Radiology

Repeatability of MR Elastography of Liver: A Meta-Analysis[1]

Suraj D. Serai, PhD
Nancy A. Obuchowski, PhD
Sudhakar K. Venkatesh, MD
Claude B. Sirlin, MD
Frank H. Miller, MD
Edward Ashton, PhD
Patricia E. Cole, MD, PhD
Richard L. Ehman, MD, PhD

Purpose: To perform a meta-analysis to generate an estimate of the repeatability coefficient (RC) for magnetic resonance (MR) elastography of the liver.

Materials and Methods: A systematic search of databases was performed for publications on MR elastography during the 10-year period between 2006 and 2015. The identified studies were screened independently and were verified reciprocally by all authors. Two reviewers independently determined the percentage RC and effective sample size from each article. A forest plot was constructed of the percentage RC estimates from the 12 studies. Bootstrap 95% confidence intervals (CIs) were constructed for the summary percentage RCs.

Results: Twelve studies comprising 274 patients met the eligibility criteria and were included for analysis. A flow diagram of studies included according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines was prepared for the inclusion and exclusion criteria. All studies included in the meta-analysis fulfilled four or more of the seven categories of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2. The estimated summary RC was 22% (95% CI: 16.1%, 28.2%). The three main sources for this heterogeneity were the trained versus untrained operator drawing contours to choose regions of interest, the time between two replicate examinations, and, finally, the field strength of the MR imaging unit. The RC estimates tended to be higher for studies that did not use a well-trained operator, those with 1.5-T field strength imaging units, and those with longer time intervals between examinations.

Conclusion: The meta-analysis results provide the basis for the following draft longitudinal Quantitative Imaging Biomarkers Alliance MR elastography claim: A measured change in hepatic stiffness of 22% or greater, at the same site and with use of the same equipment and acquisition sequence, indicates that a true change in stiffness has occurred with 95% confidence.

[1] From the Department of Radiology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, MLC 5031, Cincinnati, OH 45229 (S.D.S.); Department of Quantitative Health Sciences, the Cleveland Clinic Foundation, Cleveland, Ohio (N.A.O.); Department of Radiology, Mayo Clinic, Rochester, Minn (S.K.V., R.L.E.); Department of Radiology, UCSD Liver Imaging Group, San Diego, Calif (C.B.S.); Department of Radiology, Northwestern Memorial Hospital, Chicago, Ill (F.H.M.); Virtualscopics, Rochester, NY (E.A.); and Clinical and Translational Science–Imaging, Takeda Pharmaceuticals, Deerfield, Ill (P.E.C.). Received July 20, 2016; revision requested September 20; revision received December 10; accepted January 30, 2017; final version accepted February 24. **Address correspondence to** S.D.S. (e-mail: suraj.serai@cchmc.org).

N.A.O. supported by the National Institute of Biomedical Imaging and Bioengineering (grant XYZ). R.L.E. supported by the National Institutes of Health (grant R01 EB001981).

Chronic liver disease (CLD) is an important cause of morbidity and mortality and can lead to hepatic fibrosis, cirrhosis, portal hypertension, and hepatocellular carcinoma. CLD is a major health burden in the United States and around the world. Regardless of its etiology, when CLD is untreated, it leads to liver fibrosis, and, if progressive, to cirrhosis and its complications. Effective treatment methods for CLD are now available and can prevent progression of the liver fibrotic process or even result in regression of fibrosis when administered in the early stages of fibrosis (1). A reliable noninvasive technique is needed for detecting and staging fibrosis as well for as evaluating treatment response (2). Magnetic resonance (MR) elastography noninvasively measures tissue shear-wave stiffness, which is a potential noninvasive biomarker for quantification of liver fibrosis (3,4). Because the stiffness measurements are quantitative, MR elastography allows for comparison over time (3–5).

