



Published in final edited form as:

*J Am Soc Echocardiogr.* 2012 August ; 25(8): 842–854.e6. doi:10.1016/j.echo.2012.05.004.

## THE VENTRICULAR VOLUME VARIABILITY STUDY OF THE PEDIATRIC HEART NETWORK: STUDY DESIGN AND IMPACT OF BEAT AVERAGING AND VARIABLE TYPE ON THE REPRODUCIBILITY OF ECHOCARDIOGRAPHIC MEASUREMENTS IN CHILDREN WITH CHRONIC DILATED CARDIOMYOPATHY

Steven D. Colan, MD, FASE<sup>1,2,\*</sup>, Girish Shirali, MBBS, FASE<sup>3,\*</sup>, Renee Margossian, MD<sup>1</sup>, Dianne Gallagher, MS<sup>2</sup>, Karen Altmann, MD<sup>4</sup>, Charles Canter, MD<sup>5</sup>, Shan Chen, MS<sup>2</sup>, Fraser Golding, MD<sup>6</sup>, Elizabeth Radojewski, RN<sup>6</sup>, Jack Rychik, MD<sup>7</sup>, Mario Stylianou, PhD<sup>8</sup>, Lloyd Y. Tani, MD<sup>9</sup>, Elif Seda Selamet Tierney, MD, FASE<sup>1</sup>, Yanli Wang, MS<sup>2</sup>, Lynn A. Sleeper, ScD<sup>2</sup>, and for the Pediatric Heart Network Investigators

<sup>1</sup>Children's Hospital Boston and Harvard Medical School, Boston, MA

<sup>2</sup>New England Research Institutes, Watertown, MA

<sup>3</sup>Medical University of South Carolina, Charleston, SC

<sup>4</sup>Columbia University Medical Center, New York, NY

<sup>5</sup>St. Louis Children's Hospital, Missouri

<sup>6</sup>The Hospital for Sick Children, Toronto, Canada

<sup>7</sup>The Children's Hospital of Philadelphia, Philadelphia, PA

<sup>8</sup>National Heart, Lung, and Blood Institute, Bethesda, MD

<sup>9</sup>University of Utah, Salt Lake City, UT

### Abstract

Clinical trials often rely on echocardiographic measures of left ventricular (LV) size and function as surrogate end-points. However, the quantitative impact of factors that affect reproducibility of these measures is unknown. To address this issue, the NHLBI-funded Pediatric Heart Network designed a longitudinal observational study of children with known or suspected dilated cardiomyopathy (DCM) aged 0–22 years from 8 pediatric clinical centers.

**Methods**—Clinical data were collected together with 150 echocardiographic indices of LV size and function. Separate observers performed duplicate echocardiographic imaging. Multiple observers performed measurements from three cardiac cycles to enable assessment of intra and

---

© 2012 American Society of Echocardiography. Published by Mosby, Inc. All rights reserved.

Corresponding author: Steven D. Colan, M.D., Department of Cardiology, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, 617-355-7893 (phone), 617-739-6282 (fax), colan@alum.mit.edu.

\*Drs. Colan and Shirali are joint first authors of this work

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

interobserver variability. We studied the impact of beat averaging (BA), observer type (local vs. core) and variable type (areas, calculations, dimensions, slopes, time intervals and velocities) on measurement reproducibility. The outcome measure was %error ( $100 \times \text{difference}/\text{mean}$ )

**Results**—Of 173 enrolled subjects, 131 met criteria for DCM. BA, variable type and observer type all impacted %error ( $p < 0.0001$ ). Core inter-observer %error (median 11.4, 10.2 and 9.3% for 1-, 2- and 3-BA, respectively) was approximately twice the intra-observer %error (median 6.3, 4.9 and 4.2% for 1-, 2- and 3-BA, respectively). Slopes and calculated variables exhibited high %error despite BA. Chamber dimensions, areas, velocities and time intervals exhibited low %error.

**Conclusions**—This comprehensive evaluation of quantitative echocardiographic methods will provide a valuable resource for design of future pediatric studies. BA and a single core lab observer improve reproducibility of echo measurements in children with DCM. Certain measurements are highly reproducible, while others, despite BA, are poorly reproducible.

## Introduction

Left ventricular (LV) size and function are important independent predictors of outcome in numerous forms of cardiovascular disease. In children, echocardiography is the primary modality used to assess ventricular function, and echocardiographically derived measurements are commonly used as endpoints in pediatric clinical trials. Although there is extensive experience with this technology, there are few quantitative data concerning reproducibility of these measurements, particularly in children with dilated cardiomyopathy (DCM).<sup>(1)</sup> This issue is particularly problematic because of the wide range of factors that are known to affect this reproducibility. Patient age and habitus are known to be important, as is disease status. Less commonly appreciated is the evolution of technology over time, which requires that this issue be addressed anew with each new generation of echocardiographic equipment. Potential sources of variability in echocardiographic measurements include all of the following:

1. Interpatient variability: differences between patients related to patient-specific factors such as age, body size, variation in cardiac output, physical training, etc., and disease-specific factors such as severity of disease and treatment status;
2. Interstudy variability: longitudinal variation within an individual due to physiologic factors, variation in treatment, and change in disease status;
3. Intrastudy variability: short term (beat-to-beat) variations in the same patient secondary to respiratory effects or change in position, and minute-to-minute variation related to factors such as change in emotional state;
4. Technical factors that modulate each of the foregoing, which include:
  - a. Variability due to differences in echocardiographic and image analysis equipment and whether or not sedation was used;
  - b. Intra- and interobserver variability in data (image) acquisition;
  - c. Intra- and interobserver variability in measurements, including frame selection and structure identification.

In addition to facilitating efforts to improve reproducibility of echocardiographic data, evaluation of variability serves to identify which measured or calculated variables may be preferable and which are too poorly reproducible to be clinically reliable. Finally, design of clinical trials requires estimation of the number of patients that must be enrolled, which requires knowledge of total variability.

This report describes the design and initial findings of the Ventricular Volume Variability (VVV) study, a multicenter study in children with DCM. We present first the overall study aims, the complex study design implemented to address these aims, and the study results specific to the impact of beat averaging. There are few data that examine the impact of beat averaging and variable type on the inter- and intra-observer reproducibility of echocardiographic measurements.<sup>(2–5)</sup> We elected to perform these analyses before addressing the primary aim of the overall study, so that decisions regarding beat averaging could be applied to all subsequent study analyses.

The VVV Study was conducted to address the following aims:

Primary aim.

To determine the interstudy variability of echocardiographically-derived LV end-diastolic volume z-score, mass z-score, and ejection fraction z-score in pediatric patients with DCM; more specifically, the variance at a single point in time as well as the variance of change in measurements over time.

Secondary aims.

1. To determine the relative magnitude of the various sources of variability in echocardiographic outcomes in order to optimize operational procedures that can minimize variance.
2. To determine the interstudy variability of echocardiographically-derived indices of LV systolic and diastolic function.
3. To determine the relationship of clinical status, including treatment, to the interstudy variability and repeatability of echocardiographic measurements.

The specific analysis presented in this report had 3 aims: 1) to determine whether the choice of single-beat analysis or 2- or 3-beat averaging has an influence on inter-observer and intra-observer reproducibility of echocardiographic measurements and calculated variables; 2) to examine the effect of the type of echocardiographic variable on reproducibility as it pertains to beat averaging; and 3) to evaluate the effect of beat averaging on the reproducibility of measurements that were performed at the local centers compared to those performed at the core laboratory.

## Materials and methods

### Subjects

Pediatric patients with known or suspected DCM were enrolled at the time of echocardiographic presentation at each of the 8 study centers (see Appendix: Acknowledgements) between May 2005 and July 2007. Inclusion criteria were age <22 years, known or suspected DCM, disease duration >2 months, anticipated longitudinal follow-up to occur at the same institution, and informed consent. Exclusion criteria (listed in detail in Appendix Figure 1) included other forms of cardiomyopathy and congenital heart disease. Exclusion criteria also included non-compaction (due to an inability to reliably define LV endocardial borders), excessive non-sinus rhythm (due to excess beat-to-beat variance) and hemodynamic instability (due to the intent to assess longitudinal natural history).

### Training

The study principal investigator from each clinical site along with one or more designated site sonographers attended an in-person training session that included protocol review and demonstration of the image acquisition techniques. The measurement methods were

reviewed in detail with the site principal investigator and the designated primary sonographer for the study, one of whom performed all of the local measurements. Each center performed and submitted three practice echocardiograms to the core lab for review and feedback. After approval, enrollment commenced.

### Demographics

Height and weight were measured and body surface area (BSA) was calculated using the Haycock formula.<sup>(6)</sup> Systolic, diastolic, and mean blood pressures were recorded 4 times with the patient in a recumbent position using an automated blood pressure device. The values from the first recording were discarded and the average value of the other 3 measurements was calculated for each pressure.

### Echocardiographic acquisition

All echocardiograms were performed according to a standardized protocol with acquisition of the images listed in Appendix Table 1. Each center designated one or more “primary” sonographers who participated in hands-on study training sessions. For each echocardiographic evaluation, image acquisition was performed by a “primary” echocardiographer and then a second image acquisition was independently performed by any available experienced sonographer (who had not necessarily participated in the study training sessions). The second set of images was acquired immediately following the first.

