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

Alexandros Papachristidis, Elena Galli, Marcel L. Geleijnse, Brecht Heyde, Martino Alessandrini, Daniel Barbosa, Michael Papitsas, Gianpiero Pagnano, Konstantinos C. Theodoropoulos, Spyridon Zidros, et al.

Standardized delineation of endocardial boundaries in 3D Left Ventricular Echocardiograms

Short title: Segmentation of 3D left ventricular echocardiograms

Authors:

Alexandros Papachristidis (1), Elena Galli (2), Marcel L. Geleijnse (3), Brecht Heyde (4), Martino Alessandrini (5), Daniel Barbosa (6), Michal Papitsas (7), Gianpiero Pagnano (8), Konstantinos C. Theodoropoulos (9), Spyridon Zidros (10), Erwan Donal (11), Mark J. Monaghan (12), Olivier Bernard (13), Jan D'hooge (14) and Johan G. Bosch (15).

1. Dr. Alexandros Papachristidis, MD, King's College Hospital, Denmark Hill, SE5 9RS, London, United Kingdom. (E-mail: alexandros.papachristidis@nhs.net)

2. Dr. Elena Galli, MD, PhD, Cardiology and CIC-IT1414 CHU Rennes and LTSI, INSERM 1099, University Rennes-1, rue Henri Le Guilloux 35033 Rennes Cedex, France. (E-mail: Elena.GALLI@chu-rennes.fr)

3. Dr. Marcel L. Geleijnse, MD, PhD, Cardiology, Erasmus MC, 's-Gravendijkwal 230 3015 CE Rotterdam, Netherlands. (E-mail: m.geleijnse@erasmusmc.nl)

4. Brecht Heyde, PhD, Cardiovascular Imaging and Dynamics, KU Leuven - University of Leuven, Oude Markt 13, 3000 Leuven, Belgium. (E-mail: brecht.jk.heyde@gmail.com)

5. Martino Alessandrini, PhD, Cardiovascular Imaging and Dynamics, KU Leuven - University of Leuven, Oude Markt 13, 3000 Leuven, Belgium. (E-mail: martino.alessandrini@gmail.com),

and present address: Department of Electric Engineering, University of Bologna, Via Zamboni, 33, 40126 Bologna, Italy.

6. Daniel Barbosa, PhD, Cardiovascular Imaging and Dynamics, KU Leuven - University of Leuven, Oude Markt 13, 3000 Leuven, Belgium. (E-mail: djcbarbosa@gmail.com)

7. Michael Papitsas, MD, King's College Hospital, Denmark Hill, SE5 9RS, London, United Kingdom. (E-mail: michael.papitsas@nhs.net)

8. Gianpiero Pagnano, MD, King's College Hospital, Denmark Hill, SE5 9RS, London, United Kingdom. (E-mail: g.pagnano@nhs.net)

9. Konstantinos C. Theodoropoulos, MD, MSc, King's College Hospital, Denmark Hill, SE5 9RS, London, United Kingdom. (E-mail: konstantinos.theodoropoulos@nhs.net)

10. Spyridon Zidros, MD, MSc, King's College Hospital, Denmark Hill, SE5 9RS, London, United Kingdom. (E-mail: spyridon.zidros@nhs.net)

11. Prof. Erwan Donal, MD, PhD, Cardiology and CIC-IT1414 CHU Rennes and LTSI, INSERM 1099, University Rennes-1, rue Henri Le Guilloux 35033 Rennes Cedex, France. (E-mail: Erwan.DONAL@chu-rennes.fr)

12. Prof. Mark J. Monaghan, PhD, King's College Hospital, Denmark Hill, SE5 9RS, London, United Kingdom. (E-mail: mark.monaghan@nhs.net)

13. Associate Prof. Olivier Bernard, PhD, University of Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, University of Lyon 1, Villeurbanne, France. (E-mail: olivier.bernard@creatis.insa-lyon.fr)

14. Prof. Jan D'hooge, MSc, PhD, Cardiovascular Imaging and Dynamics, KU Leuven - University of Leuven, Oude Markt 13, herestraat 49, 3000 Leuven, Belgium (E-mail: jan.dhooge@uzleuven.be)

15. Associate Prof. Johan G. Bosch, PhD, Cardiology/Biomedical Engineering, Erasmus MC, 's-Gravendijkwal 230 3015 CE Rotterdam, Netherlands. (E-mail: j.bosch@erasmusmc.nl)

Corresponding author:

Dr. Alexandros Papachristidis, Department of Cardiology, King's College Hospital NHS Foundation Trust, Denmark Hill, SE5 9RS, London, United Kingdom. E-mail: alexandros.papachristidis@nhs.net, Fax: +442032991745, Tel: +4407550013006

ABSTRACT.

Background

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

Methods

Forty-five 3D LV echocardiographic datasets were traced by three experienced operators, from different centers, according to pre-defined guidelines. The 3D meshes were compared to each other and the endocardial areas of discrepancies were identified. A discussion and retracing protocol was used to reduce discrepancies. For each dataset an average 3D mesh was produced (reference mesh). Subsequently, 4 novice operators, divided in 2 groups, traced 20 of the datasets. Two operators followed the tracing protocol and two did not.

