resonance imaging for identifying the LV endocardial border against which to compare each ECL's tracing conventions limited the potential of harmonization and retraining efforts for improving interlaboratory reproducibility for LVEF and GLS<sub>4CH</sub> results. Nevertheless, our study sought to be one of the first to perform a detailed comparison between ECLs. Although our findings respective to parameters of LV systolic function may be affected

by using a single view assessment, an ECL study from a major cardiovascular trial demonstrated highly correlated single-plane and biplane LV volumes,<sup>10</sup> and a recent intervender study demonstrated that longitudinal strain derived from the four-chamber view alone was strongly correlated with values derived from all three apical views.<sup>25</sup> In addition, we could not assess the variability of GLS calculated using endocardial length in diastole and systole, as has recently been recommended by the American Society of Echocardiography<sup>16</sup>; software adapted to measure strain by endocardial length was not available at the time of our study.

Our study had other important limitations. First, harmonization efforts that included several group reads incorporating multiple readers, measurements, and iterations were time consuming. During multiple rounds of training, the time commitment for some interventions (e.g., averaging multiple beats and controlling image selection for Doppler measurements) was more feasible than for others (e.g., reconciling fundamental interlaboratory differences in endocardial border tracing conventions), limiting our ability to improve interlaboratory reproducibility for LV systolic functional parameters. Future studies in which ECLs try to comprehensively align measurements will require adequate time and funding to optimize corrective actions.

Second, a relatively small size of 25 echocardiograms was tested for reproducibility, although this number is similar to other reproducibility studies<sup>8,27,30</sup> and allowed large numbers of pairwise comparisons, making it statistically robust.

Third, echocardiograms selected for analysis were previously acquired in the clinical laboratories at both institutions; measurements were limited to noncontrast, 2D images for consistency, although previous studies have reported that contrast administration did not improve reproducibility of serial 2D echocardiograms in patients with breast cancer.<sup>4</sup>

Fourth, the ECLs used different software to make 2D and Doppler measurements (Penn, TomTec; DCRI, Digisonics), potentially affecting reproducibility for these parameters.

Fifth, although lack of clinical data limits the generalizability of our findings, it also minimized potential bias affecting ECL analyses. Sources of variability have been described with strain imaging.<sup>32</sup> Our GLS reproducibility findings in the context of ECL commonalities across vendor acquisition platform, image acquisition frame rates, and strain analysis software are not necessarily analogous to laboratory settings in which echocardiography studies acquired and analyzed on multiple platforms are being compared.

Finally, our study was intended to define reliability and reproducibility between ECLs rather than validate the diagnostic accuracy of measurements; the absence of a reference standard to validate accuracy limited the potential for reproducibility assessments among patient subgroups with suggestion of impaired LVEF and GLS<sub>4CH</sub>.

## CONCLUSIONS

High-quality quantitative echocardiography research requires a commitment to robust reproducibility and remediation processes with demonstrated good results. Low variability or uniformity in echocardiography measurements within or across laboratories cannot be assumed; interlaboratory agreement may be difficult even among academic ECLs that are highly internally reproducible. Comparisons of small changes in LVEF and GLS across readers and laboratories should be performed cautiously. Research designs should acknowledge these findings and consider a single ECL for individual clinical

trials and series of studies testing a single drug or device, while results of serial clinical studies should be interpreted accordingly.

## SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at <https://doi.org/10.1016/j.echo.2017.11.018>.

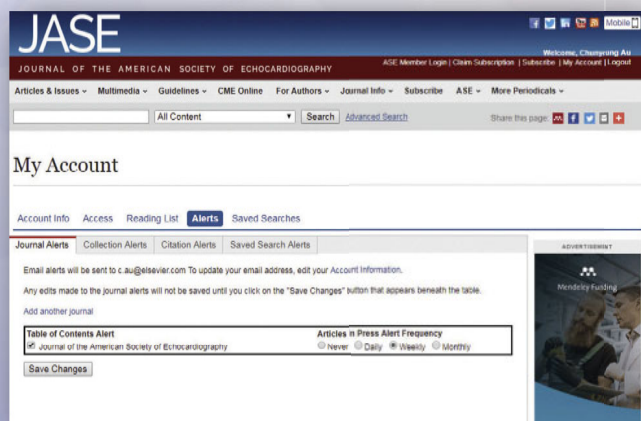
## REFERENCES

1. Khouri MC, Klein MR, Velazquez EJ, Jones LW. Current and emerging modalities for detection of cardiotoxicity in cardio-oncology. *Future Cardiol* 2015;11:471-84.
2. 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;27:911-39.
3. Douglas PS, DeCara JM, Devereux RB, Duckworth S, Gardin JM, Jaber WA, et al. Echocardiographic imaging in clinical trials: American Society of Echocardiography Standards for echocardiography core laboratories: endorsed by the American College of Cardiology Foundation. *J Am Soc Echocardiogr* 2009;22:755-65.
4. 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;61:77-84.
5. Romond EH, Jeong JH, Rastogi P, Swain SM, Geyer CE Jr., Ewer MS, et al. Seven-year follow-up assessment of cardiac function in NSABP B-31, a randomized trial comparing doxorubicin and cyclophosphamide followed by paclitaxel (ACP) with ACP plus trastuzumab as adjuvant therapy for patients with node-positive, human epidermal growth factor receptor 2-positive breast cancer. *J Clin Oncol* 2012;30:3792-9.
6. Tan-Chiu E, Yothers G, Romond E, Geyer CE Jr., Ewer M, Keefe D, et al. Assessment of cardiac dysfunction in a randomized trial comparing doxorubicin and cyclophosphamide followed by paclitaxel, with or without trastuzumab as adjuvant therapy in node-positive, human epidermal growth factor receptor 2-overexpressing breast cancer: NSABP B-31. *J Clin Oncol* 2005;23:7811-9.
7. Thavendiranathan P, Poulin F, Lim KD, Plana JC, Woo A, Marwick TH. Use of myocardial strain imaging by echocardiography for the early detection of cardiotoxicity in patients during and after cancer chemotherapy: a systematic review. *J Am Coll Cardiol* 2014;63:2751-68.
8. Negishi K, Lucas S, Negishi T, Hamilton J, Marwick TH. What is the primary source of discordance in strain measurement between vendors: imaging or analysis? *Ultrasound Med Biol* 2013;39:714-20.
9. Douglas PS, Waugh RA, Bloomfield G, Dunn G, Davis L, Hahn RT, et al. Implementation of echocardiography core laboratory best practices: a case study of the PARTNER I trial. *J Am Soc Echocardiogr* 2013;26:348-58.e3.
10. Oh JK, Pellikka PA, Panza JA, Biernat J, Attisano T, Manahan BG, et al. Core lab analysis of baseline echocardiographic studies in the STICH trial and recommendation for use of echocardiography in future clinical trials. *J Am Soc Echocardiogr* 2012;25:327-36.
11. Foster E, Wasserman HS, Gray W, Homma S, Di Tullio MR, Rodriguez L, et al. Quantitative assessment of severity of mitral regurgitation by serial echocardiography in a multicenter clinical trial of percutaneous mitral valve repair. *Am J Cardiol* 2007;100:1577-83.
12. Hole T, Otterstad JE, St John Sutton M, Froland G, Holme I, Skjaerpe T. Differences between echocardiographic measurements of left ventricular dimensions and function by local investigators and a core laboratory in a 2-year follow-up study of patients with an acute myocardial infarction. *Eur J Echocardiogr* 2002;3:263-70.
13. Sawaya H, Sebag IA, Plana JC, Januzzi JL, Ky B, Tan TC, et al. Assessment of echocardiography and biomarkers for the extended prediction of cardiotoxicity in patients treated with anthracyclines, taxanes, and trastuzumab. *Circ Cardiovasc Imaging* 2012;5:596-603.
14. Schwartz RG, McKenzie WB, Alexander J, Sager P, D'Souza A, Manatunga A, et al. Congestive heart failure and left ventricular dysfunction complicating doxorubicin therapy. Seven-year experience using serial radionuclide angiocardiology. *Am J Med* 1987;82:1109-18.
15. 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;28:1-39.e14.
16. Voigt JU, Pedrizzetti G, Lysyansky P, Marwick TH, Houle H, Baumann R, et al. Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *J Am Soc Echocardiogr* 2015;28:183-93.
17. St. John Sutton M, Plappert T. Core lab, no core lab or automated LVEF? *Eur J Echocardiogr* 2007;8:239-40.
18. Chen CH, Fetics B, Nevo E, Rochitte CE, Chiou KR, Ding PA, et al. Noninvasive single-beat determination of left ventricular end-systolic elastance in humans. *J Am Coll Cardiol* 2001;38:2028-34.
19. Narayan HK, Wei W, Feng Z, Lenihan D, Plappert T, Englefield V, et al. Cardiac mechanics and dysfunction with anthracyclines in the community: results from the PREDICT study. *Open Heart* 2017;4:e000524.
20. Zhang KW, French B, May Khan A, Plappert T, Fang JC, Sweitzer NK, et al. Strain improves risk prediction beyond ejection fraction in chronic systolic heart failure. *J Am Heart Assoc* 2014;3:e000550.
21. Barnhart HX, Yow E, Crowley AL, Daubert MA, Rabineau D, Bigelow R, et al. Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting. *Stat Methods Med Res* 2016;25:2939-58.
22. Crowley AL, Yow E, Barnhart HX, Daubert MA, Bigelow R, Sullivan DC, et al. Critical review of current approaches for echocardiographic reproducibility and reliability assessment in clinical research. *J Am Soc Echocardiogr* 2016;29:1144-54.e7.
23. Galderisi M, Benjamin EJ, Evans JC, D'Agostino RB, Fuller DL, Lehman B, et al. Intra- and interobserver reproducibility of Doppler-assessed indexes of left ventricular diastolic function in a population-based study (the Framingham Heart Study). *Am J Cardiol* 1992;70:1341-6.
24. Frikha Z, Girerd N, Huttin O, Courand PY, Bozec E, Olivier A, et al. Reproducibility in echocardiographic assessment of diastolic function in a population based study (the STANISLAS Cohort study). *PLoS ONE* 2015;10:e0122336.
25. Farsalinos KE, Daraban AM, Unlu S, Thomas JD, Badano LP, Voigt JU. Head-to-head comparison of global longitudinal strain measurements among nine different vendors: the EACVI/ASE Inter-Vendor Comparison Study. *J Am Soc Echocardiogr* 2015;28:1171-81.e2.
26. Narayan HK, French B, Khan AM, Plappert T, Hyman D, Bajulaiye A, et al. Noninvasive measures of ventricular-arterial coupling and circumferential strain predict cancer therapeutics-related cardiac dysfunction. *JACC Cardiovasc Imaging* 2016;9:1131-41.
27. Daubert MA, Yow E, Barnhart HX, Rabineau D, Crowley AL, Douglas PS. Quality improvement implementation: improving reproducibility in the echocardiography laboratory. *J Am Soc Echocardiogr* 2015;28:959-68.
28. Trastuzumab (herceptin) package insert. San Francisco, CA: Genentech.
29. Dang CT, Yu AF, Jones LW, Liu J, Steingart RM, Argolo DF, et al. Cardiac surveillance guidelines for trastuzumab-containing therapy in early-stage breast cancer: getting to the heart of the matter. *J Clin Oncol* 2016;34:1030-3.
30. 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;4:821-9.

