

VENTRICULAR FUNCTION

Quantitative assessment of regional myocardial function with MR-tagging in a multi-center study: interobserver and intraobserver agreement of fast strain analysis with Harmonic Phase (HARP) MRI

ERNESTO CASTILLO, M.D.,¹ NAEL F. OSMAN, PH.D.,¹ BOAZ D. ROSEN, M.D.,² IMAN EL-SHEHABY, M.SC.,¹ LI PAN, M.SC.,³ MICHAEL JEROSCH-HEROLD, PH.D.,⁵ SHENGHAN LAI, M.D., PH.D.,⁴ DAVID A. BLUEMKE, M.D., PH.D.,¹ and JOÃO A. C. LIMA, M.D.^{2,*}

¹The Russell H. Morgan Department of Radiology and Radiological Sciences, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

²Department of Medicine, Division of Cardiology, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

³Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

⁴Department of Radiology, University of Minnesota, Minneapolis, Minnesota, USA

⁵Department of Epidemiology, The Johns Hopkins University Bloomberg School of Hygiene and Public Health, Baltimore, Maryland, USA

Purpose. To assess the reproducibility of Harmonic Phase (HARP) analysis of myocardial MR tagged images acquired in the Multi-Center Study of Atherosclerosis (MESA). **Methods.** Using the HARP method, three independent observers performed two separate quantitative strain analyses of myocardial cine MR-tagging images blindly in 24 participants. The images were obtained in four different centers and analyzed at a single core lab. Each study comprised 3 short-axis slices subdivided in 12 segments ($24 \times 3 \times 12 = 864$ segments), each with three layers. Normal strains (circumferential [E_{cc}] and radial [E_{rr}]), principal strains (Λ_{1} , Λ_{2}), and the angle α (between E_{cc} - Λ_{2}) were calculated. Intraclass correlation coefficient (R) for peak systolic strains, and all pooled systolic and diastolic strain data were used to determine inter- and intraobserver agreement. Two observers also visually graded study quality. R values were related to the image quality in different myocardial regions and layers. **Results.** Overall, HARP yielded an excellent inter- and intraobserver agreement for peak systolic strain data (for E_{cc} , $R = 0.84$ and 0.89 , respectively) and all systolic pooled data (for E_{cc} , interobserver $R = 0.82$, intraobserver $R = 0.69$ – 0.76). Both inter and intraobserver agreement were lower for diastolic pooled data ($R = 0.69$ and 0.58 – 0.62 , respectively). There was a direct relationship between image quality and performance of the HARP analysis, with increasing inter- and intraobserver R values in studies with longer tag persistence. Both inter- and intraobserver agreement were better in the anterior and septal myocardial regions, and in the midwall layer. The intraobserver agreement was similar among the three observers. **Conclusion.** Employing the HARP method for quantitative strain analysis of myocardial MR tagged images provides a high inter- and intraobserver agreement. These good results are obtained in case of good to excellent MR image quality.

Key Words: Atherosclerosis; Magnetic resonance (MR); Myocardium; MR; Statistical analysis; Technology assessment

1. Introduction

Magnetic resonance (MR) tagging analyzed by dedicated tracking algorithms allows very precise measurements of myocardial motion and characterization of regional myocardial function (1). The majority of the methods developed previously to evaluate 3D myocardial deformation deter-

mined by myocardial strains, required time-consuming, semi-automated segmentation and tag detection requiring expert readers and multiple hours of analysis for each patient. Thus, the application of MR tagging in clinical practice has been limited despite the fact that resulting strain values could be used as sensitive indicators of regional myocardial dysfunction (2–4).

The Harmonic-Phase (HARP) method has the potential of allowing MR-tagging techniques to be useful as a clinical diagnostic tool, as it enables rapid tracking of cardiac motion with minimal manual intervention (5). In healthy subjects and in patients with coronary heart disease (CAD), HARP provides faster (up to 10 fold) and more accurate measurements

Received 8 February 2005; accepted 3 July 2005.

*Address correspondence to João A. C. Lima, M.D., The Johns Hopkins Hospital, Blalock 524D, 600 N. Wolfe St., Baltimore, MD 21287, USA; Fax: (410) 614-8222; E-mail: jlima@jhmi.edu

of myocardial strains from MR-tagged images compared to a conventional tracking technique (6). This rapid analysis has facilitated the use of MR-tagging techniques for the first time in a large-scale multicenter study (Multi-Ethnic Study of Atherosclerosis, MESA).

MESA is a prospective epidemiological study that investigates the prevalence and progression of subclinical cardiovascular disease (CVD) over a period of 7 years in a population-based sample of 6,814 men and women aged 45–84 years from different ethnic groups (White, African-American, Hispanic and Chinese) (7). The purpose of this study was to assess the inter- and intraobserver agreement for myocardial MR-tagged image analysis using the HARP technique performed in an ancillary study of myocardial MR tagging.

2. Materials and methods

The Institutional Review Board of all the participating centers approved the study, and informed consent was obtained for all participants prior to the MR exam.

2.1. Patient selection

The characteristics of subjects enrolled in MESA have been described elsewhere (7). In short, the goal of the MESA study is to investigate the mechanisms associated with the development and progression of subclinical cardiovascular disease. Thus, individuals with known cardiovascular disease were excluded. Cardiac MRI was performed on all the volunteers at enrollment. In this ancillary study, left ventricular (LV) myocardial MR-tagged cine images from 24 participants in the MESA study (10 female, 14 male; mean age \pm SD: 64 ± 8.7 years; range, 47–78 years) were randomly selected. This selection from the MESA-tagging database comprising a total of 441 consecutive exams acquired between September 2001 and October 2002 was performed by a radiologist (E. C.), who was blinded to any clinical or epidemiological data. The study included an equal number of participants ($n = 6$) from each of four centers (Columbia University, New York City, NY; Johns Hopkins University, Baltimore, MD; University of Minnesota, Minneapolis, MN; and UCLA, Los Angeles, CA).

