© 2011 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION PUBLISHED BY ELSEVIER INC. ISSN 1936-878X/\$36.00 DOI:10.1016/j.jcmg.2010.11.015

# Evaluation of Techniques for the Quantification of Myocardial Scar of Differing Etiology Using Cardiac Magnetic Resonance

Andrew S. Flett, MBBS,\*† Jonathan Hasleton, MB, CHB,‡ Christopher Cook, MBBS,\* Derek Hausenloy, MD, PHD,‡ Giovanni Quarta, MD,\*§ Cono Ariti, MSc,|| Vivek Muthurangu, MD(RES),¶ James C. Moon, MD\*† London, United Kingdom; and Rome, Italy

**OBJECTIVES** The aim of this study was to compare the reproducibility of 7 late gadolinium enhancement (LGE) quantification techniques across 3 conditions in which LGE is known to be important: acute myocardial infarction (AMI), chronic myocardial infarction (CMI), and hypertrophic cardiomyopathy (HCM).

**BACKGROUND** LGE by cardiac magnetic resonance is the gold-standard technique for assessing myocardial scar. No consensus exists on the best method for its quantification, and research in this area is scant. Techniques include manual quantification, thresholding by 2, 3, 4, 5, or 6 SDs above remote myocardium, and the full width at half maximum (FWHM) technique. To date, LGE has been linked to outcome in 3 conditions: AMI, CMI, and HCM.

**METHODS** Sixty patients with 3 LGE etiologies (AMI, n = 20; CMI, n = 20; HCM, n = 20) were scanned for LGE. LGE volume was quantified using the 7 techniques. Mean LGE volume, interobserver and intraobserver reproducibility, and impact on sample size were assessed.

**RESULTS** LGE volume varied significantly with the quantification method used. There was no statistically significant difference between LGE volume by the FWHM, manual, and 6-SD or 5-SD techniques. The 2-SD technique generated LGE volumes up to 2 times higher than the FWHM, 6-SD, and manual techniques. The reproducibility of all techniques was worse in HCM than AMI or CMI. The FWHM technique was the most reproducible in all 3 conditions compared with any other method (p < 0.001). Use of the FWHM technique for LGE quantification in paired analysis would lead to at least a 60% reduction in required sample size compared with any other method.

CONCLUSIONS Regardless of the disease under study, the FWHM technique for LGE quantification gives LGE volume mean results similar to manual quantification and is statistically the most reproducible, reducing required sample sizes by up to one-half. (J Am Coll Cardiol Img 2011;4:150–6) © 2011 by the American College of Cardiology Foundation

From the \*Department of Cardiology, The Heart Hospital, University College London Hospitals NHS Trust, London, United Kingdom; †Department of Medicine, University College London, London, United Kingdom; ‡The Hatter Cardiovascular Institute, University College London Hospitals, London, United Kingdom; \$Department of Cardiology, S. Andrea Hospital, University "La Sapienza," Rome, Italy; ||Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom; and the ¶Department of Imaging, Great Ormond Street Hospital for Children NHS Trust, London, United Kingdom. Dr. Flett holds a clinical research training fellowship with the British Heart Foundation (FS/08/028/24767). All other authors have reported that they have no relationships to disclose.

Manuscript received September 8, 2010; revised manuscript received October 26, 2010, accepted November 22, 2010.

Car as a result of acute myocardial infarction (AMI) and chronic myocardial infarction (CMI) has important prognostic implications (1-4). There is growing evidence suggesting that cardiac magnetic resonance (CMR) using the late gadolinium enhancement (LGE) technique can be considered the gold-standard modality for its assessment (5-8). LGE can also detect scar in cardiomyopathy (9-13) and in

#### See page 157

inflammatory (14,15) and infiltrative (16) conditions, and there is a growing prognostic evidence base (17). However, the optimal method for quantifying LGE remains unclear. All techniques rely on the fact that the LGE technique makes scar appear bright and as such can be defined as a signal intensity above normal myocardium, with 2 SDs being advocated by official guidelines (18). However other techniques are also used: 3, 4, 5, or 6 SDs; manual quantification (drawing regions of interest [ROIs] around scar); and the full width at half maximum (FWHM) technique, which uses half the maximal signal within the scar as the threshold. Each method results in different mean LGE volumes, confounding comparison of individual values with prognostic outcome papers. Moreover, in the research setting, where scar is used as a surrogate end point, the reproducibility of its quantification is a key determinant of sample size.

