

Fully Automated Versus Standard Tracking of Left Ventricular Ejection Fraction and Longitudinal Strain



The FAST-EFs Multicenter Study

Christian Knackstedt, MD,* Sebastiaan C.A.M. Bekkers, MD, PhD,* Georg Schummers,† Marcus Schreckenberg,† Denisa Muraru, MD, PhD,‡ Luigi P. Badano, MD, PhD,‡ Andreas Franke, MD,§ Chirag Bavishi, MD, MPH,|| Alaa Mabrouk Salem Omar, MD, PhD,|| Partho P. Sengupta, MD, DM||

ABSTRACT

BACKGROUND Echocardiographic determination of ejection fraction (EF) by manual tracing of endocardial borders is time consuming and operator dependent, whereas visual assessment is inherently subjective.

OBJECTIVES This study tested the hypothesis that a novel, fully automated software using machine learning-enabled image analysis will provide rapid, reproducible measurements of left ventricular volumes and EF, as well as average biplane longitudinal strain (LS).

METHODS For a total of 255 patients in sinus rhythm, apical 4- and 2-chamber views were collected from 4 centers that assessed EF using both visual estimation and manual tracing (biplane Simpson's method). In addition, datasets were saved in a centralized database, and machine learning-enabled software (AutoLV, TomTec-Arena 1.2, TomTec Imaging Systems, Unterschleissheim, Germany) was applied for fully automated EF and LS measurements. A reference center reanalyzed all datasets (by visual estimation and manual tracking), along with manual LS determinations.

RESULTS AutoLV measurements were feasible in 98% of studies, and the average analysis time was 8 ± 1 s/patient. Interclass correlation coefficients and Bland-Altman analysis revealed good agreements among automated EF, local center manual tracking, and reference center manual tracking, but not for visual EF assessments. Similarly, automated and manual LS measurements obtained at the reference center showed good agreement. Intraobserver variability was higher for visual EF than for manual EF or manual LS, whereas interobserver variability was higher for both visual and manual EF, but not different for LS. Automated EF and LS had no variability.

CONCLUSIONS Fully automated analysis of echocardiography images provides rapid and reproducible assessment of left ventricular EF and LS. (J Am Coll Cardiol 2015;66:1456-66) © 2015 by the American College of Cardiology Foundation.

From the *Department of Cardiology, Maastricht University Medical Centre, Maastricht, the Netherlands; †TomTec Imaging Systems GmbH, Unterschleissheim, Germany; ‡Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, Italy; §Department of Cardiology, KRH Klinikum Siloah, Hannover, Germany; and the ||Zena and Michael A. Wiener Cardiovascular Institute and the Marie-Josée and Henry R. Kravis Center for Cardiovascular Health, Mount Sinai School of Medicine, New York, New York. Dr. Knackstedt has received travel grants and research software from TomTec Imaging Systems. Mr. Schummers is an employee of TomTec Imaging Systems. Mr. Schreckenberg is an employee at TomTec Imaging Systems. Dr. Badano has received a research equipment grant from TomTec Imaging Systems. Dr. Sengupta has received research software support from TomTec Imaging Systems; has served as an advisor to Saffron Technology, TeleHealthRobotics, and Heart Test Laboratories; and has served as a consultant to Edwards Lifesciences. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose. Drs. Knackstedt and Bekkers contributed equally to this work. Pravin Shah, MD, served as the Guest Editor for this paper.

[Listen to this manuscript's audio summary by JACC Editor-in-Chief Dr. Valentin Fuster.](#)

Manuscript received April 3, 2015; revised manuscript received June 30, 2015, accepted July 27, 2015.



The quantification of left ventricular (LV) size, geometry, and function represents the most frequent indication for an echocardiographic study and is pivotal for patient evaluation (1). Although LV volumes and ejection fractions (EFs) can be measured using different imaging modalities, 2-dimensional echocardiography continues to be the most commonly utilized technique in clinical practice. Despite the existing recommendations for the use of 3-dimensional (3D) echocardiography (2) and the reported variability of the biplane disc summation method, which can be as high as 14% (3), the use of 2-dimensional echocardiography has continued to grow. However, the additional time and inconsistencies in manual tracing of the endocardial borders have led to the continued use of visual assessment of EF in busy echocardiographic laboratories (4).

SEE PAGE 1467

The inherent subjectivity in visual assessment of EF and inconsistencies in manual estimation could potentially be overcome with the use of image processing algorithms that allow fully automated measurement of EF (5,6). Moreover, automated measurements could also yield longitudinal strain (LS), a sensitive marker of cardiac function (7,8). Therefore, in the present study, we used novel, fully automated software for measuring EF and LS from biplane views of the LV and compared the values with visually-estimated EF, manually-traced EF, and manually-traced LS. We hypothesized that automated EF and LS measurements would provide values similar to manual measurements, with increased precision.

METHODS

PATIENT POPULATION. A total of 4 international cardiovascular centers (Department of Cardiology, Maastricht University Medical Centre, the Netherlands [MAA]; Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Italy [PAD]; Department of Cardiology, KRH Klinikum Siloah, Hannover, Germany; and Zena and Michael A. Wiener Cardiovascular Institute and the Marie-Josée and Henry R. Kravis Center for Cardiovascular Health, Mount Sinai School of Medicine, New York, New York) participated in this study. All of the participating centers were asked to provide datasets containing apical 4-chamber (4-C) and 2-chamber (2-C) views of patients who had undergone clinically-indicated standard transthoracic echocardiography using commercially available systems. Datasets from the 4 local centers were blinded and stored digitally in DICOM 3.0 format. In addition

to serving as a local laboratory from which basal datasets were collected, MAA also served as the reference center at which all echocardiographic studies (including those from MAA itself) were transferred and stored in a dedicated prototype image database and review platform. This platform was used to perform further analyses, such as heart cycle selection and manual tracing of endocardial contours. At MAA, all images were reviewed to confirm that 1 4-C and 1 2-C view had been recorded with at least 1 complete heart cycle.

Cardiac cycles were reviewed to avoid large variations in sinus cycle length, supraventricular or ventricular extrasystoles, and respiratory variations. The most suitable cycle was selected and analyzed visually for EF estimation, manually by Simpson's method, and by the automated software. It is important to note that to reflect the beat-to-beat variations encountered in the real world, the selection of cardiac cycles at the reference center was not tied to the cardiac cycles previously selected at the local centers.