2.2. MR imaging

The LV myocardial MR-tagged images were obtained with 1.5 T MR systems (Signa LX and CV/i, GE Medical Systems, Milwaukee, WI; Somatom Vision and Sonata, Siemens Medical Solutions, Malvern, PA) and dedicated phased-array coils. Details of the protocols among the different centers are shown on Table 1. In all participants, three double-oblique short-axis scan planes (8–10 mm thickness; gap of 5–10 mm) were obtained around the midlevel of the LV myocardium. A retrospectively ECG-gated, segmented k-space fast gradient-echo (GRE) or fast low-angle shot (FLASH) pulse sequence (repetition time (TR)/echo time (TE): 3.5–7.2 ms/2–4.2 ms) with a flip angle of 12° was used. The tagging pulse consisted of 5 (Siemens scanners) or 7 (GE scanners) nonselective radio-frequency pulses separated by spatial modulation of magnetization (SPAMM) (8) encoding gradients to achieve a parallel stripe tag pattern with a tag spacing of 5 pixels (or 7.81 mm; GE scanners) or 7 mm (Siemens scanners). The phase and frequency directions were swapped together with the tag orientation in separate breath-holds in order to make the tag orientation (0 and 90 degrees) always perpendicular to the readout direction in all cases. Image acquisition was performed with a sequential-interleaved (GE scanners) and a center-out (Siemens scanners) phase-encode order. Image reconstruction using a linear segment interpolation technique, also referred as view-sharing, was used only on the Siemens scan protocols. The matrix size in the phase-(encoding direction varied among the centers between 96–140 pixels with a three-quarter rectangular FOV (GE scanners) or full Field of View (FOV) (Siemens scanners), while all the centers used 256 pixels in the frequency-encoding direction. The views per segment (VPS) varied from 4 to 9, resulting in 19–27 phases during a breath-hold of 12–18 seconds. Therefore, the achieved temporal resolution ranged between ~ 20 –41 ms.

2.3. Image postprocessing and analysis

The acquired short-axis tagged MR-images allow tracking of myocardial strains in 2-dimensions (1). Myocardial strains are a measurement of local tissue deformation and represent myocardial regional contractile function. Strains express the fractional change in length (as percentage) from a resting state

Table 1. Pulse sequence protocol among the four different scanning centers

	TR (ms)	TE (ms)	FA ($^\circ$)	BW (kHz)	Matrix (pixels) (Freq. \times Phase)	Phase FOV	Tag spacing (mm)	VPS	Time frames	BH (s)	TRes. (ms).
A	7.2	4.2	12	62.5	256 \times 128	0.75	7.81	4–6	20	15	41
B	4.9	2.1	12	62.5	256 \times 96	0.75	7.81	4	20–26	12	20
C	3.5	2	12	49.02	256 \times 128	1	7	9	19–28	14	32
D	7	4	12	24.9	256 \times 140	1	7	7	19	18	35

TR = repetition time, TE = echo time, FA = flip angle, BW = bandwidth, PhaseFOV = rectangular field-of-view along the phase direction, VPS = views per segment, TRes = temporal resolution, BH = breath-hold duration.

(end-diastole) to one achieved following myocardial contraction (1). An scheme of the two main reference systems to calculate myocardial strains is shown in Fig. 1. Three independent readers (E.C., B.D.R., I.E.) performed the myocardial strain analysis of all image data sets within 4 weeks. Two readers were board-certified in cardiology and radiology. The third reader was an engineer familiar with the principles of myocardial MR-tagging but had little experience with quantitative analysis. The readers were blinded to clinical or epidemiological information of the participants. After this first evaluation, a second analysis of the same studies was performed in an identical fashion one week later. Each observer freely chose the presentation order and amount of data sets analyzed at each session.

The myocardial MR-Tagging images were transferred to a personal computer (Optiplex GX400, Dell Computers, Austin, TX) for postprocessing. The image analysis and strain calculations were performed with a dedicated in-house developed, interactive software tool using the HARP method (5) and coded to run on Matlab (MathWorks, Natick, MA), which is shown in Fig. 2. After importing the images, the short-axis images with both tag directions were superimposed. Endocardial and epicardial contours were manually traced on the image corresponding to a single cardiac phase, usually in late or end-systole. HARP then automatically segmented the LV-myocardium in 12 equally sized regions, each with three layers (subepicardium, midwall and sub-endocardium). This was visualized as a circular grid and also tracked automatically along the remaining cardiac phases within a few seconds. A few interactive corrections of the

contour tracking were performed when necessary for obtaining satisfactory matching. The anterior attachment of the right ventricular wall to the LV was always chosen as the landmark of reference for clockwise numbering of the segments. Circumferential shortening (E_{cc}) for every layer in each segment was computed. Four additional strain parameters (radial thickening, E_{rr} ; maximal elongation, Λ_1 ; maximal shortening, Λ_2 ; and the angle α between the direction of E_{cc} and that of Λ_2 , hereafter ‘angle’) were also calculated. The myocardial strain analysis provided from each study a total of 108 plots (12 segments \times 3 layers \times 3 slices) for each type of strain. Thus, a total of 2,592 plots were obtained from the analysis of the 24 data sets.