Our aim was to compare the 7 quantification techniques across the spectrum of disease in which LGE has been linked to outcome: AMI, CMI, and hypertrophic cardiomyopathy (HCM), assessing the LGE volumes obtained, the reproducibility of each method, and any associated effects on future study design. We hypothesized that the techniques would yield significantly different LGE volumes, that LGE quantification would vary with LGE etiology, and that the 7 techniques would have statistically different reproducibility.

## METHODS

All research was carried out at University College London Hospital NHS Trust between October 2008 and December 2009. An ethics committee of the U.K. National Research Ethics Service approved generic analysis of anonymized clinical scans. The 7 techniques (thresholding at 2, 3, 4, 5, and 6 SDs from remote myocardium; FWHM; and manual quantification) were compared in a retrospective manner in 60 consecutive patients referred for CMR with the confirmed clinical diagnosis of HCM (from a tertiary referral cardiomyopathy clinic, n = 20), AMI (n = 20), and CMI (n = 20) who had manifest LGE. Myocardial infarction was defined by clinical presentation compatible with an ST-segment elevation myocardial infarction, angiographic confirmation of coronary artery disease in the appropriate territory, and an elevated troponin level. In AMI, CMR was performed within 1 week, and in CMI, the scan was performed no less than 3 months from the AMI. No other inclusion or exclusion criteria were used. Mean LGE volume and intraobserver and interobserver variability were assessed, and implications for sample size calculations derived.

CMR protocol. Standard CMR examinations were performed in all patients (18) using a 1.5-T scanner (Avanto; Siemens Medical Imaging, Erlangen, Germany). For LGE imaging, 0.1 mmol/kg Dotarem (Guerbet, S.A., Villepinte, France) was administrated intravenously and standard breath-hold inversion recovery imaging performed (spoiled gradient-echo sequence, slice thickness 8 mm, repetition time 9.8 ms, echo time 4.6 ms, flip angle 21°, 21 lines per segment, spatial resolution  $1.4 \times 2.8 \times 8$  mm with phase swaps when appropriate, typical breath-hold 12 seconds). Inversion times are reported in Table 1 and were adjusted to optimally null normal myocardium.

Image analysis. Signal-to-noise ratio and

contrast-to-noise ratio are reported as assessments of image quality (8). Signal-to-noise ratio was calculated as mean signal intensity of enhanced area/SD of noise. Contrast-to-noise ratio was calculated as (mean signal intensity of enhanced area - mean signal intensity of unenhanced area)/1.5 imesSD air. Myocardial volume and mass analysis were carried out on cine images using standard techniques (18). Before further analysis, short-axis images were manually segmented for epicardial and endocardial borders (excluding papillary muscles) to obtain the myocardial volume. If more than 1 image of the same slice position was available because of phase swaps, the optimal image was selected for analysis. All further quantification occurred on the presegmented images, and LGE was quantified manually by tracing around its perimeter. Semiau-

#### ABBREVIATIONS AND ACRONYMS

AMI = acute myocardial infarction
<b>CMI</b> = chronic myocardial infarction
<b>CMR</b> = cardiac magnetic resonance
FWHM = full width at half maximum
HCM = hypertrophic cardiomyopathy
ICC = intraclass correlation coefficient
LGE = late gadolinium enhancement
<b>ROI</b> = region of interest