A total of 280 datasets from patients in sinus rhythm were initially retrieved; 25 were excluded because of missing DICOM data ($n = 2$), insufficient image quality ($n = 1$), repeated evaluation at 2 points in time ($n = 6$), missing visual assessment of EF ($n = 6$), missing data ($n = 5$), or missing calibration information in the DICOM file for automated analysis ($n = 5$). Accordingly, the study consisted of 255 datasets that were included in the final analysis.

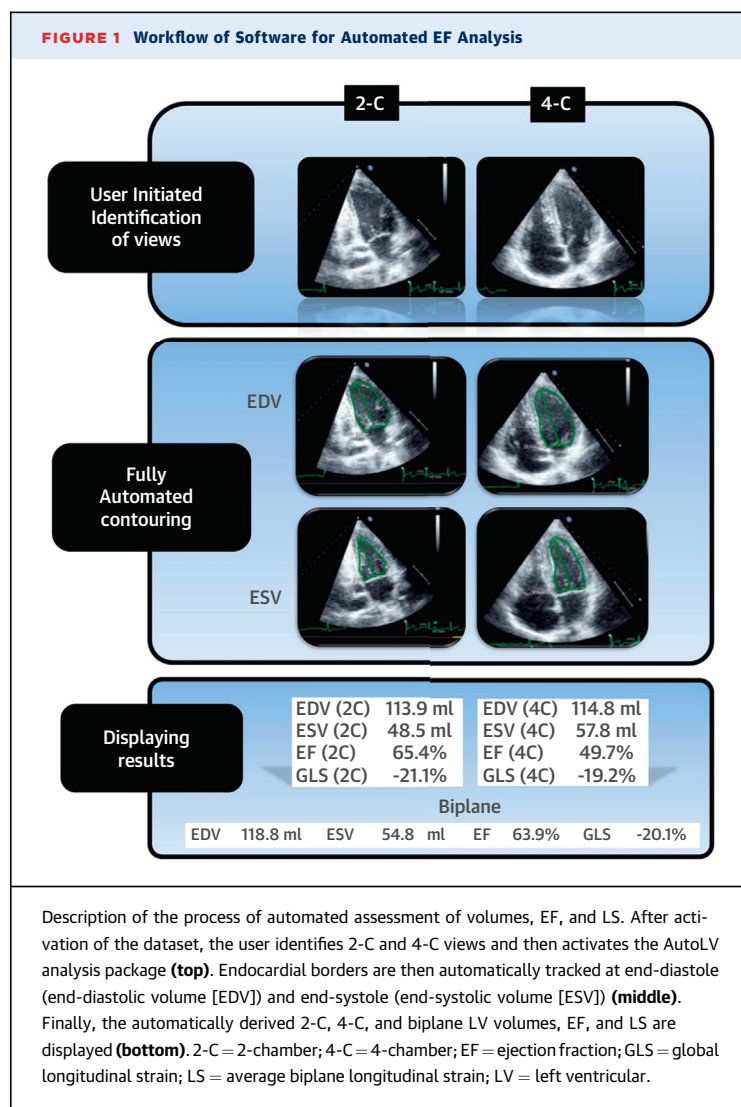
It is important to note that whereas manual and visual assessments of LV volumes and EF were performed at local centers and reassessed in the reference center, for regulatory purposes, automated volumes and EF and manual and automated LS were only performed at the reference center.

CONVENTIONAL EF ANALYSES AT THE LOCAL CENTERS. At each center, an expert investigator (level 3 training in echocardiography; C.K., A.F., L.B., P.S.) was asked to first qualitatively evaluate the EF by visual assessment (visual EF), followed by manual tracing of the endocardial border at end-diastole and -systole to measure LV volumes and EF (manual EF) by the biplane-modified Simpson's method (2). Subsequently, all deidentified images were sent to the reference center at MAA, together with the information about manually-traced LV volumes and EF, visually-estimated EF, and demographic and clinical characteristics.

ANALYSES AT THE REFERENCE CENTER. Manual and visual analyses. One of the expert investigators at MAA blinded to the analyses from the local centers reviewed all datasets and manually traced the

ABBREVIATIONS AND ACRONYMS

2-C = 2-chamber
4-C = 4-chamber
EF = ejection fraction
LS = average biplane longitudinal strain
LV = left ventricle/ventricular
MAA = Maastricht University Medical Centre
PAD = University of Padua



endocardial contours at end-systole and -diastole on 2- and 4-C views in all study patients. End-diastole was defined at the peak of the electrocardiographic R-wave and/or 1 frame before mitral valve closure. End-systole was defined as 1 frame before mitral valve opening. To ensure adequate blinding of the manual evaluation and the visual estimation of EF, values of LV volumes and EF from the manual tracing were automatically transferred to the database and were not visible to the investigator. LS was also calculated and averaged as the relative change in the length of the endocardial border from end-diastole to -systole by manually tracking the endocardial borders from 4- and 2-C views. The image quality of all studies from all centers was classified as: 1) good; 2) average; 3) poor; or 4) not analyzable.

Fully automated analysis. Fully automated measurements of LV volumes were performed using

TABLE 1 Clinical and Echocardiographic Data (n = 255)

Patient demographics	
Age, yrs	50.3 ± 17.6
Males	153 (60)
Height, cm	176.5 ± 60.9
Weight, kg	73.6 ± 14.1
Body mass index, kg/m ²	24.6 ± 4.3
Medical history	
No cardiac disease	21 (8.2)
Arterial hypertension	4 (1.6)
Coronary artery disease	82 (32.2)
Myocardial infarction	37 (14.5)
Cardiomyopathy (noncompaction)	5 (2.0)
Dilated cardiomyopathy	8 (3.1)
Ischemic cardiomyopathy	8 (3.1)
Restrictive cardiomyopathy	13 (5.1)
Constrictive pericarditis	2 (0.8)
Aortic valve pathology	2 (0.8)
Mitral valve pathology	2 (0.8)
Myocarditis	1 (0.4)
Chemotherapy-related heart failure	1 (0.4)

Values are mean ± SD or n (%).

AutoLV (TomTec-Arena 1.2, TomTec Imaging Systems, Unterschleissheim, Germany), a computer vision vendor-independent software package that applies a machine-learning algorithm for DICOM images (9). AutoLV can provide biplane end-diastolic, end-systolic, and stroke volumes and EF, as well as LS.

To measure the time spent using AutoLV, a standardized workflow was used. The dataset of a patient was uploaded to the review station and opened. Then, the investigator activated the AutoLV option, which was followed by a request to identify the 4- and 2-C views. The algorithm would then run the automated endocardial border detection and identify the end-diastole and -systole for both views (Figure 1).

