

Article

Analysis of the Accuracy and Inter-Reader Precision of Scar Quantification Techniques in Aortic Stenosis: A Comparative Cardiovascular Magnetic Resonance Imaging Study

Megan Rian Rajah ^{1,*}, Pieter-Paul Strauss Robbertse ¹, Vishesh Sood ² , Tonya Marianne Esterhuizen ³ , Anton Frans Doubell ¹  and Philip George Herbst ¹ 

¹ Division of Cardiology, Department of Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg Hospital, Cape Town 7505, South Africa

² Department of Radiology, SCP Radiology, Panorama Mediclinic Hospital, Cape Town 7500, South Africa

³ Division of Epidemiology and Biostatistics, Department of Global Health, Stellenbosch University, Cape Town 7505, South Africa

* Correspondence: 16607031@sun.ac.za

Abstract

Background: An important determinant of mortality in AS is the presence and quantity of myocardial scar. Scar quantification using late gadolinium enhancement (LGE) on cardiovascular magnetic resonance (CMR) imaging may be a useful risk-stratification tool for at-risk patients who do not meet current criteria for valve intervention. The incorporation of this tool into clinical practice is currently limited by a lack of consensus on the best LGE quantification technique to use. **Methods:** Fifteen patients with severe AS underwent LGE imaging on CMR. A reference estimate of the LGE mass was made using a semi-automatic quantitative visual method. An intensity slider (reporting software provided) was used to mark areas of enhanced signal in each short-axis slice that correlated with the reader's visual assessment of LGE, which used predetermined imaging criteria. This visual slider method (VsLM) of determining LGE mass was then used as a reference for establishing the accuracy of various semi- and fully automated methods for identifying and quantifying LGE burden. These included the signal threshold versus reference mean (STRM) method at thresholds of two-, three-, and five-standard deviations (2SD, 3SD, 5SD, respectively), the full width at half maximum (FWHM) method and the Otsu auto threshold (OAT) method. An intraclass correlation analysis was performed to establish and compare the inter-reader reliability for each method. **Results:** Three readers demonstrated 100% agreement on the presence of LGE in 12/15 (80%) of study cases. Accuracy determined by the Wilcoxon rank sum, Spearman correlation and Bland–Altman tests suggested that the 5SD method using remote myocardium reference regions of interest only in slices with visually detected LGE was best (Wilcoxon rank sum p -values ranged from 0.3 to 0.5 for the three readers, bias on Bland–Altman was <0.5 g for all three readers). This was followed by the FWHM method, but with wide minimum-maximum ranges observed. Inter-reader reliability was best for the 2SD STRM method ($ICC = 0.9$, $p < 0.001$), but accuracy using this method was clinically unacceptable. Inter-reader reliability was statistically acceptable for the VsLM ($ICC = 0.7$, $p < 0.001$). The FWHM method yielded the best balance between accuracy and reliability but may be limited by the heterogeneity of scars observed from patient to patient. **Conclusions:** The FWHM appeared to offer a reasonable balance between accuracy and precision. However, it was not always the best fit, e.g., in patients with small or non-bright scars. There may be no additional benefit to using the semi- and fully automated methods in the context of AS, and visual estimation, when performed in the manner described in this study (i.e., the VsLM), may be clinically sufficient.



Academic Editors: Giorgio Treglia,
Lucia La Mura and Ciro Santoro

Received: 26 January 2026

Revised: 2 April 2026

Accepted: 11 June 2026

Published: 28 June 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: myocardial fibrosis; fibrosis quantification; late gadolinium enhancement; aortic stenosis; cardiovascular magnetic resonance imaging; aortic valve replacement

1. Introduction

Aortic stenosis (AS) is one of the most common valvular lesions with a growing prevalence worldwide [1]. Definitive treatment by means of surgical aortic valve replacement (SAVR) or transcatheter aortic valve implantation (TAVI) is indicated in those with severe disease and symptoms or severe disease with systolic dysfunction (left ventricular ejection fraction < 50%) [2–4]. Despite the widespread implementation of these interventions, the morbidity and mortality remain high [1]. Furthermore, a mortality risk in groups not meeting current guideline recommendations for intervention has also been shown [2,3,5,6]. This raises two important questions: firstly, whether the current guideline recommendations for the timing of intervention can be improved; and secondly, whether the presence/absence of other high-risk clinical parameters warrants their incorporation into a new risk stratification strategy for reclassifying high-risk patients not qualifying for valvular intervention. These parameters include evidence of myocardial injury using biomarkers such as brain natriuretic peptide/amino-terminal B-type natriuretic peptide (BNP/NT-pro BNP, respectively), abnormal strain analysis on speckle-tracking echocardiography, and the presence and quantity of myocardial fibrosis [2,3].

Myocardial fibrosis in AS is well-described on histology and cardiovascular magnetic resonance (CMR) imaging, and it takes on two distinct patterns—diffuse interstitial fibrosis and replacement fibrosis/scar [7,8]. Replacement fibrosis/scar in AS is associated with an increased mortality risk, and this relationship has been shown to be dose-dependent [9–16]. Cardiovascular magnetic resonance imaging offers clinicians a validated approach for the assessment of replacement fibrosis through late gadolinium enhancement (LGE) imaging, which circumvents the invasive nature, required expertise, and issues of sampling bias associated with endomyocardial biopsy/histology [17–20]. The detection of LGE requires the use of inversion-recovery magnetic resonance sequences planned specifically to null the background/normal myocardium, against which hyperenhanced areas where gadolinium is concentrated can be visually assessed [21–24]. This visual identification of bright LGE areas remains the clinical mainstay for the detection of replacement fibrosis on CMR [24]. Several post-processing methods for the quantification of LGE exist. However, consensus as to which method is best for replacement fibrosis quantification in AS has yet to be reached, thus limiting its adoption into clinical practice [24,25].