Table 1. Patient and CMR Characteristics by Disease				
	AMI	СМІ	нсм	
Variable	(n = 20)	(n = 20)	(n = 20)	
Age (yrs)	58 (48–61)	60 (52–66)	50 (37–57)	
Men/women	16/4	17/3	14/6	
EDV (ml)	$147\pm27$	$172\pm51$	$136\pm33$	
ESV (ml)	67 ± 19	$82\pm41$	$34\pm13$	
SV (ml)	$80\pm21$	$90\pm25$	$102\pm27$	
EF (%)	$55\pm10$	$54\pm13$	$76\pm9$	
Mass (g)	$175 \pm 35$	175 ± 46	$236\pm74$	
Mean TI range	359–384	355-405	333–411	
SNR	$22\pm10$	$31\pm21$	$27\pm14$	
CNR	$12\pm 6$	$16 \pm 11$	$12\pm7$	
Disease-specific characteristics				
Maximal WT (mm)	_	_	$19\pm4$	
MI to scan time (days)	$2.1\pm1.2$	$132\pm14.1$	_	
Peak troponin	$7.5\pm5.2$	$\textbf{7.0} \pm \textbf{4.7}$	_	
Infarct LAD:RCA:Cx	10:7:3	10:5:5	_	

AMI = acute myocardial infarction; CMI = chronic myocardial infarction; CNR = contrast-to-noise ratio; Cx = circumflex coronary artery; EDV = end-diastolic volume; EF = ejection fraction; ESV = end-systolic volume; HCM = hypertrophic cardiomyopathy; LAD = left anterior descending coronary artery; RCA = right coronary artery; SNR = signal-to-noise ratio; SV = stroke volume; TI = inversion time; WT = wallthickness.

> tomated analyses followed using purpose-written ImageJ (National Institutes of Health, Bethesda, MD) macros to determine FWHM and 6-, 5-, 4-, 3-, and 2-SD (above the mean remote myocardial signal) derived volumes (Fig. 1). This involved the manual delineation of 2 ROIs: first in remote myocardium (defined as a region with no enhancement and normal wall motion; the size of the ROI was dictated by the boundaries of this area and avoidance of the endocardial and epicardial surfaces), used to generate mean and SD for 6-, 5-, 4-, 3-, and 2-SD thresholds. Second, an ROI was drawn around hyperintense myocardium and used to define maximal signal for the FWHM threshold. Two manual corrections were then required for all automated ROIs: first, microvascular obstruction (defined as hypointensity within an hyperintense region in patients with infarctions) was manually adjusted to be included as LGE; this was straightforward because microvascular obstruction was bounded by infarction and the endocardial margin. Second, any obvious blood pool or pericardial partial voluming and artifact (which occurred only rarely) were manually removed from the ROI. Slices with no LGE were not analyzed using the semiautomatic techniques. LGE volume as a percent of myocardium was quantified per slice, twice by 2 readers (A.S.F. and J.H., analysis blinded with 1-month temporal separation between repeat anal-

yses) in all patients, with interobserver and intraobserver reproducibility and LGE volume assessed. Statistical analysis. LGE volume by each method was compared using a 1-way repeated measure analysis of variance with Bonferroni post hoc analysis and reproducibility by intraclass correlation coefficient (ICC) (19) and the significance of reproducibility differences using a Wilcoxon rank comparison of the squared differences (20). Bland-Altman (21) testing was performed to assess systematic offsets and the relation of observed differences to LGE extent. In addition, the reproducibility data were graphically displayed, showing the variability attributed to intraobserver and interobserver effects as a proportion of total variability for each method (1 - ICC) (22). The sample size required to detect a clinically important change in LGE (5% was used illustratively) was calculated for each method for both paired and unpaired data by estimating the total patient-level variability using standard variance components analysis (23). This modeling technique includes as variables both intraobserver and interobserver effects, which accounts for any observed systematic differences in the LGE.

## RESULTS

LGE extent. LGE volume varied substantially with the quantification method used (Figs. 1 and 2). Comparing all methods against the manual technique, there was no significant difference (p = 1.0)in HCM between the manual and 6-SD or FWHM techniques, in AMI between the manual and 6-SD techniques, or in CMI between the manual and 6-SD or 5-SD techniques. The 2-SD technique generated LGE volumes up to 2 times higher than the FWHM, 6-SD, and manual techniques (p < 0.001). The difference was most marked in HCM. The most extreme mean LGE differences in each disease were from 22% to 30% (AMI), 14% to 23% (CMI), and 11% to 29% (HCM).