31. Negishi K, Negishi T, Hare JL, Haluska BA, Plana JC, Marwick TH. Independent and incremental value of deformation indices for prediction of trastuzumab-induced cardiotoxicity. *J Am Soc Echocardiogr* 2013;26:493-8.
32. Nagata Y, Takeuchi M, Mizukoshi K, Wu VC, Lin FC, Negishi K, et al. Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software. *J Am Soc Echocardiogr* 2015;28:630-41.
33. Chung ES, Leon AR, Tavazzi L, Sun JP, Nihoyannopoulos P, Merlino J, et al. Results of the Predictors of Response to CRT (PROSPECT) trial. *Circulation* 2008;117:2608-16.
34. Grundeken MJ, Ishibashi Y, Genereux P, LaSalle L, Iqbal J, Wykrzykowska JJ, et al. Inter-core lab variability in analyzing quantitative coronary angiography for bifurcation lesions: a post-hoc analysis of a randomized trial. *JACC Cardiovasc Interv* 2015;8:305-14.
35. Suinesiaputra A, Bluemke DA, Cowan BR, Friedrich MG, Kramer CM, Kwong R, et al. Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. *J Cardiovasc Magn Reson* 2015;17:63.
36. Herrman JP, Azar A, Umans VA, Boersma E, von Es GA, Serruys PW. Inter- and intra-observer variability in the qualitative categorization of coronary angiograms. *Int J Card Imaging* 1996;12:21-30.
37. Shah R, Yow E, Jones WS, Kohl LP III, Kosinski AS, Hoffmann U, et al. Comparison of visual assessment of coronary stenosis with independent quantitative coronary angiography: findings from the Prospective Multi-center Imaging Study for Evaluation of Chest Pain (PROMISE) trial. *Am Heart J* 2017;184:1-9.
38. Nallamothu BK, Spertus JA, Lansky AJ, Cohen DJ, Jones PG, Kureshi F, et al. Comparison of clinical interpretation with visual assessment and quantitative coronary angiography in patients undergoing percutaneous coronary intervention in contemporary practice: the Assessing Angiography (A2) project. *Circulation* 2013;127:1793-800.
39. Douglas PS, Chen J, Gillam L, Hendel R, Hundley WG, Masoudi F, et al. Achieving quality in cardiovascular imaging II: proceedings from the Second American College of Cardiology–Duke University Medical Center Think Tank on Quality in Cardiovascular Imaging. *JACC Cardiovasc Imaging* 2009;2:231-40.