In qualifying a biomarker for disease evaluation, it is essential to estimate the measurement error of the technique and to standardize and validate the acquisition and analysis techniques. Measurement error is commonly determined by repeatability. Repeatability represents the measurement precision in a set of conditions that include the same measurement procedure, same operators, same measuring system, same operating conditions, and same physical location, with replicated measurements made with the same or similar experimental units over a short period of time (6). Results of several studies (4,7–16) have shown that MR elastography is an accurate and highly repeatable technique and that it holds promise for use in clinical trials. MR elastography is U.S. Food and Drug Administration approved for measuring liver stiffness and can be performed by using a 1.5- or 3.0-T MR imaging unit (9,10,16). With the expanding clinical and research applications of MR elastography, the literature on MR elastography repeatability has grown, but a key repeatability metric—namely, the repeatability coefficient (RC)—has not been reported in most published articles. In this study, we performed a meta-analysis to critically evaluate the reported repeatability of liver MR elastography and to generate an estimate of the RC for use of this technique in a clinical setting. RC is a commonly used measure of repeatability and is defined here as the least significant difference between two repeated measurements in a case taken in the same conditions (17). Assuming no change in hardware and software, the RC helps establish a plausible range for a true change in measured stiffness with 95% confidence. In simple words, for stiffness comparison over time, the RC helps establish with 95% confidence the likelihood that an observed change represents a true change.

Advance in Knowledge

- With MR elastography, a measured change in hepatic stiffness of 22% or greater indicates that a true change in stiffness has occurred with 95% confidence.

Materials and Methods

This study was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (18). Because this was a meta-analysis that did not involve identifiable patient information, investigational review board approval was not necessary.

Implication for Patient Care

- The repeatability coefficient of 22% may help radiologists and clinicians interpret longitudinal changes in MR elastography–measured stiffness occurring during clinical follow-up and in research studies and will be useful in informing sample size calculations for clinical trials that use MR elastography–derived stiffness as end points.

Literature Search

A systematic search of PubMed (MEDLINE), Embase, Scopus, the Cochrane Library, the Web of Science, Cumulative Index to Nursing and Allied Health Literature, and Google Scholar databases was performed for the 10-year period prior to May 2015. An initial search strategy involving the following free-text words was performed: "hepatic fibrosis," "MR elastography," "magnetic resonance elastography," "liver," "liver parenchyma," "liver anatomy and histology," "liver physiology," "hepatic," "liver stiffness," "liver elasticity," "elastic modulus," "elasticity imaging techniques/methods," "sensitivity and specificity," "reproducibility," "repeatability," and "reliability." We also used a sensitive and precise search strategy in the PubMed database and in regular Google Search for locating any existing systematic reviews on MR elastography (to identify additional studies missed by our search), and none were identified. In addition, a manual search of reference lists from primary studies was performed to locate any potential studies missed with electronic search strategies, and consultation with experts in the field

https://doi.org/10.1148/radiol.2017161398

Content codes: GI MR

Radiology 2017; 285:92–100

Abbreviations:
CI = confidence interval
CLD = chronic liver disease
PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QIBA = Quantitative Imaging Biomarkers Alliance
QUADAS = Quality Assessment of Diagnostic Accuracy Studies
RC = repeatability coefficient
ROI = region of interest
wCV = within-subject coefficient of variation

Author contributions:
Guarantor of integrity of entire study, S.D.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, all authors; clinical studies, S.D.S., F.H.M.; statistical analysis, S.D.S., N.A.O., S.K.V., E.A., R.L.E.; and manuscript editing, all authors

Conflicts of interest are listed at the end of this article.

was performed to identify additional published studies. Studies involving human subjects and published in English or with translations available in English were included. The search was performed independently by two observers (S.D.S. and S.K.V.) using a predetermined search strategy at two major institutional libraries (the University of Cincinnati and Mayo Clinic). The identified studies were then screened independently by both observers to identify studies with repeatability. Inclusion criteria were as follows: *(a)* studies that included measurements of change in liver stiffness measured at two or more time points in similar conditions; *(b)* studies that calculated MR elastography–based stiffness as absolute values of shear modulus; and *(c)* studies that clearly reported the time between repeat measurements, mean liver stiffness, and the coefficient of variation. The exclusion criteria were as follows: *(a)* duplicate publication (based on the same primary study), *(b)* nonoriginal research, and *(c)* studies not published in English. The final list of studies that met the inclusion and exclusion criteria as defined were reviewed and corroborated by all authors.