Results

The intra-class correlation coefficients between the 3 experienced operators for end-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF) were 0.952, 0.955, and 0.932. The absolute distances between tracings were 1.11±0.45mm. The highest tracing discrepancies were at the apical cap, anterior and anterolateral walls in end-diastole and end-systole, and also at the basal anteroseptum in end-systole. The agreement to the reference meshes was better for the novice operators who followed the guidance (10.9±17.3ml, 10.2±14.7ml and -2.2±4.1% for EDV, ESV and EF), compared to those who did not (16.3±16.4ml, 17.0±16.0ml, -4.2±4.1% respectively).

Conclusion

Comparing 3D LV tracings, we identified the endocardial areas which are the most difficult to delineate. Our suggested protocol for LV tracing resulted in very good agreement between operators. The reference 3D meshes are available for on-line testing and ranking of LV tracing algorithms.

Key words: 3D echocardiography; left ventricular segmentation; endocardial tracing

Abbreviations

3D: Three-dimensional

ED: End-diastole

ES: End-systole

ICC: Intraclass Correlation Coefficient

LV: Left Ventricle

LVOT: Left Ventricular Outflow Tract

MRI: Magnetic Resonance Imaging

1. Introduction

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 geometrical 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, 6, 7]. Up to now there are no clear standards or guidelines available for 3D LV endocardial border tracing and there is no direct comparison of the actual tracings between different operators.

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

In this study, we aimed to address these issues by suggesting a protocol for LV endocardial tracing in 3D echocardiographic datasets and creating a series of clinically realistic datasets with well-established reference tracings based on manual tracings from three expert echocardiography centers. Based on this standard set, a competition for automated tracing methods was organized, associated with the Medical Image Computing and Computed Assisted Interventions (MICCAI) 2014 symposium and 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 on-line for continuous testing and ranking of fully automated or semi-automated algorithms.

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

2. Methods

2.1 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 non-ischaemic dilated cardiomyopathy. The patients were recruited in three different Institutions (Rennes University Hospital, France; University Hospital Leuven, Belgium; and Thoraxcenter Erasmus MC Rotterdam, Netherlands). Fifteen patients undergoing echocardiography and meeting the inclusion criteria were recruited in each one of the three Institutions. Exclusion criteria were left bundle branch block, visually dyssynchronous LV or unacceptable image quality. Unacceptable image quality was defined as a) significant stitching or other type of artefact affecting the tracking of endocardium or b) poor visualization of LV wall or wall out of the image sector to an extent that the image could no longer be manually analysed with good confidence in multiple segments. The image quality of the accepted datasets was graded as good, fair or poor (Figure 1). Good quality was defined when the endocardium was visible in ED and ES instances in all 17 segments throughout the cardiac cycle. Fair quality if the endocardium was not clearly seen in 1-2 segments and poor quality when the endocardial border was not clearly distinguished in ES or ED frames in more than 2 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' datasets.

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 Healthcare, Andover, MA; and SC2000 with a 4Z1c probe, Siemens, Erlangen, Germany). Machine settings were optimized to achieve the maximum quality of images while keeping

volume rate above 16Hz. The frames per cardiac cycle were 25.7±8.5 (mean ± SD). Acquired data were fully anonymized and handled within the regulations set by the local ethical committees of each hospital.

2.2 Endocardial tracing Procedure

End-diastolic (ED) and end-systolic (ES) frames were identified. Nine standard anatomical planes were defined: four longitudinal planes through the long axis under 45-degree angles and five transversal (short-axis) planes divided equally along the long axis. For the tracings, a custom non-commercial tracing package for 3D echocardiograms (Speqle3D) was used, developed by the University of Leuven [10]. A single experienced operator from each center (AP, MG, EG) with an experience of more than three hundred 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 pre-defined slices, and in the allocated ED/ES frames. All 45 datasets 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 points were to:

a) 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” of the papillary muscle to meet the endocardium at the other edge of the muscle (i.e. from the basal to apical segment or other way round). Also we suggested tracing at the level of the trough of endocardial creases in order to include trabeculations in the LV cavity.

b) Keep tissue consistency between ED and ES and between adjacent/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 regards to elevation plane artefacts. Also, the utilized 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.

c) In long axis views, draw up to the mitral valve annulus on the inside of the bright ridge up to the point where the valve leaflet is hinging. The mitral valve annulus is sometimes quite difficult to be traced 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 based on a single frame (Figure 4).

d) Partly exclude the left ventricular outflow tract (LVOT) from the cavity by drawing from the septal MV 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 LVOT and provides a smooth shape of the basal anteroseptal wall segment in order to keep the LV shape symmetrical and also avoid giving the impression of a dyskinetic segment as the LVOT expands during systole.

e) Draw apex high up near the epicardium both in ED and ES taking into consideration that there should be little displacement of the true apex point. As is known from anatomy, the left ventricular wall is actually thin at the true apex and does hardly move 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 to the epicardium at the apical region. Therefore, tracing close to epicardium can provide better anatomical ED/ES consistency than the less “visible” endocardium. This is expected to have little effect on volumes but makes the contouring more consistent.

2.4 Evaluation of correspondence

After all three experts had submitted their tracings, 3D shapes (Figure 7) were generated from the nine 2D contours in ED and ES by interpolation [10, 13]. The shapes were represented by a 3D mesh with a resolution of about 40-by-80 points (longitudinal x circumferential). Each 3D mesh consisted of about 3200 points (vertices). These vertices from the 3 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 ejection fraction (EF) 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 three-dimensional 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 to each other. To verify consensus, the following criteria were used: Hausdorff distances ≤ 5mm between each individual mesh and the reference mesh, percentage (relative) difference in LV volumes ≤ 10% between all three operators in pairs and absolute difference in LVEF ≤ 5 percentage-points between 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 between the operators. This was done by superimposing the tracings of the three operators for each one of the 9 pre-defined planes at both ED and ES (Figure 5). For the

purposes of our workshop [9], one or more of the operators would then revise their tracings in order 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 ≤ 7mm were accepted. The final averaged 3D mesh for each case, after revision process, was considered as the "reference mesh".

