Effect of Experience and Training on the Concordance and Precision of Strain Measurements

Tomoko Negishi, MD,[a] Kazuaki Negishi, MD, PhD,[a] Paaladinesh Thavendiranathan, MD,[b] Goo-Yeong Cho, MD,[c] Bogdan A. Popescu, MD, PhD,[d] Dragos Vinereanu, MD,[d] Koji Kurosawa, MD,[e] Martin Penicka, MD, PhD,[f] Thomas H. Marwick, MBBS, PhD, MPH,[a,g] on behalf of the SUCCOUR Investigators

ABSTRACT

OBJECTIVES This study sought to show the degree to which experience and training affect the precision and validity of global longitudinal strain (GLS) measurement and to evaluate the variability of strain measurement after feedback.

BACKGROUND The application of GLS for the detection of subclinical dysfunction has been recommended in an expert consensus document and is being used with increasing frequency. The role of experience in the precision and validity of GLS measurement is unknown, as is the efficacy of training.

METHODS Fifty-eight readers, divided into 4 groups on the basis of their experience with GLS, calculated GLS from speckle strain analysis of 9 cases with various degrees of image quality. Intraclass correlation coefficients (ICCs), mean difference, standard deviation (SD), and coefficient of variation (CV) were compared against the measurements of a reference group that had experience with >1,000 cases of strain measurement. Individualized feedback was distributed, and repeat measurements were performed by 40 readers. Comparisons with the baseline variation provided information about whether feedback was effective.

RESULTS The ICC for GLS was significantly greater than that for ejection fraction regardless of image quality. Experience with strain measurement affected the concordance in strain values among the readers; the group with the highest level of experience showed significantly better ICC than those with no experience, although the ICC of the inexperienced readers was still very good (0.996 vs. 0.975, p = 0.0002). As experience increased, the mean difference, SD, and CV became significantly smaller. The CV of segmental strain analysis showed significant improvement after training, regardless of experience.

CONCLUSIONS The favorable interobserver agreement of GLS makes it more attractive than ejection fraction for follow-up of left ventricular function by multiple observers. Although experience is important, the precision of GLS was high for all groups. Training appears to be of most value for the assessment of segmental strain. (J Am Coll Cardiol Img 2016;■:■-■) © 2016 by the American College of Cardiology Foundation.

Interinstitutional agreement regarding measurement of left ventricular (LV) function is vital for both clinical practice and research. Ejection fraction (EF) is used widely for this purpose, but its limitations are well known. Global longitudinal strain (GLS) is a robust marker of subclinical cardiac dysfunction that appears to be more sensitive for the detection of early change in LV function than EF [(1)]; however, the possible sources of variation in strain imaging (reader, equipment, and subject)

ABBREVIATIONS AND ACRONYMS

CV = coefficient of variation

EF = ejection fraction

GLS = global longitudinal systolic strain

ICC = intraclass correlation coefficients

LV = left ventricular

MD = mean difference

have not been fully evaluated (2). Indeed, most of the literature on strain analysis has reported on measurements performed by experienced observers, and the nature and length of the learning curve remain undefined. There is a precedent for the deployment of educational interventions to obtain better concordance and improve diagnostic accuracy in echocardiography (3,4), and similar processes to achieve an adequate level of concordance should be considered in the development of clinical trials using strain; however, little is known about whether education or feedback improves strain concordance. Accordingly, we sought to determine whether: 1) levels of experience affect precision and validity (as defined by an expert reference read); 2) GLS has better concordance than EF; and 3) whether strain concordance is improved after feedback.

METHODS

STUDY DESIGN. Echocardiograms from 9 cases with various levels of image quality were prepared for this study, 4 with good image quality and easy automated tracking, 2 with borderline quality in which strain was measurable after adjustments of tracking, and the remaining 3 with images too poor to analyze (included to determine whether observers would avoid measurement). All of the prepared images for analysis were acquired by use of standard commercial echocardiographic systems (Vivid 7 and E9, GE Medical, Milwaukee, Wisconsin). All images were recorded with the highest frame rate (55 to 80 rpm) and optimized image depth and sector width. Each observer calculated EF using the biplane method of disks (5) and obtained strain measurements using commercially available software (EchoPAC PC, GE Medical), with either version 12.0.0 or version 13.0.0 for all reads (6). Measurement of GLS has been described previously (2). GLS was obtained by averaging 3 apical views. An 18-segment model was used for segmental strain analysis.