Data Extraction and Quality Verification

The data were extracted by using a predefined form. The following data were extracted: *(a)* author, journal, and year of publication; *(b)* within-subject coefficient of variation (wCV); *(c)* number of subjects; *(d)* number of readers; and *(e)* notes on the method used to calculate the wCV. Data quality was assessed by using the Quality Assessment of Studies of Diagnostic Accuracy (QUADAS)-2 tool (19).

Statistical Analysis

Two reviewers (S.D.S. and S.K.V.) independently determined the percentage RC and effective sample size from each article. The percentage RC was defined as follows:

$$\text{Percentage RC} = 1.96 \cdot \sqrt{2 \cdot \%\text{wCV}^2},$$

where %wCV is the wCV expressed as a percentage of the mean stiffness value (ie, $\text{wSD}/\overline{Y} \cdot 100$) and wSD is the within-subject standard deviation of the measurements in a subject (17).

The effective sample size was usually the number of subjects N who participated in the test-retest design. Some studies, however, included measurements from multiple readers. For these studies, we assumed moderate correlation between readers ($r = 0.5$) and computed the design effect, deff, as $[1 + (s - 1) \times r]$, where s is the number of readers. The effective sample size is then $N \cdot s/\text{deff}$. We denote the effective sample size by M.

The 95% confidence interval (CI) for the RC for each study was computed as follows: $2.77 \cdot \sqrt{M \cdot \%\text{wCV}^2 / \chi^2_{M,\alpha}}$, where $\chi^2_{M,\alpha}$ is the αth percentile of the χ^2 distribution with M degrees of freedom. For the lower bound, α is .975, and for the upper bound, α is .025.

Because not all articles reported the percentage RC, various methods were used to extract percentage RC from each article. When the authors reported the percentage wCV, we calculated percentage RC using the above formula. When authors provided actual test-retest data, we calculated the percentage RC from the raw data. Some authors reported the limits of agreement (LOAs). Because LOAs include the bias of the measurements, we examined the Bland-Altman plot, if available, to determine the bias. We then estimated the percentage RC from the LOAs after correcting the LOAs for the bias. When the percentage RC computed by the reviewers disagreed with the estimate of percentage RC reported in the article, the estimates were discussed, and a consensus among reviewers was reached.

The heterogeneity of the percentage RC estimates was assessed by calculating the Q statistic (Cochran) along with the I^2 statistic (Higgins), which describes the percentage of the total variability across studies that is due to heterogeneity between studies rather than chance differences. On the basis of the results of the heterogeneity analysis, summary statistics were calculated for subgroups of studies.

A forest plot was constructed of the percentage RC estimates from each of the 12 articles. For each subgroup, a summary percentage RC, denoted as $\hat{\theta}$, was computed as follows by using the effective sample size as the weight, where $\widehat{\text{RC}}_i$ is the percentage RC from the ith article and M_i is the effective sample size from the ith study (20):

$$\hat{\theta} = \sqrt{\left(\sum_{i=1}^{12} \widehat{\text{RC}}_i^2 \cdot M_i\right) \Big/ \left(\sum_{i=1}^{12} M_i\right)}.$$

A bootstrap 95% percentile CI was constructed for each θ. Let N_s denote the number of studies in the sth subgroup. For each of the b bootstrap samples, N_s articles were randomly selected with replacement from the original N_s articles and $\widehat{\theta}_b$ was calculated. A total of 10 000 bootstrap samples were constructed. The 250th and 9750th largest values of $\widehat{\theta}_b$ were identified as the lower and upper confidence bounds for the CI. A funnel plot was generated to assess publication bias.