# Did you know?

You can personalize the **JASE** website  
to meet your individual needs.



Visit  
**onlinejase.com**  
today!

## APPENDIX A

### Consensus Reading Instructions

LV volumes were measured in the apical four-chamber view by tracing the LV endocardial border at end-diastole and end-systole and calculated using the Simpson method of disks, as recommended by the American Society of Echocardiography, with trabeculations and papillary muscles defined as being intracavitary.<sup>15,17</sup> These volumes were used to derive quantitative single-plane LVEF. Qualitative LVEF was assessed visually in the four-chamber apical view and recorded in 5% increments (e.g., 45%, 50%, 55%).

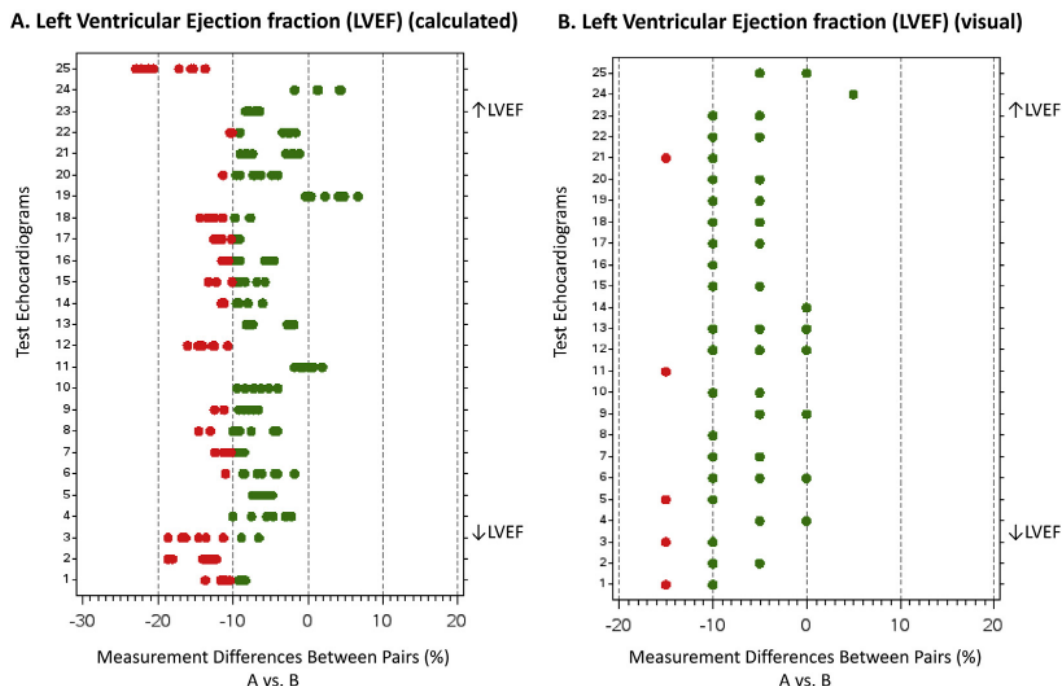
Peak early filling (E-wave) and late diastolic filling (A-wave) mitral inflow velocities were determined from pulsed-wave Doppler signals taken at the tips of the mitral valve leaflets. LV systolic timing intervals were determined from linear caliper measurements taken from the midpoint on the upslope of the electrocardiographic R wave to the LV outflow tract pulsed-wave Doppler velocity profile; R-wave flow onset determined preejection period, and R wave flow end determined total systolic period.<sup>18,26</sup>

Speckle-tracking LV GLS was performed in the apical four-chamber view, which is strongly correlated with GLS averaged across all three apical views.<sup>21</sup> The cardiac cycle with best image quality was selected by each reader. The LV endocardial border was manually traced, and a ROI was defined within the endocardial and epicardial