2.5 The tracing guidance applied to novice operators utilizing a commercially available semi-automated 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 v.3.5, Tomtec Imaging Systems GmbH, Unterschleissheim, Germany). Four operators with little experience in 3D echocardiography LV tracing were asked to trace 20 of our study datasets, which were randomly selected from the GE and Siemens datasets. The Philips datasets could not be imported to the commercially available software for analysis, due to technical limitations related to acquisition and storage process.

All four operators had an experience of 20 – 30 cases and they were from a single center (King's College Hospital, London, UK). They were divided in two groups. Operators A1 and A2 (Group A) traced based on their own discretion. The 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 3-chamber view and then automatically defines the 2-chamber and 4-chamber views (60^o and 120^o incremental views). The operator

is aligning the longitudinal axis of the LV in all three apical views in a way that the axis is crossing the LV apex and the middle of the mitral annulus 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 annulus level in end-diastole. Subsequently, the application is automatically defining the endocardial border in all 3 apical views in end-diastole. The operator can adjust the tracings manually as necessary in all views. Then, the software is tracking 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 (2, 3 and 4-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 others than the ED and ES frames. The application is continuously calculating the EDV, ESV and EF as the operator is manipulating the tracings. For the purposes of our study, once the final tracings were confirmed, the derived volumes and EF were recorded by the operator in a separate spreadsheet.

2.6 Statistical analysis

The statistical analysis was performed with IBM SPSS version 20.0.0 (IBM Corporation Software Group, Somers, NY). Difference in mean values for continuous variables was tested with a student's t-test, Mann-Whitney U test for unpaired samples or Wilcoxon signed rank test for paired samples as appropriate. Interobserver variability was tested using the intra-class correlation coefficient (ICC) for absolute agreement in a two-way mixed model. The values for continuous variables are presented as mean ± standard deviation. Statistical significance was considered for a two-tailed p-value < 0.05

3. Results

3.1 Baseline demographics and characteristics of datasets

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

3.2 Interobserver variability

The ICC for the derived clinical parameters (EDV, ESV and EF) was very high (>0.9) for the initial tracings (Table 1). This shows excellent agreement between operators. The average differences and the percentage differences between operators are shown in Table 2. The average difference for initial tracings for EDV was -3.3 ± 27.0 ml, for ESV it was -2.7 ± 23.9ml and for EF 0.5 ± 4.9 %-points. The percentage difference for EDV, ESV and EF was 10.4 ± 7.9%, 12.9 ± 10.6% and 10.4 ± 9.1%. The mean values of mean absolute distances (MAD) and Hausdorff distances (HD) are presented in Table 3.

Applying the pre-defined criteria for consensus, agreement in initial tracings was reached in 12 cases (26.7%) for ED volume, 9 cases (20.0%) for ES volume, 23 cases (51.1%) for EF and 33 cases (73.3%) for Hausdorff distances in ES and 34 (75.6%) in ED.

3.3 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 bulls-eye plot according to American Heart Association's LV segmentation guidance [15]. Four different bulls-eye plots were calculated corresponding to ES and ED frames before and after the consensus process (Figure 8).

From these bulls-eye plots of average distance we assessed the endocardial areas that show the highest distances in ED and ES frames. In the initial contours (Figure 8) the highest distances in ED were observed at the apical cap (segment 17) and also at the anterior, the mid anterolateral and apical lateral segments (1, 7, 13 and 12, 16 respectively). The best agreement was demonstrated in the infero-septum (segments 3 and 9), the apical septum (segment 14) and mid inferior wall (segment 10). In ES frame 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 apical lateral (segment 16). Additionally, a high distance error also appeared at basal anteroseptum (segment 2).

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

3.4 The tracing protocol with a commercially available semi-automated software and novice operators

3.4.a Agreement between novice operators and reference meshes.

The variation in image quality of the 20 datasets which were used in this sub-study was similar to that of the whole cohort; good/fair/poor image quality: 30%/40%/30%. The derived LV volumes from all four novice operators were lower compared to the reference meshes (Table 4). On the contrary the EF was higher (Table 4). The difference between each operator and the reference meshes was statistically significant ($p < 0.05$) for EDV, ESV and EF for all 4 operators apart from the EF by operator B1 ($p = 0.067$). The average difference for EDV, ESV and EF between Group A measurements and the reference meshes was 16.3 ±

16.4ml, 17.0 ± 16.0ml and -4.2 ± 4.1%. The relevant values for Group B were 10.9 ± 17.3ml, 10.2 ± 14.7ml and -2.2 ± 4.1 %.

3.4.b Inter-observer agreement between novice operators.

The averaged EDV and ESV of Group A operators was lower compared to Group B (136.8 ± 60.3ml vs 142.4 ± 60.0ml, p=0.097 and 80.4 ± 47.8ml vs 87.4 ± 53.0ml, p=0.014), whereas the EF was higher (44.3 ± 13.3% vs 42.2 ± 13.4%, p=0.025). The average difference (bias) for EDV, ESV and EF between operators A1 and A2 was 2.1 ± 11.2ml, 3.3 ± 9.0ml and -2.1 ± 3.1 %. In Group B (operator B1 vs B2) the relevant values were 1.1 ± 10.9ml, 0.7 ± 9.2ml and -0.2 ± 3.7%. The difference in EDV, ESV and EF between the two operators in each group was not statistically significant with the exception of EF in Group A (p<0.001). The ICC between operators in Group A for EDV, ESV and EF was 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 ICC values 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).