The methods used for LGE quantification in AS include a visual method and a variety of semi-automated ones [24–26]. As described, the mainstay of LGE detection on CMR relies on the visual identification of bright/hyperenhanced myocardial areas. This visual assessment has been further developed and used as a robust but rather blunt quantitative instrument, requiring the user to broadly classify the transmural proportion of LGE per left ventricular (LV) segment (typically, 0–25%, 25–50%, 50–75% and 75–100%) as representing the proportional extent of replacement fibrosis in those segments [24]. By tallying the proportions of each individual segment affected by replacement fibrosis, a global LV scar burden may be estimated. Although demonstrated to be robust, with data in a number of different pathologies [24,27–29], this is a time-consuming endeavour and results in a broad categorisation in terms of the scar burden. A number of semi- and fully automated methods have been developed in an attempt to refine the visual method of replacement fibrosis by narrowing the broad categories to more specific estimates that better approximate the visual assessment of the observer. In addition to improving the accuracy of the traditional

visual method, these approaches also aim to improve the speed and efficiency of LGE assessment, as well as the reproducibility of its quantification [25]. Some methods were specifically developed to reduce bias by introducing a user-selected reference region (remote myocardial reference regions representing normal myocardium or bright LGE reference regions representing scar, depending on the methodology used) that defines a slice-by-slice signal-intensity threshold above which pixels are automatically identified as LGE and quantified as such.

Several challenges with the use of these semi-automated methods in AS have previously been described in detail by our group [25] and include the persistence of bias and poor reproducibility due to a critical dependence on the size and positioning of the user-defined reference regions, as well as the coexistence of diffusely distributed interstitial fibrosis. The presence of relatively small, non-bright scars drives further complexity in interpreting the results. An interesting semi-quantitative visual method, termed the visual slider method (VslM), allows the reporting physician to manually adjust a signal-intensity threshold slider, above which pixels are highlighted and counted. The operator can therefore visually adjust the slider until relevant areas identified as true LGE are marked and counted as fibrosis. This allows the operator to manually curate areas deemed fibrosis while narrowing the scar quantity categorisation compared to the broad output of the traditional visual method, and while avoiding the challenges associated with the semi-automated methods.

A head-to-head analysis of the various LGE quantification methods in AS remains limited. A single study of this nature exists where a range of standard deviation (SD) thresholds within the signal threshold versus reference mean (STRM) method was evaluated [8]. A head-to-head analysis that includes an operator-curated visual method, such as the VslM, has not been studied, and alternatives to the STRM, such as the full width at half maximum (FWHM) and Otsu auto threshold (OAT) methods, have yet to be reported for AS. This study aimed to compare the accuracy and precision/inter-reader reproducibility of five semi-automated LGE quantification methods using an operator-curated, semi-automated estimate of replacement fibrosis, the VslM, as the reference standard. The study also aimed to illustrate the impact of different post-processing styles on the accuracy and precision of the various LGE quantification methods.

2. Methods

2.1. Study Population

The study population comprised fifteen participants selected from a larger study at Tygerberg Hospital, Cape Town, South Africa. All participants met the guideline criteria for high-gradient severe AS (aortic valve area $< 1.0 \text{ cm}^2$, mean transvalvular pressure gradient $\geq 40 \text{ mmHg}$, peak velocity $\geq 4 \text{ m/s}$) with and without systolic dysfunction as diagnosed on transthoracic echocardiography [2,3]. Patients with other haemodynamically significant valve lesions, concomitant cardiomyopathy including cardiac amyloidosis, previous myocardial infarction and/or significant coronary artery disease were excluded from the study. Significant coronary artery disease (CAD) was determined by coronary angiography (obstruction of $>50\%$ was regarded as significant). The study obtained ethical approval from the Human Research Ethics Committee of Stellenbosch University (S21/11/251 PHD), and written informed consent was obtained from all participants.

2.2. Image Acquisition

Cardiovascular magnetic resonance imaging was performed using a Magnetom Aera (Siemens, Erlangen, Germany) 1.5 Tesla scanner. A standardised protocol in accordance with the Society of Cardiovascular Magnetic Resonance's recommendations was used [23]. Late gadolinium enhancement images were acquired 10–15 min after contrast

injection (Gadovist 0.2 mmol/kg, Bayer Pharmaceuticals, Leverkusen, Germany). An electrocardiogram-gated, breath-held, inversion-recovery gradient-echo sequence was used for the image acquisition. Images were acquired in the standard cardiac planes, which included two-, three-, and four-chamber long-axis views and a short-axis stack covering from the base to the apex of the LV. The choice of inversion time (TI) was selected using a modified look-locker inversion-recovery TI scout. Typical image parameters included an 8 mm slice thickness with no gap where applicable, and TE. 1.20–1.89 ms, TR. 784–804 ms, TI 260–280 ms, and matrix size 208×256 with voxel size $\pm 1.4 \times 0.5 \times 8.0$ mm.

2.3. Image Post-Processing

The LGE images of all fifteen participants were independently post-processed by three CMR readers with six (PPRS), four (MRR) and four (VS) years of CMR experience, respectively. Two of the readers were blinded to the clinical data, to other imaging modalities, and to the LGE measurements of the other readers. Access to the cine images was available to guide the endo- and epicardial contouring. Other CMR measurements, e.g., LV volumes and parametric mapping times, however, were not made available to the two blinded readers. One reader served as the primary investigator for the larger cohort and was therefore not blinded to the clinical information or other imaging modalities and measurements.

The images were post-processed using commercially available software (CVI42, version 5.13.7, Circle Cardiovascular Imaging, Calgary, Canada). A visual analysis was initially performed to identify areas of hyperenhancement in accordance with standard practice [24]. Areas of enhancement were diagnosed if visualised in two orthogonal planes and on a phase swap of the original image. Following a suitable level of windowing, endo- and epicardial borders were contoured on every slice of the short-axis LGE stack. The VslM was chosen as a reference method and is described in further detail below. Semi-automated methods tested against the reference included (1) the signal threshold versus reference mean method (STRM) at thresholds of two-, three-, and five-standard deviations (2SD, 3SD, 5SD, respectively) above the mean signal intensity of remote myocardium, (2) the full-width at half maximum (FWHM) method and (3) the Otsu auto-threshold (OAT) method. For the STRM method, a normal myocardial region of interest (ROI) was initially drawn only in those slices identified by the reader as containing LGE (denoted by the letter “A” in the text) and then in every slice of the short-axis stack irrespective of whether LGE was identified by the visual assessment of the CMR reader (denoted by the letter “B” in the text). For method A, the extrapolation setting (a setting that applies the mean signal intensity of the ROI to the ROI-free adjacent slices) was turned off. Therefore, those slices deemed LGE-free by the CMR reader were excluded from the analysis without impacting the total myocardial volume used for the total LGE proportion (%) calculation. Care was taken to avoid the endo- and epicardial borders when drawing the myocardial ROI. For the FWHM method, an enhancement ROI was drawn in the core of the scar to minimise the impact of partial volume effects arising from the scar edges. As for the STRM method, two post-processing techniques were compared: a single ROI was drawn with extrapolation on (denoted by the letter “A” in the text), and ROIs in every scar present with extrapolation on were also drawn (denoted by the letter “B” in the text). The LGE mass (grams) and proportion (%) of total myocardial volume were recorded for analysis.