Results

The search at library 1 (Mayo Clinic) revealed 309 articles, and the search at library 2 (University of Cincinnati) revealed 350 articles. Duplicate articles were removed, and a list of 450 articles was collected in a single electronic library. After a detailed manual review of the search list provided by the libraries was performed, 12 articles met the inclusion and exclusion criteria. The 12 articles were included for meta-analysis (7,8,11,16,21–28). A flow diagram according to PRISMA guidelines of studies included was prepared for the inclusion and exclusion criteria (Fig 1). The meta-analysis fulfilled the reporting items of the PRISMA checklist. All studies included in the meta-analysis fulfilled four or more of the seven categories of the QUADAS-2 tool (Table 1) (19). A bar chart with estimates of percentage compliance was plotted (Fig 2). Weaknesses in the study of Wang et al (27) included a small sample size ($n = 5$) and ambiguity regarding the representativeness of the rest of the patient population. In the study by Shin et al (28), we were unable to find the

Figure 1

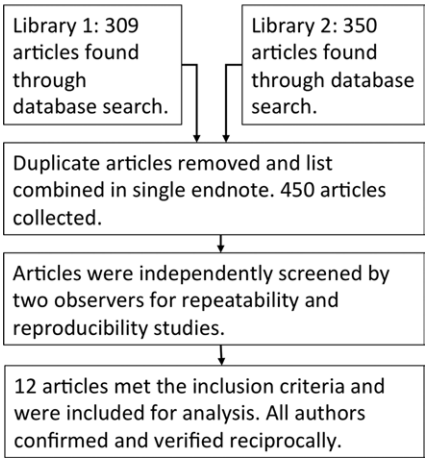

Figure 1: Flow diagram of studies included according to the inclusion and exclusion criteria.

Table 1

Assessment of Quality of Included Studies with the QUADAS Tool

		Criterion						
Study No.	Study	1	2	3	4	5	6	7
1	Wang et al (27)	?	?	S	S	?	S	S
2	Venkatesh et al (11)	S	S	S	S	S	S	S
3	Shire et al (26)	S	S	S	S	S	S	S
4	Shinagawa et al (25)	S	S	S	S	S	S	S
5	Shin et al (28)	?	?	S	S	?	S	S
6	Shi et al (8)	S	S	S	S	S	S	S
7	Lee et al (24)	?	?	S	S	?	S	S
8	Jajamovich et al (23)	S	S	S	S	S	S	S
9	Hines et al (7)	S	S	S	S	S	S	S
10	Hines et al (22)	S	S	S	?	S	S	S
11	Bohte et al (21)	S	S	S	S	S	S	S
12	Trout et al (16)	S	S	S	S	S	S	S

Note.—S = satisfactory (the study appeared to satisfy this criterion). The QUADAS criteria are as follows: 1, Risk of bias in patient selection: It assesses if the study included a consecutive or a random sample of eligible patients, otherwise there is a potential for bias. 2, Risk of bias in index test: It tests if the test results were blinded to the standard of reference. 3, Risk of bias in reference standard: It tests if there were any variations in the reference standard. 4, Risk of bias in flow and timing: It tests if the index test and reference standard test were performed in the same patient at the same time. 5, Applicability concerns in patient selection: Checks if there are concerns that the included patients do not match the review question. 6, Applicability concerns in index test: Checks if there are concerns that the interpretation differs from the review question. 7, Applicability concerns in the reference standard: Checks if the target condition as defined by the reference standard matches the question.

time interval between repeat examinations, and in the study by Lee et al (24), the selection criteria were unclear. The baseline characteristics of the included studies, including subjects, time interval, and coefficient of variation, were summarized (Table 2).

From the 12 studies included in the analysis, the estimated meta-analysis summary RC was 22% (95% CI: 16.1%, 28.2%) (forest plot shown in Fig 3). There was significant between-study heterogeneity in the RC estimates ($I^2 > 90\%$). The three main sources for this heterogeneity were trained versus untrained operators drawing contours to choose ROIs, the time between two replicate examinations, and, finally, the field strength of the MR imaging unit (Table 3). The RC estimates tended to be higher for studies that did not use a well-trained operator, with 1.5-T field strength MR imaging units, and with longer time intervals between examinations. The studies were further subgrouped as follows: *(a)* trained operator to draw ROIs (10 studies) and *(b)* MR imaging unit field strength of 1.5 T (eight studies, of which two studies had < 1 week interval between repeat examinations and six studies had > 1 week between repeat examinations) and MR imaging field strength of 3.0 T with a time interval between imaging of 1 week or less (three studies with very low heterogeneity). There remained one study with an MR imaging unit field strength of 3.0 T and more than a 2-week time interval between examinations (Table 3).