INTEROBSERVER, INTRAOBSERVER, AND BEAT-TO-BEAT VARIABILITY. To measure intraobserver and interobserver variability, a subset of 20 patient datasets was randomly selected and resubmitted to MAA and PAD to be reanalyzed using the same protocol. Physicians at both centers were blinded to the original EF and LS results. Only 1 cardiac cycle was provided to ensure that all investigators analyzed the same heartbeat.

To derive interobserver variability, 1 physician from each center was asked to recalculate manual EF and manual LS and to allocate a visual assessment of the EF for all 20 patient datasets. Two investigators (C.K. and L.B.) were asked to perform these measurements twice to assess intraobserver variability.

TABLE 2 Echocardiographic Measurements Performed With the Various Methods in the Local and the Reference Centers

	Local Centers	Reference Center
Manual assessment		
End-diastolic volume, ml*	117.1 ± 45.4	126.1 ± 48.6
End-systolic volume, ml*	52.3 ± 31.0	57.6 ± 35.7
Ejection fraction, %	57 ± 10.8	56.5 ± 12.3
Visual assessment		
Ejection fraction, %	54.3 ± 11.9	54.6 ± 11.0
Automated assessment		
End-diastolic volume, ml	—	124.4 ± 41.3
End-systolic volume, ml	—	56.1 ± 32.9
Ejection fraction, %	—	56.8 ± 12.5
LS, %	—	20.0 ± 0.6

Values are mean ± SD. *Left ventricular volumes were not provided by the local center in 38 patients.
LS = longitudinal strain.

The same subset of 20 patient datasets was also used for beat-to-beat variability. A second cardiac cycle was used to derive manual as well as automated EF and LS and was compared with measurements obtained from the original cardiac cycle.

The investigation conformed to the principles outlined in the Declaration of Helsinki, and each participating center conformed to local ethical regulations.

STATISTICAL ANALYSIS. Normal distributions of variables were checked before analysis. Continuous variables were presented as mean ± SD and categorical variables were presented as n (%). Agreement between various echocardiographic measures was performed using intraclass correlation coefficients (ICCs) and Bland-Altman analysis. The paired Student *t* test was used to compare the mean values between 2 groups.

For all statistical tests, a 2-tailed *p* value <0.05 was considered statistically significant. Statistical analysis was performed using SPSS for Windows, version 19.0 (SPSS, Inc., Chicago, Illinois) and Stata, version 11 (Stata Corp., College Station, Texas).

RESULTS

Table 1 summarizes demographic and clinical data, and **Table 2** summarizes volumes and EF measurements at all study sites by all methods. Although the average image quality was reasonably good (1.6 ± 0.7), and both manual measurement and visual estimation of EF were possible in all patients, fully automated measurements of LV volumes and LS were not feasible in 5 patients (feasibility = 98%). The average time for obtaining the EF by automated analysis was 8 ± 1 s/patient (analyzed in 20 patients).

ASSESSMENT OF DIFFERENT METHODS OF DETERMINING EF IN THE REFERENCE VERSUS LOCAL CENTERS.

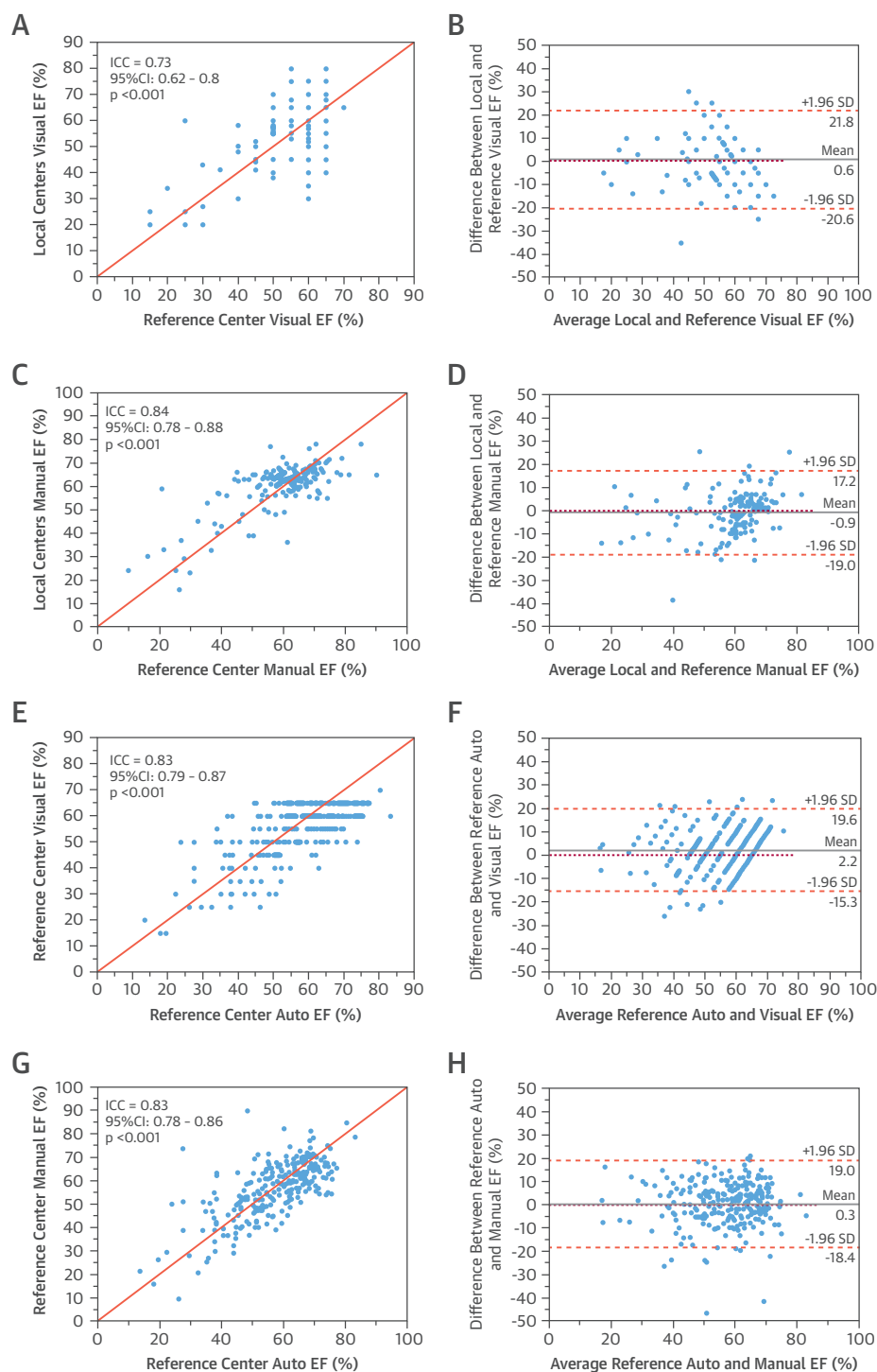
Using ICC, good correlations were seen between all conventional EF methods and automated EF (**Table 3**, **Figure 2**). Bland-Altman analysis showed that the bias and limits of agreement were relatively lower when the same methods for EF assessment were compared at the local and the reference centers (local visual EF vs. reference visual EF: mean bias −0.6%; 95% confidence interval [CI]: −2.4% to 1.2% [**Figure 2B**]; local manual EF vs. reference manual EF: mean bias 0.9%, 95% CI: −0.6% to 2.4% [**Table 3**, **Figure 2D**]), than when different methods were compared at the local and the reference centers (local visual EF vs. reference manual EF: mean bias −2.2%; 95% CI: −3.5% to −0.8%; local manual EF vs. reference visual EF: mean bias 2.4%; 95% CI: 1.5% to 3.3%) (**Table 3**). Bias