### Eligibility for longitudinal evaluation

A standardized set of measurements was performed at the study center on the primary data acquisition and the results of these measurements were used to determine eligibility for longitudinal echocardiographic evaluation. Patients who met criteria for the diagnosis of DCM based on the inclusion and exclusion criteria listed above who also were found to have both LV dilation (defined as end-diastolic dimension  $>5.5$  cm or end-diastolic dimension z-score  $>2$ ) and LV dysfunction (defined as ejection fraction  $<50\%$ , shortening fraction  $<28\%$ , or z-score for either of  $<-2$ ) were judged eligible for longitudinal assessment (Appendix Figure 2). Eligible patients had repeat echocardiographic assessment according to the study protocol on return visits up to 18 months following enrollment, providing no indication for withdrawal from the study was met. Indications for withdrawal were death, cardiac transplantation or LV reduction surgery, institution of an LV assist device including extracorporeal membrane oxygenator support, and patient or physician preference. Each subject, therefore, was expected to have 2 sets of echocardiographic images at the time of study enrollment and 2 sets from at least one follow-up visit. The target timing for the follow-up echocardiogram was 12 months, as this was considered to be a likely interval to assess change in a randomized trial. In the completed study, the mean time between paired echocardiograms using the follow-up echocardiogram closest to the 12-month target was  $9.1 \pm 3.5$  (median 9.6, range 2 to 18) months.

### Partial participation group

Subjects who had baseline image acquisition but were then found to not meet criteria for longitudinal evaluation were excluded from further participation, but their baseline images were included in the analyses of reproducibility, thereby permitting evaluation of reproducibility over a broader range of ventricular size and function in addition to enhancing the analysis of the impact of severity of dysfunction on reproducibility.

### Echocardiographic analysis

A standardized measurement protocol that included a total of 150 measurements and calculated variables was performed at the core laboratory for each of 3 cardiac cycles (450

measurements total) for each set of echocardiographic images (Appendix Table 2). Measurements were categorized as areas (9 variables), calculated variables derived from 2, 3 and 4 measured variables (19, 25 and 21 variables respectively), dimensions (16 variables), ECG time intervals and heart rates (7 and 9 variables respectively), integrals (1 variable), slopes (4 variables), Doppler and M-mode time intervals (22 variables) and Doppler velocities (17 variables). A single observer at each center performed the measurements for the primary image acquisition at the initial and follow-up visits using locally available technology, and submitted the results to the data coordinating center. The dataset constructed at the local centers is comprised of 119 of the above 150 measurements and derived indices. The images from both the primary and secondary image acquisitions from each subject visit were submitted to the data coordinating center where they were blinded to study date and patient identification and coded prior to transmission to the echocardiographic core lab for analysis (Appendix Figure 1). Image capture and transfer included video tape, analog-to-digital converted images, and DICOM, depending on the date and the center. All echocardiographic measurements were performed using custom DICOM software (EchoTrace, Marcus Laboratories, Boston, MA) At the core lab, a primary and secondary core lab observer analyzed both the primary and secondary image acquisitions by using the same measurement protocol (Appendix Table 2). In addition, the primary core lab observer repeated the measurements on blinded sets of the primary images at 1 month and 1 year following the original measurements. Primary and secondary image sets acquired at follow-up evaluations of fully eligible subjects were each analyzed by both the primary and secondary core lab observers. Altogether, there were 12 different categories of echocardiographic data sets for statistical comparison, as listed in Appendix Table 3.

### Image quality assessment

In addition to performing the measurements, the primary core lab observer performed a quality assessment for each of the images in Appendix Table 1. The image grading system was defined as:

1. Excellent: the full extent of the structure boundary of interest was clearly defined with no visible gaps, zoom-mode was activated to maximize image resolution for 2D structures and Doppler signals were scaled proportional to image size.
2. Good: the full extent of the structure boundary of interest was contained within the image sector with only brief gaps requiring interpolation and for Doppler signals there was minimal baseline artifact.
3. Fair: nearly the entire boundary of the structure of interest was contained and adequately visible but minor extrapolation beyond the imaging sector was required, boundaries had identifiable but indistinct borders and prominent baseline artifact was present on Doppler recordings.
4. Poor: significant portions of the boundary of interest required interpolation from visible but indistinct segments or more than a short segment of the boundary was outside of the imaging sector.
5. Unusable: the structure was not recorded or too poorly defined to be measured.

Additionally, each echocardiogram was also evaluated for a) trabeculations that could potentially interfere with definition of the LV apical endocardium; b) qualitative assessment of regional wall motion abnormalities, and c) septal displacement resulting in a non-circular short axis LV configuration.

## Clinical data collection

The medical records of all subjects were reviewed at the time of enrollment, at each subsequent echocardiogram, and at 18 months post-enrollment for medical history including medical therapy and changes in medical therapy, procedures, interventions, adverse events, and symptom and cardiac transplantation listing status.

## Statistical Analysis

The primary aim of this study was to determine the interstudy variability of echocardiographically-derived LV measurements in pediatric patients with DCM, and in particular, the variance of LV ejection fraction z-score at a single point in time and the variance of change in z-score between two time points. The required sample size to estimate the population standard deviation to within a prespecified tolerance is  $n = z^2_{1-\sigma/2} / (2d^2)$ , where  $d$  = allowed fractional deviation from  $\sigma$ , the population sample standard deviation.<sup>(7)</sup> To construct a two-sided 95% confidence interval for  $\sigma$  that deviates no more than 15% from the true value requires  $n = 86$ . It was estimated that 25–30% of subjects would withdraw, have echocardiograms conducted under differing sedation conditions, or have incomplete baseline/follow-up echocardiogram pairs for other reasons. Therefore, the target sample size was set to be 120 patients with qualifying baseline echocardiograms to ensure that 86 of these have paired interpretable echocardiograms (a total of 172 echocardiograms) performed under similar conditions. In actual execution, the study enrolled 173 patients, of whom 131 were eligible for longitudinal evaluation, to ensure that 86 qualifying echocardiogram pairs would be obtained. A total of 107 of these subjects had at least one follow-up echocardiogram submitted, and 97 of these formed qualifying echocardiogram pairs. The 10 subjects with non-qualifying pairs met one of the following criteria: no secondary image acquisition, incorrect sonographer, or inconsistent sedation status between the two echocardiograms.

## VVV Study Analysis Plan

Based on the study design shown in Appendix Table 3, for any given echocardiographic parameter, there were 21 or 15 sets of measurements from each study visit (3 sequential cardiac cycles  $\times$  7 different readings for a baseline visit and 5 different readings for a follow-up visit). The analyses of these data sets include the following comparisons of interest to assess:

### Interacquisition observer variability:

- 1) Images acquired by different observers and measured by the same observer

### Intraobserver variability:

- 2) Same set of images measured twice by the same person spaced by 1 month

### Intraobserver drift:

- 3) Same set of images measured twice by the same person spaced by 12 months

### Interobserver variability:

- 4) Same set of images, measurements by one core lab observer versus measurements by second core lab observer
- 5) Same set of images, measurements by core lab versus measurements by study center

### Changes over time in cardiac function holding acquisition observer and measurer constant:

- 6) Images acquired by the same observer and measured by the same observer

## Statistical Methods: Beat Averaging

The outcome measure was the percent (%) error of the mean. Echocardiographic measurements using the primary image acquisition from all baseline studies were included in analysis. For each evaluation of reproducibility between 2 measurements of the same entity, the difference ('error') between the 2 measurements was divided by the mean of those two measurements. Three settings were used to compare the following 2 measurements:

- Core laboratory inter-observer variability: Primary vs. secondary observer
- Core laboratory intra-observer variability: Primary observer immediate vs. 1-month reading
- Core vs. local laboratory inter-observer variability: Primary core laboratory observer vs. clinical center observer

We fit a mixed effects model with estimates obtained by restricted maximum likelihood with a compound symmetry covariance structure (fixed effect for beat averaging method and random effect for subject) to assess whether % error significantly differed for single- vs. 3-beat average and 2- vs. 3-beat average, for inter-observer, intra-observer and local vs. core laboratory comparisons. We fit a mixed effects model with unstructured covariance to assess whether, for each variable type, based on 3-beat-averaged measurements, inter-observer % error differed for local/core vs. core/core % error estimates, and for core lab inter- vs. intra-observer % error.

## Results

Patient evaluation and enrollment are summarized in Appendix Figure 2. A total of 275 subjects with known or suspected DCM were screened. Of those, 194 were eligible for the initial screening echocardiogram, 173 (89%) consented to participation and underwent data recording, and 131 were confirmed to have chronic DCM. There were an additional 38 subjects enrolled who did not meet criteria for LV dilation and/or dysfunction but who did not meet any exclusion criteria (partial participation). The data from these subjects were also included in the analysis of intra- and inter-observer variability.