In addition, two of the readers (E.C., B.D.R.) visually graded the study image quality. The image quality of the tags was visually assessed by consensus based on the tag persistence through the cardiac cycle. The criteria used for image quality were as follows: tag persistence on less than 50% of the frames during the cardiac cycle was considered poor, tag persistence on 50–70% of the frames was considered fair, tag persistence on 71–90% of the frames was considered good and a tag persistence on 91–100% of the images was considered very good quality.

2.4. Statistical data analysis

Intraclass correlation coefficient (R) was calculated to evaluate inter- and intraobserver agreement (this term is interchangeable with reliability or reproducibility) (10–12). R quantifies the degree of agreement between measurements

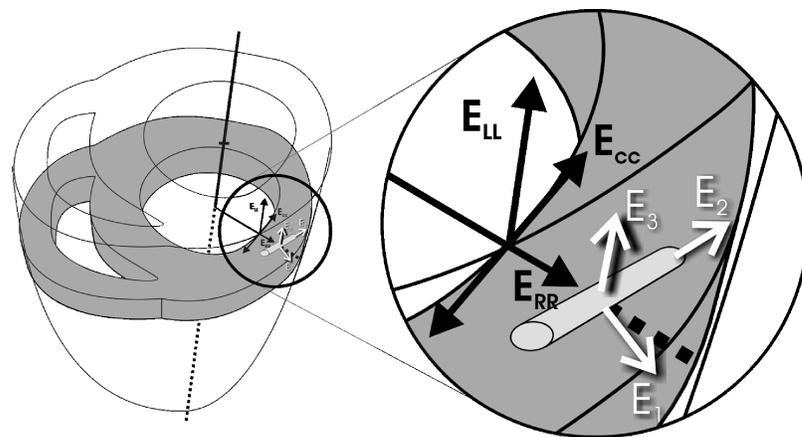


Figure 1. Scheme of coordinate systems for measuring myocardial strain defined by the finite strain tensor E . *Normal* strains (black arrows) are defined in relation to the circumferential, or short-axis, plane: Circumferential shortening (E_{cc}) occurs parallel to the tangent of the myocardium with respect to the epicardial surface (shown here as the endocardial surface for space reasons); radial thickening (E_{rr}) occurs perpendicular to the circumferential direction, toward the ventricular centroid; and longitudinal shortening (E_{ll}) occurs perpendicular to the other two components and parallel to the longitudinal axis of the left ventricle. *Principal* strains (white arrows) are defined in relation to the direction of movement in the main myocyte fiber bundles during systolic deformation. The maximal principal strain is the greatest elongation (Λ_1 , or E_1) orthogonal to the fiber direction. The minimal principal strain is the greatest shortening (Λ_2 , or E_2) parallel to the fiber direction. Principal strains are referred to the major and minor axes of an ellipse resulting from the deformation of a circle during systole because of wall shear. Strain that occurs perpendicular to these two principal strains is labeled E_3 . The angles between E_{cc} - E_2 and E_{rr} - E_1 are defined as α and β , respectively. [Reprinted with permission from Ref. (9)].

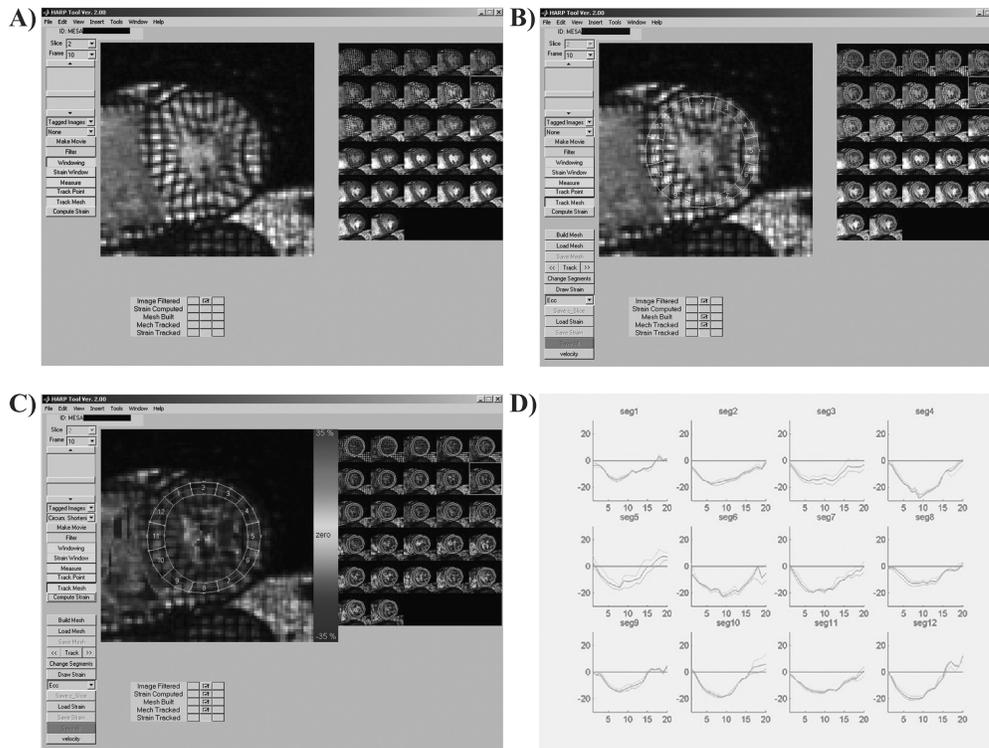


Figure 2. HARP analysis of myocardial MR-tagged images. (A) Images after importing and superimposing the 2D stripe tags in vertical and horizontal direction. (B) Endocardial and epicardial contours are manually traced on the image corresponding to a single cardiac phase. HARP then automatically segments the LV-myocardium in 12 equally sized regions, each with three layers (subepicardium = blue, midwall = red and subendocardium = green). This is visualized as a circular grid and also tracked automatically along the remaining cardiac phases within a few seconds. A few interactive corrections of the contour tracking were performed when necessary for satisfactory matching. The anterior attachment of the right ventricular wall to the LV is chosen as the landmark of reference for clockwise numbering of the segments. (C) HARP computes the strains (here shown circumferential shortening, *Ecc*) for every layer in each segment. The resulting *Ecc* curves with identical color-coding as in (B) are displayed in (D).