TABLE 3 Agreement and Bland-Altman Bias Matrix Between EFs Assessed Using Various Methods

	Local Center Visual EF	Local Center Manual EF	Reference Center Visual EF	Reference Center Manual EF	Automated EF
Agreement: ICC (95% CI)					
Local center visual EF	1.0	—	—	—	—
Local center manual EF	0.82 (0.77 to 0.86)*	1.0	—	—	—
Reference center visual EF	0.73 (0.62 to 0.80)*†	0.87 (0.83 to 0.90)*	1.0	—	—
Reference center manual EF	0.74 (0.67 to 0.80)*	0.84 (0.78 to 0.88)*†	0.86 (0.82 to 0.89)*	1.0	—
Automated EF	0.70 (0.62 to 0.77)*	0.83 (0.78 to 0.86)*	0.83 (0.79 to 0.87)*	0.83 (0.78 to 0.86)*	1.0
Bland-Altman: bias (95% CI)					
Local center visual EF	0	—	—	—	—
Local center manual EF	2.7 (1.6 to 3.8)	0	—	—	—
Reference center visual EF	−0.6 (−2.4 to 1.2)†	2.4 (1.5 to 3.3)	0	—	—
Reference center manual EF	−2.2 (−3.5 to −0.8)	0.9 (−0.6 to 2.4)†	−1.8 (−2.9 to −0.8)	0	—
Automated EF	−2.5 (−3.9 to −1.1)	0.2 (−0.9 to 1.3)	−2.2 (−3.3 to −1.1)	−0.3 (−1.5 to 0.9)	0

**p* < 0.001. †147 patients (after exclusion of cases from MAA where reference center was located to avoid same-center-related measurement bias).
CI = confidence interval; EF = ejection fraction; ICC = intraclass correlation coefficient; MAA = Maastricht.

FIGURE 2 Concordance of EF Measurements



Correlations and Bland-Altman plots between different methods. (Left) Correlation plots and (right) Bland-Altman plots between: automated and manual EF at the reference center (A and B); automated EF and visual EF at the reference center (C and D); manual EF at the reference and local centers (E and F); and visual EF at the reference and local centers (G and H). Because MAA served both as a local center and as the reference center, local site measurement bias was avoided by excluding 108 datasets provided by MAA from ICC and Bland-Altman analyses when comparing visual and manual assessments between the reference and local centers, thereby yielding 147 patient datasets (A to D). CI = confidence interval; ICC = intraclass correlation coefficient; MAA = Maastricht; other abbreviations as in Figure 1.

and levels of agreement within the local and reference centers were also relatively high (local visual EF vs. local manual EF: mean bias 2.7%; 95% CI: 1.6% to 3.8%; reference visual EF vs. reference manual EF: mean bias –1.8%; 95% CI: –2.9% to –0.8%) (Table 3).

Bland-Altman analyses between the conventional methods and the automated EF showed that the bias and levels of agreement were wider when visual EF at the local centers was compared with the automated EF (mean bias –2.5%; 95% CI: –3.9% to –1.1%) (Table 3), whereas manual EF measurements at the local centers were similar to the corresponding automated EF values, with a negligible bias (mean bias 0.2%; 95% CI: –0.9% to 1.3%) (Table 3). These observations were confirmed in the reference center, as there was a relatively higher bias for visual EF compared with the automated EF (mean bias –2.2%; 95% CI: –3.3% to –1.1%) (Table 3, Figure 2F), whereas the values for the manual and automated EF were relatively similar (mean bias of –0.3%; 95% CI: –1.5% to 0.9%) (Table 3, Figure 2H).

We also checked for >10% absolute difference between automated and manual EF at the reference center and for the manual measurements at the local centers and reference center. These observations were restricted to 147 cases, excluding the cases at MAA where the measurements were not repeated. A >10% difference between the automated and the manual measurements at the reference center occurred in 31 cases (21.1%) and between manual measurements at the local and reference centers in 35 cases (23.85; $p = 0.576$). There was no variability seen for automated measurements.

In addition, when patient datasets were classified according to image quality, correlations for automated EF were preserved for visual and manual EF in patient datasets with good (ICC = 0.83, 0.84; 95% CI: 0.76 to 0.88 and 0.76 to 0.88; respectively, $p < 0.001$) and with moderate image quality (ICC = 0.83, 0.84; 95% CI: 0.76 to 0.88 and 0.77 to 0.89; respectively, $p < 0.001$), whereas correlations marginally worsened when the image quality was poor (ICC = 0.79, 0.63; 95% CI: 0.54 to 0.9 and 0.2 to 0.83; $p < 0.001$, $p = 0.006$, respectively).

It is important to note that, because MAA served both as a local center and as the reference center, and to avoid local site measurement bias, 108 datasets provided from MAA were excluded from all ICC and Bland-Altman analyses when comparing visual and manual assessments between the reference and local centers.

LS MEASUREMENT. At the reference center, the manual tracking-derived LS was $21 \pm 6\%$ and automated LS was $20 \pm 6\%$. Although significant

differences between the absolute values was found ($p = 0.02$), ICCs and Bland-Altman analysis suggested that there was good correlation and agreement between the manual and the automated LS (ICC: 0.83; bias: 0.7%; 95% CI: 0.1% to 1.3%).

A strong correlation was seen when comparing automated LS with automated EF ($r = 0.92$; $p < 0.001$). A simplified regression equation derived from the automated measurements ($EF = 2 \times \text{automated LS} + 20$) yielded EF values that showed similar correlations with both automated and manual EF (ICC = 0.96 and 0.77, respectively; both $p < 0.001$).

