Quantification of Left Ventricular Indices From SSFP Cine Imaging: Impact of Real-World Variability in Analysis Methodology and Utility of Geometric Modeling

Christopher A. Miller, MBChB, Peter Jordan, PhD, Alex Borg, MD, Rachel Argyle, MBChB, David Clark, BSc, Keith Pearce, BSc, and Matthias Schmitt, PhD

Purpose: To assess the impact of “real-world” practice variation in the process of quantifying left ventricular (LV) mass, volume indices, and ejection fraction (EF) from steady-state free precession cardiovascular magnetic resonance (CMR) images. The utility of LV geometric modeling techniques was also assessed.

Materials and Methods: The effect of short-axis- versus long-axis-derived LV base identification, simplified versus detailed endocardial contouring, and visual versus automated identification of end-systole were evaluated using CMR images from 50 consecutive, prospectively recruited patients. Additionally, the performance of six geometric models was assessed. Repeated measurements were performed on 25 scans (50%) in order to assess observer variability.

Results: Simplified endocardial contouring significantly overestimated volumes and underestimated EF (−6.6%, P < 0.0005) compared to detailed contouring. A mean difference of −34g (P < 0.0005) was observed between mass measurements made using short-axis- versus long-axis-derived LV base positioning. A technique involving long-axis LV base identification, signal threshold-based detailed endocardial contouring, and automated identification of end-systole had significantly higher observer agreement. Geometric models showed poor agreement with conventional analysis and high variability.

Conclusion: Real-world variability in CMR image analysis leads to significant differences in LV mass, volume and EF measurements, and observer variability. Appropriate reference ranges should be applied. Use of geometric models should be discouraged.

Key Words: cardiovascular magnetic resonance imaging; left ventricle; analysis methodology; geometric modeling

LEFT VENTRICULAR (LV) mass, volume indices, and ejection fraction (EF) remain some of the most evidenced-based indicators of prognosis (1–5). Both their absolute measurement and temporal change are used to guide pharmacotherapy, device implantation, and surgical intervention. Accurate and reproducible assessment of these parameters is an important strength of cardiovascular magnetic resonance imaging (CMR).

The main components in the process of quantifying LV indices from CMR images are: LV base identification, myocardial border tracing, and identification of end-diastole and systole. While CMR is regarded as the gold standard technique for assessment of LV indices, in “real-world” CMR practice considerable variability exists in the methodology used for each of these components.

First, rather than incorporating long-axis images into the analysis, many centers use only short-axis images for analysis, despite it often being difficult to determine the basal LV slice, and how much of it to include, from short-axis images alone. Arbitrary criteria are widely used, such as inclusion of a slice within the LV when at least 50% or 75% of the cavity is surrounded by ventricular myocardium (Fig. 1) (6–8). Alternatively, if a slice is thought to include both ventricular and atrial myocardium, others advocate tracing up to the apparent junction of atrium and ventricle before joining up the contours with a straight...
The aim of this study was to assess the impact of these real-world methodological differences in CMR image analysis on LV mass, volume, and EF measurements in an unselected cohort of patients representative of clinical practice. Specifically, the effect of short-axis- versus long-axis-derived LV base identification, simplified versus detailed endocardial contouring, and manual versus automated identification of end-systole were evaluated. The effect on inter- and intraobserver variability was also assessed. In addition, the utility of LV geometric modeling techniques, advocated for analyzing CMR images and widely used by other imaging modalities (8,12–16), was assessed.

**MATERIALS AND METHODS**

**Patients**
 Fifty consecutive consenting patients undergoing clinically indicated CMR scanning at a single institution were enrolled. Patients were recruited prospectively, prior to undergoing imaging. The only exclusion criterion was known complex congenital heart disease. In order to reflect real-world clinical practice, patients were not excluded on the basis of image quality, arrhythmia, or breath-holding ability.

The study was conducted according to the Helsinki Declaration. An Ethics Committee of the UK National Research Ethics Service gave ethical approval for the study (reference number 08/H1004/153) and written informed consent was obtained from all participants before entering the study.