The comparison of the 4 groups defined by eligibility status is presented in Table 1. Because only subjects who met criteria for significant DCM were fully eligible, ventricular size was markedly larger and ventricular function was markedly worse in the fully eligible group. The severity of disease in this group is apparent, with a mean LV end-diastolic dimension 4.9 standard deviations above, and a LV ejection fraction 5.0 standard deviations below, the normal mean value. A comparison of the etiology of cardiomyopathy in the partial and full participation is presented in Table 2. For both groups, idiopathic and adriamycin-associated cardiomyopathies were the most common etiologies. A comparison of clinical status (Ross classification for children under age 5 years and New York Heart Association [NYHA] status for children over age 5) is presented in Table 3. Despite the lesser severity of echocardiographic manifestations of cardiomyopathy in the partial participation group, the distribution of clinical severity was not significantly different. However, in subjects > 5 years old, qualitatively more partial participation subjects were in NYHA class I (72%) than were full participation subjects (59%).

Appendix Table 2 presents the yield with respect to successful core laboratory measurement of the echocardiographic parameters in this study. All parameters were measurable at least 94% of the time, with the exception of the time interval between mitral regurgitation velocity of 1 and 3 msec (58%), spectral tissue Doppler diastolic summation wave velocity (left, 88%; septal, 80%; right, 93%), and peak early diastolic velocity/peak mitral inflow

(left, 64%; septal 79%; average, 61%). This information may be a useful consideration in endpoint selection for future studies.

### Beat averaging: Analyses of core laboratory measurements

**Beat averaging and reproducibility**—Summary statistics for the % error for single-beat, 2- and 3-beat averaging methods for inter- and intra-observer reproducibility for all 150 variables are reported in Table 4. Overall, the magnitude of intra-observer % error was approximately half that of inter-observer % error ( $p < .001$ ). As the number of beats that were averaged was increased, both inter- and intra-observer % error decreased.

**Variable type and inter-observer reproducibility**—Inter-observer reproducibility was significantly affected by the type of echo variable and by beat-averaging method ( $p < 0.0001$ ). Figure 1 displays plots of mean inter-observer reproducibility by type of echo variable and by beat-averaging method. Measurements of slope and calculations involving 2, 3 and 4 measured variables had the lowest inter-observer reproducibility (highest % error), e.g., the median of median % error was 39% for slopes and 14% for variables calculated from 4 measurements.

**Variable type and intra-observer reproducibility**—Intra-observer reproducibility was significantly affected by the type of echo variable and by beat-averaging method ( $p < 0.0001$ ). Figure 2 displays plots of median intra-observer reproducibility by type of echo variable and by beat-averaging method. Measurements of slope and calculations involving 4 measured variables had the lowest intra-observer reproducibility (highest % error), e.g., the median of median % error was 16% for slopes and 5% for measurements calculated from 4 measured variables.

**Variables with best and worst reproducibility**—Even when 3-beat averaging was used, a wide range of % error was found for both inter- and intra-observer analyses (Figure 3). The 20 variables with the highest % error had inter-observer mean % error ranging from 17.7 to 46.7%, while intra-observer mean % error ranged from 8.6 to 23.2% (Table 5). Of these 20 variables, 13 were common to both inter- and intra-observer categories. Calculated variables and tissue Doppler measurements of isovolumic acceleration consistently exhibited large % errors.

Of the variables with the lowest inter- and intra-observer % error, many were measurements of heart rate. These were used as ‘internal controls’ and were not analyzed further. Table 6 lists the 20 variables with the lowest % error after excluding heart rate measurements (range 0.9% to 4.8% for inter-observer and 0.33% to 2.1% for intra-observer). Of these 20 variables, 9 were common to both inter- and intra-observer categories. Measurements of LV internal chamber dimensions and traced areas consistently exhibited low % errors.

**Comparison of core laboratory to local measurements**—Summary statistics for the % error for single-beat, 2- and 3-beat averaging methods for core vs. local-observer reproducibility for all 119 variables are reported in Table 7. Overall % error decreased as the number of averaged beats increased. Figure 3 and 4 display the reproducibility of variables between a single observer at the core laboratory, between 2 observers at the core laboratory, and between a core lab observer and a single observer at the local center, plotting % error by type of echo variable for the 3 beat-averaging method. As previously noted, measurements of slope and calculated variables exhibited the greatest % error. Despite 3-beat averaging, many variables exhibited high % error (Table 8, left column). Mean % error among the 20 variables with the highest % error ranged from 23.0% to 56.4%. Calculated variables, tissue Doppler measurements of isovolumic acceleration and M-mode measurements of LV wall



thickness exhibited high % errors. Of these 20 variables, 11 also appeared in the list of 20 variables with the highest % inter-observer error (core laboratory, Table 5). These consisted of tissue Doppler measurements of isovolumic acceleration, and calculated variables derived from 4 variables. Of the variables with the lowest inter- and intra-observer % error, many were measurements of heart rate. Table 8 (right column) lists the 20 variables with the lowest % error after excluding heart rate measurements. Of these 20 variables, 18 also appeared in the list of 20 variables with the lowest % inter-observer error (core laboratory, Table 6). Measurements of LV internal chamber dimensions and traced areas consistently exhibited low % errors.

**Comparison of inter- to intra-observer reproducibility**—Table 9 summarizes the 3-beat averaged median % error by class of echo variable, using the classification of variables that was detailed earlier. For all categories of variables, core lab intra-observer % error was significantly lower than core lab inter-observer % error ( $p=0.004$  for slopes and  $p<0.001$  for all others). For all categories of variables, core lab inter-observer % error was significantly lower than local vs. core lab % error ( $p=0.018$  for the ‘integral’ variable type and  $p<0.001$  for all others).

**Optimization of laboratory design**—Having observed that a single observer design improves reproducibility for all variables, we evaluated whether inter-observer variability was significantly less for core-versus-core comparison (figure 3) compared to core-versus-local comparison (figure 4). Overall, core lab, same reader %error is 52% that of the %error for core lab, 2 reader model (6.37 vs. 12.27). Core lab, 2-reader %error is 78% that of core-versus-local model (12.27 versus 15.69). Thus, the greatest improvement in reproducibility was achieved with a single-reader core lab model with more modest benefits associated with a multi-reader core lab model.

## Discussion

### Selection of patients

The primary motivation for this observational study was to provide the information required to design a medical interventional trial for DCM in children. Because of the high frequency of early death or complete recovery in children with acute onset DCM (8,9), we targeted enrollment of children with chronic, relatively stable disease, excluding acute onset disease and patients listed or likely to be listed for transplantation. In order to optimize the likelihood of detecting improved outcomes, we excluded patients with known confounders of assessment of ventricular function, including paced rhythm, noncompaction, and complex congenital heart disease. We also excluded co-morbid conditions likely to limit survival independent of cardiomyopathy. We chose to include the echocardiograms performed at initial evaluation in the subjects whose ventricular function was too normal to justify inclusion in the longitudinal evaluation. While these echocardiograms did not contribute to the study aim of assessing interstudy variability (within-subject change over time), including these subjects expanded the range of ventricular size and function that is included in the reproducibility analysis. This allows us to address the issue of reproducibility over the full clinical spectrum, and strengthens the analysis of the impact of disease severity on reproducibility.

### Selection of endpoints

Because the annual incidence of DCM in children is 10-fold lower than in adults with a 4-year freedom from death or transplant of 55% (10), it would be difficult to recruit a sufficiently large study population to detect a significant reduction in these hard endpoints in a reasonable time frame. Indeed, despite including 8 clinical centers it required 2.25 years to

enroll 131 subjects into the current observational study. Echocardiographic measures that are correlated with clinical events can be valuable surrogate outcomes that provide enhanced statistical power, and are particularly useful in shorter-term studies. However, published information critical to achieving accurate sample size calculations is lacking. The literature contains some estimates of the variance of change in LV ejection fraction, but they were based on adult data. There are estimates of the variance of LV ejection fraction in children, but little from longitudinal data that would provide information on the correlation of measurements over time. We also recognized that the standard deviation estimates ranged widely depending on whether ejection fraction was measured locally or centrally. Hence, the current study was designed to obtain an estimate of variance for change in LV ejection fraction in children with DCM, and to investigate in detail the relative contributions of numerous factors to the variance. The goal was to then use this information in the design of future pediatric trials.

### **Reproducibility of echocardiographic endpoints**

Any variance in end-point assessment impedes detection of treatment effect, emphasizing the importance of maximizing reproducibility of echocardiographic endpoints. It is common practice in clinical trials to make use of core laboratories, standardized protocols, and personnel training to maximize reproducibility. Despite this, available adult data indicate that the magnitude of the technique-related variance imposes a significant penalty in terms of sample size requirements for studies based on echocardiographic endpoints. For example, echocardiographic assessment of m-mode LV mass is reported to have intra-sonographer variance of 10%, inter-sonographer variance of 10%, intra-observer variance of 10% that increases to 16% when the same observer repeats the interpretation 5 years later, inter-observer variance of 14%, and a frequency of non-measurable echocardiograms of 33%.<sup>(11)</sup> The reliability of the measurements is inversely related to age and body mass index, predicting a potentially better performance of the technology in children. Although studies comparing core laboratory versus clinical center interpretation of echocardiograms in children are available, studies examining the comparative performance of the full spectrum of indices of ventricular function and the relative importance of the various sources of variance have not been performed. (1) In addition, the technical advances in echocardiographic equipment and the general adoption of DICOM image storage might well diminish the relevance of these older studies.