The 10 studies that used a trained operator to draw contours had a summary percentage RC estimate of 18.4% (bootstrap 95% CI: 14.2%, 22.2%), and the two studies that used an untrained operator to draw the contours had a summary percentage RC of 34.5%.

The estimated RCs ranged from 10% (magnet field strength of 3.0 T with < 1 day between examinations) to 34.1% (magnet field strength of 1.5 with 3 weeks between examinations), with a mean of 22%. The estimated summary RC from eight studies that used MR imaging units with a field strength of 1.5 T was 25.2% (bootstrap 95% CI: 17.4%, 31.9%). The estimated summary RC from four studies that used MR imaging units with field strengths of 3.0 T with an interval between examinations of 1 week or less was 12.7% (bootstrap 95% CI: 10.0%, 15.9%). The one remaining study with an MR imaging unit field strength of 3.0 T and an average 17-day interval between examinations reported an estimated RC of 22.2%.

The funnel plot showed that the studies with the largest sample sizes fall near the summary value of 22% at the top of the plot, and studies with smaller sample sizes fall fairly symmetrically on either side at the bottom of the plot (Fig 4). One study with a large sample size that fell out of the funnel was the study by Hines et al (7); the study that is borderline outside of the funnel is the study by Lee et al (24).

Discussion

Two key aspects of precision are repeatability and reproducibility. Repeatability refers to test conditions that are as constant as possible, where using the same equipment within a "short time interval" obtains independent test results by the same method with identical set-up (MR elastography hardware) in the same MR imaging unit. Reproducibility refers to test conditions in which results are obtained with the same technique and identical set-up but in different MR imaging units with

Figure 2

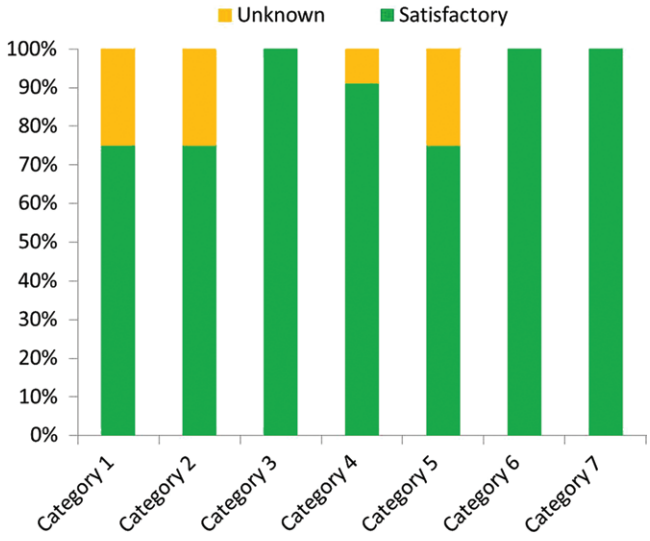

Figure 2: Bar chart shows estimates of the studies' percentage compliance with the QUADAS-2 tool.

different operators (eg, in a cross-sectional or cross-vendor comparison [16,29]). When a new technique is to be used in a clinical setting, the repeatability of the method should always be estimated. MR elastography has been shown to have the highest combination of sensitivity, specificity, repeatability, and reproducibility as compared with other noninvasive imaging-based techniques in the evaluation of liver fibrosis (4,12,29–31). In our study, we addressed the importance of repeatability in a longitudinal claim.