2.4. Reference Method

A visual slider method (denoted by the abbreviation “VslM”) was selected as a reference against which other methods were compared. Rather than the traditional visual method (Figure 1a) of user-estimated transmural extent/proportional area LGE (%) per LV segment, a semi-quantitative approach was employed using the “manual per slice” option

offered by CVI42 (Figure 1b,c). Endo- and epicardial borders were contoured on every slice of the short-axis stack. A slider function, guided by the reader's visual assessment, was then manually adjusted (slice-by-slice) to mark pixels considered as hyperenhanced. This yielded an objective estimate of the LGE mass (grams) and the proportion of LGE (%) of total myocardial volume. This pixel-wise quantification enabled a more objective visual estimate than the traditional visual approach (Figure 1).

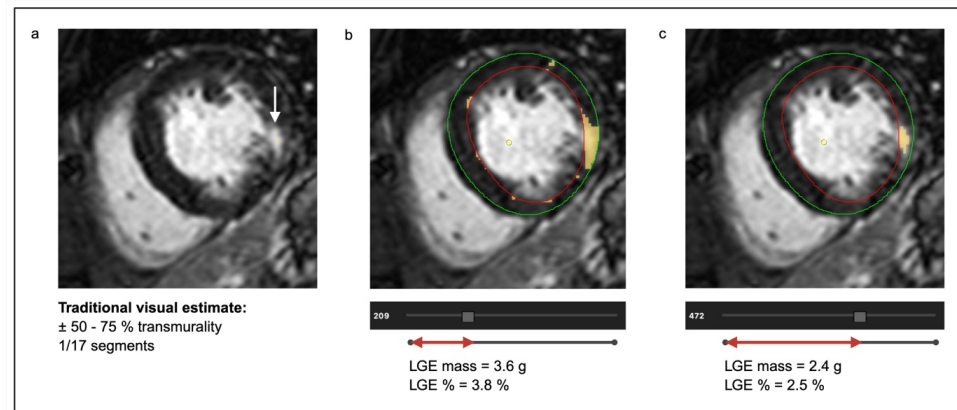


Figure 1. Illustration of the visual slider method (VslM) of LGE quantification. (a) Inferolateral subendocardial LGE is shown by the white arrow. The traditional visual assessment yields a transmural extent of ± 50 –75% in one of seventeen LV segments. (b,c) The semi-quantitative VslM approach. The yellow-shaded region illustrates the pixels counted as enhanced, based on the manual adjustment of a slider function/tool (depicted below each short-axis image) by the CMR reader. This function is available on CVI42 software, which provides a cumulative, quantitative output in the form of LGE mass (grams) and proportion LGE (%) for each slice analysed. VslM—visual slider method, LGE—late gadolinium enhancement, LV—left ventricular.

2.5. Statistical Analysis

Statistical analysis was performed using GraphPad Prism (version 10.0.2, GraphPad, Boston, TX, USA) and SPSS (version 29.0.2.0, IBM, New York, NY, USA). Normality tests were performed using the Shapiro–Wilk test, histograms, and quantile–quantile plots. Accuracy was tested using paired Wilcoxon rank-sum tests, Spearman's correlation (reported as rho with a 95% confidence interval and *p* value) and Bland–Altman analyses (reported as bias $\pm$ standard deviation). For the Bland–Altman analysis, a natural log transformation was performed, and the normality of the differences was confirmed using the Shapiro–Wilk test. For the inter-reader reliability evaluation, intraclass correlation (ICC) analysis was performed on the normalised data. The images of all fifteen participants were read by the same three, non-specific readers. A two-way random effects model was therefore selected [30]. The ICC for the absolute agreement of single measurements between readers for each method was reported together with its *p*-value.

3. Results

3.1. Study Population

The baseline characteristics of the study population are shown in Table 1. Fifteen patients with high-gradient severe AS were analysed for this study. The mean aortic valve area was 0.51 ± 0.15 cm² with a mean gradient of 61 ± 18 mmHg. Patients with and without systolic dysfunction were included, with a mean LV ejection fraction of $47 \pm 18\%$. The prevalence of hypertension was high [13 (87%)], but with well-controlled blood pressure readings (systolic blood pressure 126 ± 22 mmHg and diastolic blood pressure

73 ± 12 mmHg). There were no cases of CAD (determined by coronary angiography) or cardiac amyloidosis (determined by CMR).

Table 1. Baseline characteristics of the study population.

Parameter	Population (n = 15)
Age (years)	61 ± 14
Sex	
Female n (%)	12 (80)
Male n (%)	3 (20)
Hypertension n (%)	13 (87)
Diabetes Mellitus n (%)	5 (33)
Coronary artery disease n (%)	0 (0)
AVA (cm ²)	0.51 ± 0.15
Mean gradient (mmHg)	61 ± 18
Peak gradient (mmHg)	96 ± 28
LVEF (%)	47 ± 18
SBP (mmHg)	126 ± 22
DBP (mmHg)	73 ± 12

Continuous variables are presented as mean ± standard deviation. Categorical variables are presented as absolute numbers (percentages). AVA—aortic valve area, LVEF—left ventricular ejection fraction, SBP—systolic blood pressure, DBP—diastolic blood pressure.

3.2. Accuracy by Method Used

Of fifteen severe AS cases, three were found to have no LGE (determined by visual assessment) with 100% agreement amongst the three readers. As shown in Figure 2, the median LGE mass varied significantly according to the method used. Compared to the reference method (VsLM), the largest absolute differences were obtained using the OAT method [reader one: 3.8 (0.9–6.2 g) vs. 35.5 (28.0–45.4 g), reader two: 3.6 (1.2–9.7 g) vs. 30.6 (26.9–37.8 g) and reader three: 3.0 (1.1–6.5 g) vs. 52.0 (42.2–69.6 g) respectively]. Using paired Wilcoxon rank-sum tests and the VsLM as a reference, significant differences ($p > 0.05$) were found for the 2SD (A), 3SD (A) and OAT methods across all three readers (Supplementary Table S1). No statistically significant differences were found for readers two and three when using the 5SD (A) method and for readers one and two when using the FWHM (A) method (Supplementary Table S1). Significant, near-perfect linear relationships were found for all methods when compared to the VsLM except for the OAT method (Spearman's rho ranged from 0.1 to 0.4 across readers) (Supplementary Table S1). The smallest bias was observed using the 5SD (A) method, and this finding was consistent across all three readers (reader one: -0.03 ± 1.37 , reader two: 0.03 ± 0.76 and reader three: 0.32 ± 0.69) (Supplementary Table S2). The largest bias was observed using the OAT method (reader one: 2.81 ± 1.60 , reader two: 2.20 ± 0.94 and reader three: 2.77 ± 1.08) (Supplementary Table S2).

