Standardized Delineation of Endocardial Boundaries in Three-Dimensional Left Ventricular Echocardiograms

Alexandros Papachristidis, MD, Elena Galli, MD, PhD, Marcel L. Geleijnse, MD, PhD, Brecht Heyde, PhD, Martino Alessandrini, PhD, Daniel Barbosa, PhD, Michael Papitsas, MD, Gianpiero Pagnano, MD, Konstantinos C. Theodoropoulos, MD, MSc, Spyridon Zidros, MD, MSc, Erwan Donal, MD, PhD, Mark J. Monaghan, PhD, Olivier Bernard, PhD, Jan D'hooge, MSc, PhD, and Johan G. Bosch, PhD, *London, United Kingdom; Rennes and Villeurbanne, France; Rotterdam, The Netherlands; Leuven, Belgium; and Bologna, Italy*

Background: Three-dimensional (3D) echocardiography is fundamental for left ventricular (LV) assessment. The aim of this study was to determine discrepancies in 3D LV endocardial tracings and suggest tracing guidance.

Methods: Forty-five 3D LV echocardiographic data sets were traced by three experienced operators, from different centers, according to predefined guidelines. The 3D meshes were compared with one another, and the endocardial areas of discrepancies were identified. A discussion and retracing protocol was used to reduce discrepancies. For each data set, an average 3D mesh was produced (reference mesh). Subsequently, four novice operators, divided into two groups, traced 20 of the data sets. Two operators followed the tracing protocol and two did not.

Results: The intraclass correlation coefficients among the three experienced operators for end-diastolic volume, end-systolic volume, and ejection fraction were 0.952, 0.955, and 0.932. The absolute distances between tracings were 1.11 ± 0.45 mm. The highest tracing discrepancies were at the apical cap and anterior and anterolateral walls in end-diastole and end-systole and also at the basal anteroseptum in end-systole. Agreement with the reference meshes was better for the novice operators who followed the guidance (10.9 ± 17.3 mL, 10.2 ± 14.7 mL, and −2.2 ± 4.1% for end-diastolic volume, end-systolic volume, and ejection fraction) compared with those who did not (16.3 ± 16.4 mL, 17.0 ± 16.0 mL, and −4.2 ± 4.1%, respectively).

Conclusions: Comparing 3D LV tracings, the endocardial areas that are the most difficult to delineate were identified. The suggested protocol for LV tracing resulted in very good agreement among operators. The reference 3D meshes are available for online testing and ranking of LV tracing algorithms. (J Am Soc Echocardiogr 2017;30:1059-69.)

Keywords: 3D echocardiography, Left ventricular segmentation, Endocardial tracing

Three-dimensional (3D) echocardiography provides significant advantages over two-dimensional (2D) echocardiography and is currently applied in several aspects of cardiology.[1,2] The most common indication for performing echocardiography in adults is the evaluation of left ventricular (LV) size and function.[3] The use of 3D echocardiographic imaging eliminates geometric assumptions and misinterpretation errors caused by foreshortened views in 2D mode.[2,4] Several trials have demonstrated the reproducibility of 3D-derived LV measurements.[5-7] At present there are no clear standards or guidelines available for 3D LV endocardial border tracing, and there is no direct comparison of actual tracings among different operators.

Automated tracing of the left ventricle in 3D cardiac ultrasound data sets has been a subject of scientific research for the past 20 years,[8] but there has hardly been any comparison of different methods on the same data sets.[6]

In this study, we aimed to address these issues by suggesting a protocol for LV endocardial tracing in 3D echocardiographic data sets and

From the King's College Hospital, London, United Kingdom (A.P., M.P., G.P., K.C.T., S.Z., M.J.M.); Cardiology and CIC-IT1414 CHU Rennes and LTSI, INSERM 1099, University Rennes-1, Rennes, France (E.G., E.D.); Cardiology, Erasmus MC, Rotterdam, The Netherlands (M.L.G.); Cardiovascular Imaging and Dynamics, KU Leuven – University of Leuven, Leuven, Belgium (B.H., D.B., J.D.); Department of Electric Engineering, University of Bologna, Bologna, Italy (M.A.); University of Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, University of Lyon 1, Villeurbanne, France (O.B.); and Cardiology/Biomedical Engineering, Erasmus MC, Rotterdam, Netherlands (J.G.B.).

Conflicts of Interest: None.

Reprint requests: Alexandros Papachristidis, MD, Department of Cardiology, King's College Hospital NHS Foundation Trust, Denmark Hill, London SE5 9RS, United Kingdom (E-mail: *alexandros.papachristidis@nhs.net*).

0894-7317/$36.00

Copyright 2017 by the American Society of Echocardiography. All rights reserved.

http://dx.doi.org/10.1016/j.echo.2017.06.027

Abbreviations

2D = Two-dimensional

3D = Three-dimensional

ED = End-diastolic

EDV = End-diastolic volume

ES = End-systolic

ESV = End-systolic volume

ICC = Intraclass correlation coefficient

LV = Left ventricular

LVEF = Left ventricular ejection fraction

LVOT = Left ventricular outflow tract

creating a series of clinically realistic data sets with well-established reference tracings on the basis of manual tracings from three expert echocardiography centers. On the basis of this standard set, a competition for automated tracing methods was organized, associated with the Medical Image Computing and Computed Assisted Interventions 2014 symposium, which has been published previously.[9] The purpose of this competition was to provide reference 3D LV meshes for testing LV endocardial tracing algorithms. The reference meshes remain available online for continuous testing and ranking of fully automated or semiautomated algorithms.

Finally, we evaluated the usefulness of our tracing protocol in a clinically relevant setting, in which commercially available software was used by novice operators.

METHODS

Acquisition Protocol

We included 45 individuals: 15 healthy individuals, 15 patients with previous myocardial infarction at least 3 months before the time of echocardiography, and 15 patients with nonischemic dilated cardiomyopathy. The patients were recruited at three different institutions (Rennes University Hospital, Rennes, France; University Hospital Leuven, Leuven, Belgium; and Thoraxcenter, Erasmus MC, Rotterdam, The Netherlands). Fifteen patients undergoing echocardiography and meeting the inclusion criteria were recruited at each institution. Exclusion criteria were left bundle branch block, visually dyssynchronous left ventricle, and unacceptable image quality. Unacceptable image quality was defined as (1) significant stitching or other types of artifacts affecting the tracking of endocardium or (2) poor visualization of the LV wall or wall out of the image sector to an extent that the image could no longer be manually analyzed with good confidence in multiple segments. The image quality of the accepted data sets was graded as good, fair, or poor (Figure 1). Good quality was defined when the endocardium was visible in end-diastolic (ED) and end-systolic (ES) instances in all 17 segments throughout the cardiac cycle, fair quality if the endocardium was not clearly seen in one or two segments, and poor quality when the endocardial border was not clearly distinguished in ES or ED frames in more than two segments, but the operator could still define the border with confidence by tracking the endocardium throughout the cardiac cycle and also by considering adjacent segments. The variation in image quality was a result of recruitment of cases in a real-life setting and was not intentional. The image quality variation was similar in all three hospitals' data sets.

We used echocardiography machines from three different vendors: Vivid E9 with a 4V probe (GE Vingmed Ultrasound, Horten, Norway), iE33 with an X3-1 or X5-1 probe (Philips Medical Systems, Andover, MA), and SC2000 with a 4Z1c probe (Siemens Healthcare, Erlangen, Germany). Machine settings were optimized to achieve maximum image quality while keeping the volume rate above 16 Hz. The mean number of frames per cardiac cycle was 25.7 ± 8.5. Acquired data were fully anonymized and handled within the regulations set by the local ethical committee of each hospital.