### **Sources of measurement variability**

Some sources of variability are nearly impossible to quantify or control, such as variance secondary to differences between machines, transducers, and machine settings, due to the variety of ways in which these factors can be combined. In contrast, there are other sources of variability that can potentially be reduced through study design modifications, such as use of a core laboratory or expanded data collection to yield beat-averaged measurements. These design modifications invariably increase costs, and the cost versus benefit analysis that is needed involves comparing the reduction in cost related to patient enrollment versus the cost escalation related to end-point determination. To that end, we designed this study to assess inter-observer, intra-observer, and inter-acquisition reproducibility, in addition to the impact of controllable factors such as beat averaging and potentially controllable factors such as image quality, as well as non-controllable factors such as age, body mass index, and severity of disease.

### **Beat averaging**

The impact of beat averaging on the reproducibility of echocardiographic measurements has not been evaluated comprehensively in either adults or in children. While it is intuitive that averaging multiple beats should lead to improved reproducibility, this has logistic

implications that are increasingly relevant in practice. Studies in adult patients have demonstrated beat-to-beat variations in echocardiographic measurements of cardiac dimensions, volumes and Doppler waveforms based on technologies for image acquisition, storage and analysis which, while appropriate to the era, would be considered dated today. (12–16) These studies demonstrated that respiratory effects on ventricular filling were an important physiologic source of beat-to-beat variation and concluded that in order to minimize the impact of beat-to-beat variability, measurements should be made using breath holding, or be performed at end-expiration, or that beat averaging is necessary. Neither breath holding nor acquisitions that coincide with end-expiration are practical alternatives in multi-center studies, particularly in children who may be unable to cooperate. We elected to evaluate the impact of averaging up to 3 beats because at rest, the respiratory cycle generally encompasses three cardiac cycles. In recent years, the wide availability of echocardiography as a bedside tool has been coupled with an expansion of the number and type of measurements that can be made on an echocardiogram. Interestingly, two recent randomized clinical trials have demonstrated that the effects of medications such as carvedilol and enalapril on echocardiographic measures of cardiac function in children vary significantly from those seen in adults. (17,18) Together, these studies point to the need for evaluating whether, in the contemporary era, factors such as the number of observers, beat averaging and variable type have an impact on the reproducibility of echocardiographic measurements that are in clinical use in children with DCM.

The current study revealed a wide range of % error for echocardiographic measurements, with high % errors for measurements of slope and variables that are derived from 2, 3 or 4 measured quantities. The high % error of local versus core laboratory measurements for all types of variables points to the importance of a core laboratory for performing measurements from the perspective of reproducibility of measurements in multicenter studies. In addition, even within the standardized framework of a core laboratory, in order to enhance reproducibility, a single core lab observer is preferable to multiple core lab observers. Similarly, single-beat measurement leads to high % error, which decreases with the use of beat averaging. These findings have important implications for the structure, design and logistics of core laboratory analysis of echocardiographic measurements. The reproducibility of measurements of slopes and calculated echocardiographic variables in the current study is remarkably lower than that reported from single-center studies. (19,20) One potential explanation is that in the current study, all core laboratory measurements were performed using a single echocardiographic measurement program but the image acquisition and the measurements performed at the clinical sites were performed using whatever platform was in use at their location. This approach was used because the use of vendor-specific technologies for image acquisition and analysis imposes fundamental limitations on their applicability to multi-center clinical trials. In an era of increasing logistic demands, the automation of echocardiographic measurements using vendor-independent computerized algorithms may provide for more efficient beat averaging and, eventually, potentially improve their reproducibility. (21–23) While measurements of slopes and calculated variables may play an important role in clinical practice, these findings sound a cautionary note in the selection of such variables as echocardiographic endpoints in clinical trials.

### Limitations

Modalities such as 3D echocardiography and myocardial deformation imaging were not studied because they were not widely available at the time that the study commenced (2005). We did not examine other sources of variability; these will be the topic of future analyses of this project.

## Conclusions

In a multi-center study, 3-beat averaging, the use of a core laboratory and a single observer yield better reproducibility for echo measurements in children with DCM. Despite beat averaging, measurements of slope and some calculated variables remain poorly reproducible. In contrast, measurements of ventricular chamber dimensions, traced areas and time intervals are highly reproducible. These findings have implications for study design and power, choice of endpoint and core lab structure in clinical trials of pediatric heart disease.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

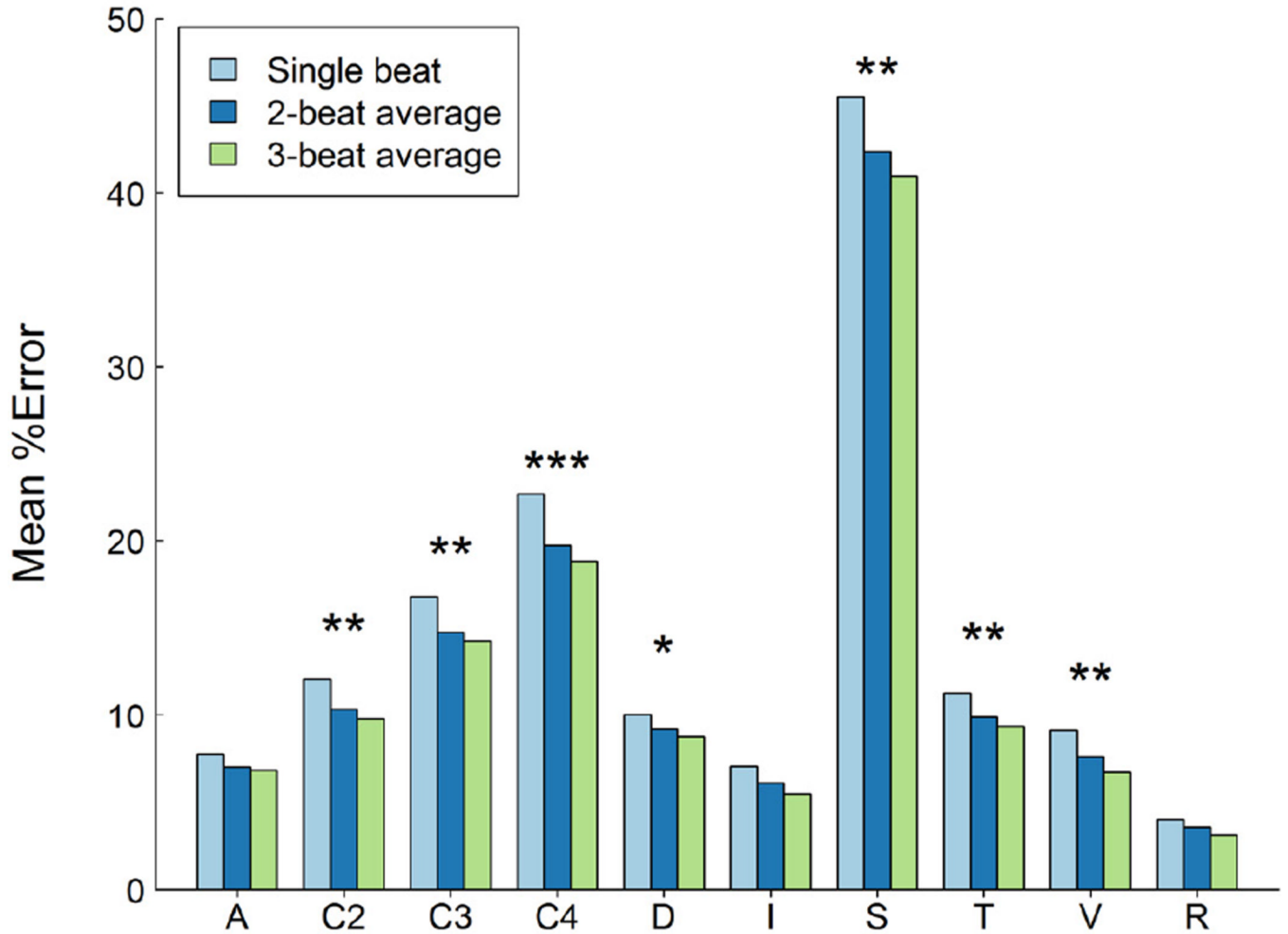
Supported by U01 grants from the National Heart, Lung, and Blood Institute (HL068269, HL068270, HL068279, HL068281, HL068285, HL068292, HL068290, HL068288). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NHLBI.