In our meta-analysis results, the estimated percentage RC and 95% CIs from the 12 studies ranged from 12% to 37%. In the study by Bohte et al (21), they reported an intraimage wCV of 7.0% ± 2.3, a within-day wCV of 16.8% ± 5.5, and a within-weeks wCV of 22.2% ± 7.3. However, there was no information on whether patients were imaged in fasting conditions. It is possible that the larger within-day and within-weeks wCV may be attributable to the inconsistency of examinations performed in a fasting or a postprandial state. It has been reported that in patients with CLD, liver stiffness is markedly increased after food intake because of postprandial effects (32). No restrictions with regard to fasting among others could explain the variation in wCV from 7% to 22% in the same subject population. In a similar study by Jajamovich et al (23) in which patient fasting status was tightly controlled, a wCV of 3.8% was observed. However, when examinations performed after fasting were compared with those performed after eating, the wCV increased to 6.8% because of postprandial effects. The largest wCV range (37.0%) was observed by Hines et al (7). The wCV estimates from that study involved different examinations on different days, different examinations on the same day, multiple readers (interreader variability), multiple readings (intrareader variability), and a combined subject population of healthy volunteers and patients with CLD. Such a study explores many precision components to vary, and a higher wCV is thus expected. Calculation of mean shear stiffness of the liver involves manually specifying ROIs in the liver parenchyma in which shear waves are visible, while excluding major blood vessels that are wider than 3 mm and within the contour of the liver as seen on the MR elastography magnitude images. Areas close to the liver margins and immediately close to the driver are to be avoided because they have the highest wave reflections, which can attribute to artificially increased values. To avoid these areas, it is recommended to stay approximately 1 cm away from the liver margins (14,33,34). Additionally, careful attention should be paid to the placement of ROIs over the left lobe of the liver, which should in general be avoided when possible. The left lobe of the liver typically has low confidence levels because of motion artifacts from cardiac pulsations. It is also important that the ROIs should be drawn with reference to the magnitude, wave, and elastography images. If needed, the ROI should be modified to exclude areas with low wave amplitude, areas of incoherent waves seen in the wave images, and areas of low confidence as seen by the checkerboard pattern in the masked elastography images. For an experienced reader, the ROI can be drawn in a single step, keeping the guiding principles in mind (34). In our meta-analysis report, in addition to sample size and number of days between repeat examinations, we observed that one of the sources of heterogeneity arose from reader experience in drawing ROIs. In the study by Jajamovich et al (23) (sample size, n = 30), the ROI was drawn by a postdoctoral fellow with 2 years of image analysis experience and with the supervision of a single radiologist with 10 years of experience, resulting in a percentage RC of 10.5%. Also, in the studies by Shinagawa et al (25) and by Shi et al (8), ROIs were drawn by two radiologists, resulting in percentage RCs of 10% and 15.9%, respectively. However, in the two studies by Hines et al (7,22), the ROIs were drawn by the operator, probably without avoiding vessels and liver margins as the details are missing, resulting in a relatively higher percentage RCs of 23.5% and 37%, respectively. MR elastography in its current form relies on the expertise of the user to draw the ROI on the liver. No standards for ROI shape and size have as yet been established, and various techniques have been described in the literature, ranging from one to three circular ROIs, multiple free-hand