**Image Acquisition**
 All patients underwent CMR imaging using a 1.5 T scanner (Avanto; Siemens Medical Imaging, Erlangen, Germany) with a 32-element phased-array coil. Steady-state free precession (SSFP) end-expiratory breath-hold cines were acquired in three long-axis planes (horizontal long-axis, vertical long-axis, and three-chamber long-axis). Using the horizontal and vertical long axis images to ensure perpendicularity to the ventricular septum, short-axis cines were then acquired from the atroventricular ring to the apex (SA stack). Typical parameters included repetition time 2.9 msec, echo time 1.2 msec, flip angle 80°, matrix 256 × 208, in-plane pixel size 1.4 × 1.4 mm, slice thickness 8 mm (interslice gap 2 mm for the SA stack), temporal resolution 30–50 msec depending on heart rate (25 acquired phases). The number of short-axis cines acquired was patient-specific, depending on ventricular size (mean 11). Retrospective gating was used in 48 patients. Prospective gating was used in two patients due to arrhythmia.

**Image Analysis**
 LV mass and volume indices were quantified for each scan using the three methods described below. Analysis was performed by a Level 2 accredited operator with over 3 years experience of CMR and of each analysis method.
Method 1

The epicardial border was manually traced at end-diastole in successive short-axis slices (Argus Syngo MR software, Siemens Medical Imaging). The “compacted” endocardial border was manually traced on the short-axis slices at end-diastole and end-systole such that papillary muscles and trabeculae were included in mass and excluded from volumetric measurements (Fig. 2A). At the base of the heart, slices were considered to be within the LV if 50% or more of the cavity was surrounded by ventricular myocardium. If the basal slice contained both ventricular and atrial myocardium, contours were drawn up to the junction and joined by a curved line through the blood pool (Fig. 1A,B), drawn appropriate to the part of the slice that appeared ventricular after observing the through-plane motion of the LV on the short-axis cine images. The end-diastolic and end-systolic frames were identified as the frames in which the LV cavity visually appeared largest and smallest, respectively. In cases where there was a discrepancy between slices, the frames in which the more basal slices had the largest cavity (for end-diastole) and smallest cavity (for end-systole) were used, as it was felt these would more closely represent the largest and smallest cavity volumes, respectively.

Method 2

Analysis was identical to Method 1 except that manual detailed tracing of the endocardium was performed such that papillary muscles and trabeculae were included in mass and excluded from volumetric measurements (Fig. 2B).

Method 3

The epicardial border was manually traced at end-diastole in successive short-axis slices (CMRtools, Cardiovascular Imaging Solutions, UK). A second contour was placed within the myocardium on the short-axis slices at end-diastole and systole allowing the signal intensity between the two contours (i.e., the signal intensity of myocardium) to be automatically determined. The endocardial border was then determined using a signal intensity-based “thresholding” tool, which delineated individual voxels as blood or myocardium and as such papillary muscles and trabeculae were included in mass and excluded from volumetric measurements (Fig. 2C). The thresholding was adjusted manually until appearances correlated with visual assessment. The LV base was identified using the long-axis images; the mitral valve position was identified visually at end-diastole and systole on the long-axis images, which allowed the valve plane to be tracked through the cardiac cycle. This was then integrated into the SA stack analysis. As a result, short-axis data basal to the mitral valve plane were excluded from mass and volume calculations (if the valve plane cut through a short axis slice, the part of the slice that was basal to the valve plane was excluded; Fig. 1C–F). Finally, end-systole was automatically determined as being the frame with the smallest cavity volume by calculating cavity volume at each frame.

For all methods, end-diastolic (EDV) and end-systolic (ESV) volumes were calculated as the sum of the endocardial areas in each short-axis slice at end-diastole and end-systole, respectively, multiplied by the interslice distance (i.e., slice thickness plus interslice gap). Stroke volume (SV) and EF were derived from EDV and ESV. End-diastolic myocardial mass was determined as the sum of the areas between the endocardial and epicardial borders in each short-axis slice at end-diastole, multiplied by the interslice distance to give the myocardial volume, which was then multiplied by 1.05 g/cm$^3$, the specific density of myocardium, to give myocardial mass. Short axis slices were numbered with slice 1 being the most basal, and the most basal slice included in the analysis at end-diastole and end-systole was recorded for each technique, no matter how little of the slice was included (the number of short-axis slices included in the analysis at end-diastole and end-systole potentially differed due to the excursion of the mitral valve plane during

Figure 2. Methodological differences in endocardial contouring. a: Simplified contouring; the “compacted” endocardial border was traced resulting in papillary muscles and trabeculae being included in volumetric and excluded from mass measurements. b: Detailed manual contouring; detailed tracing of the endocardium was performed manually such that papillary muscles and trabeculae were included in mass and excluded from volumetric measurements. c: Threshold-based detailed contouring; detailed tracing of the endocardium was performed using a signal intensity-based “thresholding” tool, which delineated individual voxels as blood or myocardium. The thresholding level was manually adjusted until endocardial appearances correlated with visual assessment.
the cardiac cycle). End-systolic frame used and the time taken for analysis was also recorded.