PROTOCOL 1. GLS for 4 cases with adequate measurement quality was measured by 58 readers with various levels of strain experience from North America, Europe, Asia, and Oceania. Readers were divided into 4 groups by strain experience: no experience (0 cases), limited experience (1 to 20 cases), intermediate experience (21 to 100 cases), and highly experienced (>100 cases). Average strain measurements from 5 highly experienced readers with >1,000 cases of experience (the reference group) were compared with those from these 4 groups for assessment of precision.

PROTOCOL 2. GLS, segmental strain, and 2-dimensional EF of all 9 cases were assessed by a subgroup of 40 readers with different levels of experience (less experienced, ≤100 cases; experienced, >100 cases) from 22 different institutes. Readers were instructed not to measure EF or strain if they thought image quality was inadequate. Each reader received personalized feedback from expert review of the strain tracings in the core laboratory. After training, a set of 6 cases (excluding those with inadequate images) were remeasured by all readers in a blinded manner. Peak longitudinal strains from each segment were compared with average segmental values from the reference group. Protocol 2 was part of an international multicenter trial of the incremental value of myocardial strain for the detection of cardiotoxicity (SUCCOUR [Strain Surveillance During Chemotherapy for Improving Cardiovascular Outcomes], ANZCTR [Australian New Zealand Clinical Trials Registry] number ACTRN12614000341628, approved by the institutional review board of each institution).

All authors had full access to and take responsibility for the integrity of the data.

STATISTICAL ANALYSIS. Intraclass correlation coefficients (ICCs) of GLS and EF were used to determine concordance and improvement of agreement. The difference in GLS and segmental strain between each reader and the reference reads was calculated. Mean difference (MD), standard deviation (SD), and coefficient of variance (CV) were compared with the values from the reference group. Student *t* test and paired *t* tests were used to compare continuous variables when appropriate. The Kruskal-Wallis test was used for comparisons among groups, followed by pairwise comparisons, with the p value adjusted for multiple comparisons. The Jonckheere-Terpstra test was used to test the trend among the groups. Statistical analyses were performed with IBM SPSS statistics version 20.0.0 (SPSS Inc., Chicago, Illinois) and R version 3.1.0. (R Foundation for Statistical Computing, Vienna, Austria) with the "cocron" package. All p values reported are from 2-sided tests, and $p < 0.05$ was considered statistically significant.

RESULTS

PATIENT CHARACTERISTICS. Indications for echocardiography were predominantly for the detection of subclinical LV dysfunction in asymptomatic patients; case descriptions are summarized in Online Table S1. Cardiac volumes were within the normal range (end-diastolic volume 106 ± 21 ml, end-systolic volume 44 ± 12 ml), and patients had preserved EF (59 ± 4%).

Some case subjects had decreased GLS because of their underlying medical conditions. Reference standards for EF and GLS were the average of measurements made by the reference group (Online Table S1).

PROTOCOL 1. Reader experience with strain measurement and echocardiography training levels are shown in Online Table S2. On the basis of their strain measurement experience level, 58 readers were divided into 4 groups: no experience (n = 13), limited experience (n = 12), intermediate experience (n = 10), and highly experienced (n = 23). All highly experienced readers had completed level 3 echocardiography training. In the group with no experience, most of the readers were at level 1 in general echocardiography training, with 1 reader who had finished level 3 echocardiography training without any experience in strain analysis.

IMPACT OF EXPERIENCE IN STRAIN MEASUREMENT (GLS). Although the ICC of the group with no experience was very good (0.975; 95% confidence interval: 0.912 to 0.998), that of the highly experienced group was even better (0.996; 95% confidence interval: 0.988 to 1.000; p = 0.0002). With the accumulation of GLS experience, compared with the expert readers, the MD, SD, and CV became significantly smaller (**Figure 1**). Subsequent pairwise comparisons showed that the MD of the highly experienced group was significantly less than that of the group with no experience (adjusted p value [p_{Adj}] = 0.003), with the same findings for SD (p_{Adj} = 0.012), and CV (p_{Adj} = 0.014). No statistically significant differences were seen between the other groups.

PROTOCOL 2. In protocol 2, the 40 readers were categorized into 2 groups based on the number of cases with strain measurement, with 17 characterized as less experienced and 23 as experienced. Most of the less experienced readers had completed level 3 echocardiography training. One in the less experienced group had never experienced strain measurement before this study (Online Table S2). In initial measurements, 2 observers in the less experienced group failed to avoid measuring the 3 cases with inadequate image quality for strain analysis, whereas the rest of the readers successfully excluded them from the analysis.