obtained on the same subject by several observers and is expressed as a number between 0 and 1. In accordance with other authors (13, 14), inter- and intraobserver agreement was considered as poor when R was < 0.4 ; fair, $R = 0.40$ – 0.59 ; good, $R = 0.60$ – 0.74 ; and excellent $R \geq 0.75$. R was determined for all pooled data points of the strain plots before and after sorting the studies according to their image quality. The pooled data were subsequently divided into systolic and diastolic phases and analyzed separately. The end-systolic phase was chosen visually by consensus in every participant using the criterion of the smallest ventricular cavity volume. Peak strain values of the strain curves were used for the calculation of R as well.

R was also determined for four separate LV myocardial wall regions of each slice determined as follows: anterior wall (segments 1–3), lateral wall (segments 4–6), inferior wall (segments 7–9), and septum (segments 10–12). For each region, there were 3 layers. For each layer, there were unequal numbers of data points due to the different number of time frames among the studies. An unbalanced repeated measures analysis-of-variance (ANOVA) was used to determine R and its statistical significance. Denote Y_{ijklmn} as the observed value of the n th repeated measurement (of the m th frame, of the l th

layer, of the k th segment, of the j th slice on the i th individual. According to a linear model, $Y_{ijklmn} = \mu + t_n + r_m + e_{ijklmn}$, $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$; $l = 1, \dots, L$; $m = 1, \dots, m_{ijk}$; $n = 1, 2, \dots, n$, where μ is the grand mean of the error-free measurements (the “true” value) Y in the population of interest; t_n is the effect of the n th observed measurements, which is assumed to be normally distributed with mean 0 and variance σ_t^2 ; r_m reflects the effect of the m th frame, which is assumed to be normally distributed with mean 0 and variance σ_r^2 ; and e_{ijklmn} represents the random error associated with rating, which is also assumed to be normally distributed with mean 0 and variance σ_e^2 . All random variables $\{t_n, r_m, e_{ijklmn}\}$ are assumed to be mutually independent. A p value less than .05 was considered to indicate a statistically significant difference. SAS statistical software was used to perform the analyses (version 8.2, SAS Institute, Cary, NC).

3. Results

From each MR tagging dataset, 108 curves (3 slices \times 12 segments \times 3 layers) of every type of strain were obtained, each with (mean \pm SD) 21.4 ± 2.5 (range: 19–27) data points

Table 2. Interobserver and intraobserver variability for all peak strain values (n = 2,592) related to the tag persistence

Tag persistence					
Strain	Very good		Good	Fair	
Interobserver <i>R</i> *					p value
Ecc	0.84 (0.83, 0.86)		0.8 (0.78, 0.82)	0.74 (0.7, 0.78)	.0269
Err	0.71 (0.67, 0.75)		0.69 (0.67, 0.71)	0.57 (0.52, 0.62)	.048**
L1	0.72 (0.68, 0.75)		0.73 (0.72, 0.73)	0.59 (0.56, 0.61)	.041**
L2	0.85 (0.85, 0.86)		0.81 (0.79, 0.84)	0.75 (0.71, 0.78)	.246
Angle α	0.66 (0.64, 0.68)		0.76 (0.73, 0.79)	0.64 (0.61, 0.67)	.207
Intraobserver <i>R</i> *					p value
Ecc	0.89 (0.87, 0.89)		0.81/0.84 (0.8, 0.9)	0.79/0.82 (0.72, 0.84)	.111
Err	0.77 (0.77, 0.78)		0.74 (0.74, 0.76)	0.69/0.67 (0.65, 0.76)	.095
L1	0.78 (0.75, 0.8)		0.79 (0.75, 0.84)	0.7 (0.63, 0.74)	.078
L2	0.88 (0.88, 0.89)		0.84 (0.81, 0.89)	0.81 (0.73, 0.84)	.045**
Angle α	0.74 (0.73, 0.74)		0.8 (0.76, 0.83)	0.77/0.79 (0.68, 0.82)	.336

Ecc, circumferential shortening; Err, radial thickening; L1, lambda 1 or maximal elongation; L2, lambda2 or maximal shortening.
 **R* = intraclass correlation coefficient for peak strain data; expressed as mean/median of *R*, respectively unless very close or identical. An *R* was considered poor when < 0.4; fair, *R* = 0.40–0.59; good, *R* = 0.60–0.74; and excellent, *R* ≥ 0.75. Numbers in parentheses are 5 and 95 percentiles, respectively.
 **Statistical significant inter- or intraobserver differences between observers for that strain measurement (p < 0.05).

(observations) over time. Each curve contained 10.7 ± 1.7 (range: 7–13) systolic data points and 10.7 ± 2.3 (range: 8–17) diastolic data points. Thus, from each participant’s data set 2,308 ± 273.6 (range: 2,052–2,808) observations for each type of strain were obtained; 1,156.5 ± 180.4 (range: 756–1404) during systole and 1,152 ± 246.8 (range: 864–1836) during diastole. As a result, the analysis output of the total 72 slices yielded 2,592 curves consisting of 55,404 observations (27,756 systolic, 27,648 diastolic) from each of the 5 types of strain. These pooled values and the peak measurements of each curve were compared with the results obtained by each

reader in both readings. The average analysis time per study was about 9 minutes, with a range of 7–11 minutes.