INTRAOBSERVER, INTEROBSERVER, AND BEAT-TO-BEAT VARIABILITY.

Table 4 summarizes intraobserver and interobserver variability measurements for visual, manual, and automated EF, as well as for manual LS derived from MAA and PDA. In general, the intraobserver variability was not statistically significantly different, except for visual estimations at MAA ($p < 0.001$) and manual LS estimations at PAD ($p = 0.03$) (Table 4). Bland-Altman analysis showed that the bias and levels of agreement for intraobserver variability in both centers were larger for visual EF than for manual EF or manual LS. In contrast, interobserver variability was significantly different for both visual ($p = 0.001$) and manual EF ($p < 0.001$), whereas it was not different for LS ($p = 0.539$) (Table 4). Bland-Altman analysis also showed that the bias and levels of agreement for interobserver variability were larger for visual and manual EF compared with manual LS (Table 4). Automated EF and LS measurements had no variability because machine analysis revealed the same pattern recognition and similar measurement on repeated assessments. Finally, beat-to-beat variability was $-0.96 \pm 3.52\%$ for automated EF, $2.7 \pm 8.16\%$ for manual EF, $-0.19 \pm 1.31\%$ for automated LS, and $1.09 \pm 3.29\%$ for manual LS.

DISCUSSION

The main finding of this study was that a fully automated measurement of EF is technically feasible, can be performed within a few seconds, and is comparable to manual tracking (Central Illustration). Furthermore, the automated image analysis yields information beyond LV volumes and EF, such as LS, and (unlike all conventional methods of EF and LS assessment) does not include any variability.

THE CLINICAL NEED FOR AUTOMATED ASSESSMENT OF LV FUNCTION.

As with most techniques, echocardiography is associated with a steep learning curve (10), and the overall proficiency and expertise of the practitioner is directly related to the volume

TABLE 4 Interrater and Intrarater Reliability of EF and LS Measurements in MAA and PAD

	Intraobserver Variability					Interobserver Variability				
	Reading 1	Reading 2	p Value	ICC	Bland-Altman Bias (LOA)	MAA	PAD	p Value	ICC	Bland-Altman Bias (LOA)
Visual EF, %						43.8 ± 14.8	51.4 ± 17.4	0.001	0.87	–7.6 (–25.6 to 10.4)
MAA	43.8 ± 14.8	50.5 ± 13.6	<0.001	0.96	–6.8 (–9.4 to –4.1)					
PAD	51.4 ± 17.4	49.5 ± 15.9	0.18	0.97	1.9 (–1.0 to 4.7)					
Manual EF, %						49.6 ± 14.8	62.0 ± 16.1	<0.001	0.78	–12.4 (–29.8 to 5.0)
MAA	49.6 ± 14.8	52.2 ± 14.3	0.19	0.90	–2.6 (–6.7 to 1.4)					
PAD	62.0 ± 16.1	62.1 ± 17.2	0.93	0.96	–0.1 (–3.0 to 2.8)					
Manual LS, %						19.7 ± 6.0	20.3 ± 6.0	0.539	0.88	–0.6 (–9.2 to 8.0)
MAA	18.3 ± 7.0	19.7 ± 6.0	0.14	0.90	–1.4 (–3.3 to 0.5)					
PAD	21.8 ± 7.0	20.3 ± 6.0	0.03	0.95	1.6 (0.1 to 3.0)					
Automated LS, %	19.1 ± 6.0	19.1 ± 6.0	–	1.00	0	19.1 ± 6.0	19.1 ± 6.0	NS	1.00	0

Values are mean ± SD.
LOA = levels of agreement; PAD = Padua; other abbreviations as in Tables 2 and 3.

performed at a given center (11). Overall, 100 echocardiographic examinations are required for attaining expertise, including accurate determination of EF (12). However, the dramatic increase in the use of echocardiography and the accompanying growth in new echocardiographers have exceeded the ability for adequate training.

EF is an important diagnostic and prognostic echocardiographic marker that is used to decide and monitor treatment options in patients with heart failure. However, because of the large variability in EF measurements that can occur at different centers, it is possible that therapies in one-fifth of patients may be confounded when decisions are made on the basis of EF (13). Moreover, newer modalities, such as strain imaging, impose more challenges in the echocardiography field (14). Consequently, there has been renewed interest in automated software tools that can facilitate assessment of LV function with the least possible variability among echocardiographers and between different centers (14,15).

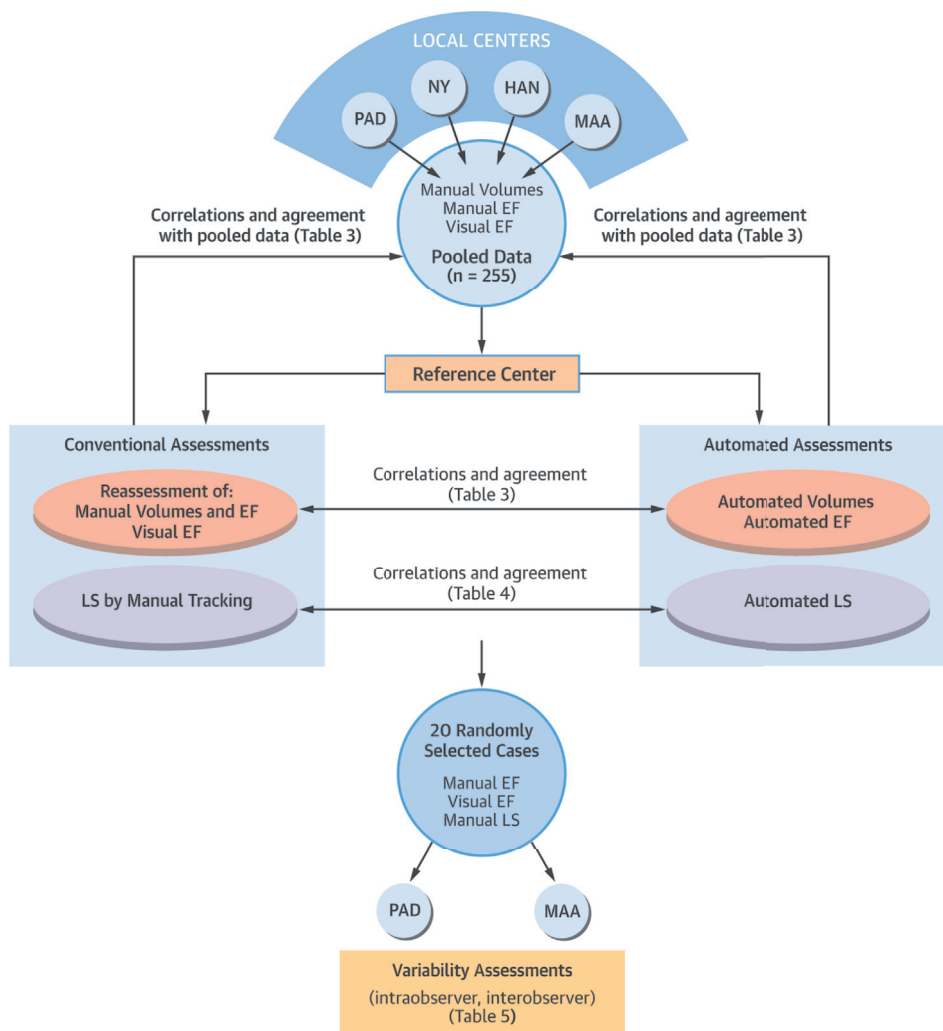
AUTOMATION SHORTENS TIME REQUIRED FOR ASSESSING LVEF. Automated determination of EF is not a new idea and has been applied to different imaging modalities, including echocardiography (6,16–19). However, technical limitations have prevented automated EF from being adopted in clinical practice (20). Recent improvements in automated boundary detection have circumvented some of these limitations (6,17–19). In contrast to older methods, which were dependent on image quality and gain settings (21–23), newer algorithms that rely primarily on speckle tracking and artificial intelligence in tracking techniques were recently reported to be feasible in