Endocardial Tracing Procedure

ED and ES frames were identified. Nine standard anatomic planes were defined: four longitudinal planes through the long axis under 45° angles and five transverse (short-axis) planes divided equally along the long axis. For the tracings, a custom noncommercial tracing package for 3D echocardiograms (Speqle3D) was used, developed at the University of Leuven.[10] A single experienced operator from each center (A.P., M.L.G., E.G.) with experience of >300 3D LV tracing analyses was appointed to perform the tracings. Each operator independently traced the endocardial border in the nine predefined planes, in both ED and ES instances. To guarantee direct comparisons, the operators were only allowed to contour in the nine predefined slices and in the allocated ED and ES frames. All 45 data sets were traced by all three operators.

A set of guidelines for performing the LV tracing was defined at the beginning of the project and revised subsequently by comparing the tracing conventions of the different centers. Basic aims were as follows:

Include trabeculae and papillary muscles in the LV cavity (Figure 2). A suggestion was for the operator to take as a reference point the endocardial border that is free of papillary muscle and then trace "outside" the papillary muscle to meet the endocardium at the other edge of the muscle (i.e., from the basal to the apical segment or vice versa). Also, we suggested tracing at the level of the trough of endocardial creases to include trabeculations in the LV cavity.

Keep tissue consistency between end-diastole and end-systole and between adjacent and intersecting planes. The operator was asked to play the cine loop forward and backward to ensure that the traced endocardial border in end-diastole was corresponding to the same tissue line in end-systole by tracking the endocardium throughout the cardiac cycle. During this process, special consideration was taken with regard to elevation plane artifacts. Also, the prototype software showed the projection of the intersection points between tracings of orthogonal planes (Figure 3). The operator therefore ensured tissue consistency between transverse and longitudinal planes.

In long-axis views, draw up to the mitral valve annulus on the inside of the bright ridge up to the point at which the valve leaflet is hinging. The mitral valve annulus is sometimes quite difficult to trace with consistency. For this reason, we suggested that the operator should trace at the ventricular side of the annulus and pay special attention to identify the leaflet hinge point by reviewing the cine loop instead of judging on the basis of a single frame (Figure 4).

Partly exclude the LV outflow tract (LVOT) from the cavity by drawing from the septal mitral valve hinge point to the septal wall to create a smooth shape (Figure 5). The LVOT is one of the most challenging parts of LV endocardial tracing. We proposed to trace in a way that partially excludes the LVOT and provides a smooth shape of the basal anteroseptal wall segment to keep the LV shape symmetric and also avoid giving the impression of a dyskinetic segment as the LVOT expands during systole.

Draw the apex high up near the epicardium in both end-diastole and end-systole, taking into consideration that there should be

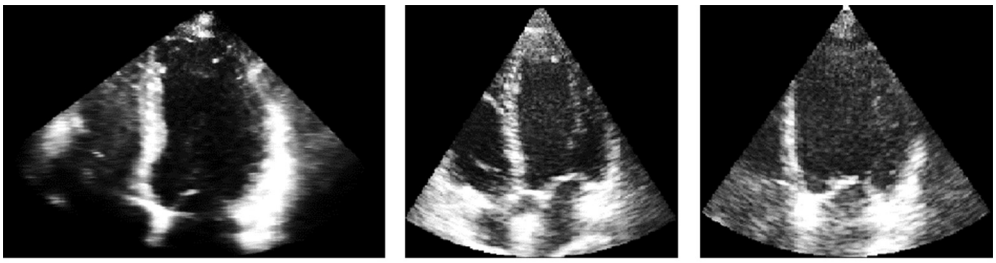

Figure 1 Examples of variability in image quality of data sets.

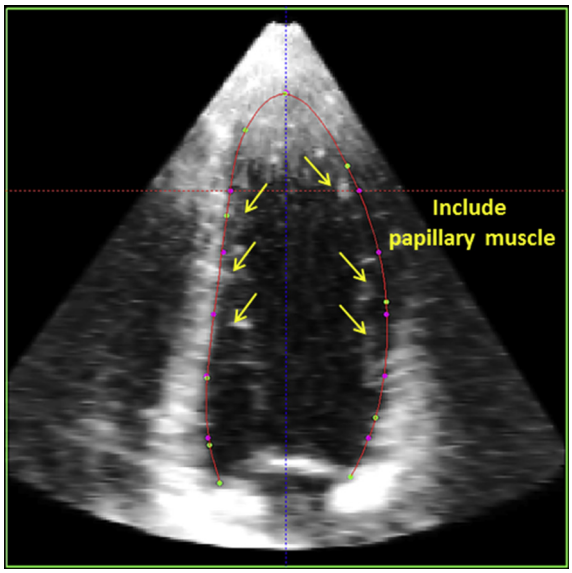

Figure 2 Example of tracing including trabeculae and papillary muscles (*yellow arrows*) in the LV cavity.

little displacement of the true apex point. As is known from anatomy, the LV wall is actually thin at the true apex and hardly moves during the cardiac cycle.[11,12] Apparent apical wall thickness and motion in cross-sectional images is due to foreshortening and/or blurred trabeculations (Figure 6). Also, the endocardium is more difficult to visualize compared with the epicardium at the apical region. Therefore, tracing close to epicardium can provide better anatomic ED and ES consistency than the less "visible" endocardium. This is expected to have little effect on volumes but makes the contouring more consistent.

Evaluation of Correspondence

After all three experts had submitted their tracings, 3D shapes (Figure 7) were generated from the nine 2D contours in end-diastole and end-systole by interpolation.[10,13] The shapes were represented by a 3D mesh with a resolution of about 40 by 80 points (longitudinal × circumferential). Each 3D mesh consisted of about 3,200 points (vertices). These vertices from the three operators' tracings were averaged to produce the reference mesh. The 3D meshes of the three operators were compared pairwise and mean absolute distances, the maximal perpendicular distance between all points of two meshes (the so-called Hausdorff distance[14]), LV volumes, and LV ejection fraction (LVEF) differences were calculated. The surface distance between the surface of an operator's 3D mesh and the surface of the reference mesh was represented in a 3D model using a color-coded approach (Figure 7). This allowed visualization of endocardial border discrepancies between a single operator's tracing and the reference mesh. For the purposes of the workshop,[9] the meshes of the three operators had to be consistent with one another. To verify consensus, the following criteria were used: Hausdorff distance ≤ 5 mm between each individual mesh and the reference mesh, percentage (relative) difference in LV volumes ≤ 10% among all three operators in pairs, and absolute difference in LVEF ≤ 5 percentage points among all three operators in pairs. The percentage difference was calculated as the absolute value of the difference between two operators divided by the mean. If consensus criteria were not met, the tracings would be discussed among the operators. This was done by superimposing the tracings of the three operators for each one of the nine predefined planes at both end-diastole and end-systole (Figure 5). For the purposes of our workshop,[9] one or more of the operators would then revise their tracings to reach consensus following discussion. Then the evaluation process would be repeated and slightly milder consensus criteria were applied: the average of the three pairwise observer differences was evaluated and Hausdorff distances ≤ 7 mm were accepted. The final averaged 3D mesh for each case, after revision process, was considered the "reference mesh."