3.1. Inter- and intraobserver agreement of peak strain values

Inter- and intraobserver results for the peak values of the five strain parameters (Ecc, Err, Lambda 1, Lambda 2, and angle α) related to their image quality are listed in Table 2. Overall, the intraobserver *R* values were higher than the interobserver *R* values for all strains and all image quality groups. Among

Table 3. Interobserver and intraobserver variability for all pooled Ecc data (n = 27,756 systolic, 27,648 diastolic) in the different myocardial regions related to the tag persistence

Region	Tag persistence								S	D		
	Very good		Good		Fair		S	D				
	S	D	S	D	S	D						
Interobserver <i>R</i> *											p value	
Anterior	0.75 (0.74, 0.75)	0.65 (0.64, 0.65)	0.74 (0.73, 0.75)	0.62 (0.61, 0.62)	0.57 (0.54, 0.61)	0.54 (0.51, 0.56)	.023	.081†				
Lateral	0.66 (0.64, 0.68)	0.54 (0.53, 0.54)	0.63 (0.61, 0.64)	0.48 (0.46, 0.48)	0.46 (0.44, 0.47)	0.32 (0.31, 0.33)	.034	.009				
Inferior	0.69 (0.68, 0.69)	0.55 (0.54, 0.56)	0.62 (0.61, 0.63)	0.5 (0.49, 0.50)	0.56 (0.55, 0.56)	0.4 (0.38, 0.41)	.017	.013				
Septum	0.73 (0.71, 0.74)	0.62 (0.61, 0.63)	0.76 (0.75, 0.76)	0.64 (0.64, 0.64)	0.75 (0.74, 0.77)	0.53 (0.52, 0.53)	.356†	.01				
Intraobserver <i>R</i> *											p value	
Anterior	0.79 (0.72, 0.83)	0.66 (0.66, 0.66)	0.77 (0.71, 0.77)	0.63 (0.59, 0.65)	0.58 (0.54, 0.65)	0.55 (0.51, 0.57)	.002	.016				
Lateral	0.69 (0.64, 0.73)	0.56 (0.50, 0.64)	0.69 (0.62, 0.71)	0.48 (0.46, 0.5)	0.46 (0.44, 0.49)	0.33 (0.3, 0.34)	.004	.007				
Inferior	0.70 (0.67, 0.71)	0.56 (0.53, 0.57)	0.64 (0.59, 0.65)	0.48 (0.46, 0.49)	0.62 (0.6, 0.64)	0.41 (0.38, 0.42)	.002	.002				
Septum	0.73 (0.7, 0.75)	0.62 (0.6, 0.64)	0.77 (0.71, 0.79)	0.64 (0.62, 0.65)	0.75 (0.74, 0.78)	0.53 (0.51, 0.57)	.227†	.006				

S, systole; D, diastole.
 **R* = intraclass correlation coefficient; mean and median of *R* were identical. An *R* was considered poor when < 0.4; fair, *R* = 0.40–0.59; good, *R* = 0.60–0.74; and excellent, *R* ≥ 0.75. Numbers in parentheses are 5 and 95 percentiles, respectively.
 †No statistical significant inter- or intraobserver differences between observers were observed in the corresponding myocardial regions (p > 0.05).

all strain parameters, Ecc and Lambda 2 provided the highest results of the inter- and intraobserver R related to the tag persistence. These were very similar and considered as 'excellent' in the groups with good and very good tag persistence (Ecc: interobserver $R = 0.8$ and 0.84 , intraobserver $R = 0.84$ and 0.89 ; Lambda 2: interobserver $R = 0.81$ and 0.85 , intraobserver $R = 0.84$ and 0.88 , respectively). In the group with fair tag persistence, R was either 'good' or 'excellent' (Ecc: interobserver $R = 0.74$, intraobserver $R = 0.82$; Lambda 2: interobserver $R = 0.75$, intraobserver $R = 0.81$, respectively). Moreover, the results of R showed no statistical significance ($p > .05$) between the different image quality groups for interobserver R of Ecc, lambda 2, and angle. Intraobserver R yielded also no statistically significant differences among the different image quality groups for Ecc and angle.

For Err and Lambda 1, the interobserver R were 'good' or 'fair' depending on the tag persistence, with a statistically significant difference ($p < .05$) between the different image quality groups. The intraobserver R 's for Err and lambda 1 were higher, and even considered as 'excellent' in the groups with good and very good image quality, while the differences among these groups were not statistically significant ($p > .05$).

3.2. Inter- and intraobserver agreement among different myocardial regions

Both inter- and intraobserver R were higher in the anterior and septal regions, compared to the lateral and inferior regions (Table 3). R values ranged between 'good' and 'excellent' for good and very good image qualities, and again intraobserver and systolic values were higher than interobserver and diastolic values, respectively. The differences between the R values for the different image quality groups were statistically significant ($p < .05$) for all regions except for the systolic

data in the septal region and the diastolic data in the anterior region.

3.3. Inter- and intraobserver reproducibility among different myocardial layers

Among the three layers within the myocardial wall, the highest R values were achieved in the midwall layer (Table 4). Interobserver and intraobserver R values were nearly identical, and increased from fair to excellent with increasing image quality. R values for systole values were higher than for diastole. There was a statistically significant difference ($p < .05$) between the R values obtained for each image quality group on each layer with exception of the systolic results on the epicardial layer.