several studies (5,6,17,18,24–27). However, the time needed for the automated analysis and the levels of agreement with reference methods in these studies were variable (Table 5). In addition, some studies either used a semiautomated EF determination or required manual corrections that resulted in variable results and increased measurement time (Table 5).

In comparison, the new software algorithm used in the current study differs from the previous approaches, as the user only identifies 2- and 4-C views followed by activation of the automated EF evaluation (Figure 1). The program subsequently detects and contours the endocardium and cardiac cycle and determines the EF and corresponding measurements. Importantly, this process only took 8.1 ± 1 s/patient dataset, thus showing greater time efficiency than all previously described approaches (Table 5).

REPRODUCIBILITY OF TECHNIQUES FOR ASSESSING LVEF. Overall, the decision to rely on either manual or visual determination of EF is a matter of long-standing debate (28,29). More recently, even with the advent of better images due to improved software and hardware, there are still conflicting results. Variability is a particularly important issue when considering the use of manual EF, and is of greatest concern when considering visual EF because of the subjective nature of the assessment. McGowan et al. (30) recently reviewed results from studies that evaluated EF using different techniques. They reported that interobserver variability in previous studies ranged from 9% to 21% for manually-tracked EF and from 8% to 17% for visually-assessed EF, with an intraobserver variability that ranged from 6% to 13% for Simpson’s method and from 11% to 13% for

CENTRAL ILLUSTRATION Automated Versus Standard Tracking of Left Ventricular EF and LS



Knackstedt, C. *et al.* J Am Coll Cardiol. 2015; 66(13):1456-66.

From 4 international centers (PAD, NY, HAN, and MAA), 255 echocardiographic datasets containing manual calculations of volumes and EF using the biplane Simpson's method and visual assessment of EF were collected and stored (pooled data). MAA also served as the reference center to which all echocardiographic studies were transferred. At the reference center, 2 different types of analyses were done: 1) the conventional assessment composed of manual calculation of LV volumes and EF and visual assessment of EF, in addition to calculation of LS by manual tracking of endocardial borders; and 2) fully automated LV volumes, EF, and LS performed using the AutoLV software package. Correlations and agreement were checked between the reference center and the pooled data for manual, visual, and automatic volumes and EF (data derived from MAA were excluded in this comparison to avoid bias). Correlations and agreement were also checked between the manual and automated volumes, EF, and LS calculated at the reference center. Finally, a subset of 20 patients' datasets were randomly selected and resubmitted to MAA and PAD to reassess manual calculation and visual estimation of EF, in addition to manual LS to determine intraobserver and interobserver variability. EF = ejection fraction, HAN = Hannover; LS = average biplane longitudinal strain; LV = left ventricular; MAA = Maastricht; NY = New York; PAD = Padua.

the visual assessment. More recent studies have reported similar variability for manual and visual EF (6,17,27,31,32). In addition, Kaufmann *et al.* (13) demonstrated that EF, measured either manually or

by visual assessment, has a good correlation between 2 centers ($r^2 = 0.63$), but with a wide level of agreement (bias = 0.2, levels of agreement = -17.4% and +17.8%) and relatively large variability between

TABLE 5 Time Needed for and Variability in Automated EF Measurements in Different Studies

First Author, Year (Ref. #)	N	Feasibility	Modality	Time Required for Measurement (s)	Interobserver Variability (%)	Intraobserver Variability (%)	Correlation With Reference Methods
Kühl, 2004 (44)	24	100%	Semiautomated 3D	720 ± 300	−1.5 ± 7.0	−0.6 ± 4.1	
Cannesson et al., 2007 (6)	218	92%	2D	48 ± 26	1.3 ± 1.7	0.5 ± 1.2	0.98
Rahmouni et al., 2008 (18)	92	—	2D	—	—	3.7 ± 3.5	0.80
Maret et al., 2008* (17)	60	100%	2D with and without manual correction	159 ± 46	3.5 ± 6.7	2.2 ± 4.7	0.89
Szulik et al., 2011 (27)	81	90%	2D	54 ± 22	—	—	0.77
Barbosa et al., 2013 (5)	24	100%	3D	30.7	6.06 ± 4.85	—	0.853
Shibayama et al., 2013 (19)	41	93.2%	Automated 3D	37 ± 8	−6.6 to 7.7†	−0.1 to 0.2†	0.54
			Semiautomated 3D	371 ± 116	−7.9 to 15†	−7.8 to 13.1†	0.90
Current study	255	98%	2D	8 ± 1	0	0	0.83

*Results presented represent corrected AutoEF; time needed for uncorrected AutoEF is 79 ± 5 s. †Levels of agreement.
2D = 2-dimensional; 3D = 3-dimensional; other abbreviations as in Table 2.

the centers ($14.1 \pm 10.9\%$). They also mentioned that these differences were not primarily due to varying image quality (13). Overall, the intraobserver and interobserver differences, bias, and levels of agreement reported in our study for visual and manual EF are consistent with the published data, confirming the problem of variability between centers in the visual and manual EF assessments. In contrast, automated EF had no data variability, which is consistent with most intelligent tracking techniques and which gives it a great advantage over conventional methods in addressing the current problem of EF variability between sonographers and centers.