The Tracing Guidance Applied to Novice Operators Using a Commercially Available Semiautomated Software

We additionally explored the usefulness of our tracing protocol using a vendor-independent commercially available software for 3D LV endocardial tracing (4D LV Analysis Image Arena version 3.5; TomTec Imaging Systems, Unterschleissheim, Germany). Four operators with little experience in 3D echocardiographic LV tracing were asked to trace 20 of our study data sets, which were randomly selected from the GE and Siemens data sets. The Philips data sets could not be imported to the commercially available software for analysis, because of technical limitations related to acquisition and storage process.

All four operators had experience of 20 to 30 cases and were from a single center (King's College Hospital, London, United Kingdom). They were divided into two groups. Operators A1 and A2 (group A) traced according to their own discretion. Operators B1 and B2 (group B) were provided with our tracing protocol and recommendations. All operators were blinded to other operators' results and also to the reference meshes.

In the TomTec platform, the software initially detects the ED and ES instances. The operator is required to define the LVOT in a short-axis plane as a reference point. From that reference point, the software identifies the apical three-chamber view and then automatically defines the two-chamber and four-chamber views (60° and 120° incremental views).

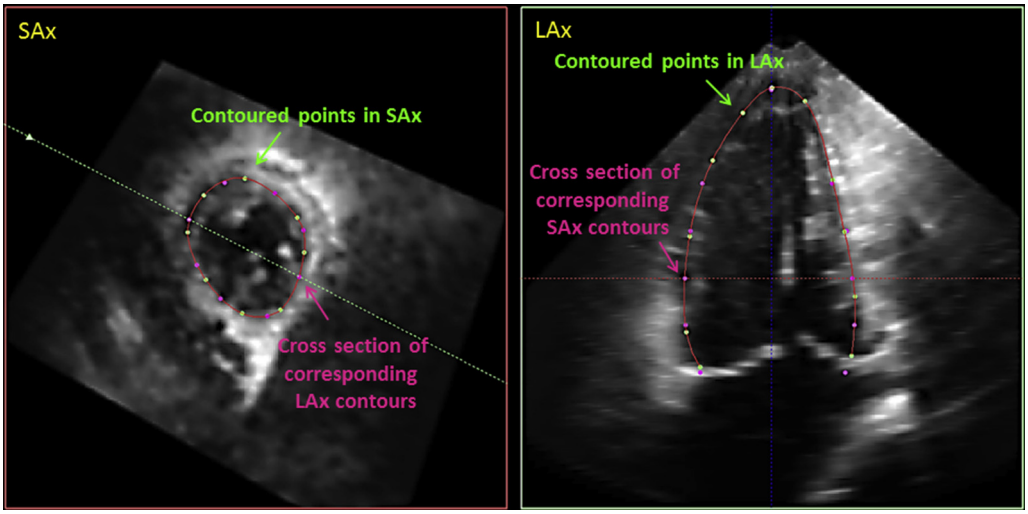

Figure 3 Example of manual endocardial drawing in Speqle3D software. (*Left*) Transverse plane (short axis, SAx). (*Right*) Longitudinal plane (long axis, LAx). The *green dots* in both images represent the points set on the endocardial border by the operator on the actual plane. The *red line* represents the endocardial contour created by b-spline interpolation of the *green dots*. The *pink dots* represent the cross-sectional points of the contours in the orthogonal planes.

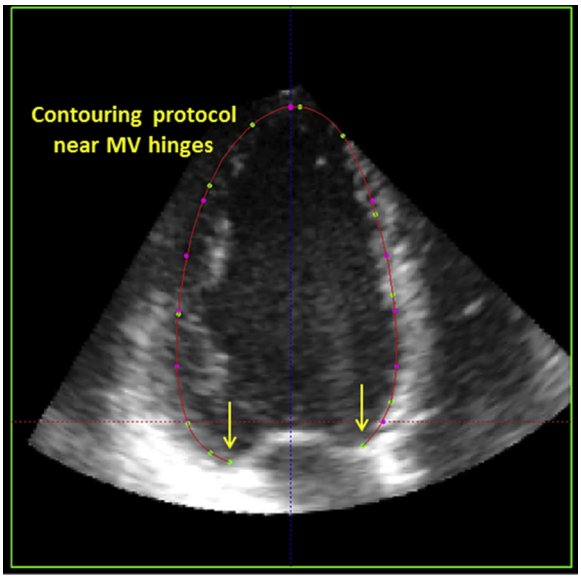

Figure 4 Example of manual tracing pointing out the mitral valve (MV) hinge points (*yellow arrows*).

The operator aligns the longitudinal axis of the left ventricle in all three apical views such that the axis crosses the LV apex and the middle of the mitral annular plane in all views. Two markers at the two ends of the long-axis line are placed by the operator at the apex and the mitral annular level in end-diastole. Subsequently, the application automatically defines the endocardial border in all three apical views in end-diastole. The operator can adjust the tracings manually as necessary in all views. Then the software tracks the endocardium throughout the cardiac cycle. Once this is completed, the operator can modify the tracings in ES and ED instances in all apical views (two-, three-, and four-chamber views) and also in a short-axis plane, which can be manually swept along the full length of the LV long axis. The operator is not allowed to modify the endocardial tracing in frames other than the ED and ES frames. The application continuously calculates the ED volume (EDV), ES volume (ESV) and LVEF as the operator manipulates the tracings. For the purposes of our study, once the final tracings were confirmed, the derived volumes and LVEF were recorded by the operator in a separate spreadsheet.

Statistical Analysis

Statistical analysis was performed using SPSS version 20.0.0 (IBM, Somers, NY). Differences in mean values for continuous variables were tested using Student's *t* test, the Mann-Whitney *U* test for unpaired samples, or the Wilcoxon signed rank test for paired samples as appropriate. Interobserver variability was tested using the intraclass correlation coefficient (ICC) for absolute agreement in a two-way mixed model. Values for continuous variables are presented as mean ± SD. Statistical significance was considered for a two-tailed P value $< .05$.

RESULTS

Baseline Demographics and Characteristics of Data Sets

The mean age of the tested population was 53.9 ± 18.8 years, and 88.9% of patients were men. The mean values and SDs of volumes and LVEF of the reference meshes are shown in Table 1. Fourteen cases (31%) were of good quality, 16 (36%) of fair quality, and 15 (33%) of poor quality.

Interobserver Variability

The ICCs for the derived clinical parameters (EDV, ESV, and LVEF) were very high (>0.9) for the initial tracings (Table 1). This shows excellent agreement among operators. The average differences and the percentage differences among operators are shown in Table 2. The average difference for initial tracings for EDV was −3.3 ± 27.0 mL, for ESV was −2.7 ± 23.9 mL, and for LVEF was 0.5 ± 4.9 percentage points. The percentage differences for EDV, ESV, and LVEF were 10.4 ± 7.9%, 12.9 ± 10.6%, and 10.4 ± 9.1%, respectively. The mean values of mean absolute distances and Hausdorff distances are presented in Table 3.