Geometric Modeling

LV volumetric analysis was performed on each scan using six geometric models (Fig. 3) (12,17–19). Appropriate contours were drawn for each model using open-source OsiriX Imaging Software (Pixmeo, Switzerland, v. 3.8). EDV, ESV, and EF were calculated using an OsiriX software plug-in, with the exception of the triplane model where contours were drawn using OsiriX but calculations were performed manually, as no appropriate plug-in was available. Papillary muscles and trabeculae were included in volumetric measurements. In the absence of a gold standard comparator, results obtained from each model were compared with the method found to have the highest observer agreement of the three described above, as prespecified. LV internal diameter, defined as the distance between the endocardial border of the septum and the inferolateral wall just distal to the mitral valve leaflet tips in the three-chamber long-axis view, was measured at end-diastole and end-systole and compared to EDV and ESV, respectively.

Observer Variability

To assess interobserver variability, 25 randomly selected scans (50%) were independently reanalyzed by a second experienced observer using the methods described. To assess intraobserver variability, 25 scans (50%) were reanalyzed by the first observer with a 1-month temporal separation between analyses.

Statistics

Values are expressed as mean ± standard deviation (SD) unless stated otherwise. Agreement between each method was evaluated using Bland–Altman testing by calculating mean difference (bias) and 95% limits of agreement (ie, mean difference ± 2 SD). The significance of the differences between each method was assessed using Wilcoxon rank testing. Correlation was assessed using Spearman’s rank correlation coefficient (r). Inter- and intraobserver variability were evaluated using the repeatability coefficient (defined as $1.96 \times \sqrt{\text{sum of the squares of the differences between observer measurements divided by n}}$) (20). The variability of each method was compared using a Wilcoxon rank comparison of the squared differences between observer measurements (an estimate of the within-subject variance for that method multiplied by two) for each technique (20,21). P < 0.05 was considered significant. Statistical analysis was performed using SPSS Statistics (v. 19, IBM).

RESULTS

Patient characteristics and scan indications are displayed in Table 1. Patients exhibited a variety of cardiovascular pathology and a wide range of mass, EDV, ESV, and EF (Table 2).

Volumetric, EF, and Mass Measurement

Compared to the other two techniques, Method 1 significantly overestimated both EDV and ESV, but ESV to a greater degree (Tables 2, 3; Fig. 4B,C,E,F). As a

<table>
<thead>
<tr>
<th>Patient Characteristics, Scan Indications, and Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number (%)</strong></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Arrhythmia</td>
</tr>
<tr>
<td>Frequent univentricular ectopics</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
</tr>
<tr>
<td>Atrial flutter</td>
</tr>
<tr>
<td>Scan indication</td>
</tr>
<tr>
<td>Myocardial perfusion assessment</td>
</tr>
<tr>
<td>Viability assessment*</td>
</tr>
<tr>
<td>Valvular assessment</td>
</tr>
<tr>
<td>Aortic assessment</td>
</tr>
<tr>
<td>Main scan diagnosis</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>Valvular heart disease</td>
</tr>
<tr>
<td>Pericardial disease</td>
</tr>
<tr>
<td>Ascending aortic aneurysm</td>
</tr>
<tr>
<td>Intracardiac mass</td>
</tr>
</tbody>
</table>

*Includes assessment for cardiomyopathy, pericardial disease, intracardiac masses, and myocardial viability.
result, Method 1 significantly underestimated SV and EF (Tables 2, 3; Fig. 4H,I). As a consequence, six patients (12%) classified as having a normal EF by Method 3 (using a nominal normal EF cutoff of ≥55% (9)) were classified as having a reduced EF by Method 1. Furthermore, six patients (12%) classified as having mild to moderate LV impairment by Method 3 (EF <55% but >35%) were classified as having severe LV impairment (EF ≤35%) by Method 1. There was no significant difference between Method 2 and Method 3 for EDV, ESV, SV, and EF, but the 95% limits of agreement were wide for each parameter, suggesting these methods were not interchangeable (Tables 2, 3; Fig. 4A,D,G).

While Method 1 significantly underestimated mass compared with to Method 2, both underestimated mass compared to Method 3 (Tables 2, 3; Fig. 4J–L).