3.4. Influence of scanning protocol on image quality and interobserver agreement

Table 5 summarizes the effects of the number of phase-encoding steps and phase-encode order used in the different scanning protocols on the interobserver analysis and tag persistence. Among all the five strains evaluated, the results shown correspond to Ecc, as it is the primary value computed. The R values and the number of slices with longer tag persistence increased as a larger number of phase-encoding steps were used. For all pooled systolic data, R improved from the category of 'good' (mean and median of $R = 0.70$ and 0.72 , respectively) using 96 phase-encoding steps up to the 'excellent' category (all mean and median of $R > 0.8$) using 128 or more phase encoding steps. The diastolic pooled data showed lower R values than the systolic pooled data. However, it also presented the same increase with a larger number of phase-encoding steps (mean and median R values from $0.62/0.63$ up to a maximum of $0.73/0.75$, respectively).

Table 4. Interobserver and intraobserver variability for all pooled Ecc data ($n = 27,756$ systolic, $27,648$ diastolic) in the different myocardial layers related to the tag persistence

Layer	Tag persistence								p value
	Very good		Good		Fair		S	D	
	S	D	S	D	S	D			
Interobserver R^*									
Endocardium	0.66 (0.64, 0.68)	0.58 (0.56, 0.6)	0.68 (0.67, 0.68)	0.57 (0.55, 0.59)	0.52 (0.5, 0.55)	0.45 (0.43, 0.47)			.015 .001
Midwall	0.79 (0.78, 0.79)	0.66 (0.65, 0.66)	0.76 (0.75, 0.76)	0.6 (0.59, 0.61)	0.61 (0.60, 0.63)	0.47 (0.45, 0.49)			.012 .026
Epicardium	0.69 (0.68, 0.69)	0.52 (0.50, 0.53)	0.64 (0.62, 0.65)	0.48 (0.47, 0.50)	0.58 (0.56, 0.6)	0.4 (0.40, 0.41)			.101 [†] .038 [†]
Intraobserver R^*									
Endocardium	0.71 (0.62, 0.77)	0.62 (0.55, 0.61)	0.74 (0.66, 0.78)	0.56 (0.55, 0.61)	0.57 (0.55, 0.59)	0.45 (0.44, 0.47)			.013 .013
Midwall	0.80 (0.78, 0.80)	0.66 (0.65, 0.67)	0.76 (0.74, 0.79)	0.6 (0.58, 0.63)	0.61 (0.58, 0.66)	0.47 (0.46, 0.47)			.006 .006
Epicardium	0.70 (0.68, 0.70)	0.53 (0.51, 0.53)	0.65 (0.60, 0.66)	0.49 (0.47, 0.52)	0.60 (0.55, 0.60)	0.42 (0.39, 0.45)			.023 .023

S, systole; D, diastole.

* R = intraclass correlation coefficient; mean and median of R were identical. An R was considered poor when < 0.4 ; fair, $R = 0.40-0.59$; good, $R = 0.60-0.74$; and excellent, $R \geq 0.75$. Numbers in parentheses are 5 and 95 percentiles, respectively.

[†]No statistical significant inter- or intraobserver differences between observers were observed in the corresponding myocardial layers ($p > 0.05$).

Table 5. Comparison of interobserver variability for all pooled *Ecc* data ($n = 27,756$ systolic, 27,648 diastolic) and tag persistence between the different centers

Center	PS	PO	Systole	Diastole	Tag persistence [§]		
			R^*	R^*	Fair	Good	Very good
A	96	S-I	0.7/0.72 (0.56, 0.85)	0.62/0.63 (0.51, 0.74)	8/18 (44)	5/18 (28)	5/18 (28)
B	128	S-I	0.81/0.82 (0.61, 0.92)	0.67/0.69 (0.62, 0.72)	5/18 (27.8)	9/18 (50)	4/18 (22.2)
C	128	C	0.84/0.86 (0.77, 0.9)	0.69/0.7 (0.62, 0.76)	0/18 (0)	9/18 (50)	9/18 (50)
D	140	C	0.82/0.85 (0.61, 0.95)	0.73/0.75 (0.48, 0.89)	1/18 (5.6)	2/18 (11.1)	15/18 (83.3)
All centers			0.8/0.82 (0.61, 0.92)	0.68/0.69 (0.51, 0.85)	14/72 (19.4)	25/72 (34.7)	33/72 (45.8)

PS = number of phase-encoding steps of the image matrix; PO = phase-encode order; S = sequential scheme; I = interleaved scheme; C = center-out scheme.

[§]Numbers are the slices of the total obtained at each center, and all combined. Numbers in parentheses indicate percentage of slices within each center.

* R = intraclass correlation coefficient expressed as mean/median, respectively. An R was considered poor when < 0.4 ; fair, $R = 0.40-0.59$; good, $R = 0.60-0.74$; and excellent, $R \geq 0.75$. Numbers in parentheses are 5 and 95 percentiles, respectively.

Yet, the diastolic values from all centers remained in the category of ‘good’ agreement. Considering the pooled data from all centers, the interobserver agreement achieved for *Ecc* was ‘excellent’ for systole (mean and median of $R = 0.8/0.82$) and ‘good’ for diastole (mean and median of $R = 0.68/0.69$). The number of slices with longer tag persistence also increased with larger number of phase-encoding steps. The majority of all obtained slices (58/72, 80.5%) had good or very good tag persistence and, hence, image quality. There were no slices with poor tag persistence.