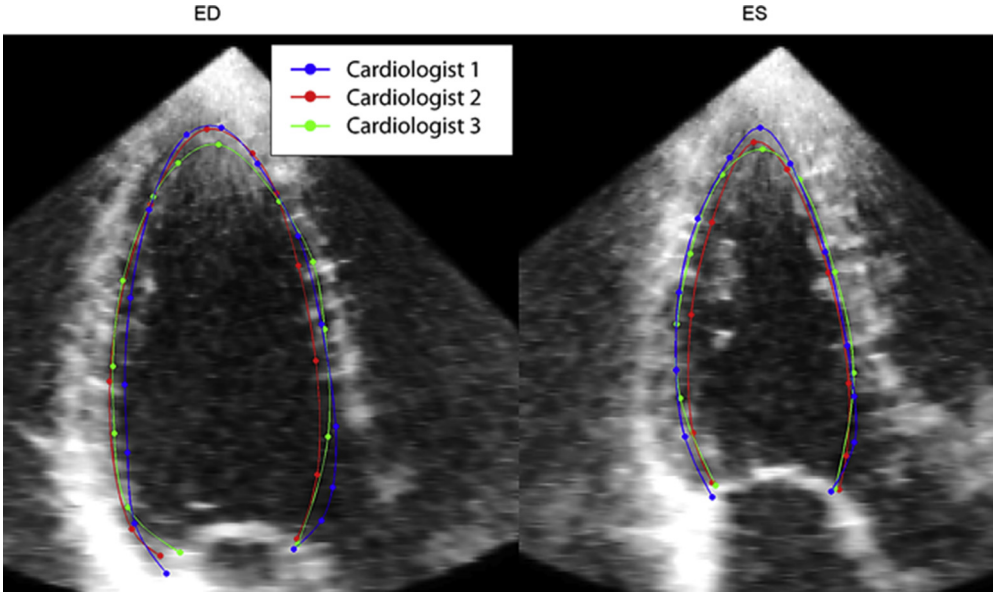

Figure 5 Superimposition of the three experienced operators' manual tracings in a long-axis view. (*Left*) End-diastole. (*Right*) End-systole. The tracing of the LVOT (bottom right end of contours) has been performed in a way that partially excludes LVOT, and a smooth curved line is drawn from the mitral valve hinge point to the septal wall curve.

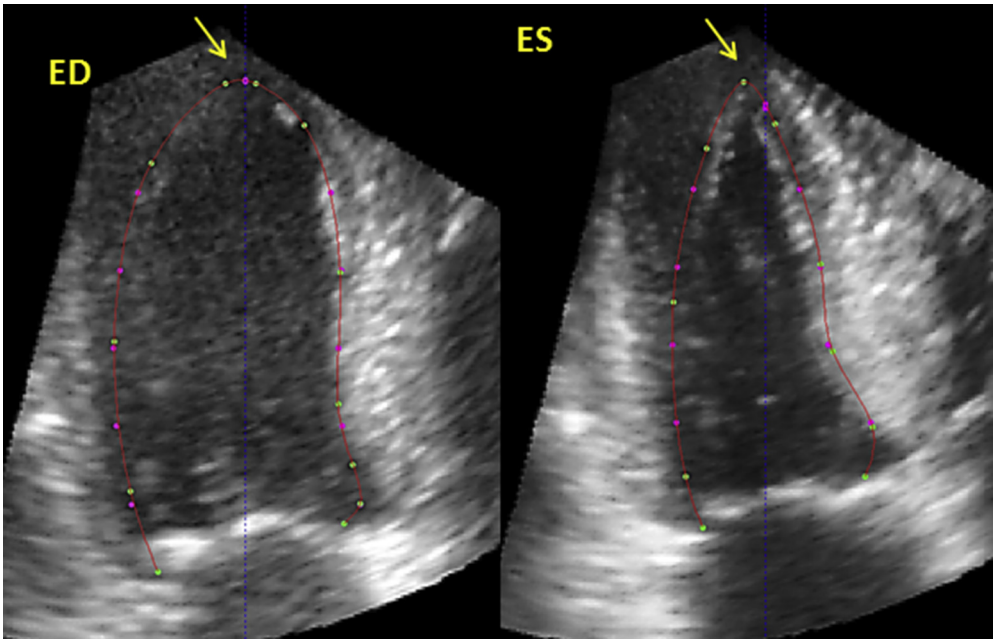

Figure 6 Example of tracing of the apex (*yellow arrows*), which is close to epicardium with little displacement in end-systole.

Applying the predefined criteria for consensus, agreement in initial tracings was reached in 12 cases (26.7%) for EDV, nine cases (20.0%) for ESV, 23 cases (51.1%) for LVEF, and 33 cases (73.3%) for Hausdorff distances in end-systole and 34 (75.6%) in end-diastole.

Distance Differences in Tracing

The surface distances of all three operators were averaged over the 45 cases. The resulting distances are shown in a 17-segment bull's-eye plot according to American Heart Association's LV segmentation guidance.[15] Four different bull's-eye plots were calculated corresponding to ES and ED frames before and after the consensus process (Figure 8).

From these bull's-eye plots of average distance, we assessed the endocardial areas that showed the highest distances in ED and ES frames. In the initial contours (Figure 8), the highest distances in end-diastole were observed at the apical cap (segment 17) and also at the anterior, the mid anterolateral, and apical lateral segments (segments 1, 7, 13, 12, and 16, respectively). The best agreement was demonstrated in the inferoseptum (segments 3 and 9), the apical septum (segment 14), and the mid inferior wall (segment 10).

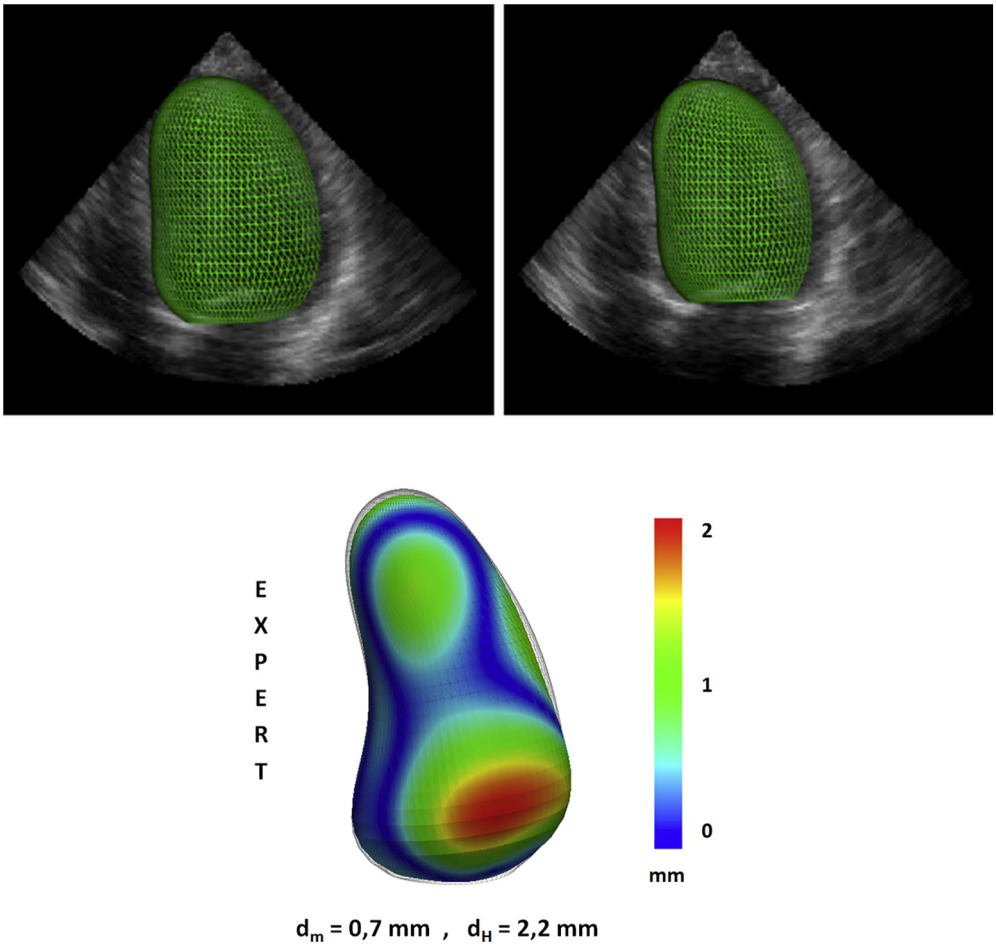

Figure 7 (*Top*) Three-dimensional meshes were generated for each individual tracing in ED (*left*) and ES (*right*) instances. (*Bottom*) An example of a 3D mesh derived by one experienced operator's tracings with the endocardial distances from reference mesh illustrated in color code.