4. Discussion

In this study: a) We suggest a guidance for LV endocardial tracing in 3D echocardiographic datasets that results in good agreement between experienced operators from different centers. Good agreement between 3D echocardiography and Magnetic Resonance Imaging (MRI) has been previously reported, but currently there is no standardized guidance for LV tracing in 3D echocardiographic datasets. b) We identified the endocardial areas which show the most significant distance error in 3D LV manual tracings of experienced operators and we provide additional recommendation to improve agreement. c) The provided tracing protocol can be useful in a clinical setting and can improve accuracy of novice

operators' tracings, using a vendor-independent semi-automated commercially available software.

The ICC for EDV, ESV and EF between the three experienced operators was 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] where semi-automated methods were used. In our study the operators were asked to trace manually 18 planes (9 planes in ED and 9 planes in ES), whereas in commercially available software the operator's input is usually required in only 3 to 6 planes. The challenge of fully manual tracing is also depicted in the results of our sub-study. The semi-automated software overcomes the inexperience of the operators providing better inter-observer agreement compared to a fully manual protocol. Given this advantage of a semi-automated software, one may wonder why we elected the laborious and less reproducible fully manual workflow to create the reference meshes. This was done because we aimed to provide reference 3D meshes to test LV tracing algorithms. Using a semi-automated algorithm to create the reference tracings would bias the measurements towards this particular algorithm. Therefore, a tested algorithm of similar infrastructure would be expected to perform better compared to other algorithms. What we wanted was truly unbiased reference meshes and a fair comparison between algorithms, hence the fully manual workflow.

For the same reason, the reference 3D meshes were created by averaging the manual tracings of 3 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 inter-observer variability in 3D LV measurements and also it is not possible for any operator to reproduce exactly the same tracing of a specific 3D dataset, especially in a fully manually workflow like ours.

Therefore, we believe that averaging the variability of tracings, especially when they come from more than one experienced operators, 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 3 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 used protocol with the discussion-revision process resulted in excellent agreement between experienced operators' tracings, as demonstrated by both clinical (LV volumes and EF) and anatomical criteria (distance errors). Therefore, the produced reference tracings are expected to be "accurate" in terms of LV volumes and EF (thought this has not been tested versus another modality, i.e. MRI, 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 each other's tracings and also to derived clinical values (LV volumes and EF), when they initially traced the datasets. In all commercially available 3D LV software the operator can review the derived ED and ES volumes as well as the EF. It is not uncommon in clinical practice, especially in datasets of poor image quality, for the operator to adjust the endocardial tracing based on his visual estimation of volumes and EF, posing significant bias on endocardial tracing. In our study this bias was completely eliminated as the used prototype software did not allow calculation of the LV volumes and EF. Therefore, the tracing was made based only on anatomical evaluation of LV endocardium and structures. This does not apply to our sub-study, where the commercially available software allowed visualization of the LV volumes and EF.

This is a multi-center 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] report somehow lower variability, but they evaluated the

interobserver variability in their own center. Tsang et al. [17] have studied the inter-institutional measurements of LV volumes between operators from two different centers and they additionally assessed the effect of a short common training period. They report an ICC for EDV, ESV and EF as 0.75, 0.69 and 0.79 respectively utilizing a semi-automated software. These values are considerably lower compared to our study. After the intervention of common training, they report improvement in ICC's which are comparable to the values in our study. The mean percentage difference by Tsang et al. [17] was 13.6%, 15.9%, and 12.2% for EDV, ESV, and EF. In our study the percentage difference was 10.4%, 12.9%, and 10.4% respectively. Mor-Avi et al. [16] report percentage difference of 8±8% for EDV and 13±14% for ESV in a study held in 4 different institutions.

Nikitin et al. [7] included patients with good acoustic window only. Soliman et al. [6] report that they excluded patients if more than two LV segments were not well visualized. They estimate that their cohort represents 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 2 segments not well visualized (poor quality images, Figure 1). Therefore, there is significant difference in image quality comparing these studies. We elected to include those patients as they represent a significant proportion of the daily echocardiography practice. Despite that difference, Soliman et al. [6] report an average percentage difference for EDV of 8.2 ± 11.4% for a multi-plane 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=0.021, Table 2).

An interesting finding in our study is that despite excellent agreement in quantification of the left ventricle based on ICC, the percentage difference in LV volumes between operators was more than 10% in 73.3% of the cases for EDV and in 80.0% of the cases

for ESV. Also, the absolute difference in EF was greater than 5 percentage points in 48.9% of the cases. Mor-Avi et al. [16] also mention that in their study the variability levels in individual patients far exceeded the acceptable 10-15% levels for LV volumes, though they do 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 based on their individual training. Though the motion of endocardium from ED to ES 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 EF.