3.3. Accuracy and Precision by Post-Processing Method Used

Two different post-processing methods were tested for the 2SD, 3SD, 5SD and FWHM techniques. For the STRM method, myocardial ROIs drawn in every slice of the short-axis stack (method B) consistently yielded higher LGE masses compared to method A across all three standard deviation thresholds and for all three readers (Figure 3). Similarly, method B for FWHM, which utilised multiple scar ROIs rather than a single scar, also yielded higher LGE masses for all three readers. The magnitude of the absolute differences for FWHM was, however, small (Supplementary Table S3). Bias for all semi-automated methods (2SD, 3SD, 5SD, FWHM) was higher using method B compared to method A (Supplementary Table S2).

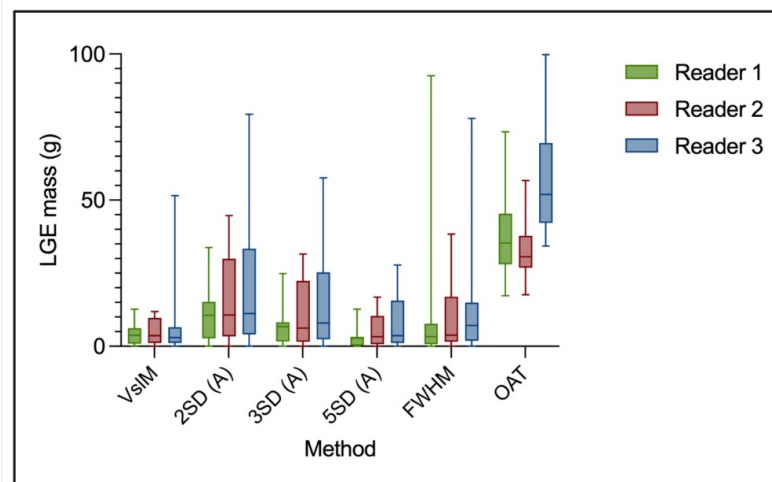


Figure 2. Column graph displaying the median and interquartile range of LGE mass determined using six different methods amongst three different readers. The median LGE mass, when compared to that of the VslM, varied significantly for the 2SD (A), 3SD (A) and OAT methods amongst all three readers. The 5SD (A) method for readers two and three and the FWHM method for readers one and two yielded non-significant differences in the median LGE mass when compared with the VslM. Significance was determined using the Wilcoxon rank-sum test, and a p -value < 0.05 was considered significant. LGE—late gadolinium enhancement, VslM—visual slider method, 2SD (A)—two standard deviations method A, 3SD (A)—three standard deviations method A, 5SD (A)—five standard deviations method A, FWHM—full width at half maximum, OAT—Otsu auto threshold.

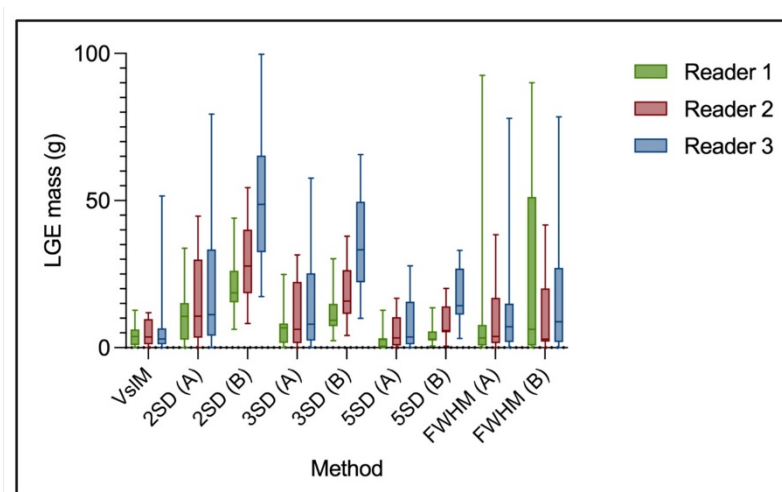


Figure 3. Column graph comparing post-processing method (A) vs. (B) for the STRM and FWHM techniques. In STRM method A, a normal myocardial ROI was drawn only in slices containing LGE compared to a normal myocardial ROI drawn in every single slice for method B. Using the VslM as the reference, method B consistently produced higher absolute LGE masses compared to method A across all standard deviation thresholds. Similarly, method B (multiple ROIs) for the FWHM technique yielded higher absolute values compared to method A (a single ROI), but these differences were smaller than for the STRM methods. LGE—late gadolinium enhancement, VslM—visual slider method, 2SD (A)—two standard deviations method A, 2SD (B)—two standard deviations method B, 3SD (A)—three standard deviations method A, 3SD (B)—three standard deviations B, 5SD (A)—five standard deviations method A, 5SD (B)—five standard deviations method B, FWHM (A)—full width at half maximum method A, FWHM (B)—full width at half maximum method B, ROI—region of interest.

3.4. Inter-Reader Reliability/Precision

The ICCs evaluating reliability amongst the three readers for each method are shown in Table 2. The ICC for VslM was acceptable at 0.7 ($p < 0.001$). The 2SD (A) method yielded the best reliability (ICC = 0.9 with $p < 0.001$) while the fully automated OAT method yielded unacceptable reliability (ICC = 0.5, $p < 0.001$). Both the 3SD (A) and FWHM (A) methods yielded clinically acceptable reliability (ICCs = 0.8 and 0.8, respectively). Amongst the (B) methods, the FWHM (B) method yielded a good ICC of 0.8 ($p < 0.001$), but the remainder of the (B) methods yielded poor reliability results (ICCs < 0.4 with p values < 0.001) (Supplementary Table S4).

Table 2. Intraclass correlation coefficients for inter-reader reliability using a two-way random effects model.