Bland-Altman analysis. Before reproducibility testing, Bland-Altman analysis (see the Online Appendix for the 42 Bland-Altman plots performed) confirmed that there was minimal interobserver or intraobserver bias (offset) for all methods, except in 5 HCM assessments (2 SDs through 6 SDs), for which a small systematic offset appears to have arisen from the 2 observers' drawing their remote myocardium ROIs differently. Additionally, it was confirmed that observed differences in LGE were not related to the mean.



LGE reproducibility. Compared with any other technique, for both intraobserver and interobserver reliability, the FWHM technique was the most reproducible in all 3 conditions (p < 0.001). The reproducibility of all techniques was worse in HCM than in AMI and CMI. In CMI and HCM, manual quantification was the least reproducible technique. For HCM, the FWHM technique was the only statistically acceptable quantification method (ICC >0.7) (24) (Fig. 3).

Impact of reproducibility on sample size calculations. The enhanced reproducibility of the FWHM technique has a large impact on the sample size needed to demonstrate a change in the number of patients required for a trial with the same power. On the basis of a difference in paired comparisons of slices between 2 observers, the FWHM technique's interobserver reproducibility compared with the next best method reduces required sample size by 69% (CMI), 60% (AMI), and 91% (HCM) for equivalent power. On the basis of unpaired analysis, taking into account an estimate

of the total variability of each method on a patient basis, the calculated sample sizes using the FWHM technique compared with the next best method is reduced by 17% (AMI), 7% (CMI), and 56% (HCM). For example, the sample size required in patients with HCM for the FWHM technique to detect a 5% change in LGE is halved (from 126 to 56 patients). This contrasts with the worst technique (2 SDs), which would require 296 patients.

## DISCUSSION

As previously shown by other groups, these data confirm that in 3 conditions, 7 commonly used techniques for LGE quantification produce widely differing results for fibrosis quantification. Here, the officially recommended 2-SD technique can double the LGE volume compared with the manual, FWHM, and 6- or 5-SD techniques. In the absence of a gold standard, one might assume that the manual technique, repre-



senting thresholding by the human eye, could be used as the gold standard. However, we found that the manual technique is not reproducible, particularly in HCM, for which it is statistically unacceptable (ICC <0.7). In the absence of a gold standard with which the assessment of accuracy would be possible, this study provides a robust analysis of reproducibility. Comparing diseases, HCM LGE was less reproducible than LGE in infarction, and all techniques were statistically unacceptable except FWHM. Comparing techniques, the FWHM technique was most reproducible regardless of etiology. Any of the alternative LGE quantification techniques would more than double the size of a clinical trial.

Although it is the most reproducible, the FWHM technique does have potential limitations relating to its accuracy (25). The threshold of half the maximal of the volume of LGE necessarily assumes a bright infarct core, implying that homogenously gray infarcts or multiple patchy infarcts with separate islands of necrosis may be less accurately delineated. In our experience, however, cases such as this were rare, and when they did occur, all methods were more prone to errors. In HCM, LGE is frequently less well defined than in infarction, and as such, the delineation of myocardium with normal signal (for use in the thresholding techniques based on the mean and SD of signal intensity) was often difficult. The identification of an area of high signal, however, was more straightforward for the FWHM technique. This is reflected in our results showing impaired reproducibility of all techniques in HCM compared with infarcts, but with the FWHM technique relatively preserved.

Prior studies of LGE quantification are limited. The FWHM method was reported to be the most accurate in 1 animal infarct model using the gold standard of histology (26). Beek et al. (27) studied mean  $\pm$  2- to 8-SD techniques with FWHM and correlated infarct size in CMI with viability after revascularization. The most predictive technique was 6 SDs, but it was not statistically superior to any other method, despite a wide range of infarct sizes produced. These investigators made no reproducibility assessments. Despite studying similar techniques as we have (limited to HCM only) with comparable SD ratios for their quantification methods, our conclusions somewhat contradict those of Spiewak et al. (28). They too found that FWHM was the most reproducible method (in a substudy of 5 patients), but because quantification did not corStudy limitations. The biggest limitation of this study is that there is no gold standard for the assessment of LGE, so accuracy cannot be assessed. However, this is an issue that cannot be addressed. Using postmortem data in animal studies is inadequate because of several differences between human and animal infarction patterns, and autopsy study is confounded by nongated (or simulated gating) scanning methods. In addition, intraobserver and interobserver reproducibility but not interstudy reproducibility have been assessed. Assessments of interstudy reproducibility are rarely done but are likely to demonstrate lower ICCs than found here and consequently require increased sample sizes for clinical trials. We would hypothesize, however, that the FWHM method would retain the optimal reproducibility. Analysis here was limited to a single dose of gadolinium at 0.1 mmol/kg. Other groups have used 0.2 mmol/kg and a variety of time points after contrast to acquire images. There is the potential for these issues to affect reproducibility, although this would be an issue of contrast-to-noise and signal-to-noise ratios, which we believe is outside of the remit of a study based on post-processing techniques. All the LGE images used in this study were optimized, but our consecutive patient recruitment still reflects real-world scanning.