Table 2

Baseline Characteristics of Included Studies

Study No.	Study	Year	Study Design	Sample Size	Age Range (y)	Male Subjects (%)	MR Imaging Unit(s)	Field Strength (T)	Frequency (Hz)	Property Measured*	Subjects	No. of Readers	Time Interval	CV Reported (%)	RC (%)	RC 95% CI (%)	Fasting vs Feeding	ROI Drawn by	ROI Size
1	Wang et al (27)	2011	Prospective	5	NA	NA	Siemens Espree	1.5	60	G*	Healthy subjects	2	2 Weeks	9–12	23	14.3, 56.4	No information	Two radiologists	Three round or oval
2	Venkatesh et al (11)	2014	Prospective	41	23–63	44	GE HDx	1.5	60	G*	Healthy subjects	2	4–6 Weeks	8.4	18.8	13.5, 31.0	4–6 Hours fasting	One experienced reader (5 y of experience)	Two or three round or oval ROIs
3	Shire et al (26)	2011	Prospective	9	20–57	44	GE HDx	1.5	60	G*	Five healthy subjects, four patients	3	1–2 Weeks	6–11	17	12.2, 28.0	8 Hours fasting	Software	One large
4	Shinagawa et al (25)	2014	Prospective	10	27–63	90	GE 750 W	3.0	60	G*	Healthy subjects	1	1 Week	NA	10	7.0, 17.5	No information	Two radiologists	One large
5	Shin et al (28)	2014	Retro	15	57 (Mean)	NA	GE HDx	1.5	60	G*	Patients	2	2 Weeks	NA	14	10.3, 21.7	No information	One experienced reader (3 y of experience)	Three round ROIs
6	Shi et al (8)	2014	Prospective	22	18–56	41	GD HD	3.0	60	G*	Healthy subjects	2	1 Week (short term); 27–30 weeks (long term)	5.75	15.9	12.7, 21.4	8 Hours fasting	Two radiologists	One large
7	Lee et al (24)	2014	Retro	47	27–82	68	GE HDx	1.5	60	G*	Patients	2	8–10 Minutes	13	25.3	21.0, 31.7	No information	Two radiologists	Three round or oval
8	Jajamovich et al (23)	2014	Prospective	30	55.8 (Mean)	77	GE 750	3.0	60	G*	11 Healthy subjects, 19 patients	2	20 Minutes	3.8	10.5	8.6, 13.4	6 Hours fasting and then repeated after feeding	One radiologist (10 y of experience)	One large
9	Hines et al (7)	2010	Prospective	30	21–68	53	GE HDx	1.5	60	G*	20 Volunteers, 10 patients	2	2–4 Weeks	17.4	37	30.4, 47.3	No information	Operator	One large
10	Hines et al (22)	2011	Prospective	11	23–39	75	GE HDx	1.5	60	G*	Healthy subjects	1	5 Weeks	8.5	23.5	16.6, 39.9	Feeding between examinations	Operator	One large
11	Bohte et al (21)	2013	Prospective	30	19–59	60	Philips	3.0	50	G*	16 volunteers, 14 patients	1	1–4 Weeks	10.1	22.2	17.7, 29.7	No information	One radiologist	One large
12	Trout et al (16)	2016	Prospective	24	22–55	21	GE and Philips	1.5 And 3.0	60	G*	Healthy subjects	1	Same day	10.7	16.6	13.0, 23.1	6–8 Hours fasting	One reviewer (6 y of experience)	One large

Note.—CV = coefficient of variation, NA = not applicable, Retro = retrospective, ROI = region of interest.

* Food and Drug Administration–approved, commercially available implementations of MR elastography calculate the magnitude of the complex shear modulus of tissue (|G*|), often called "stiffness." The complex shear modulus (G*) consists of a real part called the storage modulus (G') and an imaginary part called the loss modulus (G").