	VslM	2SD (A)	3SD (A)	5SD (A)	FWHM (A)	OAT
ICC	0.7	0.9	0.8	0.6	0.8	0.5
p value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

A p -value < 0.05 is considered statistically significant. ICC—intraclass correlation coefficient, VslM—visual slider method, 2SD (A)—two standard deviations method A, 3SD (A)—three standard deviations method A, 5SD (A)—five standard deviations method A, FWHM (A)—full width at half maximum method A, OAT—Otsu auto threshold.

4. Discussion

This study, comparing different LGE quantification techniques in the context of AS, showed that the choice of method (VslM/STRM/FWHM/OAT) and post-processing style significantly influenced the accuracy and reproducibility of the LGE burden as determined by three independent CMR readers.

The most accurate methods emerging from this study included the 5SD STRM and FWHM methods. These findings are consistent with data from other studies that have performed a head-to-head analysis of different LGE quantification techniques in non-ischaemic cardiovascular diseases (hypertrophic and dilated cardiomyopathy) [29,31,32]. The scarring in these non-ischaemic pathologies share similarities with AS, which include the presence of relatively non-bright scars with concomitant diffuse interstitial fibrosis [29,31,32]. In two of these studies [29,31], a visual method was also used as a reference. The 5SD STRM, 6SD STRM and FWHM methods were found to be most accurate [29,31]. In the third study, histology from whole explanted hearts was used as a reference, and the 6SD STRM method emerged as most accurate, while the FWHM technique was not evaluated [32]. The finding that higher thresholds (5 or 6SDs) were more accurate than 2 or 3SD thresholds likely reflects the impact of high background diffuse interstitial fibrosis, which broadens the range of signal intensities across the myocardium. The STRM method is based on standard deviations. Therefore, a broader range of signal intensities increases the range of pixels that would be considered as enhanced, and to avoid scar overestimation, a higher standard deviation threshold is required. This challenge also exists in AS, where replacement fibrosis often coexists with diffuse interstitial fibrosis and plausibly explains our findings, where the high standard deviation thresholds were found to be more accurate compared to the lower (2–3 SDs) thresholds.

In the context of AS specifically, a direct comparison of different LGE quantification techniques remains exceedingly rare. Research that compared different STRM thresholds against histology found the 3SD STRM method to be most accurate [8]. The difference in the accuracy results between their study and ours may have resulted from the different reference methods used (histology vs. the VslM in our study). The histological examination by Treibel et al. described the presence of microscars measuring smaller than the typical voxel size on CMR [8]. Since each voxel represents an averaged signal intensity, voxels

occupied by both normal myocardial tissue and one or more microscars may reflect a relatively lower signal intensity than expected of scar. The implication of this is that microscars on CMR are potentially missed by both the eye of the reader (visual method) and a high standard-deviation threshold (e.g., 5SD) [21,22,28]. This may explain the high accuracy of the 3SD method when histology is used as the reference (where microscars are more easily diagnosed), compared with the better accuracy of the 5SD method in our study, which used a visually guided estimate as the reference.

In our study, accuracy was assessed by comparing median LGE quantities for each semi-automated method to those of the VsLM assessment rather than histology, which is considered the gold standard for fibrosis evaluation [19,20]. Although there is no consensus recommendation on which LGE quantification technique is optimal for non-ischaemic pathologies, including AS, the guideline recommendation for LGE quantification in general emphasises the initial use of a visual assessment, even when a semi-automated approach is used [24]. Thus, in clinical practice, an informed evaluation is usually made using the information from both quantification strategies. The traditional visual method requires the CMR reader to estimate scar burden according to the cumulative transmural LGE extent in each of the seventeen AHA segments [24]. This visual assessment is, by its nature, subjective. The visual assessment performed in this study (the VsLM approach), however, avoided the need for a user-derived estimate of transmural extent (thereby limiting some of the subjectivity/bias) by use of a slider function in the post-processing software. This function provided an automated LGE mass, determined quantitatively by summing the pixels in a manually defined region of scar (Figure 1), guided by the visual assessment of the reader. Although the semi-automated methods were developed with the aim of improving objectivity, they still rely on visual input from the user who manually adjusts contours and ROIs until satisfied with the visual feedback from the software, which marks/highlights the areas it has counted as scar. For these reasons, we considered the VsLM a reasonable reference in the absence of histology for comparing the accuracy of the semi-automated methods.

While accuracy is an important factor to consider when agreeing on an LGE quantification method, reproducibility is of equal importance if we are to adopt LGE quantification into widespread clinical practice. The 2SD STRM method had the best inter-reader reproducibility in this study. However, a low threshold masks the more subtle impact of inherent signal heterogeneity on the final outcome. This allows for better inter-reader reproducibility but at the expense of significant overestimation of the scar burden, as shown by the clinically unacceptable accuracy of the 2SD STRM method in our data. Good reproducibility was also shown for the 3SD STRM and FWHM methods. In the case of the FWHM method, it has often presented itself as one of the most reproducible techniques in other non-ischaemic pathologies, including hypertrophic cardiomyopathy [29,31]. This is in part related to the fact that the ROI is drawn in a region of scar, therefore negating the challenge of drawing a normal myocardial ROI in myocardium that is also burdened with diffuse interstitial fibrosis. The combination of its good reproducibility and statistically acceptable accuracy makes the FWHM method a potentially well-balanced candidate for scar quantification in AS, but it is not without limitations.

Although the means of the LGE mass differences between the FWHM and VsLM methods for all three readers were small (<1 g), the ranges of the mass differences were wide (Figure 2). On closer inspection, large discrepancies in median values between the FWHM and VsLM methods mostly arose in cases where scars were small, non-bright, or both. In our cohort of AS, scar has been found to be heterogeneous from patient to patient. While non-ischaemic bright-signal-intensity scars that are suitable for quantification by the FWHM method (which assumes a bright core) [25,28] are seen, several patients also

present with small, non-bright scars. A single unifying semi- or fully automated method that accounts for this heterogeneity in the nature of the scarring while remaining both accurate and precise may not realistically exist.