4. Discussion

The results of our study can be summarized as follows: 1) for both peak and pooled systolic data HARP yielded an excellent inter- and intraobserver agreement; 2) inter- and intraobserver agreement showed a direct relationship to the image quality of the myocardial MR-tagged images; 3) for diastolic pooled data R values were lower due to tag fading; and 4) both inter- and intraobserver agreement were better in the anterior and septal myocardial regions, and in the mid-wall layer.

Previously reported data on inter- and intraobserver agreement of quantitative analysis of myocardial MR-tagged images are scarce. Early studies showed that the interobserver and intraobserver agreement of *Ecc* measured in the LV of normal and hypertensive subjects with a conventional tracking technique is good (correlation coefficient, $r = 0.92$) (15, 16). Using HARP, the interobserver reproducibility of strain measurements was previously assessed in a smaller number of 8 exams (6). For pooled data in normal and dysfunctional myocardium, HARP yielded highly reproducible results for both *Ecc* and Λ_2 ($r = 0.98$ and 0.99 , respectively). Our study presents **three** main differences compared with these studies. First, we used intraclass correlation coefficient (R) instead of correlation coefficient (r ; occasionally specified as the product-moment correlation)

to evaluate the performance of HARP. Although the r measure can be used to assess reliability of continuous measurements, it is not the appropriate method to assess inter- and intraobserver agreement (or reliability) (10–12). The correlation coefficient (r) is a measure of linear association rather than reliability. It also does not take into account systematic error between observers. This test can be used in a test–retest situation, where the systematic error is not due to a lack of reliability but to a learning effect (10–12). The intraclass correlation coefficient (R) based on ANOVA does take into account the amount of systematic error, and is more applicable for assessing inter- and intraobserver agreement, as demonstrated in quantitative MR perfusion analysis (17). Second, we evaluated studies obtained from different centers and different types of scanners within a large-scale study rather than single-center studies. Third, not only peak strain values have been compared, but also all pooled data from the systolic and diastolic phases. To our knowledge, this is the first study that evaluates inter- and intraobserver agreement of strain measurements obtained by MR-tagging, and uses all these three components.

Our agreement results are related with cardiac geometry. The strain parameters with the best agreement results for peak values were *Ecc* and Λ_2 . An important feature of *Ecc* was that its accuracy did not change significantly regardless of the image quality. This feature reflects its robustness and is in good agreement with previous studies. Moore et al. studied 31 healthy volunteers and demonstrated that *Ecc* is one of the strain parameters with highest precision. The accuracy of *Ecc* is explained by the larger number of tags around the myocardial circumference than across the wall. This large number of tags provides a high density of displacement data (18). In contrast, radially oriented *Err* has relatively low precision because only two or three tags span the wall. The influence of the cardiac geometry can be observed also in the agreement results of Λ_1 and Λ_2 , which were better for the latter. Λ_1 , or principal strain of maximal systolic thickening (or diastolic shortening), is close to radial

strain in the normal heart, and has lower measurement precision relative to magnitude than Λ_2 , principal strain of greatest shortening (18). The analysis of the inter- and intraobserver R in the myocardial individual regions and layers provides further insight into the differences in the results between pooled systolic and diastolic data. Overall, in all myocardial regions, the R values of the systolic phase are higher than the diastolic ones. Among the four regions, both inter- and intraobserver R 's are higher in the anterior and septal regions, followed by the inferior and lateral regions. These results are most likely due to the higher SNR usually present in the septal and anterior regions, and are related to the distance from the surface coil elements (19). In addition, the larger through-plane motion component in the lateral region compared to the septum may contribute to the worse results in that region. For the three myocardial layers analyzed, the midwall layer yielded better results than the epicardial and endocardial layers since it has less influence from tracing and segmentation errors that may be present in the other layers. When comparing the differences in the intraobserver R values, it is noteworthy that the two experienced readers showed almost identical performance, and their reading were only slightly different from the readings of the inexperienced observer. This reflects the robustness and simplicity of the strain analysis with HARP, which may facilitate its clinical application.

There are some limitations in this study, mainly related to technical issues that shaped the study design. The MR-tagging scanning protocol of this study was set up with the premise of high temporal resolution and longest tag persistence as possible. One of the major limitations of MR-tagging, however, is tag fading in diastole due to the exponential decay of myocardial T_1 -relaxation (8, 20). This hampers the visualization and detection of the tags by semi-automatic and automatic algorithms such as HARP, as it has been demonstrated in our study. Among the different techniques developed to overcome this limitation, highlights the complimentary-SPAMM technique (CSPAMM), which was not available in our MR systems (20). As an alternative, SPAMM tagging with a parallel stripe tag pattern instead of a grid pattern was used in this study. This approach allows an efficient k-space reduction by acquiring close to a single line of central k-space and, therefore, a reduction in the number of phase-encode views without losing precision in estimating the position of the tag lines (21). The tags are adequately sampled as they are oriented perpendicular to the frequency-encoding direction with a 256 readout resolution. In addition, there is no interference between the magnetization of orthogonal tagging planes, and the higher signal-noise-ratio (SNR) of the parallel lines enables a longer sampling into the cardiac cycle (22). The drawbacks of this approach are the need of swapping the frequency-encoding direction and changing the tagging angle to obtain two orthogonal sets of tags, and, therefore, doubling the number of breath-holds that in our study resulted in an average total scan time of 5–7

minutes. There is also a possibility of changing the short axis planes (through-plane motion) during acquisition of the horizontal and the vertical tags.