ASSESSMENT OF LS. Assessment of LS is a robust method for assessing LV systolic function (7). Several studies have shown that LS provides incremental diagnostic and prognostic value, and is a more sensitive marker of LV systolic function than EF (33–39). However, variability in LS values has been a source of concern (13), and has been largely attributed to the use of different software from different vendors (40). The European Association of Cardiovascular Imaging–American Society of Echocardiography Industry Task Force therefore initiated a standardization process, which initially focused on standardization of LS. In their recent report, LS measurements using software packages from 7 different ultrasound machine vendors and 2 software-only vendors were comparable and had better reproducibility than EF (41). The current study similarly shows that manually measured LS is more robust and less variable than manually traced and visual EF. Our study also confirms the feasibility of the automated measurement of LS (42,43), which has the distinct advantage of having no interobserver and intraobserver variability.

STUDY LIMITATIONS. First, studies have reported better estimation of LV function using 3D echocardiography (44). There are also 3D tracking methods that apply automated algorithms (5,19,25,26,45), and 3D strain is a feasible method that can provide both global and regional evaluations with 1 analysis (45). The 3D techniques were not evaluated in this study and may have the potential for more accurate assessments (45). Second, in the present study, the interinstitutional variability of EF was evaluated; however, the variability of LS was only assessed at the reference center because the automated LS assessment protocol has not been approved for clinical use by the Food and Drug Administration and, therefore, could not be implemented at all centers. Third, because visual estimations of EF were done in the same settings as the manual assessments at local centers, it is possible that visual EF might have been biased by the manual measurements. However, the evaluation of intraobserver and interobserver variability was completely blinded and no adjustment was possible, as the measured EF was not displayed when the investigators tracked the images manually. Similarly, at the central laboratory, visual assessment was performed without knowledge of the Simpson measurement. Fourth, patients with atrial fibrillation and other arrhythmias were excluded from the study. Thus, the performance of the software in patients with arrhythmias and irregular heartbeats remains to be tested in further studies.

Finally, although intraobserver, interobserver, and beat-to-beat variability were measured, normal interstudy variability was not addressed in the current study, and thus, future studies should take this into consideration.

CONCLUSIONS

Machine learning-enabled echocardiography image analysis for fully automated assessment of LV volumes and EF is feasible and gives precise results within seconds that are comparable to manual determination. Furthermore, this new technique provides additional information on quantitative variables, such as LS, which provide further incremental assessment of LV systolic function beyond EF.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Partho Sengupta, Zena and Michael A. Wiener Cardiovascular Institute, Marie-Josée and Henry R. Kravis Center for Cardiovascular Health, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029-6574. E-mail: Partho.Sengupta@mountsinai.org.

PERSPECTIVES

COMPETENCY IN SYSTEMS-BASED PRACTICE: Echocardiographic assessment of LVEF and longitudinal strain may be useful to guide clinical management, but manual measurements are time-consuming and variable. The development of automated methods to assess LV function could standardize these measurements to facilitate clinical research and enhance patient care.

TRANSLATIONAL OUTLOOK: Future studies should evaluate whether automated LV assessments in longitudinal patient care and clinical trials of therapeutic interventions improve the consistency and value of serial echocardiographic assessments compared with visual assessment and manual measurements.

REFERENCES

1. Wong M, Johnson G, Shabetai R, et al., for the V-HeFT VA Cooperative Studies Group. Echocardiographic variables as prognostic indicators and therapeutic monitors in chronic congestive heart failure. Veterans Affairs cooperative studies V-HeFT I and II. *Circulation* 1993;87:VI65-70.
2. 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.
3. Hoffmann R, Barletta G, von Bardeleben S, et al. Analysis of left ventricular volumes and function: a multicenter comparison of cardiac magnetic resonance imaging, cine ventriculography, and unenhanced and contrast-enhanced two-dimensional and three-dimensional echocardiography. *J Am Soc Echocardiogr* 2014;27:292-301.
4. Thavendiranathan P, Popović ZB, Flamm SD, et al. Improved interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images. *J Am Soc Echocardiogr* 2013;26:1267-73.
5. Barbosa D, Heyde B, Dietenbeck T, et al. Quantification of left ventricular volume and global function using a fast automated segmentation tool: validation in a clinical setting. *Int J Cardiovasc Imaging* 2013;29:309-16.
6. Cannesson M, Tanabe M, Suffoletto MS, et al. A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. *J Am Coll Cardiol* 2007;49:217-26.
7. Reisner SA, Lysyansky P, Agmon Y, et al. Global longitudinal strain: a novel index of left ventricular systolic function. *J Am Soc Echocardiogr* 2004;17:630-3.
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. *Eur Heart J Cardiovasc Imaging* 2015;28:183-93.
9. Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.
10. Khan AM, Wieggers SE. The importance of being expert: is it time to revisit the concept? *J Am Soc Echocardiogr* 2012;25:218-9.
11. Akinboboye O, Sumner J, Gopal A, et al. Visual estimation of ejection fraction by two-dimensional echocardiography: the learning curve. *Clin Cardiol* 1995;18:726-9.
12. Picano E, Lattanzi F, Orlandini A, et al. Stress echocardiography and the human factor: the importance of being expert. *J Am Coll Cardiol* 1991;17:666-9.
13. Kaufmann BA, Min SY, Goetschalckx K, et al. How reliable are left ventricular ejection fraction cut offs assessed by echocardiography for clinical decision making in patients with heart failure? *Int J Cardiovasc Imaging* 2013;29:581-8.
14. Blondheim DS, Friedman Z, Lysyansky P, et al. Use of an automatic application for wall motion classification based on longitudinal strain: is it affected by operator expertise in echocardiography? A multicentre study by the Israeli Echocardiography Research Group. *Eur Heart J Cardiovasc Imaging* 2012;13:257-62.
15. Liel-Cohen N, Tsadok Y, Beeri R, et al. A new tool for automatic assessment of segmental wall motion based on longitudinal 2D strain: a multicenter study by the Israeli Echocardiography Research Group. *Circ Cardiovasc Imaging* 2010;3:47-53.
16. Seng CH, Demirli R, Amin MG, et al. Automatic left ventricle detection in echocardiographic images for deformable contour initialization. *Conf Proc IEEE Eng Med Biol Soc* 2011;2011:7215-8.
17. Maret E, Brudin L, Lindstrom L, et al. Computer-assisted determination of left ventricular endocardial borders reduces variability in the echocardiographic assessment of ejection fraction. *Cardiovasc Ultrasound* 2008;6:55.
18. Rahmouni HW, Ky B, Plappert T, et al. Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence-assisted border detection. *Am Heart J* 2008;155:562-70.
19. Shibayama K, Watanabe H, Iguchi N, et al. Evaluation of automated measurement of left ventricular volume by novel real-time 3-dimensional echocardiographic system: validation with cardiac magnetic resonance imaging and 2-dimensional echocardiography. *J Cardiol* 2013;61:281-8.
20. Thomas JD, Adams DB, Devries S, et al. Guidelines and recommendations for digital echocardiography. *J Am Soc Echocardiogr* 2005;18:287-97.
21. Marsousi M, Ahmadian A, Kocharian A, et al. Active ellipse model and automatic chamber detection in apical views of echocardiography images. *Ultrasound Med Biol* 2011;37:2055-65.
22. Jacob M, Blu T, Unser M. Efficient energies and algorithms for parametric snakes. *IEEE Trans Image Process* 2004;13:1231-44.
23. Mignotte M, Meunier J, Tardif JC. Endocardial boundary estimation and tracking in echocardiographic images using deformable templates and Markov Random Fields. *Pattern Anal Appl* 2001;4:256-71.
24. Karagiannis SE, Roelandt J, Qazi M, et al. Automated coupled-contour and robust myocardium tracking in stress echocardiography. *Eur J Echocardiogr* 2007;8:431-7.