### CONCLUSIONS

The FWHM technique for LGE quantification is the most reproducible, regardless of underlying etiology, across the spectrum of cardiac disease in which LGE quantification is known to be important. The officially recommended technique of 2 SDs from the mean performed worst in our analysis. The enhanced reproducibility of the FWHM method allows for study sample size on the basis of scar as an end point to be reduced (halved in HCM) for equivalent power. In HCM, the only technique with satisfactory reproducibility is the FWHM technique.





Interobserver and intraobserver variability (calculated as 1 - intraclass correlation coefficient [ICC]) of the 7 techniques is greatest in HCM. The FWHM method performs most reproducibly regardless of etiology. Intraobserver variability is less marked than interobserver variability, as would be expected. Abbreviations as in Figures 1 and 2.

**Reprint requests and correspondence:** Dr. James C. Moon, The Heart Hospital, 16–18 Westmoreland Street, London W1G 8PH, United Kingdom. *E-mail: james.moon@ uclb.nbs.uk.* 

#### REFERENCES

- Wu E, Ortiz JT, Tejedor P, et al. Infarct size by contrast enhanced cardiac magnetic resonance is a stronger predictor of outcomes than left ventricular ejection fraction or endsystolic volume index: prospective cohort study. Heart 2008;94:730-6.
- Larose E, Rodés-Cabau J, Pibarot P, et al. Predicting late myocardial recovery and outcomes in the early hours of ST-segment elevation myocardial infarction traditional measures compared with microvascular obstruction, salvaged myocardium, and necrosis characteristics by cardiovascular magnetic resonance. J Am Coll Cardiol 2010;55:2459-69.
- Choi KM, Kim RJ, Gubernikoff G, Vargas JD, Parker M, Judd RM. Transmural extent of acute myocardial infarction predicts long-term improvement in contractile function. Circulation 2001;104:1101–7.
- Krittayaphong R, Laksanabunsong P, Maneesai A, Saiviroonporn P, Udompunturak S, Chaithiraphan V. Comparison of cardiovascular magnetic resonance of late gadolinium enhancement and diastolic wall thickness to predict recovery of left ventricular function after coronary artery bypass surgery. J Cardiovasc Magn Reson 2008;10:41.
- Hendel RC, Patel MR, Kramer CM, Poon M. ACCF/ACR/SCCT/ SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging. J Am Coll Cardiol 2006;48:1475–97.
- 6. Wu E, Judd RM, Vargas JD, Klocke FJ, Bonow RO, Kim RJ. Visualisation of presence, location, and transmural extent of healed Q-wave and non-Qwave myocardial infarction. Lancet 2000;357:21–8.
- Wagner A, Mahrholdt H, Holly TA, et al. Contrast-enhanced MRI and routine single photon emission computed tomography (SPECT) perfusion imaging for detection of subendocardial myocardial infarcts: an imaging study. Lancet 2003;1:359–60.
- Simonetti OP, Kim RJ, Fieno DS, et al. An improved MR imaging technique for the visualization of myocardial infarction. Radiology 2001;218: 215–23.