**Basal Slice Selection, End-Systolic Frame Selection, and Analysis Time**

Method 3 determined the base of the LV to be “more basal” than Methods 1 and 2 at both end-diastole and end-systole. At end-diastole, the most basal short-axis slice included in Method 3 was 1.4 ± 0.5 (where short-axis slice number 1 is the most basal), compared to 2.0 ± 0.6 for Methods 1 and 2; P < 0.0005. Likewise at end-systole, the most basal short-axis slice included in Method 3 was 2.1 ± 0.6, compared to 3.0 ± 0.7 for Methods 1 and 2; P < 0.0005.

The frame used as end-systole was also significantly different between Method 3 (frame number 10.2 ± 1.5) and Methods 1 and 2 (frame number 9.9 ± 1.5; P = 0.003). However, in the 21 patients where there was a difference in end-systolic frame selection, mean difference in ESV was only 2 ± 1 mL.

Analysis time for Method 1 (7.0 ± 1 minutes) and Method 3 (7.5 ± 1 minutes) were both significantly shorter than for Method 2 (11.0 ± 2 minutes; P < 0.0005 for both). There was no significant difference between analysis time for Method 1 and Method 3.

**Observer Variability**

Method 3 had higher interobserver and intraobserver agreement than Methods 1 and 2 for measurement of EF (interobserver; P < 0.0005 compared to Method 1 and P = 0.001 compared to Method 2; intraobserver; P = 0.04 compared to Method 1 and Method 2) but there was no difference in observer agreement between Methods 1 and 2 for measurement of EF (Fig. 5). Both Method 2 and Method 3 had significantly higher observer agreement than Method 1 for measurement of EDV and ESV. However, there was no

### Table 2

<table>
<thead>
<tr>
<th></th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDV (mL)</td>
<td>177±53 (77-324)</td>
<td>169±52 (69-315)</td>
<td>170±54 (61-328)</td>
</tr>
<tr>
<td>ESV (mL)</td>
<td>92±52 (24-258)</td>
<td>78±50 (19-245)</td>
<td>76±47 (23-240)</td>
</tr>
<tr>
<td>SV (mL)</td>
<td>85±26 (21-142)</td>
<td>92±30 (19-154)</td>
<td>94±28 (13-150)</td>
</tr>
<tr>
<td>EF (%)</td>
<td>50±15 (19-72)</td>
<td>57±17 (20-78)</td>
<td>57±16 (21-81)</td>
</tr>
<tr>
<td>Mass (g)</td>
<td>96±29 (47-158)</td>
<td>105±33 (52-176)</td>
<td>140±39 (74-233)</td>
</tr>
</tbody>
</table>
difference in observer agreement between the Method 2 and Method 3 for EDV and ESV measurements. Method 3 had higher interobserver agreement than Methods 1 and 2 for measurement of mass ($P = 0.02$ compared to Method 1 and $P < 0.0005$ compared to Method 2), although intraobserver agreement was not significantly different.

**Geometric Modeling**

Geometric models were compared with Method 3. EDV, ESV, and EF varied substantially between geometric models (Table 4). The biplane model showed the highest agreement with Method 3, with no significant difference in EDV, ESV, and EF, although limits of agreement were wide for each parameter. For the majority of the other models, volume and EF measurements were significantly different from the semiautomated technique, with very wide limits of agreement. (Comparisons of geometric models with Methods 1 and 2 for measurement of EF are included in Table 5.) There was a moderate correlation between LV internal diameter in diastole and EDV ($r_s 0.65$) and a good correlation between LV internal diameter in systole and ESV ($r_s 0.83$).

Interobserver and intraobserver agreement for measurement of EF was low for all models and was significantly lower than that of Method 3 (Fig. 6). Interobserver agreement for measurement of EDV using the Teicholz, triplane, modified Simpson’s and biplane models was not significantly different from that of Method 3. However, all models also had significantly lower interobserver agreement than Method 3 for measurement of ESV. All models also had significantly lower intraobserver agreement for measurement of EDV and ESV than Method 3.

**DISCUSSION**

This study assessed the impact of real-world practice variation in the process of quantifying LV indices from...
SSFP CMR images. Principally, different methods of LV base identification and endocardial contouring led to significant differences in LV mass, volume and EF measurements, and observer variability, findings that have direct implications for clinical and academic CMR practice. In addition, despite their widespread clinical use, particularly in the field of echocardiography, this study questions the value of geometric models.