Table 1 EDV, ESV, and LVEF on the basis of the references meshes, which were computed from initial tracings, and the relevant ICCs among the three experienced operators

Variable	Reference mesh, mean ± SD	ICC (95% CI)
EDV (mL)	173.5 ± 83.5	0.952 (0.918–0.972)
ESV (mL)	114.5 ± 78.5	0.955 (0.927–0.974)
LVEF (%)	38.5 ± 13.6	0.932 (0.861–0.959)

For ES frames, the same trends can be seen, as the highest distance remained at the apex (segment 17), the mid and apical anterior wall (segments 7 and 13), the mid anterolateral (segment 12), and the apical lateral (segment 16). Additionally, a high distance error also appeared at the basal anteroseptum (segment 2).

After the revision process, the overall distance errors improved for all segments (Table 3, Figure 8). Nevertheless, the trends observed before consensus still held. At end-diastole, the highest distance errors were again at segments 17, 1, 7, 13, 6, 12, and 16 and at end-systole at the apical cap (segment 17), the mid anterior wall (segment 7), and the basal anteroseptum (segment 2).

Tracing Protocol with Commercially Available Semiautomated Software and Novice Operators

Agreement between Novice Operators and Reference Meshes. The variation in image quality of the 20 data sets that were used in this substudy was similar to that of the whole cohort: good, fair, and poor image quality of 30%, 40%, and 30%. The derived LV volumes from all four novice operators were lower compared with the reference meshes (Table 4). On the contrary, LVEF was higher (Table 4). The difference between each operator and the reference meshes was statistically significant ($P < .05$) for EDV, ESV, and LVEF for all four operators, apart from LVEF by operator B1 ($P = .067$). The average differences for EDV, ESV, and LVEF between group A measurements and the reference meshes were 16.3 ± 16.4 mL, 17.0 ± 16.0 mL, and −4.2± 4.1%. The relevant values for group B were 10.9 ± 17.3 mL, 10.2 ± 14.7 mL, and −2.2 ± 4.1%.

Interobserver Agreement between Novice Operators. The averaged EDV and ESV of group A operators were lower compared with group B (136.8 ± 60.3 vs 142.4 ± 60.0 mL [$P = .097$] and 80.4 ± 47.8 vs 87.4 ± 53.0 mL [$P = .014$]), whereas LVEF was higher (44.3 ± 13.3% vs 42.2 ± 13.4%, $P = .025$). The average differences

Table 2 Average differences among the three experienced operators for the initial and revised (after consensus) tracings

	EDV (mL)		ESV (mL)		LVEF (%)	
	Mean ± SD	Percentage difference (%)	Mean ± SD	Percentage difference (%)	Mean ± SD	Percentage difference (%)
All (*N* = 45)						
Initial tracings	−3.3 ± 27.0	10.4 ± 7.9	−2.7 ± 23.9	12.9 ± 10.6	0.5 ± 4.9	10.4 ± 9.1
After consensus	1.5 ± 13.9	5.9 ± 4.5	1.1 ± 9.5	5.8 ± 4.9	0.1 ± 3.2	6.8 ± 5.7
Good/fair image quality (*n* = 30)						
Initial tracings	−1.2 ± 17.7	9.6 ± 6.6*	−1.6 ± 17.2	8.4 ± 10.8	0.9 ± 4.9	10.5 ± 8.3
After consensus	0.8 ± 10.5	5.5 ± 3.5*	0.4 ± 8.6	3.3 ± 4.0	0.5 ± 3.1	7.2 ± 6.0
Poor image quality (*n* = 15)						
Initial tracings	−6.0 ± 35.6	13.2 ± 8.9*	−4.7 ± 33.4	8.3 ± 10.1	−0.3 ± 4.9	10.2 ± 10.6
After consensus	1.7 ± 17.2	7.7 ± 5.9*	2.5 ± 11.0	4.7 ± 6.1	−0.6 ± 3.2	6.1 ± 5.0

Comparisons were made between good- or fair- and poor-quality image data sets for all variables.
*$P < .05$.

Table 3 Average differences between individual operators' and reference meshes for the initial and revised (after consensus) tracings

	HD in end-diastole (mm)	HD in end-systole (mm)	MAD in end-diastole (mm)	MAD in end-systole (mm)
All (*N* = 45)				
Initial tracings	3.6 ± 1.2	3.7 ± 1.2	1.1 ± 0.5	1.1 ± 0.5
After consensus	2.8 ± 0.8	2.9 ± 0.8	0.8 ± 0.2	0.8 ± 0.2
Good/fair quality (*n* = 30)				
Initial tracings	3.5 ± 1.0	3.7 ± 1.4	1.0 ± 0.3	1.1 ± 0.3
After consensus	2.8 ± 0.5	2.9 ± 0.8	0.8 ± 0.1	0.8 ± 0.1
Poor quality (*n* = 15)				
Initial tracings	3.8 ± 1.5	3.7 ± 1.2	1.2 ± 0.3	1.2 ± 0.4
After consensus	2.9 ± 0.7	2.9 ± 0.8	0.9 ± 0.2	0.8 ± 0.1

HD, Hausdorff distance; *MAD*, mean absolute distance.
Data are expressed as mean ± SD.

(bias) for EDV, ESV, and LVEF between operators A1 and A2 were 2.1 ± 11.2 mL, 3.3 ± 9.0 mL, and −2.1 ± 3.1%. In group B (operator B1 vs B2), the relevant values were 1.1 ± 10.9 mL, 0.7 ± 9.2 mL, and −0.2 ± 3.7%. The differences in EDV, ESV, and LVEF between the two operators in each group were not statistically significant, with the exception of LVEF in group A ($P < .001$). The ICCs between operators in group A for EDV, ESV, and LVEF were 0.983 (95% CI, 0.959–0.993), 0.981 (95% CI, 0.952–0.992), and 0.963 (95% CI, 0.862–0.987). The relevant ICCs in group B were 0.984 (95% CI, 0.960–0.994), 0.986 (95% CI, 0.965–0.994), and 0.966 (95% CI, 0.915–0.986).

DISCUSSION

We suggest guidance for LV endocardial tracing in 3D echocardiographic data sets that results in good agreement among experienced operators from different centers. Good agreement between 3D echocardiography and magnetic resonance imaging has been previously reported, but currently there is no standardized guidance for LV tracing in 3D echocardiographic data sets. We identified the endocardial areas that show the most significant distance error in 3D LV manual tracings of experienced operators, and we provide additional recommendation to improve agreement. The provided tracing protocol can be useful in a clinical setting and can improve the accuracy of novice operators' tracings, using vendor-independent, semiautomated, commercially available software.