- Choudhury L, Mahrholdt H, Wagner A, et al. Myocardial scarring in asymptomatic or mildly symptomatic patients with hypertrophic cardiomyopathy. J Am Coll Cardiol 2002;40:2156–64.
- Keller DI, Osswald S, Bremerich J, et al. Arrhythmogenic right ventricular cardiomyopathy: diagnostic and prognostic value of the cardiac MRI in relation to arrhythmia-free survival. Int J Cardiovasc Imaging 2003;19:537–43.
- Moon JC, McKenna WJ, McCrohon JA, Elliott PM, Smith GC, Pennell DJ. Toward clinical risk assessment in hypertrophic cardiomyopathy with gadolinium cardiovascular magnetic resonance. J Am Coll Cardiol 2003;41:1561–7.
- Dumont C, Monserrat L, Soler R, et al. Clinical significance of late gadolinium enhancement on cardiac magnetic resonance in patients with hypertrophic cardiomyopathy. Rev Esp Cardiol 2007;60:15–23.
- Adabag SA, Maron B, Appelbaum E, et al. Occurrence and frequency of arrhythmias in hypertrophic cardiomyopathy in relation to delayed enhancement on cardiovascular magnetic resonance. J Am Coll Cardiol 2008;51:1369–74.
- Smedema JP, Snoep G, van Kroonenburgh MP, et al. Evaluation of the accuracy of gadolinium-enhanced cardiovascular magnetic resonance in the diagnosis of cardiac sarcoidosis. J Am Coll Cardiol 2005;45:1683–90.
- Wagner A, Schulz-Menger J, Dietz R, Friedrich MG. Long-term follow-up of patients with acute myocarditis by magnetic resonance imaging. MAGMA 2003;16:17–20.
- Maceira AM, Joshi J, Prasad SK, et al. Cardiovascular magnetic resonance in cardiac amyloidosis. Circulation 2005; 111:186–93.
- Flett AS, Westwood MA, Davies LC, Mathur A, Moon JC. The prognostic implications of cardiovascular magnetic resonance. Circ Cardiovasc Imaging 2009;2:243–50.
- Kramer C, Barkhausen J, Flamm SD, Kim R, Nagel E. Standardized cardiovascular magnetic resonance imaging (CMR) protocols, Society for Cardiovascular Magnetic Resonance: Board of Trustees Task Force on Standardized Protocols. J Cardiovasc Magn Reson 2008;10:35.
- McGraw K, Wong S. Forming inferences about some intraclass correlation coefficients. Psychol Meth 1996;1:30–46.

- Moon JC, Lorenz CH, Francis JM, Smith GC, Pennell DJ. Breath-hold FLASH and FISP cardiovascular MR imaging: left ventricular volume differences and reproducibility. Radiology 2002;223:789–97.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
- 22. Bellenger N, Davies LC, Francis JM, Coats AJ, Pennell DJ. Reduction in sample size for studies of remodeling in heart failure by the use of cardiovascular magnetic resonance. J Cardiovasc Magn Reson 2000;4:271–8.
- 23. Rabe-Hesketh S, Skrondal A. Multi-Level and Longitudinal Modeling Using Stata. College Station, TX: Stata Press, 2008.
- Vincent WJ. Statistics in Kinesiology. 2nd ed. Champaign, IL: Human Kinetics, 1999.
- 25. Kim HW, Farzaneh-Far A, Kim RJ. Cardiovascular magnetic resonance in patients with myocardial infarction current and emerging applications. J Am Coll Cardiol 2010;55:1–16.
- 26. Amado LC, Gerber B, Gupta SN, et al. Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model. J Am Coll Cardiol 2004;44:2383–9.
- 27. Beek AM, Bondarenko O, Afsharzada F, van Rossum AC. Quantification of late gadolinium enhanced CMR in viability assessment in chronic ischemic heart disease: a comparison to functional outcome. J Cardiovasc Magn Reson 2009;11:6.
- Spiewak M, Malek LA, Misko J, et al. Comparison of different quantification methods of late gadolinium enhancement in patients with hypertrophic cardiomyopathy. Eur J Radiol 2010;74:e149–53.

Key Words: cardiac magnetic resonance • cardiomyopathy • imaging • myocardial infarction • SVCs.

#### ► A P P E N D I X

For the 42 Bland-Altman plots performed, please see the online version of this article.