## References

1. Lipshultz SE, Easley KA, Orav J, Kaplan S, Starc TJ, Bricker JT, et al. Reliability of multicenter pediatric echocardiographic measurements of left ventricular structure and function - The prospective P<sup>2</sup>C<sup>2</sup>HIV study. *Circulation*. 2001; 104:310–316. [PubMed: 11457750]
2. Fleming TR. Surrogate end points in cardiovascular disease trials. *Am Heart J*. 2000; 139(4):S193–S196. [PubMed: 10740130]
3. Gheorghiade M, Adams KF Jr, Gattis WA, Teerlink JR, Orlandi C, O'Connor CM. Surrogate end points in heart failure trials. *Am Heart J*. 2003; 145(2):S67–S70. [PubMed: 12594457]
4. Fogel MA. Use of ejection fraction (or lack thereof), morbidity/mortality and heart failure drug trials: a review. *Int J Cardiol*. 2002; 84(2–3):119–132. [PubMed: 12127364]
5. Norgård G, Johannessen KA. Variability of digitized left ventricular M-mode echocardiography: a study in healthy subjects and patients with repaired tetralogy of Fallot. *Clin Physiol*. 1993 Jul; 13(4):373–383. [PubMed: 8370237]
6. Haycock GB, Schwartz GJ, Wisotsky DH. Geometric method for measuring body surface area: a height-weight formula validated in infants, children, and adults. *J Pediatr*. 1978; 93:62–66. [PubMed: 650346]
7. Thompson WA, Andriess J. The required sample size when estimating variances. *The American Statistician*. 1961; 15:22–23.
8. Matitiau A, Perez-Atayde A, Sanders SP, Sluysmans T, Parness IA, Spevak PJ, et al. Infantile dilated cardiomyopathy: Relation of outcome to left ventricular mechanics, hemodynamics, and histology at the time of presentation. *Circulation*. 1994; 90:1310–1318. [PubMed: 8087940]
9. Lewis AB. Late recovery of ventricular function in children with idiopathic dilated cardiomyopathy. *Am Heart J*. 1999; 138:334–338. [PubMed: 10426848]
10. Towbin JA, Lowe AM, Colan SD, Sleeper LA, Orav EJ, Clunie S, et al. Incidence, causes, and outcomes of dilated cardiomyopathy in children. *J Am Med Assoc*. 2006; 296:1867–1876.
11. Gardin JM. How reliable are serial echocardiographic measurements in detecting regression in left ventricular hypertrophy and changes in function? *J Am Coll Cardiol*. 1999; 34:1633–1636. [PubMed: 10551716]
12. Assmann PE, Slager CJ, van der Borden SG, Dreyse ST, Tijssen JG, Sutherland GR, Roelandt JR. Quantitative echocardiographic analysis of global and regional LV function: a problem revisited. *J Am Soc Echocardiogr*. 1990 Nov-Dec;3(6):478–487. [PubMed: 2278713]
13. Wallerson DC, Devereux RB. Reproducibility of echocardiographic LV measurements. *Hypertension*. 1987 Feb; 9(2 Pt 2):II6–II18. [PubMed: 3542819]

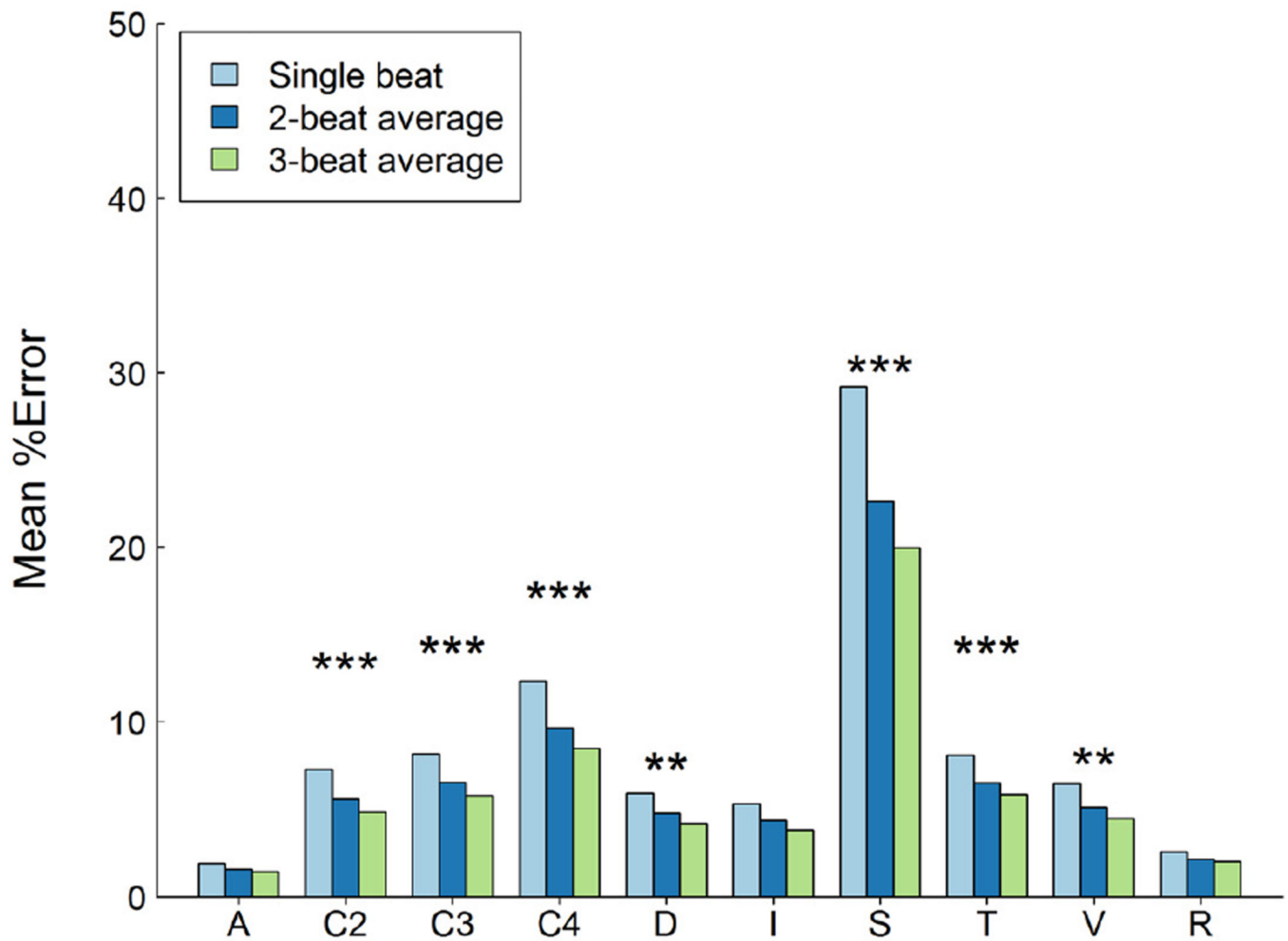
14. Gordon EP, Schnittger I, Fitzgerald PJ, Williams P, Popp RL. Reproducibility of LV volumes by two-dimensional echocardiography. *J Am Coll Cardiol*. 1983 Sep; 2(3):506–513. [PubMed: 6875114]
15. Rijsterborgh H, Mayala A, Forster T, Vletter W, van der Borden B, Sutherland GR, Roelandt J. The reproducibility of continuous wave Doppler measurements in the assessment of mitral stenosis or mitral prosthetic function: the relative contributions of heart rate, respiration, observer variability and their clinical relevance. *Eur Heart J*. 1990 Jul; 11(7):592–600. [PubMed: 2373094]
16. Morin DP, Cottrill CM, Johnson GL, Wilson HD, Vine DL, Noonan JA. Effect of respiration on the echocardiogram in children with cystic fibrosis. *Pediatrics*. 1980 Jan; 65(1):44–49. [PubMed: 7355035]
17. Shaddy RE, Boucek MM, Hsu DT, Boucek RJ, Canter CE, Mahony L, Ross RD, Pahl E, Blume ED, Dodd DA, Rosenthal DN, Burr J, LaSalle B, Holubkov R, Lukas MA, Tani LY. Pediatric Carvedilol Study Group. Carvedilol for children and adolescents with heart failure: a randomized controlled trial. *JAMA*. 2007 Sep 12; 298(10):1171–1179. [PubMed: 17848651]
18. Hsu DT, Zak V, Mahony L, Sleeper LA, Atz AM, Levine JC, Barker PC, Ravishankar C, McCrindle BW, Williams RV, Altmann K, Ghanayem NS, Margossian R, Chung WK, Border WL, Pearson GD, Stylianou MP, Mital S. Pediatric Heart Network Investigators. Enalapril in infants with single ventricle: results of a multicenter randomized trial. *Circulation*. 2010 Jul 27; 122(4):333–340. [PubMed: 20625111]
19. Cheung MM, Smallhorn JF, Vogel M, Van Arsdell G, Redington AN. Disruption of the ventricular myocardial force-frequency relationship after cardiac surgery in children: noninvasive assessment by means of tissue Doppler imaging. *J Thorac Cardiovasc Surg*. 2006 Mar; 131(3):625–631. [PubMed: 16515915]
20. Lyseggen E, Rabben SI, Skulstad H, Urheim S, Risoe C, Smiseth OA. Myocardial acceleration during isovolumic contraction: relationship to contractility. *Circulation*. 2005 Mar 22; 111(11):1362–1369. [PubMed: 15753217]
21. Zwehl W, Levy R, Garcia E, Haendchen RV, Childs W, Corday SR, Meerbaum S, Corday E. Validation of a computerized edge detection algorithm for quantitative two-dimensional echocardiography. *Circulation*. 1983 Nov; 68(5):1127–1135. [PubMed: 6616792]
22. Park J, Zhou SK, Jackson J, Comaniciu D. Automatic mitral valve inflow measurements from Doppler echocardiography. *Med Image Comput Comput Assist Interv*. 2008; 11(Pt 1):983–990. [PubMed: 18979841]
23. Melo SA Jr, Macchiavello B, Andrade MM, Carvalho JL, Carvalho HS, Vasconcelos DF, Berger PA, da Rocha AF, Nascimento FA. Semi-automatic algorithm for construction of the LV area variation curve over a complete cardiac cycle. *Biomed Eng Online*. 2010 Jan 15; 9:5. [PubMed: 20078864]