The ICCs for EDV, ESV, and LVEF among the three experienced operators were similar and very high in initial and final individual tracings (>0.9). The variability for initial tracings (Table 2) is similar to previously published studies.[16,17] However, our study has some unique features. First of all, the tracing process was fully manual and therefore more challenging in terms of variability with respect to previous studies[5,7,16,17] in which semiautomated methods were used. In our study, the operators were asked to trace manually 18 planes (nine planes in end-diastole and nine planes in end-systole), whereas in commercially available software the operator's input is usually required in only three to six planes. The challenge of fully manual tracing is also depicted in the results of our substudy. The semiautomated software overcomes the inexperience of the operators, providing better interobserver agreement compared with a fully manual protocol. Given this advantage of semiautomated software,

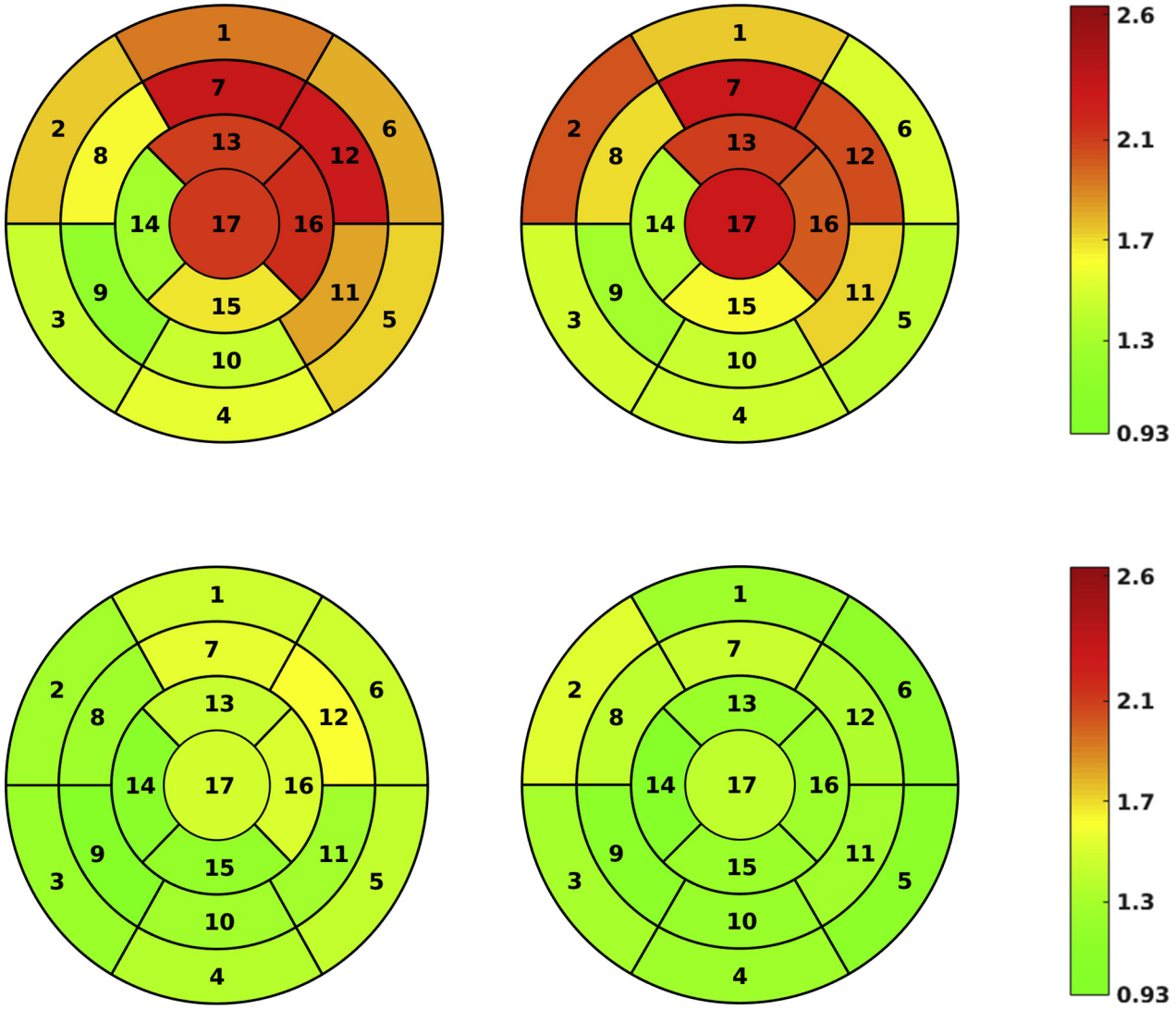

Figure 8 (*Top*) Seventeen-segment bull's-eye plot showing the average distances between the 3D meshes derived by individual experienced operators' initial tracings and the reference meshes in end-diastole (*left*) and end-systole (*right*). (*Bottom*) The same bull's-eye plot as above after the consensus process.

one may wonder why we chose the laborious and less reproducible fully manual work flow to create the reference meshes. This was done because we aimed to provide reference 3D meshes to test LV tracing algorithms. Using a semiautomated algorithm to create the reference tracings would bias the measurements toward this particular algorithm. Therefore, a tested algorithm of similar infrastructure would be expected to perform better compared with other algorithms. What we wanted was truly unbiased reference meshes and a fair comparison between algorithms, hence the fully manual work flow.

For the same reason, the reference 3D meshes were created by averaging the manual tracings of three experienced operators. We did not use a single operator's tracings, as this would result in biased reference meshes. There is always some degree of interobserver variability in 3D LV measurements, and also it is not possible for any operator to reproduce exactly the same tracing of a specific 3D data set, especially in a fully manually work flow such as ours. Therefore, we believe that averaging the variability of tracings, especially when they come from more than one experienced operator, can result in a more realistic reference 3D mesh. Additionally, the reference 3D meshes in our study were produced only if the tracings of the three operators met the strict agreement criteria we had set. If they failed, the tracings were discussed between the operators with the aforementioned superimposition method and retraced as necessary. The protocol with the discussion-revision process resulted in excellent agreement among experienced operators' tracings, as demonstrated by both clinical (LV volumes and LVEF) and anatomic (distance errors) criteria. Therefore, the produced reference tracings are expected to be "accurate" in terms of LV volumes and LVEF (though this was not tested against another modality, such as magnetic resonance imaging, in our study) and also "realistic" in terms of tracing distance errors.

A strong advantage of our study is the fact that the operators were blind to one another's tracings and also to derived clinical values (LV volumes and LVEF) when they initially traced the data sets. In all commercially available 3D LV software, the operator can review the derived EDV and ESV as well as the LVEF. It is not uncommon in clinical practice, especially in data sets of poor image quality, for

Table 4 EDV, ESV, and LVEF by four operators (A1, A2, B1, and B2) using commercially available software

	A1	A2	B1	B2	Reference
EDV (mL)	137.8 ± 59.0	135.7 ± 62.2	142.7 ± 59.5	141.7 ± 60.3	153.0 ± 68.3
ESV (mL)	82.0 ± 46.3	78.69 ± 49.6	87.5 ± 52.0	86.9 ± 54.5	97.4 ± 61.2
LVEF (%)	43.2 ± 12.5	45.3 ± 14.2	42.1 ± 13.6	42.4 ± 13.7	40.0 ± 13.2

Data are expressed as mean ± SD.

the operator to adjust the endocardial tracing on the basis of his or her visual estimation of volumes and LVEF, imposing significant bias on endocardial tracing. In our study this bias was completely eliminated because the prototype software did not allow calculation of LV volumes and LVEF. Therefore, the tracing was made entirely on the basis of an anatomic evaluation of LV endocardium and structures. This did not apply to our substudy, in which the commercially available software allowed visualization of LV volumes and LVEF.