Use of the semi- and fully automated methods was expected to limit the subjectivity associated with the visual method. These methods, however, still require significant user input as mentioned. To illustrate, the STRM and FWHM methods require the user to start with the visual identification of the scar. The ROIs drawn are also user-guided, and in this cohort, the differences in the ROIs drawn contributed to some of the wide variation observed using the semiautomated methods. For example, when readers visually estimated an LGE mass within 1 g of each other, the 2SD STRM method in the same patients yielded LGE mass differences that were nearly double one another. In these cases, systematic bias alone could not explain these differences, as proportional differences would have been expected. The variation, therefore, must have arisen from differences in the location and/or size of the ROI. The three readers in this study were selected from different institutions with different levels of education and experience with regard to CMR. Despite this, inter-reader reproducibility for the VslM estimations was statistically acceptable. An argument can be made that, if a single unifying semi-automated method in the context of AS does not exist, the visual method, when performed in the manner described in this article (VslM), may be sufficient for scar quantification in AS.

The impact of post-processing style has previously been speculated to impact the accuracy of LGE quantification, yet this remains to be systematically investigated in the existing AS literature. In alignment with prior speculation, our data showed that the style of post-processing impacted both the accuracy and reproducibility of LGE quantification in AS. The STRM (B) method, which utilised a myocardial ROI in every slice, including those without visible LGE, was the least accurate and least reproducible across standard deviation thresholds. The STRM (B) method consistently overestimated scar burden relative to the VslM. This was expected since every single slice would have yielded an enhanced proportion by the mathematical nature of the technique at the range of thresholds tested (the magnitude of the threshold is determined by the standard deviation of pixels within the ROI—therefore, for as long as there is a ROI, there will always be a standard deviation and unless an extremely high threshold is selected, there will always be a proportion of pixels in the rest of the slice that exceeds the chosen threshold). One argument is that we should include a myocardial ROI on every slice (including those without visible LGE) to be truly objective in quantifying LGE with the STRM method. The argument here is that the software is objectively identifying pixels with intermediate signal intensity that the eye of the CMR reader would otherwise ignore and is therefore offering a less biased estimate. Though this is only true if the studies that validated a given threshold against histology analysed their LGE in this manner (i.e., using the STRM B method). Unfortunately, histological validation studies of the various methods are already rare, and this vital information is rarely reported on.

In real-world practice, AS often coexists with CAD. Myocardial infarction from obstructive CAD manifests in the form of subendocardial LGE, and evidence suggests that infarct scar in AS carries a similarly elevated risk of mortality as non-infarct scar [11]. In our cohort, patients with significant CAD and prior myocardial infarction were excluded. Therefore, the impact of the infarct scar on the results of this study is unknown. This is important given that the infarct scar is often bright on LGE, in contrast with the relatively non-bright, heterogeneous signal-intensity scarring observed in AS. Previous studies, including the ReLate Study, which analysed one of the largest cohorts in the literature, showed that the FWHM technique was the most reproducible for the quantification of infarction-associated LGE [31,33]. The performance of manual thresholding, on the other

hand, was least reproducible [31,33]. This highlights the heterogeneity of myocardial scarring in cardiovascular disease and illustrates the complexity of recommending a single LGE quantification technique that uniformly accounts for this heterogeneity in clinical practice, where dual pathology may be encountered.

5. Limitations

A histological analysis was not performed in this study, therefore limiting the ability to test accuracy against what is considered the gold standard. It must be noted, however, that endomyocardial tissue is typically sampled from the basal septum. This represents one of seventeen LV myocardial segments, compared with CMR, which visualises the entire ventricle. This is an important consideration given strong evidence that a myocardial fibrosis gradient exists in AS. In our own cohort, septal predominance in fibrosis burden was also observed, in addition to the basal-to-apical gradient [34]. Therefore, while endomyocardial biopsy is considered the gold standard for fibrosis detection and quantification, it is not without relevant limitations, including the lack of complete representation of the LV.

An inter-reader intraclass correlation analysis based on myocardial regions and/or myocardial segments was not performed in this study. Given the observation of a fibrosis gradient in patients with severe AS, this analysis may offer further insights in future studies.

6. Conclusions

Scar quantification may serve as a useful parameter for risk stratification of high-risk AS patients who do not meet current guideline recommendations for valve intervention. However, no consensus exists on which quantification method is optimal for this population of patients. This study compared various scar quantification techniques in the context of severe AS. In addition to previously published studies on scar quantification in AS, it also included the FWHM and OAT methods, as well as an additional analysis on how post-processing styles may impact the accuracy and reproducibility of LGE quantification. We found that the visual method, when performed in the manner described in this article (i.e., the VslM approach), may be reasonably sufficient. Where a semi-automated method is preferred, the FWHM method utilising a single scar without ROI extrapolation appears to offer a reasonable balance in terms of accuracy and reproducibility. Importantly, the heterogeneous nature of myocardial scarring in AS may limit the use of the FWHM method in some cases, including those with small or non-bright scars. Larger studies designed with histological or clinical outcome-linked validation are still needed.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics16132017/s1>, Table S1: Wilcoxon rank-sum tests and Spearman rho correlation analyses; Table S2: Bland-Altman Analyses; Table S3: Wilcoxon rank-sum tests and Spearman correlation analyses for the B methods; Table S4: Intraclass correlation coefficient analyses for the B methods.

Author Contributions: M.R.R.: Conceptualisation, Methodology, Project Administration, Funding acquisition, Investigation, Formal analysis, Visualisation, Writing—original draft, review and editing, Visualisation. P.-P.S.R.: Methodology, Investigation, Validation, Writing—review and editing. V.S.: Investigation, Validation, Writing—review and editing. T.M.E.: Supervision, Methodology, Writing—review and editing. A.F.D.: Supervision, Conceptualisation, Methodology, Funding Acquisition, Resources, Writing—review and editing. P.G.H.: Supervision, Conceptualisation, Methodology, Funding Acquisition, Resources, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was enabled by funding from the SUNHEART centre of the Division of Cardiology, Department of Medicine at Tygerberg Hospital/Stellenbosch University and the Harry Crossley Foundation. The degree from which this study emanated was funded by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Clinician Researcher Programme. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the SAMRC.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and was approved by the Health Research Ethics Committee of Stellenbosch University (Reference: S21/11/251 PHD, approved on 24 February 2022).

Informed Consent Statement: Written informed consent, including that of publication, was obtained from each participant in the study.