Recruitment was prospective and consecutive, without exclusion on grounds of image quality, arrhythmia, or breath-holding ability, with prospective gating required in 4% of patients. Patients exhibited a wide range of LV mass, volumes, and EF and a variety of pathology. It is therefore felt that the study is representative of real-world CMR practice.

Methodological differences in endocardial border tracing had the greatest impact on LV volume and EF measurements. Simplified endocardial contouring, where papillary muscles and trabeculae are included within volumes and excluded from mass, led to a significant overestimation of volumes compared to detailed contouring (manual or threshold-derived). However, in keeping with other studies, ESV was overestimated to a greater degree (EDV overestimated by 5%, ESV by 19%, compared to detailed manual endocardial contouring) (22,23). Consequently, simplified contouring led to a significant underestimation of EF. The degree of EF underestimation in this study (6% difference between simplified and detailed manual contouring) is larger than that found by Weinsaft et al (23) (3%), although the current study includes patients with a wider range of volumes and EF. In a study assessing the impact of simplified versus detailed tracing of trabeculae alone, Papavassiliu et al (22) found the simplified technique underestimated EF by 2% (papillary muscles were included within mass measurements in all subjects). The greater underestimation of ESV relative to EDV by the simplified technique may result from it being more difficult to differentiate the compacted endocardial border from papillary muscles and trabeculae at end-systole.

The technique of simplified endocardial contouring is widely practiced (10,11), despite detailed contouring (manual or threshold-based) having been used to establish normal reference ranges for SSFP imaging (6,9,24) and the greater accuracy of detailed contouring in ex vivo validatory studies (25,26). In this study, the use of simplified contouring had potential clinical implications, with nearly a quarter of patients being assigned to a different EF “category” compared to detailed contouring (ie, abnormal EF rather than normal, or severely impaired EF rather than mild-moderately impaired), thus highlighting the importance of using reference ranges appropriate to the analysis method used. Some software packages allow independent tracing of papillary muscles, which has the advantage of allowing volumetric (and mass) data to be presented both inclusive and exclusive of the papillary muscles. However, independent tracing of trabeculae is often more difficult, and the fundamental methodology of such software, ie, manual border tracing, is the same as that assessed in the current study. One of the reasons why simplified endocardial contouring is commonly used is in order to reduce analysis time; indeed, simplified contouring reduced analysis time by over a third compared to detailed manual contouring in the current study.

**Figure 5.** Interobserver and intraobserver variability of each analysis method for measurement of volume indices, EF and mass, expressed as repeatability coefficients. The repeatability coefficient is the range within which measurements by two different observers (interobserver) and two measurements by the same observer (intraobserver) are expected to lie. *A significant difference, assessed using a Wilcoxon rank comparison of the squared differences ($P < 0.05$). Abbreviations as in Table 2.
Long-axis-derived LV base identification led to highly significantly more basal positioning of the LV base compared to the short-axis-derived method. This finding demonstrates that the LV cavity frequently extends beyond the slice determined to be the most basal using the widely used criteria whereby short-axis slices are included within the LV if 50% or more of the cavity is surrounded by ventricular myocardium. Despite this finding, however, variability in base positioning did not have a significant impact on volumetric measurements, with no significant differences in volume and EF measurements seen between Methods 2 and 3. It is possible that the more basal positioning of the LV base using the long-axis method was offset by including a greater amount of cavity from the basal short axis slice using the short-axis method, due to the technique of joining up the endocardial contours with a curved line through the blood pool, as is widely practiced (27).

Although there was a significant difference between automated and visually determined end-systole, the absolute difference in frame count was small, and the effect on ESV minimal (2 mL).

Unlike with volumetric measurements, methodological differences in LV base identification appeared to be more important than method of endocardial tracing for measurement of mass, with a mean difference of 34g seen between Methods 2 and 3, which had comparable detailed endocardial contouring, but different base positioning. This large difference is in accordance with the difference seen between studies by Maceira et al (24) and Hudsmith at al (6) that defined normal values for LV mass (and volumes and EF) using different analysis techniques. Using Method 3,
Maceira et al found mean myocardial mass to be 156g in healthy males and 128g in healthy females but using Method 2, Hudsmith et al found mean mass to be 123g and 96g in healthy males and females, respectively. Of note, there were no discernable differences in epicardial contouring between Methods 2 and 3. This finding further highlights the importance of using normal reference ranges appropriate to analysis method used. The method of simplified endocardial border tracing did lead to a significant underestimation of mass compared to the detailed method; however, the difference was small in comparison (9g).