This was a multicenter study involving three operators from three different centers and different countries, as opposed to previously published single-center studies. Both Nikitin *et al.*[7] and Jenkins *et al.*[5] reported somehow lower variability, but they evaluated the interobserver variability at their own centers. Tsang *et al.*[17] studied interinstitutional measurements of LV volumes between operators from two different centers, and they additionally assessed the effect of a short common training period. They reported ICCs for EDV, ESV, and LVEF of 0.75, 0.69, and 0.79, respectively, using semiautomated software. These values are considerably lower compared with our study. After the intervention of common training, they reported improvement in ICCs comparable with the values in our study. The mean percentage differences obtained by Tsang *et al.*[17] were 13.6%, 15.9%, and 12.2% for EDV, ESV, and LVEF. In our study, the percentage differences were 10.4%, 12.9%, and 10.4%, respectively. Mor-Avi *et al.*[16] reported percentage differences of 8 ± 8% for EDV and 13 ± 14% for ESV in a study conducted at four different institutions.

Nikitin *et al.*[7] included patients with good acoustic windows only. Soliman *et al.*[6] reported that they excluded patients if more than two LV segments were not well visualized. They estimate that their cohort represented the best half of the patients investigated in their echocardiography laboratory in terms of image quality. In our study, 33% of patients had more than two segments not well visualized (poor quality images; Figure 1). Therefore, there is a significant difference in image quality between these studies. We elected to include those patients because they represent a significant proportion of daily echocardiography practice. Despite that difference, Soliman *et al.* reported an average percentage difference for EDV of 8.2 ± 11.4% for a multiplanar interpolation method, whereas in our study it was 10.4 ± 7.9%. In the subgroup of cases with good or fair image quality in our study, the average percentage difference was 9.6 ± 6.6% as opposed to 13.2 ± 8.9% in the subgroup of cases with poor image quality ($P = .021$; Table 2).

An interesting finding in our study is that despite excellent agreement in quantification of the left ventricle on the basis of the ICC, the percentage differences in LV volumes between operators were >10% in 73.3% of the cases for EDV and in 80.0% of the cases for ESV. Also, the absolute difference in LVEF was greater than 5 percentage points in 48.9% of the cases. Mor-Avi *et al.*[16] also mentioned that in their study, the variability levels in individual patients far exceeded the acceptable 10% to 15% levels for LV volumes, though they did not provide more details. Reviewing this finding by direct comparison of contours in the context of our tracing protocol in predefined planes, it seems that the operators were tracing in a specific manner on the basis of their individual training. Although the motion of endocardium from end-diastole to end-systole was well tracked by all three operators, the actual endocardial points were different among the operators in the ED and ES frames, resulting in higher discrepancy in volumes as opposed to LVEF.

By analyzing the bull's-eye plots it seems that the areas of highest distance errors were the apical cap, the anterior, and the anterolateral walls (with the exception of basal anterolateral segment) in end-diastole and the same segments plus the basal anteroseptum in end-systole. These findings may be explained by the fact that the apex is always difficult to visualize and identify because of near-field artifacts and the presence of trabeculations. The anterolateral papillary muscle may cause some confusion with regard to the exact endocardial border of the anterolateral wall, and for this reason the basal anterolateral segment, which is not in close proximity to the papillary muscle, may be more easily visualized compared with the mid and apical segments. The presence of the papillary muscle in the anterior wall seems to cause similar difficulties in recognizing the actual endocardial border. Also, the orientation of the anterior wall relative to the ultrasound beam and its proximity to the lung tissue, which causes dropout more frequently, makes the anterior and anterolateral walls more difficult to trace accurately. In end-systole there is an additional high distance error in the basal anteroseptal segment (segment 2), which anatomically corresponds to the LVOT. Indeed, this area is generally difficult to trace in a consistent manner.

The best agreement was shown in segments 3, 9, 14, and 10 in end-diastole. The absence of a papillary muscle in the inferoseptum (segments 3, 9, and 14) along with the usually good-quality imaging of this wall in the four-chamber view contributed to the lowest distance error that was observed in these segments. The tracing of the mid inferior segment (segment 10) showed good agreement despite the presence of a papillary muscle, probably because in the two-chamber view, the inferior wall is usually more easily visualized compared with the anterior wall. In end-systole, the best agreement in tracings was seen again in the inferoseptal wall, probably for the same reasons. After the consensus process, the highest distance errors remained in the same segments. This points out the inherent difficulties in identifying the endocardial border of the anterior and anterolateral wall as well as the apical cap.

The absence of ventricular myocardium and clear-cut endocardial border in the LVOT makes the tracing of segment 2 (basal anteroseptum) quite challenging. In this setting, one might suggest that tracing of the LVOT should follow the level of the aortic valve annulus from the mitral valve hinge point to the aortic valve cusp hinge points and then to the septum. This might result in better agreement among operators. However, the LVOT expands during systole, as opposed to other segments, and the aortic annulus moves superiorly. This may give a false impression of dyskinesia, especially when using software that detects LV subvolume changes to assess LV dyssynchrony.

We paid significant attention to these anatomic pitfalls while establishing the tracing protocol, and we managed to achieve individual tracings within very strict limits of agreement. However, to address this issue, we came up with additional recommendations for these particular areas, and we propose that operators be more cautious when tracing at the apex, anterior, and anterolateral wall as well as at the basal anteroseptum. For the anterior and anterolateral walls, we suggest using the short-axis plane to ensure that a smooth, nearly circular contour is created. This may be particularly useful in the ED frame. In the ES instance it would be difficult to make a relevant recommendation because of possible regional wall motion abnormalities. The apical cap and the basal anteroseptum remain the Achilles' heel of endocardial tracing on the basis of our findings. For the apical cap, we would further suggest following the endocardial tracings of mid and apical segments so that in a normal heart the tracing would result in a bullet-shaped apex, while it is expected to be more rounded in dilated ventricles. However, some variability in tracings may be inevitable, and this is probably related to poor visualization of the apex with ultrasound. An additional suggestion for the LVOT tracing would be to ensure a symmetric shape of the tracings at the basal parts of the left ventricle in the three-chamber apical view. This practically means that the curve of the basal anteroseptum should be similar to that of the basal posterior segment, or at least not significantly irregular, in end-diastole. This would provide some symmetry and consistency in basal segments tracing. For the ES frame, such a correlation cannot be recommended.

The knowledge of areas of higher distance errors may be helpful in the interpretation of wall motion abnormalities or other echocardiographic modalities, such as segmental strain or 3D segmental displacement analysis. We would suggest that operators be more cautious in interpretation of the segmental analyses of the aforementioned areas, as visual evaluation or measurements may be less reproducible and less accurate compared with other areas. Interestingly enough, Marwick[18] reported that the anterolateral wall is a frequent site of false-negative results on stress echocardiography.