Data Availability Statement: The dataset analysed for this publication is available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AS	aortic stenosis
SAVR	surgical aortic valve replacement
TAVI	transcatheter aortic valve implantation
CMR	cardiovascular magnetic resonance
LGE	late gadolinium enhancement
LV	left ventricle/left ventricular
VslM	visual slider method
SDs	standard deviations
STRM	signal threshold versus reference mean method
FWHM	full-width at half-maximum method
OAT	Otsu auto threshold method
CAD	coronary artery disease
TI	inversion time
TE	time to echo
TR	time to repeat
ROI	region of interest
ICC	intraclass correlation coefficient

References

1. Coffey, S.; Roberts-Thomson, R.; Brown, A.; Carapetis, J.; Chen, M.; Enriquez-Sarano, M.; Zühlke, L.; Prendergast, B.D. Global epidemiology of valvular heart disease. *Nat. Rev. Cardiol.* **2021**, *18*, 853–864. [\[CrossRef\]](#)
2. Vahanian, A.; Beyersdorf, F.; Praz, F.; Milojevic, M.; Baldus, S.; Bauersachs, J.; Capodanno, D.; Conradi, L.; De Bonis, M.; De Paulis, R.; et al. 2021 ESC/EACTS Guidelines for the management of valvular heart disease: Developed by the Task Force for the management of valvular heart disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). *Eur. Heart J.* **2022**, *43*, 561–632. [\[CrossRef\]](#)
3. Otto, C.M.; Nishimura, R.A.; Bonow, R.O.; Carabello, B.A.; Erwin, J.P.; Gentile, F.; Jneid, H.; Krieger, E.V.; Mack, M.; McLeod, C.; et al. 2020 ACC/AHA Guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines. *Circulation* **2021**, *143*, e27–e227.
4. Lee, G.; Chikwe, J.; Milojevic, M.; Wijeyesundera, H.C.; Biondi-Zoccai, G.; Flather, M.; Gaudino, M.F.L.; Fremes, S.E.; Tam, D.Y. ESC/EACTS vs. ACC/AHA guidelines for the management of severe aortic stenosis. *Eur. Heart J.* **2023**, *44*, 796–812. [\[CrossRef\]](#)
5. Bohbot, Y.; de Meester de Ravenstein, C.; Chadha, G.; Rusinaru, D.; Belkhir, K.; Trouillet, C.; Pasquet, A.; Marechaux, S.; Vanoverschelde, J.-L.; Tribouilloy, C. Relationship between left ventricular ejection fraction and mortality in asymptomatic and minimally symptomatic patients with severe aortic stenosis. *JACC Cardiovasc. Imaging* **2019**, *12*, 38–48. [\[CrossRef\]](#)