Another limitation of the study is the acquisition of only short-axis images due to time-constraints in the scanning protocol, and lack of capability of analysis of long-axis images by HARP. This shortfall can be surmounted by a 3D-HARP analysis, currently under development. Also, although generally acceptable in the literature, the definitions of R as either high or low are arbitrary, a value of $R > 0.75$ is used generally to indicate a good agreement (23). On the other hand, reproducibility of any measurement is considered clinically acceptable only when $R \geq 0.60$ (13). Nevertheless, our results, particularly those for peak and pooled systolic data with tag persistence $\geq 71\%$ of the cardiac cycle, were consistently above both threshold values.

5. Conclusion

In conclusion, our results demonstrate the robustness and accuracy of myocardial tagged image analysis by HARP method. This analysis yields an 'excellent' inter and intra-observer agreement both for the peak strain values and pooled systolic data, and 'good' agreement in the diastolic phase strains. There is a strong relationship between image quality, tag persistence and the number of phase-encoding steps. Thus, when stripe tags are used, a matrix with 128 phase-encoding or more should be used. The results of this ancillary study may facilitate the clinical application of HARP for the evaluation of regional left ventricular function.

Acknowledgments

This research was supported by grant RO1-HL66075-01 and contracts N01-HC-9519 through N01-HC-95163, and N01-HC-95168 from the National Heart, Lung, and Blood Institute. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at <http://www.mesa-nhlbi.org>. E.C. is supported by a research grant from the Fundación Ramón Areces, Madrid, Spain.

References

1. McVeigh ER. MRI of myocardial function: motion tracking techniques. *Magn Reson Imaging* 1996; 14:137–150.
2. Croisille P, Moore CC, Judd RM, Lima JA.C, Arai M, McVeigh ER, Becker LC, Zerhouni EA. Differentiation of viable and nonviable myocardium by the use of three-dimensional tagged MRI in 2-day-old reperfused canine infarcts. *Circulation* 1999; 99:284–291.
3. Stuber M, Scheidegger MB, Fischer SE, Nagel E, Steinemann F, Hess OM, Boesiger P. Alterations in the local myocardial motion pattern in

- patients suffering from pressure overload due to aortic stenosis. *Circulation* 1999; 100:361–368.
4. Gotte MJ, van Rossum AC, Twisk JWR, Kuijper JPA, Marcus JT, Visser CA. Quantification of regional contractile function after infarction: strain analysis superior to wall thickening analysis in discriminating infarct from remote myocardium. *J Am Coll Cardiol* 2001; 37:808–817.
 5. Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. *Magn Reson Med* 1999; 42:1048–1060.
 6. Garot J, Bluemke DA, Osman NF, Rochitte CE, McVeigh ER, Zerhouni EA, Prince JL, Lima JAC. Fast determination of regional myocardial strain fields from tagged cardiac images using harmonic phase MRI. *Circulation* 2000; 101:981–988.
 7. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, David R, Kronmal R, Liu K, Nelson JC, O’Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* 2002; 156:871–881.
 8. Axel L, Dougherty L. MR imaging of motion with spatial modulation of magnetization. *Radiology* 1989; 171:841–845.
 9. Castillo E, Lima JAC, Bluemke DA. Regional myocardial function: advances in MR imaging and analysis. *Radiographics* 2003; 23:S127–S140.
 10. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966; 19:3–11.
 11. Schouten HJ. Measuring pairwise interobserver agreement when all subjects are judged by the same observers. *Stat Neerl* 1982; 36:45–61.
 12. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med* 2002; 21:3431–3446.
 13. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 1981; 86:127–137.
 14. Oppo K, Leen E, Angerson WJ, Cooke TG, McArdle CS. Doppler perfusion index: an interobserver and intraobserver reproducibility study. *Radiology* 1998; 208:453–457.
 15. Clark NR, Reichek N, Bergey P, Hoffman EA, Brownson D, Palmon L, Axel L. Circumferential myocardial shortening in the normal human left ventricle. Assessment by magnetic resonance imaging using spatial modulation of magnetization. *Circulation* 1991; 84:67–74.
 16. Palmon LC, Reichek N, Yeon SB, Clark NR, Brownson D, Hoffman E, Axel L. Intramural myocardial shortening in hypertensive left ventricular hypertrophy with normal pump function. *Circulation* 1994; 89:122–131.
 17. Muhling OM, Dickson ME, Zenovich A, Huang Y, Wilson BV, Wilson RF, Anand IS, Seethamraju RT, Jerosch-Herold M, Wilke NM. Quantitative magnetic resonance first-pass perfusion analysis: inter- and intraobserver agreement. *J Cardiovasc Magn Reson* 2001; 3:247–256.
 18. Moore CC, Lugo-Olivieri CH, McVeigh ER, Zerhouni EA. Three-dimensional systolic strain patterns in the normal human left ventricle: characterization with tagged MR imaging. *Radiology* 2000; 214:453–466.
 19. Constantinides CD, Westgate CR, O’Dell WG, Zerhouni EA, McVeigh ER. A phased array coil for human cardiac imaging. *Magn Reson Med* 1995; 34:92–98.
 20. Fischer SE, McKinnon GC, Maier SE, Boesiger P. Improved myocardial tagging contrast. *Magn Reson Med* 1993; 30:191–200.
 21. McVeigh ER, Atalar E. Cardiac tagging with breath-hold cine MRI. *Magn Reson Med* 1992; 28:318–327.
 22. Salido TB, Hundley WG, Link KM, Epstein FH, Hamilton CA. Effects of phase encode order and segment interpolation methods on the quality and accuracy of myocardial tags during assessment of left ventricular contraction. *J Cardiovasc Magn Reson* 2002; 4:245–254.
 23. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med* 1989; 19:61–70.