By analysing the bulls-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 ED, and the same segments plus the basal anteroseptum in ES. These findings may be explained by the fact that the apex is always difficult to visualize and identify because of near-field artefacts and the presence of trabeculations. The anterolateral papillary muscle may cause some confusion with regards 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 to 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 is causing a drop out more frequently, makes the anterior and anterolateral walls more difficult to trace accurately. In ES 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 ED. The absence of a papillary muscle in the infero-septum (segments 3, 9, 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 easier visualized compared to the anterior wall. In ES the best agreement in tracings was seen again in the infero-septal 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 LVOT makes the tracing of segment 2 (basal anteroseptum) quite challenging. In this setting one might suggest that the tracing of LVOT should follow the level of the aortic valve annulus from the MV hinge point, to the aortic valve cusps hinge points and then to the septum. This might result in better agreement between 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 which detects LV sub-volume changes to assess LV dyssynchrony.

We have paid significant attention to these anatomical pitfalls while establishing the tracing protocol and we managed to achieve individual tracings within very strict limits of agreement. However, in order to address this issue, we came up with additional recommendations for these particular areas and we propose that the operators should 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 to utilize the short axis plane to ensure that a smooth, nearly circular contour, is created. This may be particularly

useful in end-diastolic frame. In end-systolic 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 based on 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 symmetrical shape of the tracings at the basal parts of the LV in the 3-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 end-systolic frame such a correlation cannot be recommended.

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

We evaluated the usefulness of the proposed guidelines and the aforementioned additional recommendations in operators with little experience in 3D echocardiography using a semi-automated vendor-independent commercially available software. The group that followed the tracing protocol had better agreement to the reference meshes. Additionally, the bias between operators who followed the tracing protocol was lower compared to 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 pre-defined guidelines. The absence of significant difference between the two groups can be explained by the fact that the 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's interaction is related to higher inter-observer variability.

Limitations

The current results are obtained on image data of acceptable image quality. However, in everyday clinical practice cardiologists and echocardiographers face the challenge of sub-optimal image quality. As it has been demonstrated previously, the 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 pre-specified 2D planes derived from the 3D dataset. In most of 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 sub-study where a semi-automated software was used and the operators were required to identify the 3-chamber apical view and align the long axis of the LV.

Due to technical limitations, we could not test the distance errors of endocardial tracings in our sub-study, where 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) is using the time-volume curve to calculate the EDV and ESV, whereas in the manual workflow used by the experienced operators the ED and ES frames were selected manually before the datasets were provided to the operators for tracing. This may be a confounder in the volume and EF differences noted between the experienced and novice operators. However, this is a limitation that we had to accept in our attempt to test the proposed tracing guidance with a commercially available software (TomTec), using as reference meshes the ones produced with a custom software (Speqle 3D).

5. 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 EF) was very high. We identified that the apical cap, the anterior and antero-lateral walls, as well as the basal anteroseptum are correlated with the highest distance errors between operators. The used 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 proven useful to less experienced operators to achieve better agreement to the reference tracings using a semi-automated commercially available software, when compared to operators of similar experience who traced based on their own discretion. Our reference tracings have been used to validate algorithms for LV automatic quantification [9] and the datasets are available on-line for continuous evaluation of LV tracing algorithms, fostering innovation in algorithmic development for new automated tools.

Acknowledgments

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

Alexandros Papachristidis performed the manual tracings of the 3D datasets, performed the statistical analysis, and wrote the manuscript.

Marcel L. Geleijnse performed the manual tracings of the 3D datasets and revised the manuscript.

Elena Galli performed the manual tracings of the 3D datasets and revised the manuscript.

Brecht Heyde provided expert opinion on data processing and workshop setup, processed the submitted 3D datasets to generate the predefined planes for tracing, modified the Speqle 3D software to meet participants' needs, and revised the manuscript.

Martino Alessandrini provided expert opinion on data processing and workshop setup, modified the Speqle 3D software and revised the manuscript.

Daniel Barbosa provided expert opinion on data processing and workshop setup and revised the manuscript.

Michal Papitsas, performed the tracings with the commercially available software and revised the manuscript.

Gianpiero Pagnano, performed the tracings with the commercially available software and revised the manuscript.

Konstantinos C. Theodoropoulos performed the tracings with the commercially available software and revised the manuscript.

Spyridon Zidros, performed the tracings with the commercially available software and revised the manuscript

Erwan Donal and Mark J. Monaghan provided expert opinion on tracing conventions, facilitated the project setup and revised the tracings and the manuscript.

Jan D'hooge and Prof. Olivier Bernard supervised the whole project, provided expert opinion, and revised the manuscript.

Johan G. Bosch supervised the project, provided expert opinion, and wrote the manuscript.

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 Echocardiography 2012;25(1):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(10):2053–69.
3. Monaghan MJ. Role of real time 3D echocardiography in evaluating the left ventricle. Heart 2006;92(1):131-136.
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(1):114-117.
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(4):878-86.
6. Soliman OII, 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(9):1042-1049.
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(5):365-372.
8. Leung KYE, Bosch JG. Automated border detection in three-dimensional echocardiography: principles and promises. Eur J Echocardiography 2010;11(2):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(4):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 JWB, Beck G, Vecht RJ. Left ventricular apical thin point. Br Heart J. 1977;39:806-809.

12. Johnson KM, Johnson HE, Dowe DA. Left Ventricular Apical Thinning as Normal Anatomy. J Comput Assist Tomogr. 2009;33(3):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, 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. Cardiovascular imaging. Jul 2008;1(4):413-423.

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 Nov;26(11):1253-7

18. Marwick TH. Stress echocardiography. Heart. 2003;89:113–118.

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(10):1073-80.

Table 1. EDV, ESV and EF based on the references meshes, which were computed from initial tracings, and the relevant intra-class correlation coefficients (ICC) between the three experienced operators.