6. Jacquemyn, X.; Strom, J.B.; Strange, G.; Playford, D.; Stewart, S.; Kutty, S.; Bhatt, D.L.; Bleiziffer, S.; Grubb, K.J.; Pellikka, P.A.; et al. Moderate aortic valve stenosis is associated with increased mortality rate and lifetime loss: Systematic review and meta-analysis of reconstructed time-to-event data of 409,680 patients. *J. Am. Heart Assoc.* **2024**, *13*, e033872. [[CrossRef](#)] [[PubMed](#)]
7. Newton, N.; Liu, C.Y.; Croisille, P.; Bluemke, D.; Lima, J.A.C. Assessment of myocardial fibrosis with cardiac magnetic resonance. *J. Am. Coll. Cardiol.* **2011**, *57*, 891–903. [[CrossRef](#)] [[PubMed](#)]
8. Treibel, T.A.; López, B.; González, A.; Menacho, K.; Schofield, R.S.; Ravassa, S.; Fontana, M.; White, S.K.; DiSalvo, C.; Roberts, N.; et al. Reappraising myocardial fibrosis in severe aortic stenosis: An invasive and non-invasive study in 133 patients. *Eur. Heart J.* **2018**, *39*, 699–709. [[CrossRef](#)]
9. Weidemann, F.; Herrmann, S.; Störk, S.; Niemann, M.; Frantz, S.; Lange, V.; Beer, M.; Gattenlöhner, S.; Voelker, W.; Ertl, G.; et al. Impact of myocardial fibrosis in patients with symptomatic severe aortic stenosis. *Circulation* **2009**, *120*, 157–184. [[CrossRef](#)]
10. Dweck, M.R.; Joshi, S.; Murigu, T.; Alpendurada, F.; Jabbour, A.; Melina, G.; Banya, W.; Gulati, A.; Roussin, I.; Raza, S.; et al. Midwall fibrosis is an independent predictor of mortality in patients with aortic stenosis. *J. Am. Coll. Cardiol.* **2011**, *58*, 1271–1279. [[CrossRef](#)] [[PubMed](#)]
11. Musa, T.A.; Treibel, T.A.; Vassiliou, V.S.; Captur, G.; Singh, A.; Chin, C.; Dobson, L.E.; Pica, S.; Loudon, M.; Malley, T.; et al. Myocardial scar and mortality in severe aortic stenosis. *Circulation* **2018**, *138*, 1935–1947. [[CrossRef](#)]
12. Rajesh, G.N.; Thottian, J.J.; Subramaniam, G.; Desabandhu, V.; Sajeev, C.G.; Krishnan, M.N. Prevalence and prognostic significance of left ventricular myocardial late gadolinium enhancement in severe aortic stenosis. *Indian Heart J.* **2017**, *69*, 742–750. [[CrossRef](#)]
13. Azevedo, C.F.; Nigri, M.; Liguchi, M.L.; Pomerantz, P.M.; Spina, G.S.; Sampaio, R.O.; Tarasoutchi, F.; Grinberg, M.; Rochitte, C.E. Prognostic significance of myocardial fibrosis quantification by histopathology and magnetic resonance imaging in patients with severe aortic valve disease. *J. Am. Coll. Cardiol.* **2010**, *56*, 278–287. [[CrossRef](#)] [[PubMed](#)]
14. Papanastasiou, C.A.; Kokkinidis, D.G.; Kampaktis, P.N.; Bikakis, I.; Cunha, D.K.; Evangelos, K.; Greenwood, J.P.; Garcia, M.J.; Karamitsos, T.D. The prognostic role of late gadolinium enhancement in aortic stenosis: A systematic review and meta-analysis. *JACC Cardiovasc. Imaging* **2020**, *13*, 385–392. [[CrossRef](#)]
15. Balciunaite, G.; Skorniakov, V.; Rimkus, A.; Zaremba, T.; Palionis, D.; Valeviciene, N.; Aidietis, A.; Serpytis, P.; Rucinskas, K.; Sogaard, P.; et al. Prevalence and prognostic value of late gadolinium enhancement on CMR in aortic stenosis: Meta-analysis. *Eur. Radiol.* **2020**, *30*, 640–651. [[CrossRef](#)] [[PubMed](#)]
16. Thornton, G.D.; Vassiliou, V.S.; Musa, T.A.; Aziminia, N.; Craig, N.; Dattani, A.; Davies, R.H.; Captur, G.; Moon, J.C.; Dweck, M.R.; et al. Myocardial scar and remodelling predict long-term mortality in severe aortic stenosis beyond 10 years. *Eur. Heart J.* **2024**, *45*, 2019–2022. [[CrossRef](#)]
17. Calin, A.; Rosca, M.; Beladan, C.C.; Enache, R.; Mateescu, A.D.; Ginghina, C.; Popescu, B.A. The left ventricle in aortic stenosis—Imaging assessment and clinical implications. *Cardiovasc. Ultrasound* **2015**, *13*, 22. [[CrossRef](#)]
18. Calin, A.; Mateescu, A.D.; Popescu, A.C.; Bing, R.; Dweck, M.R.; Popescu, B.A. Role of advanced left ventricular imaging in adults with aortic stenosis. *Heart* **2020**, *106*, 962–969. [[CrossRef](#)]
19. Bing, R.; Cavalcante, J.L.; Everett, R.J.; Clavel, M.A.; Newby, D.E.; Dweck, M.R. Imaging and impact of myocardial fibrosis in aortic stenosis. *JACC Cardiovasc. Imaging* **2019**, *12*, 283–296. [[CrossRef](#)] [[PubMed](#)]
20. Bohbot, Y.; Renard, C.; Manrique, A.; Levy, F.; Maréchaux, S.; Gerber, B.L.; Tribouilloy, C. Usefulness of cardiac magnetic resonance imaging in aortic stenosis. *Circ. Cardiovasc. Imaging* **2020**, *13*, e010356. [[CrossRef](#)]
21. Ridgway, J.P. Cardiovascular magnetic resonance physics for clinicians: Part I. *J. Cardiovasc. Magn. Reson.* **2010**, *12*, 71. [[CrossRef](#)] [[PubMed](#)]
22. Biglands, J.D.; Radjenovic, A.; Ridgway, J.P. Cardiovascular magnetic resonance physics for clinicians: Part II. *J. Cardiovasc. Magn. Reson.* **2012**, *14*, 78. [[CrossRef](#)]
23. Kramer, C.M.; Barkhausen, J.; Flamm, S.D.; Kim, R.J.; Nagel, E. Standardized cardiovascular magnetic resonance (CMR) protocols 2013. *J. Cardiovasc. Magn. Reson.* **2013**, *15*, 91. [[CrossRef](#)]
24. Schulz-Menger, J.; Bleumke, D.A.; Bremerich, J.; Flamm, S.D.; Fogel, M.A.; Friedrich, M.G.; Kim, R.J.; von Knobelsdorff-Brenkenhoff, F.; Kramer, C.M.; Pennell, D.J.; et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance—2020 update: Society for Cardiovascular Magnetic Resonance (SCMR): Board of trustees task force on standardized post-processing. *J. Cardiovasc. Magn. Reson.* **2020**, *22*, 19. [[CrossRef](#)]
25. Rajah, M.R.; Doubell, A.F.; Herbst, P.G. Quantification of replacement fibrosis in aortic stenosis: A narrative review on the utility of cardiovascular magnetic resonance imaging. *Diagnostics* **2024**, *14*, 2435. [[CrossRef](#)] [[PubMed](#)]
26. Hoffmann, R.; Altiok, E.; Friedman, Z.; Becker, M.; Frick, M. Myocardial deformation imaging by two-dimensional speckle-tracking echocardiography in comparison to late gadolinium enhancement cardiac magnetic resonance for analysis of myocardial fibrosis in severe aortic stenosis. *Am. J. Cardiol.* **2014**, *114*, 1083–1088. [[CrossRef](#)]
27. Gräni, C.; Eichhorn, C.; Bière, L.; Kaneko, K.; Murthy, V.L.; Agarwal, V.; Aghayev, A.; Steigner, M.; Blankstein, R.; Jerosch-Herold, M.; et al. Comparison of myocardial fibrosis quantification methods by cardiovascular magnetic resonance imaging for risk stratification of patients with suspected myocarditis. *J. Cardiovasc. Magn. Reson.* **2019**, *21*, 14. [[CrossRef](#)] [[PubMed](#)]

28. Kwong, R.Y.; Farzaneh-Far, A. Measuring myocardial scar by CMR. *JACC Cardiovasc. Imaging* **2011**, *4*, 157–160. [[CrossRef](#)]
29. Spiewak, M.; Malek, L.A.; Misko, J.; Chojnowska, L.; Milosz, B.; Klopotoski, M.; Petryka, J.; Dabrowski, M.; Kepka, C.; Ruzyllo, W. Comparison of different quantification methods of late gadolinium enhancement in patients with hypertrophic cardiomyopathy. *Eur. J. Radiol.* **2010**, *74*, e149–e153. [[CrossRef](#)]
30. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)]
31. Flett, A.S.; Hasleton, J.; Cook, C.; Hausenloy, D.; Quarta, G.; Ariti, C.; Muthurangu, V.; Moon, J.C. Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. *JACC Cardiovasc. Imaging* **2011**, *4*, 150–156. [[CrossRef](#)]
32. Iles, L.M.; Ellims, A.H.; Llewellyn, H.; Hare, J.L.; Kaye, D.M.; McLean, C.A.; Taylor, A.J. Histological validation of cardiac magnetic resonance analysis of regional and diffuse interstitial myocardial fibrosis. *Eur. Heart J. Cardiovasc. Imaging* **2015**, *16*, 14–22. [[CrossRef](#)] [[PubMed](#)]
33. Gavrysh, J.; Reisdorf, P.; Hadler, T.; Mayr, T.; Ammann, C.; Gröschel, J.; Kuhnt, J.; von Knobelsdorff-Brenkenhoff, F.; Muehlberg, F.; Schwenke, C.; et al. Reproducibility of late gadolinium enhancement quantification techniques in ischemic and non-ischemic heart diseases (ReLate Study). *Front. Cardiovasc. Med.* **2025**, *12*, 1621292. [[CrossRef](#)] [[PubMed](#)]
34. Rajah, M.R.; Doubell, A.F.; Herbst, P.G. High afterload rather than myocardial fibrosis predicts reduced ejection fraction in severe aortic stenosis with afterload mismatch. *Open Heart* **2025**, *12*, e003345. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.