We evaluated the usefulness of the proposed guidelines and the aforementioned additional recommendations in operators with little experience in 3D echocardiography using semiautomated, vendor-independent, commercially available software. The group that followed the tracing protocol had better agreement with the reference meshes. Additionally, the bias between operators who followed the tracing protocol was lower compared with those who did not. The limits of agreement were comparable in the two groups. One would expect more narrow limits of agreement between operators who followed the predefined guidelines. The absence of significant difference between the two groups can be explained by the fact that operators who are trying to follow specific guidance tend to interact more with the tracings. As shown in our and other studies,[19] a higher degree of operator interaction is related to higher interobserver variability.

Limitations

The present results were obtained on image data of acceptable image quality. However, in everyday clinical practice, cardiologists and echocardiographers face the challenge of suboptimal image quality. As it has been demonstrated previously, image quality is related to bias in assessing 3D LV volumes.[20]

Furthermore, the manual tracing of the LV endocardium to produce the reference meshes in our study was performed in a way that does not reflect the actual process in everyday practice. In particular, the operators were provided with prespecified 2D planes derived from the 3D data set. In most currently available 3D echocardiography software, the operator is expected to align the image so that the longitudinal axis of the left ventricle crosses the apex and the middle of mitral valve annulus in all planes. Thus, the discrepancies related to different plane orientations by the individual operators were not tested in this study, though this effect is probably negligible. This limitation, however, was abolished in our substudy, in which semiautomated software was used and the operators were required to identify the three-chamber apical view and align the long axis of the left ventricle.

Because of technical limitations, we could not test the distance errors of endocardial tracings in our substudy, in which the commercially available software was used. Therefore, we cannot comment on tracing distance errors between novice operators or the efficacy of our additional recommendations to improve the tracing agreement in the noted areas of highest discrepancy.

Finally, the commercially available software (TomTec) uses the time-volume curve to calculate EDV and ESV, whereas in the manual work flow used by the experienced operators, the ED and ES frames were selected manually before the data sets were provided to the operators for tracing. This may be a confounder in the volume and LVEF differences noted between the experienced and novice operators. However, this is a limitation we had to accept in our attempt to test the proposed tracing guidance with commercially available software (TomTec), using as reference meshes the ones produced with custom software (Speqle3D).

CONCLUSIONS

The described protocol produces LV endocardial tracings with small variability. The level of agreement between operators as measured by differences in tracing distances and clinical calculations (LV volumes and LVEF) was very high. We identified that the apical cap, the anterior and anterolateral walls, as well as the basal anteroseptum are correlated with the highest distance errors between operators. The protocol and tracing guidance resulted in well-established reference 3D LV meshes and may serve as a convention for 3D LV endocardial tracing. It has been proved useful to less experienced operators to achieve better agreement with reference tracings using semiautomated commercially available software, compared with operators of similar experience who traced on the basis of their own discretion. Our reference tracings have been used to validate algorithms for LV automatic quantification,[9] and the data sets are available online for continuous evaluation of LV tracing algorithms, fostering innovation in algorithmic development of new automated tools.

ACKNOWLEDGMENTS

We thank the sonographers and cardiologists of the cooperating centers for their contributions, as well as Prof. Piet Claus for providing the basis of the software tool for volumetric data analysis (Speqle3D).

REFERENCES

1. Lang RM, Badano LP, Tsang W, Adams DH, Agricola E, Buck T, et al. EAE/ASE recommendations for image acquisition and display using three-dimensional echocardiography. J Am Soc Echocardiogr 2012;25:3-46.

2. Lang RM, Mor-Avi V, Sugeng L, Nieman PS, Sahn DJ. Three-dimensional echocardiography. The benefits of the additional dimension. J Am Coll Cardiol 2006;48:2053-69.
3. Monaghan MJ. Role of real time 3D echocardiography in evaluating the left ventricle. Heart 2006;92:131-6.
4. Lang RM, Mor-Avi V, Dent JM, Kramer CM. Three-dimensional echocardiography: is it ready for everyday clinical use? JACC Cardiovasc Imaging 2009;2:114-7.
5. Jenkins C, Bricknell K, Hanekom L, Marwick TH. Reproducibility and accuracy of echocardiographic measurements of left ventricular parameters using real-time three-dimensional echocardiography. J Am Coll Cardiol 2004;44:878-86.
6. Soliman OI, Krenning BJ, Geleijnse ML, Nemes A, Bosch JG, van Geuns RJ, et al. Quantification of left ventricular volumes and function in patients with cardiomyopathies by real-time three-dimensional echocardiography: a head-to-head comparison between two different semiautomated endocardial border detection algorithms. J Am Soc Echocardiography 2007;20:1042-9.
7. Nikitin NP, Constantin C, Loh PH, Ghosh J, Lukaschuk EI, Bennett A, et al. New generation 3-dimensional echocardiography for left ventricular volumetric and functional measurements: comparison with cardiac magnetic resonance. Eur J Echocardiogr 2006;7:365-72.
8. Leung KY, Bosch JG. Automated border detection in three-dimensional echocardiography: principles and promises. Eur J Echocardiogr 2010;11:97-108.
9. Bernard O, Bosch JG, Heyde B, Alessandrini M, Barbosa D, Camarasu-Pop S, et al. Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography. IEEE Trans Med Imaging 2016;35:967-77.
10. Heyde B, Barbosa D, Claus P, Maes F, Dhooge J. Three-dimensional cardiac motion estimation based on non-rigid image registration using a novel transformation model adapted to the heart. Proc STACOM 2012; LNCS 7746:142150.
11. Bradfield JW, Beck G, Vecht RJ. Left ventricular apical thin point. Br Heart J 1977;39:806-9.
12. Johnson KM, Johnson HE, Dowe DA. Left ventricular apical thinning as normal anatomy. J Comput Assist Tomogr 2009;33:334-7.
13. Claus P, Choi HF, D'hooge J, Rademakers FE. On the calculation of principle curvatures of the left-ventricular surfaces. Conf Proc IEEE Eng Med Biol Soc 2008;2008:961-4.
14. Huttenlocher DP, Kedem K. Computing the minimum Hausdorff distance for point sets under translation. Proc. of 6th Annual ACM Symp. on Comp. Geom. 1990 (SCG'90, Berkeley, CA), p. 340-349.
15. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the American Heart Association. Circulation 2002;105:539-42.
16. Mor-Avi V, Jenkins C, Kuhl HP, Nesser H-J, Marwick T, Franke A, et al. Real-time 3-dimensional echocardiographic quantification of left ventricular volumes: multicenter study for validation with magnetic resonance imaging and investigation of sources of error. JACC Cardiovasc Imaging 2008;1:413-23.
17. Tsang W, Kenny C, Adhya S, Kapetanakis S, Weinert L, Lang RM, et al. Interinstitutional measurements of left ventricular volumes, speckle-tracking strain, and dyssynchrony using three-dimensional echocardiography. J Am Soc Echocardiogr 2013;26:1253-7.
18. Marwick TH. Stress echocardiography. Heart 2003;89:113-8.
19. Hansegård J, Urheim S, Lunde K, Malm S, Rabben SI. Semi-automated quantification of left ventricular volumes and ejection fraction by real-time three-dimensional echocardiography. Cardiovasc Ultrasound 2009;7:18.
20. Tighe DA, Rosetti M, Vinch CS, Chandok D, Muldoon D, Wiggin B, et al. Influence of image quality on the accuracy of real time three-dimensional echocardiography to measure left ventricular volumes in unselected patients: a comparison with gated-SPECT imaging. Echocardiography 2007;24:1073-80.