	Reference mesh	**ICC**
EDV (ml)	173.5 ± 83.5	0.952 (0.918-0.972)
ESV (ml)	114.5 ± 78.5	0.955 (0.927-0.974)
EF (%)	38.5 ± 13.6	0.932 (0.861-0.959)

EDV: End-Diastolic Volume, ESV: End-Systolic Volume, EF: Ejection Fraction

Table 2. Average differences (mean ± SD) between the three experienced operators for the initial and revised (after consensus) tracings.

		ED Volume (ml)		ES Volume (ml)		EF (%)	
		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/fair and poor quality image datasets for all variables. * indicates p value < 0.05

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

		HD in ED (mm)	HD in ES (mm)	MAD in ED (mm)	MAD in ES (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, ED: End-diastole, ES: End-systole

Table 4. EDV, ESV and EF of four operators (A1, A2, B1, and B2) using a commercially available software

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

EDV: End-Diastolic Volume, ESV: End-Systolic Volume, EF: Ejection Fraction

Figures Legends.

Figure 1. Examples of variability in image quality of datasets.

Figure 2. Example of tracing including trabeculae and papillary muscles in the LV cavity.

Figure 3. Example of manual endocardial drawing in Speqle3D software.

Figure 4. Example of manual tracing pointing out the mitral valve hinge points.

Figure 5. Superimposition of the three experienced operators' manual tracings in a long axis view.

Figure 6. Example of tracing of the apex which is close to epicardium with little displacement in end-systole.

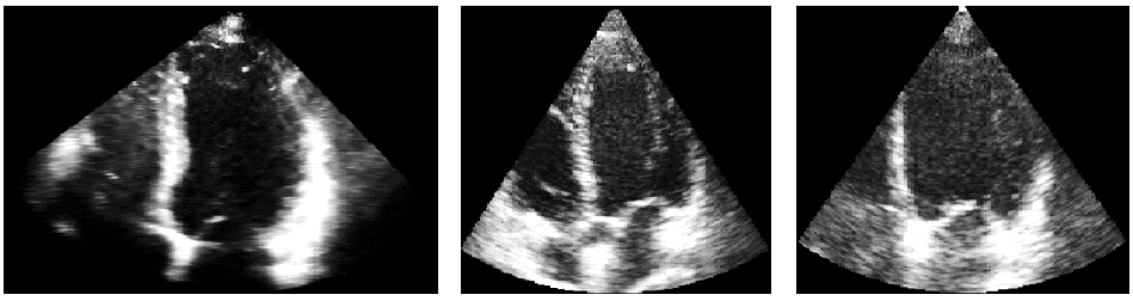
Figure 7. Upper Panel: Three-dimensional meshes were generated for each individual tracing in end-diastolic (left panel) and end-systolic (right panel) instances.

Lower Panel: An example of a 3D mesh derived by one experienced operator's tracings with the endocardial distances from reference mesh illustrated in color code.

Figure 8. Upper Panel: 17-segment bulls-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).

Lower panel: The same bulls-eye plot as above after the consensus process.

Figure 1. Examples of variability in image quality of datasets.



(a) good (b) fair (c) poor

Figure 2. Example of tracing including trabeculae and papillary muscles in the LV cavity

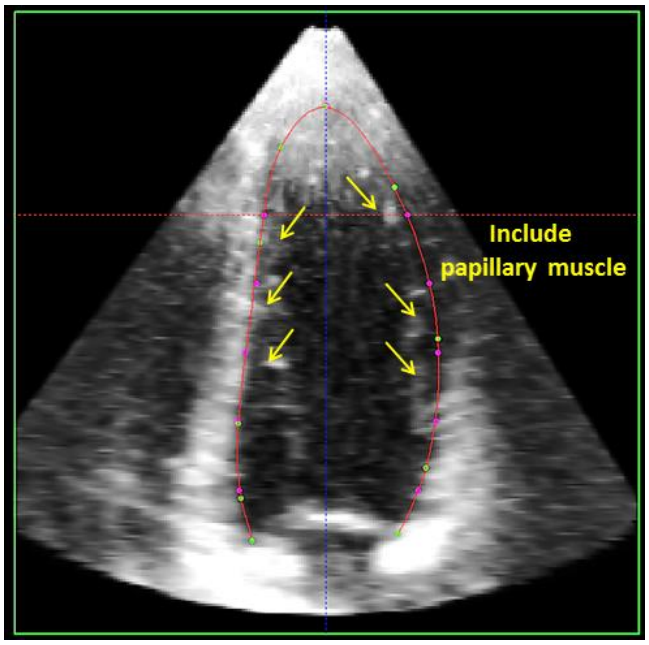
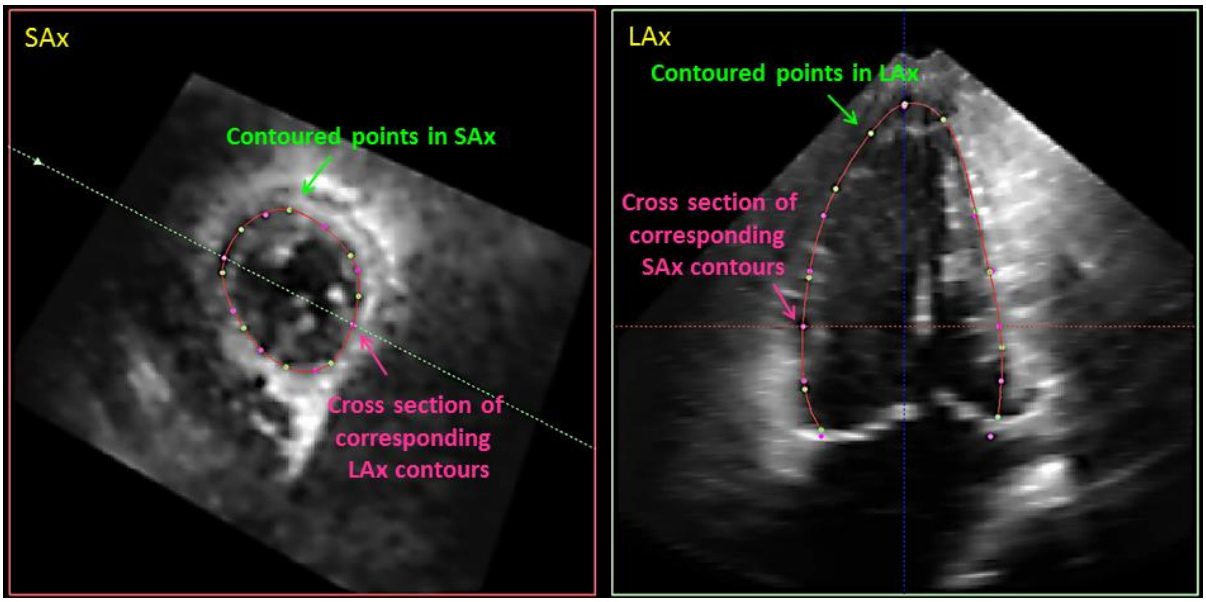