### Inter-reader Reproducibility

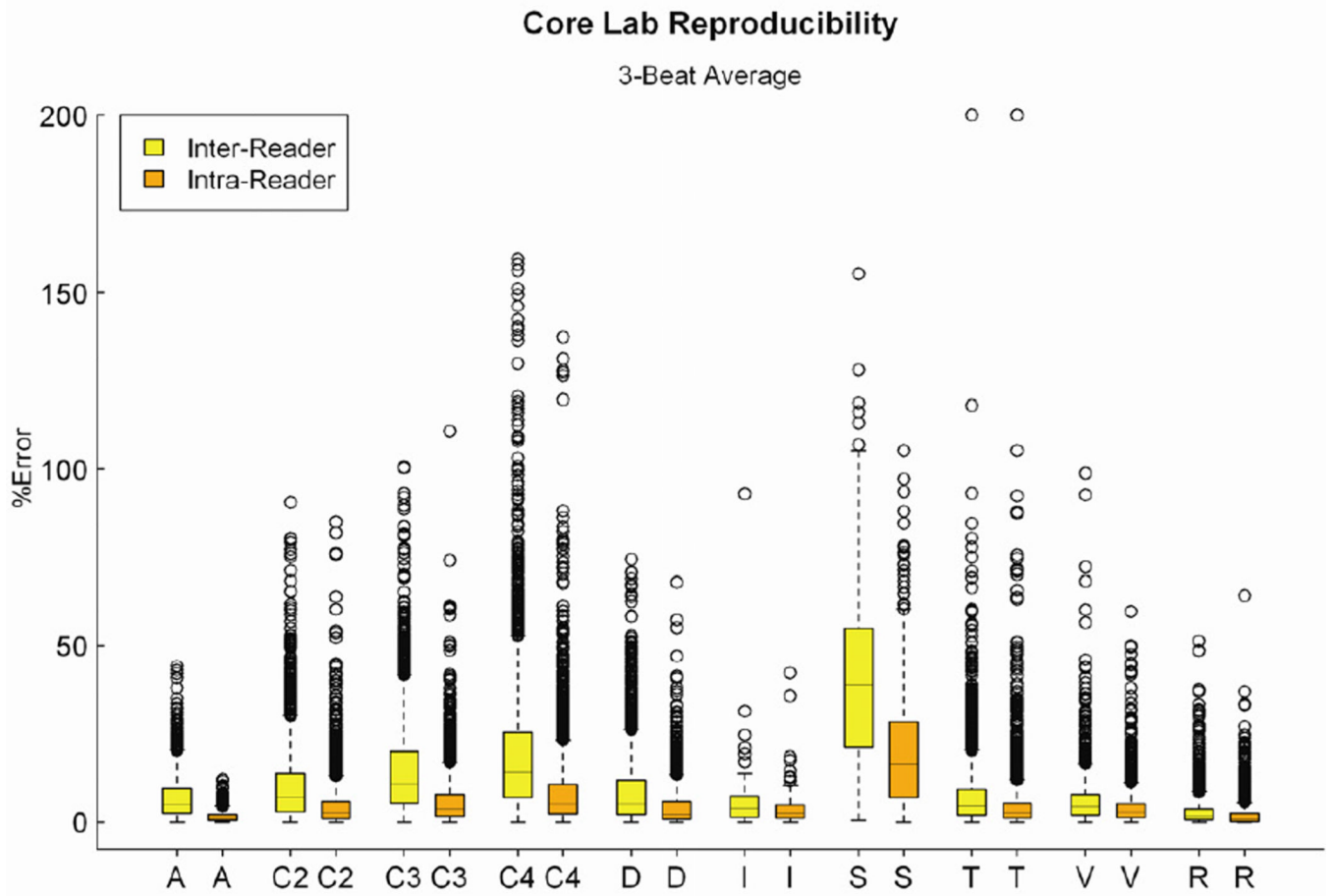


**Figure 1.** Comparison of the mean core laboratory inter-observer % error (on the Y axis) for single-beat, 2- and 3-beat averages for all 150 variables. These have been classified using the schema detailed in the text, where A = area; C2, C3, C4 = calculated from 2, 3 and 4 variables respectively; D = dimension; I = integral; S = slope; T = Doppler time interval; V = velocity, and R = heart rate. Measurements of slope and calculated variables exhibit the greatest % error. A single asterisk means that only 1 versus 3 beat analysis yields a statistically significant difference (Bonferroni testing), two asterisks mean 1 versus 2 and 1 versus 3 are different, and 3 asterisks means that 1 versus 2, 2 versus 3, and 1 versus 3 are all different.

### Intra-reader Reproducibility

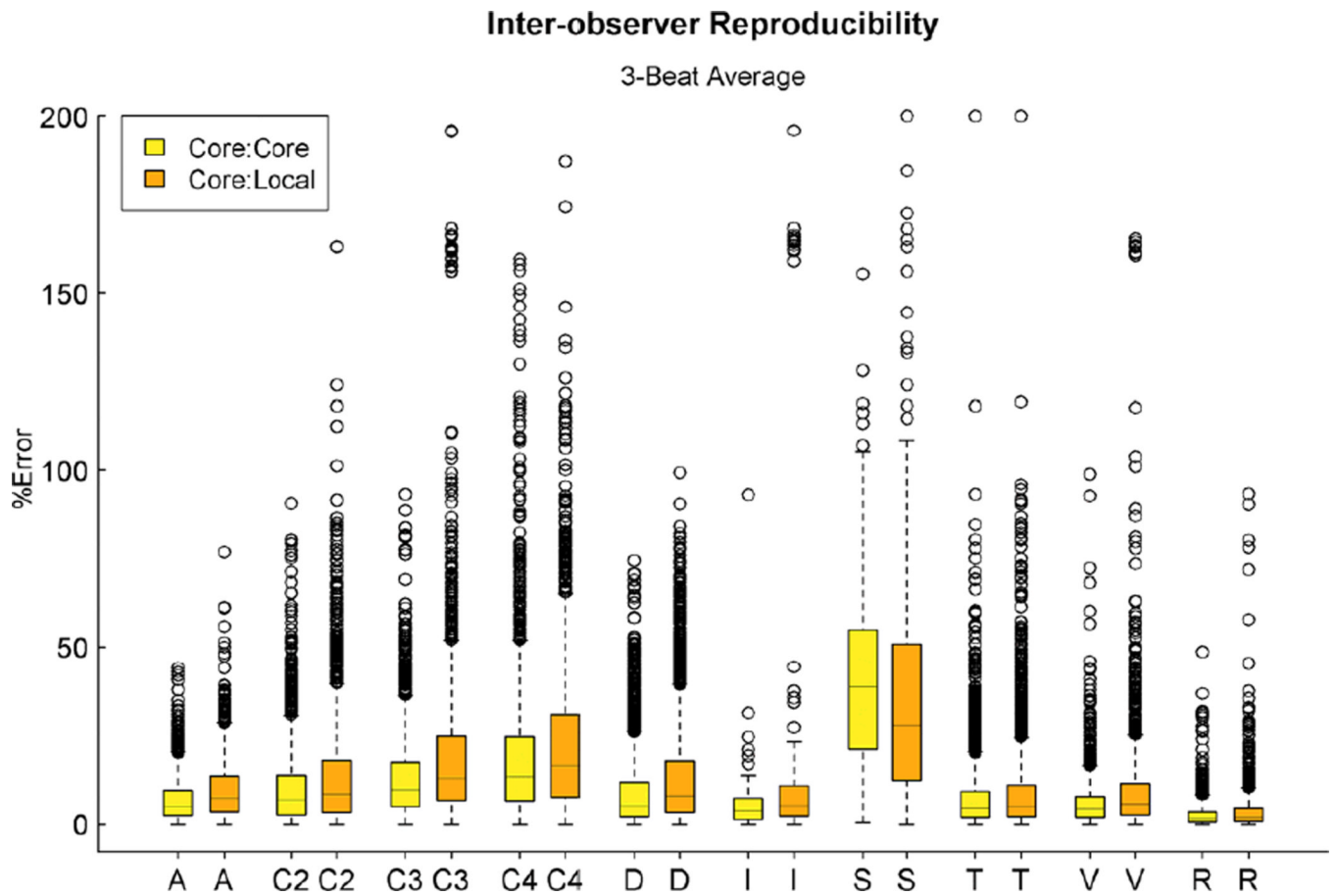


**Figure 2.** Comparison of the mean core laboratory intra-observer % error (on the Y axis) for single-beat, 2- and 3-beat averages for all 150 variables, classified using the schema detailed in Figure 1. The pattern of percent error was parallel to that seen in inter-observer % error shown in Figure 1. Significance is denoted using asterisks as described in figure 1. The relative magnitude of the inter- versus intra-observer % error is illustrated in Figure 3.