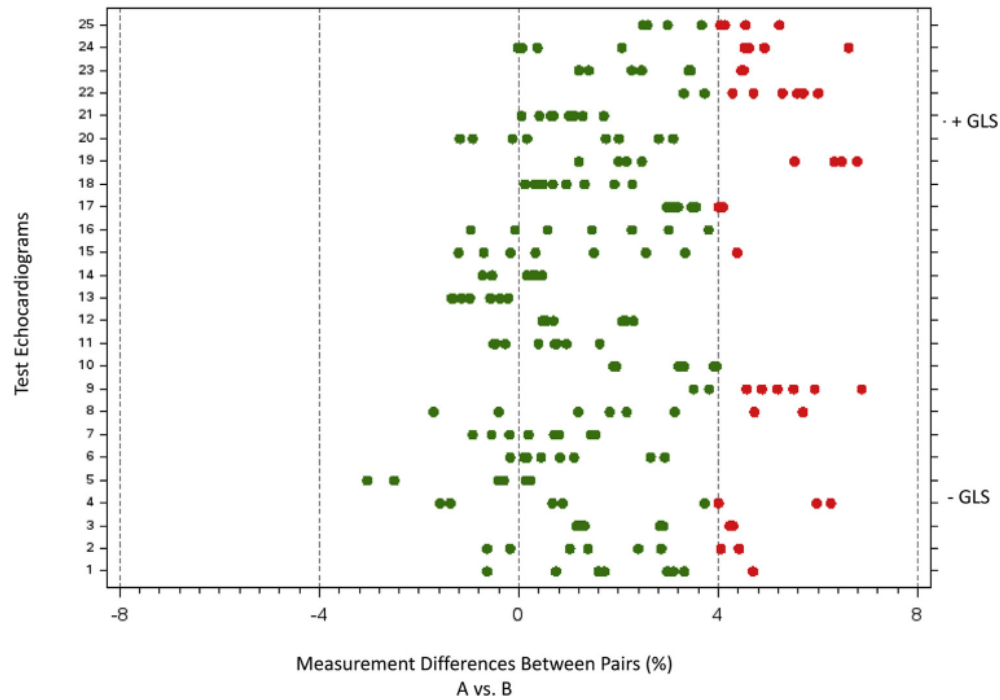
borders. The ROI was defined at either the end-systolic frame (i.e., the frame before the aortic valve completely closes) or end-diastolic frame (i.e., the frame before the mitral valve completely closes) of one cardiac cycle, for end-systolic and end-diastolic strain, respectively. After visual verification of segmental tracking results throughout the cardiac cycle, tracings with poor tracking were manually adjusted in individual myocardial segments to obtain the best possible tracking in all segments. No segments were excluded. Peak systolic endocardial GLS<sub>4CH</sub> values were computed automatically and averaged across all segments.<sup>15,16</sup>

Of note, members of our group have found that image quality is more frequently inadequate in the two-chamber view (~25% of echocardiograms had inadequate apical two-chamber views in a recent study).<sup>19</sup> In addition, LV end-diastolic volume and LV end-systolic volume measured by the single-plane and biplane Simpson method were highly correlated ( $r > 0.95$ ) in ECL analysis of 182 randomly selected patients from the Surgical Treatment for Ischemic Heart Failure trial.<sup>10</sup> Prior studies from members of our group and others have also demonstrated a strong correlation between longitudinal strain obtained from the apical four-chamber view with values derived from the apical two-chamber view<sup>20,26</sup> and from all three apical views.<sup>21</sup> As such, LV volumes, LVEF, and GLS were derived from the four-chamber view to ensure consistency.





**Supplemental Figure 1** LVEF reproducibility results between ECLs. **(A)** Interlaboratory pairwise comparisons on testing of calculated LVEF (CP = 0.675) illustrate that calculated LVEF values were lower at laboratory A compared with laboratory B. **(B)** Interlaboratory pairwise comparisons on testing of visual LVEF (CP = 0.900) illustrate that visual LVEF values were acceptably reproducible. LVEF was determined in the apical four-chamber view. *Red dots* indicate reader pairs exceeding the minimum acceptable difference of 10%; *green dots* indicate those pairs within the acceptable difference. A, Laboratory A; B, laboratory B; ↑ LVEF, higher LVEF values; ↓ LVEF, lower LVEF values (range of LVEF values depicted is referenced to laboratory A results).



**Supplemental Figure 2** LV GLS<sub>4CH</sub> reproducibility results between ECLs. Interlaboratory pairwise comparisons on testing of GLS<sub>4CH</sub> (CP = 0.805) illustrate a trend toward positive measurement differences for laboratory A versus laboratory B consistent with more negative (higher absolute) GLS<sub>4CH</sub> values at laboratory B. GLS<sub>4CH</sub> was determined in the apical four-chamber view. *Red dots* indicate reader pairs exceeding the minimum acceptable difference of 4%; *green dots* indicate those pairs within the acceptable difference. *A*, Laboratory A; *B*, laboratory B; *+GLS*, more positive GLS<sub>4CH</sub> values; *-GLS*, more negative GLS<sub>4CH</sub> value (range of GLS<sub>4CH</sub> values depicted is referenced to laboratory A results).