ICC OF GLS AND EF (INITIAL MEASUREMENT). GLS and EF of the 6 appropriate cases by the reference readers were −18.1 ± 0.8% and 59 ± 6%, respectively. The ICCs for GLS were significantly better than those for 2-dimensional EF (p < 0.001) regardless of image quality (Online Table S3). The ICC of GLS from good-quality images was significantly

higher than that from borderline-quality images (p = 0.01), although it still had excellent concordance (0.993). Also, the ICC of EF in good-quality images was significantly higher than that for borderline-quality images (p < 0.001).

EFFECT OF FEEDBACK. The interval between the first and second measurements was 71 ± 8 days, which was similar between the less experienced and experienced groups (63 ± 33 days vs. 78 ± 40 days, p = 0.21). The ICC of GLS from the second measurement remained high (0.99) regardless of the level of experience. The ICC of less experienced readers was significantly smaller than that of the experienced group at the first measurement (p = 0.02) and remained significant after feedback was provided (p = 0.009). MD, SD, and CV did not show any significant improvement after feedback in GLS measurements, although those values in the second measurements tended to be smaller than the first measurement (Online Table S4).

In the segmental strain analysis, the SD (p = 0.046) and CV (p = 0.003) of all readers improved significantly after feedback, but MD did not (p = 0.15). Improvements in CV were seen regardless of experience (Online Table S5).

DISCUSSION

The results of this study showed that: 1) experience in strain imaging affects the concordance in strain values among readers; 2) GLS is less variable than EF, which indicates it is a suitable marker for multicenter trials with many readers; 3) image quality affects strain concordance; and 4) feedback from the core laboratory improves concordance in segmental strain measurement regardless of the experience of the readers, but not in GLS.

EXPERIENCE IN STRAIN MEASUREMENT. In the training statement, level 2 training includes knowledge of strain echocardiography, and level 3 training is required for independent interpretation and proper use (7). Although there were no significant differences among the 3 groups, highly experienced readers showed less variation than the other groups. This implies that not only knowledge but also experience in tracking is required for better concordance. Greater strain experience reduces GLS variability and increases precision. Our results should encourage readers with limited experience to undertake a quality control process before involvement in multicenter trials (and likely also for clinical practice).

ROBUSTNESS OF GLS. In some situations, such as monitoring a patient during chemotherapy,

FIGURE 1 Impact of Experience on Strain Measurements
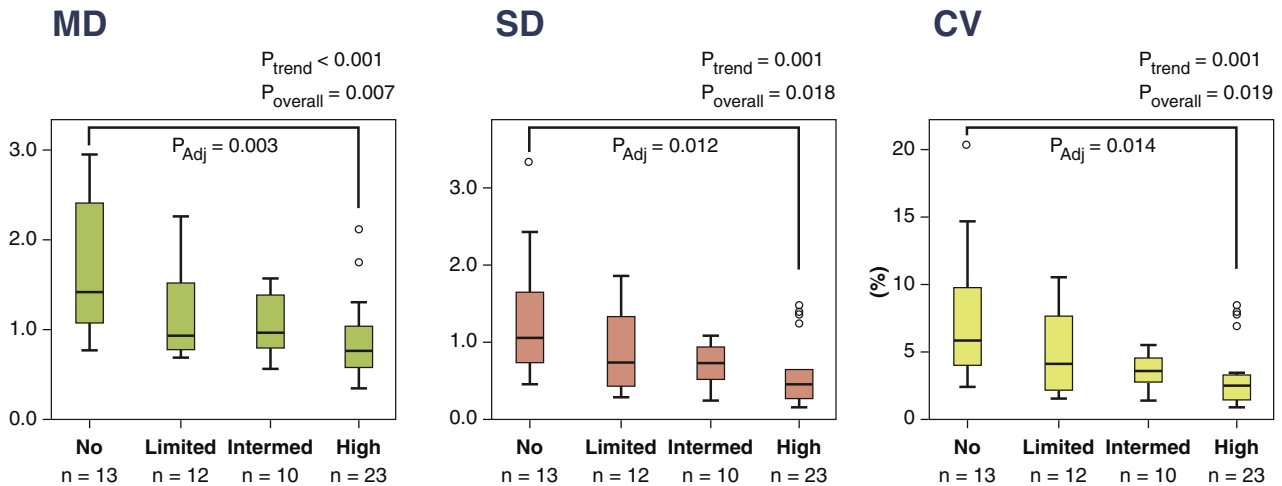
With increasing experience, there was a significant trend toward lower mean difference (MD), standard deviation (SD), and coefficient of variation (CV) in GLS measurements. High = highly experienced; Intermed = intermediate experience; Limited = limited experience; No = no experience; p_{Adj} = adjusted p value.

sequential follow-up is needed, and reliability is essential. As shown in this study, GLS is more robust than EF, even in images of borderline quality, which indicates it is a more appropriate measure by which to follow LV function longitudinally. A higher level of concordance would permit trialists to reduce the necessary sample size and increase the ability to identify small effect sizes.