**Figure 3.** Distributions for the raw data for core laboratory inter- and intra-observer % error for 3-beat averages of all 150 variables, classified into the categories described in figure 1. Each ‘box and whiskers’ plot displays the median, upper and lower quartiles and upper and lower extreme values of the % error. There was a significant upward skew of the distribution for the majority of the variable types. There were wide ranges of % error for individual variables, of the difference in intra and inter-observer percent error, and of the relative frequency and magnitude of the outliers.





**Figure 4.** Distributions for the raw data for local center versus corelab inter-observer variability and corelab versus corelab interobserver variability for the 119 variables measured at the local center, using the classification schema detailed in Figure 1. Each ‘box and whiskers’ plot displays the median, upper and lower quartiles and upper and lower extremevalues of the % error.

**Table 1**

Patient Characteristics by Eligibility and Participation Status

Characteristic	Not Eligible*	Not Enrolled	Partial Participation	Full Participation	P-value**
<b>N</b>	85	21	38	131	
<b>Mean age at screening echo, yr</b>	9.2±6.0	12.6±4.8	9.8±5.4	9.3±6.0	0.05
Male	10.0±5.8	11.6±5.5	10.7±4.8	9.9±6.1	0.65
Female	8.1±6.2	13.6±4.0	8.8±6.1	8.7±6.0	0.04
<b>Male</b>	58.8%	47.6%	52.6%	44.3%	0.65
<b>Race</b>					0.55
White	78.4%	66.7%	68.4%	65.6%	
Black or African American	14.9%	19.0%	23.7%	27.5%	
Asian	6.8%	4.8%	5.3%	5.3%	
Other	0.0%	9.5%	2.6%	1.5%	
<b>Hispanic</b>	6.5%	0.0%	8.6%	15.2%	0.19
<b>Echocardiographic Inclusion Criteria (local measurements)</b>					
LV end-diastolic dimension (EDD)					
Mean z-score			1.7±1.0	4.9±2.8	<.001
Median (Q1, Q3) z-score			1.6(0.9,2.1)	4.1(2.8,6.3)	<.001
EDD > 5.5 cm			10.5%	47.3%	<.001
LV shortening fraction (SF)					
Mean z-score			-2.7±2.9	-8.2±5.7	<.001
Median (Q1, Q3) z-score			-2.0(-4.3,-1.0)	-6.9(-11, -4.4)	<.001
Mean SF (%)			28.6±7.1	19.1±6.7	<.001
SF < 28%			44.7%	93.1%	<.001
LV ejection fraction (EF)					
Mean z-score			-2.1±1.8	-5.0±2.3	<.001
Median (Q1, Q3) z-score			-1.8(-3.2, -0.7)	-4.7(-6.6, -3.1)	<.001
Mean EF (%)			52.2±9.1	37.5±11.7	<.001
EF <50%			34.2%	85.4%	<.001

\* Includes 81 patients ineligible by chart review and 4 patients ineligible by baseline echocardiogram

\*\*# p-value denotes comparison of Not Enrolled vs. Partial Participation vs. Full Participation

LV=LV; Q1 = first quartile; Q3=third quartile

**Table 2**

Primary cause of DCM in 169 enrolled subjects.

<b>Variable</b>	<b>Partial Participation</b>	<b>Full Participation</b>	<b>P-value</b>
<b>Etiology</b>			0.05
Idiopathic	44.7%	66.4%	
Adriamycin-associated	28.9%	10.7%	
Neuromuscular disease	5.3%	3.1%	
Single gene defect	2.6%	3.1%	
Metabolic disorder	5.3%	1.5%	
Mitochondrial disorder	0.0%	1.5%	
Other	13.2%	13.7%	
Known or suspected familial Cardiomyopathy	18.4%	16.0%	0.81

**Table 3**

Congestive heart failure classification at baseline in 169 enrolled subjects.

	<b>Partial Participation</b>	<b>Full Participation</b>	<b>P- value</b>
<b>Ross Classification (age &lt; 5 years)</b>	<b>9</b>	<b>45</b>	1.00
I. No limitation or symptoms	77.8%	77.8%	
II. Symptoms but no growth failure	11.1%	13.3%	
III. Growth failure & prolonged feeding time in infants	11.1%	8.9%	
IV. Growth failure & symptomatic at rest	0.0%	0.0%	
<b>NYHA Classification (Age 5 years)</b>	<b>29</b>	<b>86</b>	<b>0.46</b>
I. No limitation of activities; no symptoms from ordinary activities	72.4%	59.3%	
II. Slight, mild limitation of activities; comfortable with rest or mild exertion	27.6%	38.4%	
III. Marked limitation of activity; comfortable only at rest	0.0%	2.3%	
IV. At complete rest; any physical activity brings on discomfort and symptoms occur at rest	0.0%	0.0%	

NYHA = New York Heart Association

**Table 4**

Percent (%) error for single-beat and 2- and 3-beat averaging methods for both inter- and intra-observer reproducibility for all 150 variables.

Averaging Method	Range (Median)	
	Inter-observer % error	Intra-observer % error
1-beat	1.1 – 56.9 (11.4)	0.53 – 32.3 (6.3)
2-beat average	0.97 – 49.4 (10.2)	0.41 – 25.8 (4.9)
3-beat average	0.9 – 46.7 (9.3)	0.33 – 23.2 (4.2)

Table 5

Core laboratory measurements: Twenty variables with the highest mean inter- and intra-observer % error with 3-beat averaging.

Variable name	Inter-observer mean % error	Variable type	% error	Variable name	Intra-observer mean % error	Variable type	% error
Septal STD isovolumic contraction period acceleration *		Tissue Doppler, slope	46.7	Left STD isovolumic contraction period acceleration *		Tissue Doppler, slope	23.2
Mitral valve Tei index *		M-mode (4)	44.6	Duration of pulmonary vein flow reversal during atrial systole *		Doppler, time	22.4
Left STD isovolumic contraction period acceleration *		Tissue Doppler, slope	43.9	Septal STD isovolumic contraction period acceleration *		Tissue Doppler, slope	21.3
Right STD isovolumic contraction period acceleration *		Tissue Doppler, slope	38.7	Mitral valve Tei index *		M-mode (4)	20.6
Duration of pulmonary vein flow reversal during atrial systole *		Doppler, time	37.0	LV flow propagation velocity *		Color M-mode, slope	17.8
LV flow propagation velocity *		Color M-mode, slope	34.7	Right STD isovolumic contraction period acceleration *		Tissue Doppler, slope	17.6
Systemic resistance, biapical *		2D (4)	22.8	Mitral valve early deceleration time *		Doppler, time	15.8
Biplane apical cardiac output *		2D (3)	22.3	Mitral and aortic valve separate Tei index *		Doppler, time (4)	15.5
Biplane apical cardiac index		2D (3)	22.3	Right STD Tei index *		Tissue Doppler, time (4)	15.3
Biplane apical LV mass : volume ratio		2D (4)	22.1	Left STD Tei index *		Tissue Doppler, time (4)	13.3
Mitral and aortic valve separate Tei index *		Doppler, time (4)	21.0	LV velocity of circumferential fiber shortening *		2D (4)	11.5
Right STD Tei index *		Tissue Doppler, time (4)	21.0	Mitral and aortic valve simultaneous Tei index		Doppler, time (4)	11.1
Mitral valve early deceleration time *		2D (4)	20.9	LV fractional shortening		2D (2)	10.3
Biplane apical LV stroke volume		Doppler, time	20.8	Septal STD Tei index		Tissue Doppler, time (4)	9.6
LV velocity of circumferential fiber shortening *		2D (4)	18.8	Time between mitral regurgitation velocity of 1 and 3 m/sec		Doppler, time	9.5
LV end-systolic fiberstress		M-mode (3)	18.7	Biplane apical systemic vascular resistance *		2D (4)	9.4
Mod-Simpsons LV mass : volume ratio		2D (4)	18.3	LV velocity of circumferential fiber shortening		M-mode (4)	9.1
Mod-Simpsons systemic vascular resistance		2D (4)	18.2	LV thickness to dimension ratio		M-mode (2)	8.9
Left STD Tei index *		Tissue Doppler, time (4)	18.1	LV end-systolic meridional wall stress		2D (3)	8.9
Area-length systemic vascular resistance		2D (4)	17.7	Biplane apical cardiac output *		2D (3)	8.6

2D = 2-dimensional; LV = left ventricle; STD = spectral tissue Doppler. For calculated variables, the number of variables used in the calculation is in parentheses.

An asterisk '\*' marks the variables that feature among the 20 variables for both the highest inter-observer and the highest intra-observer % error at the core laboratory.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



Table 6

Core laboratory measurements: Twenty variables with the lowest inter- and intra-observer % error with 3-beat averaging.