Methodological differences in image analysis led to substantial disparity in observer agreement, a finding that has important implications for patient follow-up and for determining research study sample sizes (28). Method 3 had the highest observer agreement for measurement of EF and mass. This technique used the long-axis-derived method of identifying the LV base, which allows straightforward identification of the mitral valve plane from long-axis images, and hence LV base, with minimal scope for observer variability.

Furthermore, end-systole was determined automatically. Therefore, the main variability for this technique stemmed from the manual adjustment required to set the thresholding level appropriate to the endocardial border. However, this involves substantially less user input than the manual endocardial tracing techniques, and our experience is that accustomed observers perform this consistently. Indeed, likely related to the minimal manual input required, these results demonstrate that the threshold-determined method of endocardial contouring is considerably more consistent than manual tracing. Epicardial border tracing was another potential source of variability, although this would only affect mass measurements, and due to its typically circular shape, epicardial contouring is generally straightforward. In light of these findings, it is suggested that the methods of long-axis-derived LV base identification, detailed endocardial contouring using a signal threshold (or similar)-based technique and automated determination of end-systole, which involve minimal manual input, should be incorporated into all analysis software.

Measurement of mass showed greater variability than all other parameters, which is in keeping with the findings of other studies (28). A number of studies have advocated the use of geometric models as a timesaving alternate to conventional CMR image analysis (12–16). Indeed, these models are incorporated into state-of-the-art CMR software analysis packages, and are widely used by other imaging modalities, particularly echocardiography. The findings of this study, the largest to assess the utility of geometric modeling, show that the agreement between geometric modeling techniques and conventional analysis is poor, with wide limits of agreement for all models. In addition, the variability of all models was too high for clinical or academic use, with repeatability coefficients of 13% or greater for measurement of EF.

These findings are in apparent contrast to a smaller study (n = 25) by Thiele et al (12), which concluded that the level of agreement between the Triplane model and conventional analysis for measurement of EF was such that this model could be used clinically. However, agreement was assessed using “mean relative difference,” defined as the difference of two techniques divided by their mean value, and the 95% confidence interval of the mean difference, which represents the confidence interval for the mean difference derived from the sample of individuals and, importantly, does not represent the confidence interval for the difference for an individual, and by definition would result in substantially narrower “limits of agreement” than the method used in the present study (i.e., absolute mean difference and 95% limits of agreement, calculated as the mean difference ± 2 SDs of the mean difference, as described by Bland and Altman (20)). Indeed, using the method described by Thiele et al, the mean relative difference in the present study between the Triplane model and Method 3 is 0.1%, with a 95% confidence interval of ± 1.6%, which is similar to that reported by Thiele (3.7 ± 2.3% for patients with no regional dysfunction and 6.1 ± 5.7%)

![Figure 6. Interobserver and intraobserver variability of each geometric model for measurement of volume indices and EF, expressed as repeatability coefficients. *A significant difference (P < 0.05) compared to conventional analysis (Method 3). Geometric models showed significantly greater variability than conventional analysis for all parameters, with the exception of the interobserver variability of triplane, biplane, modified Simpson’s, and Teicholz model-derived measurements of EDV. Abbreviations as in Tables 2 and 4.](image-url)
for patients with regional dysfunction). Agreement data for measurement of EDV and ESV were not presented and patients with suboptimal image quality were excluded. It is suggested that geometric models devalue CMR and their use should be discouraged.

Future methods for quantification of LV mass, volume, and EF may include deformable models, active shape and appearance modeling, and atlas-based methods, in which models are derived from anatomical or statistical properties of large databases of patients and then applied to individuals (29). Such techniques have the potential to allow fully automated, and hence highly accurate and reproducible analysis, although they are currently not sufficiently developed for clinical use.

The main limitation of this study is that no comparison is made to an externally validated gold standard, so accuracy cannot be assessed. In addition, this is a single-site, single-vendor study. Interobserver variability was assessed using only two observers and intraobserver variability was assessed using only a single observer. Reproducibility of measurements, requiring two separate scans, was not assessed.

In conclusion, real-world variability in the process of quantifying LV indices from SSSF CMR images leads to significant differences in LV mass, volume and EF measurements, and observer variability. Reference ranges appropriate to analysis method used must be applied. Use of geometric models for analyzing CMR images should be discouraged.

REFERENCES