IMAGE QUALITY OF ECHOCARDIOGRAPHY. Tracking quality can be suboptimal if images are of poor quality (8), and good image quality is an essential feature for accurate strain tracking. In our study, we selected different qualities of images for analysis. The ICC of good-quality images was significantly greater than that of borderline images, which substantiates the importance of image quality for concordance. However, even when the image quality was borderline, the ICC of GLS was very high and better than EF. These results further support the use of GLS in the clinical setting.

EFFECT OF FEEDBACK FOR STRAIN MEASUREMENT. The sharing of interpretations is a useful process to provide feedback and facilitate quality improvement in echocardiography. Previous studies have shown significant improvement in echocardiography after a training process (3,4,9). Our results show an improvement of concordance (measured as CV) of segmental strain analysis after training. The personalized feedback used in this study identified a number of points for improvement of strain measurements, with the 2 main sources of discordance of

measurements being the width and location of regions of interest, especially at the mitral annulus and apex (10). We emphasize the importance of working on recognizing the adequacy of tracking instead of the strain numbers alone. We propose that the training process should comprise an initial evaluation of concordance, followed by detailed feedback based on actual strain tracing and tracking, then repeated sessions if necessary.

We were unable to show an impact of training on global strain analysis for a number of reasons. Perhaps the main one was that the ICC of GLS was very high at baseline. In addition, the limited number of readers with no experience who were involved in protocol 2 could have affected the results.

STUDY LIMITATIONS. First, the numbers of prepared cases were small because we collected interpretations from more than 20 different institutes around the world. Second, this study was conducted with a single version of software, which might not reflect the performance of other vendors; however, current differences among vendors are less than in the past, and we think that the superiority of GLS over EF is likely to be generic to GLS. Our results might not be readily applicable to a population with a wider range of EF and GLS; however, no previous report has shown any systematic bias in interobserver or intraobserver variabilities, so we would expect the effect of experience on strain variability to be similar throughout ranges of EF or GLS.

CONCLUSIONS

The favorable interobserver agreement of GLS makes it more attractive than EF for follow-up of LV function by multiple observers. Although experience is important, the precision of GLS was high for all groups. Training appears to be of the most value for the assessment of segmental strain.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Thomas H. Marwick, Baker IDI Heart & Diabetes Institute, 75 Commercial Road, Melbourne, Victoria 3004, Australia. E-mail: tom.marwick@bakeridi.edu.au.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Although experience in strain measurement affects the precision of GLS, this precision was high for every reader, regardless of training. Training appears to be of the most value for the assessment of segmental strain.

TRANSLATIONAL OUTLOOK: GLS has more favorable interobserver agreement than EF. A training process is an important step in multicenter clinical trials, especially for regional strain. Further studies among larger numbers of readers, especially nonexperts (e.g., cardiology, internal medicine, and nursing trainees) will confirm the precision of GLS in the general community and define the exact requirements of the learning curve.

REFERENCES

1. 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.

2. 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.

3. Pérez de Isla L, Moreno F, Garcia Saez JA, et al. Efficacy and learning curve of a hand-held echocardiography device in an oncology outpatient clinic: expanding the use of echoscopic heart examination beyond cardiology. Mol Clin Oncol 2015;3:820–4.

4. 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.

5. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015;28:1–39.e14.

6. Yang H, Marwick TH, Fukuda N, et al. Improvement in strain concordance between two major vendors after the strain standardization initiative. J Am Soc Echocardiogr 2015;28:642–8.e7.

7. Ryan T, Berlacher K, Lindner JR, Mankad SV, Rose GA, Wang A. COCATS 4 Task Force 5: Training in Echocardiography. J Am Coll Cardiol 2015;65: 1786–99.

8. Voigt JU, Pedrizzetti G, Lysyansky P, 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.

9. Dahiya A, Bolen M, Grimm RA, et al. Development of a consensus document to improve multireader concordance and accuracy of aortic regurgitation severity grading by echocardiography versus cardiac magnetic resonance imaging. Am J Cardiol 2012;110: 709–14.

10. Negishi K, Negishi T, Kurosawa K, et al. Practical guidance in echocardiographic assessment of global longitudinal strain. J Am Coll Cardiol Img 2015;8:489–92.

KEY WORDS concordance, experience, global longitudinal strain

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