Figure 3. Example of manual endocardial drawing in Speqle3D software



Left: Transverse plane (short axis). Right: Longitudinal plane (long axis). The green dots in both images represent the points set on endocardium 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 section points of the contours in the orthogonal planes.

Figure 4. Example of manual tracing pointing out the mitral valve hinge points.

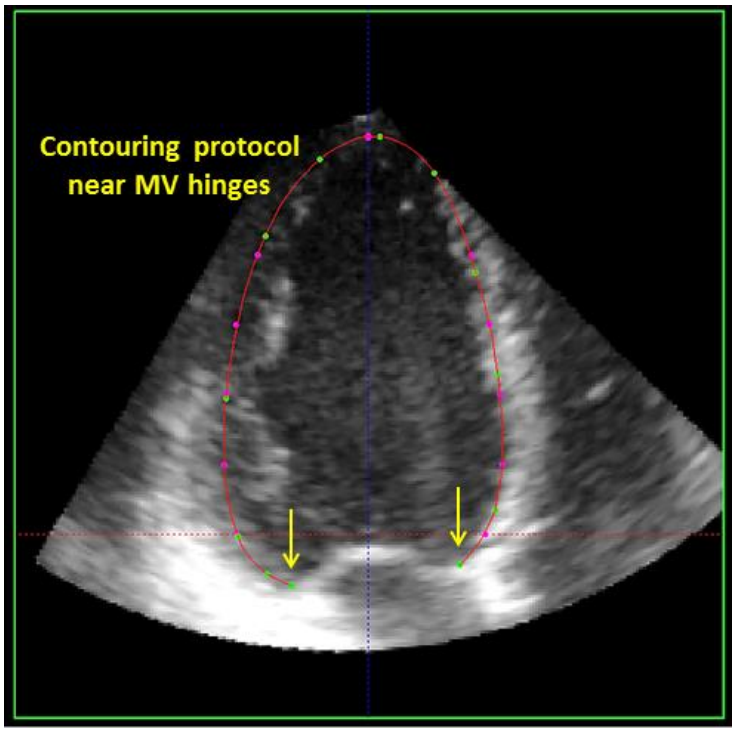
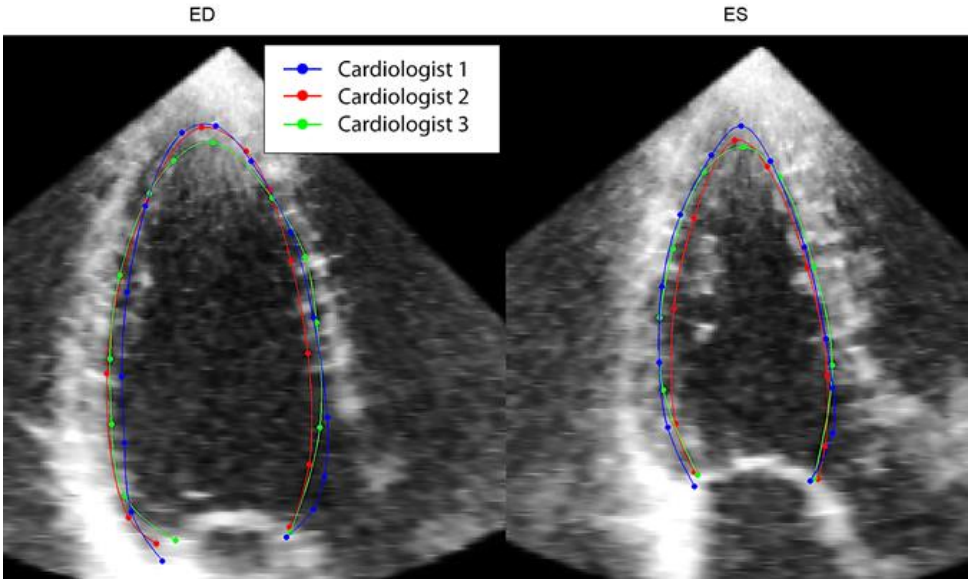


Figure 5. Superimposition of the three experienced operators' manual tracings in a long axis view.