25. Barbosa D, Dietenbeck T, Heyde B, et al. Fast and fully automatic 3-d echocardiographic segmentation using B-spline explicit active surfaces: feasibility study and validation in a clinical setting. *Ultrasound Med Biol* 2013;39:89-101.
26. Barbosa D, Dietenbeck T, Schaerer J, et al. B-spline explicit active surfaces: an efficient framework for real-time 3-D region-based segmentation. *IEEE Trans Image Process* 2012;21:241-51.
27. Szulik M, Pappas CJ, Jurcut R, et al. Clinical validation of a novel speckle-tracking-based ejection fraction assessment method. *J Am Soc Echocardiogr* 2011;24:1092-100.
28. Gudmundsson P, Rydberg E, Winter R, et al. Visually estimated left ventricular ejection fraction by echocardiography is closely correlated with formal quantitative methods. *Int J Cardiol* 2005;101:209-12.
29. van 't Hof AW, Schipper CW, Gerritsen JG, et al. Comparison of radionuclide angiography with three echocardiographic parameters of left ventricular function in patients after myocardial infarction. *Int J Cardiac Imaging* 1998;14:413-8.
30. McGowan JH, Cleland JG. Reliability of reporting left ventricular systolic function by echocardiography: a systematic review of 3 methods. *Am Heart J* 2003;146:388-97.
31. Hoffmann R, von Bardeleben S, ten Cate F, et al. Assessment of systolic left ventricular function: a multi-centre comparison of cineventriculography, cardiac magnetic resonance imaging, unenhanced and contrast-enhanced echocardiography. *Eur Heart J* 2005;26:607-16.
32. Shahgaldi K, Gudmundsson P, Manouras A, et al. Visually estimated ejection fraction by two dimensional and triplane echocardiography is closely correlated with quantitative ejection fraction by real-time three dimensional echocardiography. *Cardiovasc Ultrasound* 2009;7:41.
33. Nesbitt GC, Mankad S, Oh JK. Strain imaging in echocardiography: methods and clinical applications. *Int J Cardiovasc Imaging* 2009;25 Suppl 1: 9-22.
34. Kalam K, Otahal P, Marwick TH. Prognostic implications of global LV dysfunction: a systematic review and meta-analysis of global longitudinal strain and ejection fraction. *Heart* 2014;100: 1673-80.
35. Spethmann S, Rieper K, Riemekasten G, et al. Echocardiographic follow-up of patients with systemic sclerosis by 2D speckle tracking echocardiography of the left ventricle. *Cardiovasc Ultrasound* 2014;12:13.
36. Bière L, Donal E, Terrien G, et al. Longitudinal strain is a marker of microvascular obstruction and infarct size in patients with acute ST-segment elevation myocardial infarction. *PLoS One* 2014; 9:e86959.
37. Rhea IB, Uppuluri S, Sawada S, et al. Incremental prognostic value of echocardiographic strain and its association with mortality in cancer patients. *J Am Soc Echocardiogr* 2015;28: 667-73.
38. Saito M, Negishi K, Eskandari M, et al. Association of left ventricular strain with 30-day mortality and readmission in patients with heart failure. *J Am Soc Echocardiogr* 2015;28: 652-66.
39. Dahou A, Bartko PE, Capoulade R, et al. Usefulness of global left ventricular longitudinal strain for risk stratification in low ejection fraction, low-gradient aortic stenosis: results from the multicenter True or Pseudo-Severe Aortic Stenosis study. *Circ Cardiovasc Imaging* 2015;8:e002117.
40. Nagata Y, Takeuchi M, Mizukoshi K, et al. Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software. *J Am Soc Echocardiogr* 2015;28:630-41.
41. Farsalinos K, Daraban AM, Ünü S, Thomas JD, Badano LP, Voigt JU. Variability in global longitudinal strain measurements among nine different vendors: the EACVI/ASE Inter-Vendor Comparison Study. *J Am Soc Echocardiogr* 2015 Jul 23 [E-pub ahead of print].
42. Wierzbowska-Drabik K, Hamala P, Roszczyk N, et al. Feasibility and correlation of standard 2D speckle tracking echocardiography and automated function imaging derived parameters of left ventricular function during dobutamine stress test. *Int J Cardiovasc Imaging* 2014;30:729-37.
43. Belghithia H, Brette S, Lafitte S, et al. Automated function imaging: a new operator-independent strain method for assessing left ventricular function. *Arch Cardiovasc Dis* 2008;101:163-9.
44. Kühl HP, Schreckenberger M, Rulands D, et al. High-resolution transthoracic real-time three-dimensional echocardiography: quantitation of cardiac volumes and function using semi-automatic border detection and comparison with cardiac magnetic resonance imaging. *J Am Coll Cardiol* 2004;43:2083-90.
45. Pérez de Isla L, Balcones DV, Fernández-Golfin C, et al. Three-dimensional-wall motion tracking: a new and faster tool for myocardial strain assessment: comparison with two-dimensional-wall motion tracking. *J Am Soc Echocardiogr* 2009;22: 325-30.

KEY WORDS agreement, automated function, echocardiography, observer variation, software