Variable name	Inter-observer mean % error	Variable type	% error	Variable name	Intra-observer mean % error	Variable type	% error
LV function: Eccentricity index		2D	0.90	LV function: Eccentricity index		2D	0.33
Septal STD: Isovolumic contraction onset to isovolumic relaxation end*		Tissue Doppler, time	3.2	LV ED 4-ch epicardial long axis dimension*		2D	0.67
Mitral valve peak early diastolic velocity		Doppler, velocity	3.2	LV ED 4-ch endocardial long axis dimension*		2D	0.71
Aortic valve ejection time		Doppler, time	3.6	LV ES 4-ch endocardial long axis dimension*		2D	0.80
LV ED 4-ch long axis epicardial dimension*		2D	3.7	LV ED short axis epicardial area*		2D	0.89
LV ED short axis dimension*		2D	3.8	LV ED short axis endocardial area		2D	0.93
LV ED short axis dimension*		M-mode	3.8	LV ES short axis endocardial area		2D	1.0
Mitral valve closure time		Doppler, time	3.9	Area-length end-diastolic volume		2D (2)	1.2
LV ES short axis dimension*		2D	4.0	LV ED 4-ch long axis epicardial area*		2D	1.3
LV ED 4-ch endocardial long axis dimension*		2D	4.0	Area-length end-systolic volume		2D, 2	1.3
Simultaneous mitral and aortic valve Doppler: Ejection time		Doppler, time	4.2	LV ED 4-ch long axis endocardial area		2D	1.4
Left STD isovolumic contraction onset to isovolumic relaxation end		Tissue Doppler, time	4.2	LV ES 4-ch long axis endocardial area (2D)		2D	1.5
Aortic valve: Peak velocity		Doppler, velocity	4.3	LV ED short axis dimension*		M-mode	1.5
LV ES 4-ch endocardial long axis dimension*		2D	4.4	LV ED short axis dimension*		2D	1.6
Septal STD ejection time		Tissue Doppler, time	4.5	LV ED 2-ch long axis epicardial area		2D	1.7
Aortic annulus diameter		2D	4.5	LV ES short axis dimension		M-mode	1.8
Right STD isovolumic contraction onset to isovolumic relaxation end		Tissue Doppler, time	4.6	LV sphericity index		2D, 2	1.9
Right STD peak systolic velocity		Tissue Doppler, velocity	4.7	LV ES short axis dimension*		2D	1.9
LV ED short axis epicardial area*		2D	4.7	Mod-Simpsons end-diastolic volume		2D, 2	2.0
LV ED 4-ch long axis epicardial area*		2D	4.8	LV ED 2-chamber long axis endocardial area		2D	2.1

2D = two-dimensional; 2-ch = 2-chamber; 4-ch = 4-chamber; ED = end-diastolic; ES = end-systolic; LV = left ventricle; M = M-mode; STD: spectral tissue Doppler. For calculated variables, the number of variables used to calculate these is in parentheses. Measurements of heart rate have been excluded from this table.

An asterisk "\*" marks the variables that feature among the 20 variables for both the lowest inter-observer and the lowest intra-observer % error at the core laboratory.

**Table 7**

Comparison of local to core laboratory measurements. Percent (%) error for single-beat and 2- and 3-beat averaging methods for reproducibility for all 119 variables.

<b>Averaging Method</b>	<b>Range (median)</b>
1-beat	1.7 – 58.4 (14.7)
2-beat average	1.6 – 56.9 (12.8)
3-beat average	1.6 – 56.4 (12.0)

**Table 8**

Comparison of local to core laboratory measurements, listing the 20 variables with the highest and lowest % error with 3-beat averaging.

Variable name	Variable type	% error	Variable name	Variable type	% error
Duration of pulmonary vein flow reversal during atrial systole*	Doppler, time	56.4	LV function: Eccentricity index	2D	1.6
Mitral plus aortic valve: Tei index*	M-mode (4)	39	LV ED short axis dimension**	M-mode	4.3
Left STD isovolumic contraction acceleration*	Tissue Doppler, slope	37.4	Aortic annulus diameter**	2D	4.5
Septal STD isovolumic contraction acceleration*	Tissue Doppler, slope	36.6	Aortic valve peak velocity**	Doppler velocity	5.1
Right STD isovolumic contraction acceleration*	Tissue Doppler, slope	35.3	LV ED short axis dimension**	2D	5.3
LV flow propagation velocity*	Color M-mode, slope	34.5	LV ES short axis dimension	M-mode	5.5
LVES meridional wall stress	2D (3)	31.6	LV ED short axis epicardial area**	2D	5.5
LVES fiberstress*	M-mode (3)	31.6	Mitral valve peak early velocity**	Doppler velocity	5.6
LV thickness to dimension ratio	M-mode (2)	29.2	LV ED 4-ch endocardial long axis dimension**	2D	5.8
Mitral and aortic valve separate Tei index*	Doppler, time (4)	26.8	Right STD: isovolumic contraction onset to isovolumic relaxation end**	Tissue Doppler, time	5.8
LVED wall thickness	M-mode	26.2	Aortic valve ejection time**	Doppler, time	5.9
LV thickness to dimension ratio	2D (2)	25.6	Mitral and aortic valve simultaneous Doppler: ejection time**	Doppler, time	5.9
Right STD Tei index*	Tissue Doppler time (4)	25.5	Septal STD: isovolumic contraction onset to isovolumic relaxation end**	Tissue Doppler, time	5.9
Left STD Tei index*	Tissue Doppler, time (4)	24.3	Left STD: isovolumic contraction onset to isovolumic relaxation end**	Tissue Doppler, time	6
Mitral and aortic valve simultaneous Tei index	Doppler time (4)	23.9	LV ED 4-ch epicardial long axis dimension**	2D	6.1
LV velocity of circumferential fiber shortening*	2D (4)	23.5	Left STD: Peak diastolic summation wave velocity	Tissue Doppler, velocity	6.1
LVES fiberstress	2D (3)	23.2	Septal STD: Ejection time**	Tissue Doppler, time	6.1
Aortic valve systemic resistance	Doppler (3)	23.2	LV ES 4-ch endocardial long axis dimension**	2D	6.4
LVES wall thickness	M-mode	23.1	LVES short axis dimension**	2D	6.6
Aortic valve stroke volume	Doppler (3)	23.0	Mitral valve closure time	Doppler time	6.6

For calculated variables, the number of variables used to calculate these is in parentheses. 2D = 2-dimensional; 4-ch = 4-chamber; ED = end-diastolic; ES = end-systolic; STD = spectral tissue Doppler.

In the left columns (20 variables with the highest inter-observer error), an asterisk (\*) marks the variables that also feature among the 20 variables with the highest inter-observer % error at the core laboratory (Table 6).

In the right columns (Variables with the lowest inter-observer error), a double asterisk \*\* marks the variables that also feature among the 20 variables with the lowest inter-observer % error at the core laboratory (Table 6).

**Table 9**

Comparison of 3-beat averaged mean and median of median % error for all variables.

Variable type	Mean (median) of % error					P-value (core inter- vs. core intra-observer % error) <sup>‡</sup>
	Local vs. core observer % variables	Core lab inter-observer % (119 variables)	Core lab inter-observer % error	Core lab intra-observer % error	P-value (local/core vs. core inter-observer % error) <sup>*</sup>	
A	9.7 (7.3)	6.8 (5.1)	6.8 (5.1)	1.5 (0.9)	<.001	<.001
C2	13.6 (8.4)	10.1 (6.7)	10.0 (7.0)	4.6 (2.5)	<.001	<.001
C3	20.8 (12.9)	13.1 (9.7)	14.4 (10.9)	5.8 (3.7)	<.001	<.001
C4	22.2 (16.5)	18.8 (13.4)	18.8 (14.0)	8.4 (5.2)	<.001	<.001
D	13.2 (7.9)	8.7 (5.2)	8.7 (5.2)	4.2 (2.1)	<.001	<.001
I	19.1 (5.2)	5.5 (3.9)	5.5 (3.9)	3.9 (2.4)	<.001	0.018
S	35.9 (27.5)	40.5 (38.6)	40.5 (38.6)	19.9 (16.3)	0.004	<.001
T	10.8 (5.1)	8.9 (4.4)	8.9 (4.4)	5.5 (2.5)	<.001	<.001
V	9.9 (5.6)	6.2 (4.3)	6.2 (4.3)	4.2 (2.7)	<.001	<.001
R	4.0 (1.9)	2.9 (1.6)	3.2 (1.6)	2.0 (0.8)	<.001	<.001

These have been classified using the schema detailed in the text, where A = area; C2, C3, C4 = calculated from 2, 3 and 4 variables respectively; D = dimension; I = integral; S = slope; T = Doppler time interval; V = velocity, and R = heart rate. Local vs. core, core lab inter-observer and core lab intra-observer % error have been calculated for all classes of variables. Percent error statistics are based on data from 150 variables unless otherwise noted.

<sup>\*</sup> Local vs. Core lab observer %error vs. Core lab inter-observer %error based on 119 variables

<sup>‡</sup> Core lab inter-observer vs. Core lab intra-observer %error based on 150 variables