Left panel: end-diastole. Right panel: 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 which is close to epicardium with little displacement in end-systole.

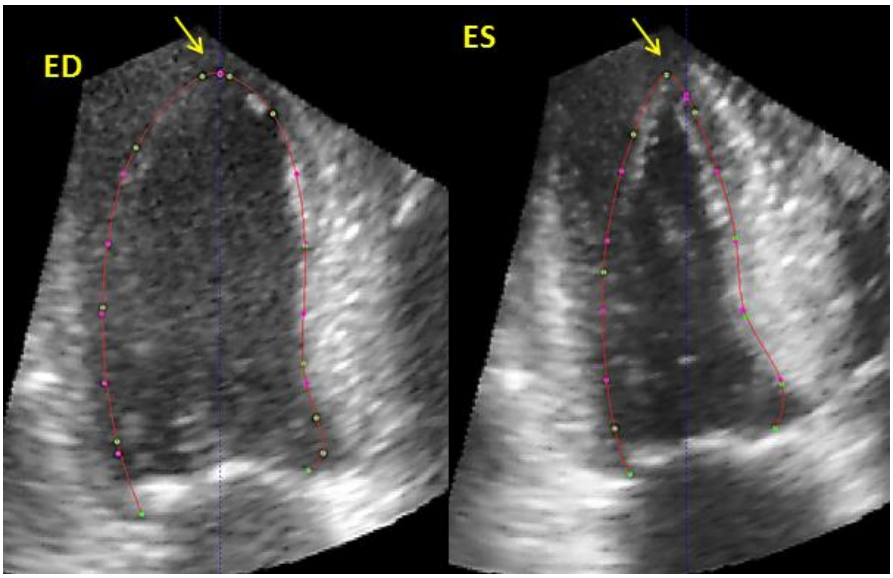
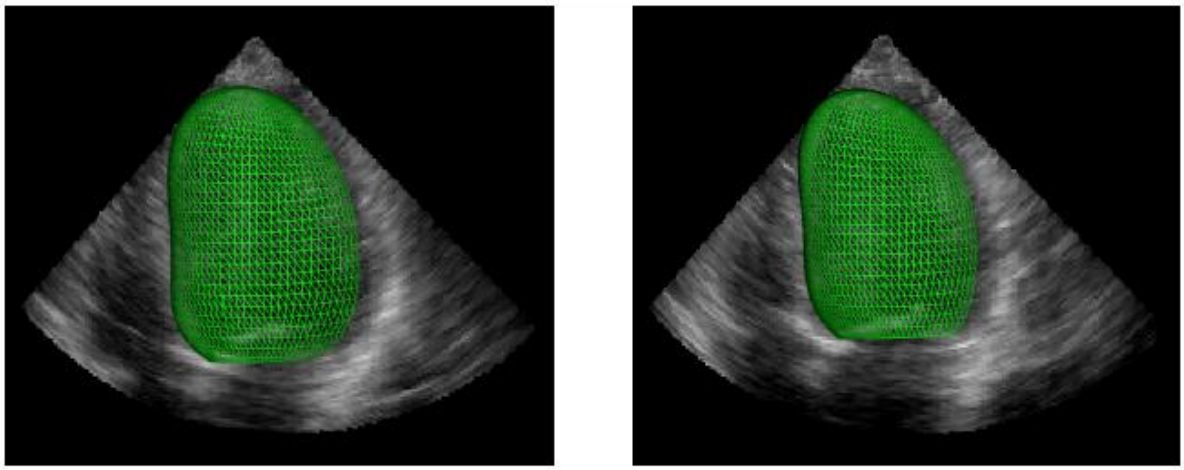
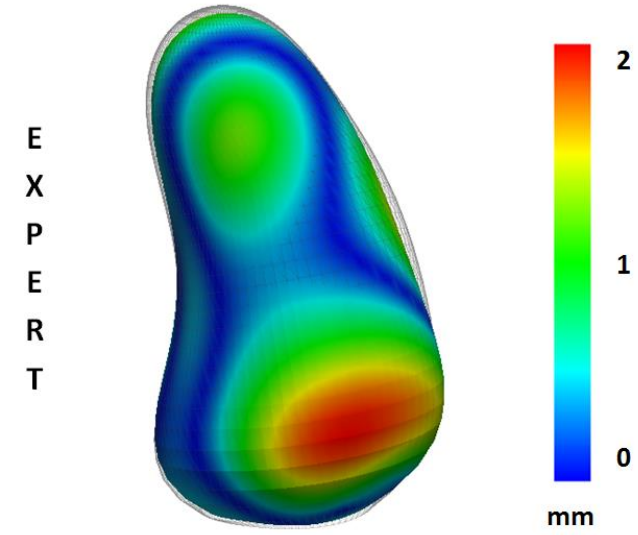


Figure 7. Upper Panel: Three-dimensional meshes were generated for each individual tracing in end-diastolic (left panel) and end-systolic (right panel) instances.

Lower Panel: An example of a 3D mesh derived by one experienced operator's tracings with the endocardial distances from reference mesh illustrated in color code.





d_m = 0,7 mm , d_H = 2,2 mm

Figure 8. Upper Panel: 17-segment bulls-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).

Lower panel: The same bulls-eye plot as above after the consensus process.