Figure 3

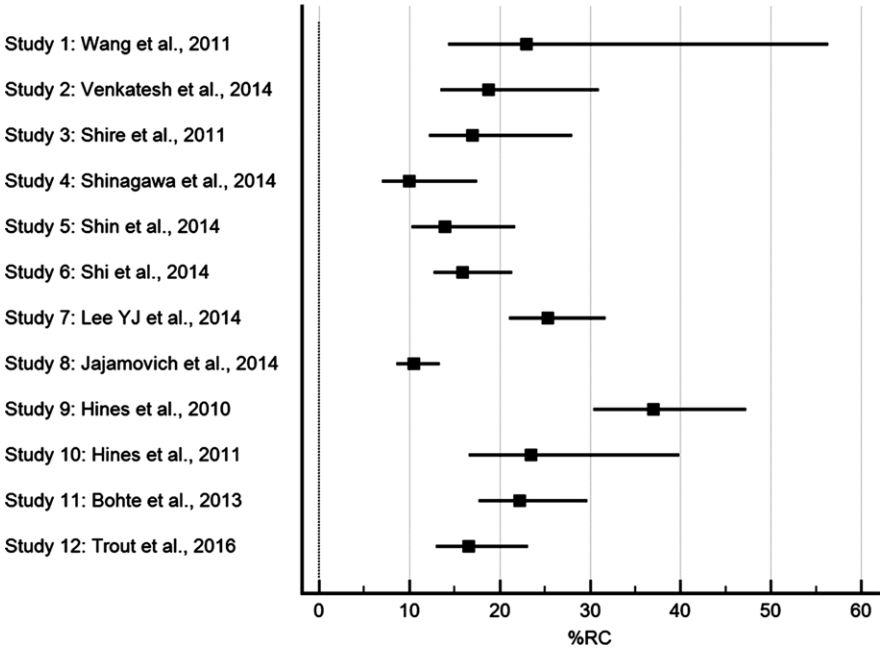

Figure 3: Forest plot shows results from 12 studies (7,8,11,16,21–28). Summary RC = 22% (range, 16.1%–28.2%).

Table 3

Summary of Subgroup Analyses

Parameter	No. of Studies	Summary RC (%)	Bootstrap 95% CI
Trained operator to draw ROI	10	18.4	14.2, 22.21
≤1 Week between examinations	5	17.5	11.6, 23.41
1 Week between examinations	5	19.3	15.6, 21.81
Untrained operator to draw ROI	2	34.5	…
1.5-T field strength	8	25.2	17.4, 31.91
≤1 Week between examinations	2	21.7	…
1 Week between examinations	6	26.0	16.7, 34.21
3.0-T field strength	4	15.6	10.5, 20.81
≤1 Week between examinations	3	12.7	10.0, 15.91
1 Week between examinations	1	22.2	…
All 12 studies	12	22.0	16.1, 28.21

ROIs, one large freehand ROI on a single section or all sections, to ROIs being selected automatically by software (16,23,26,34). In general, the sites try to follow the recommendations of "placing the ROI as large as possible in a way that excludes large vessels." However, if a free-hand ROI tool is not available, it is often quite difficult to obtain large ROIs with circular ROIs, as the measureable regions are often complex in shape. One of the strategies to remove or reduce measurement variability would be to have dedicated personnel trained to draw contours with the supervision of an experienced radiologist or to use semiautomated or automated liver elasticity calculation software, as described by Shire et al (26) and Dzyubak et al (35,36).

Our work is motivated by the activities of the Radiological Society of North America (RSNA) Quantitative Imaging Biomarkers Alliance (QIBA) (37). The mission of QIBA is to improve the value and practicality of quantitative imaging biomarkers by promoting standardization and reducing the variability across devices (hardware) and vendor software platforms, thereby facilitating the qualification of biomarkers with sufficient repeatability and reproducibility for use in clinical care and as end points in clinical trials (17,37). This is done by preparing a QIBA profile document that is intended to span multiple vendor imaging unit platforms, such that variability is included to the maximum extent possible. QIBA intends to use the RC and the 95% CI estimated from this study within the MR elastography profile document to define a plausible range of reliable detection of biologic change. Assuming no change in hardware and software, any values obtained outside [−RC, +RC] can be considered a true change over time. The RC, along with its 95% CI, is then included in the profile claim by considering the clinical requirements for its performance. This profile document is intended for a broad audience, including imaging unit and third-party device manufacturers, pharmaceutical companies, diagnostic agent manufacturers, medical imaging sites, imaging contract research organizations, physicians, technologists, researchers, professional organizations, educational institutions, and various accreditation and regulatory authorities. To this purpose, it is important for the RC claimed in the profile to be realistic and reasonably achievable across imaging centers and readers spanning a relevant range of technical expertise.

Limitations of our study included the fact that this was a retrospective analysis, with inhomogeneity owing to lack of standardized performance of MR elastography studies and lack of biopsy validation in all cases. The wCV used in our analysis was used as reported in the articles. We tried to minimize the impact by recalculating and validating the reported wCV for the data available in the article whenever possible. The small number of studies limited our ability to test

Figure 4

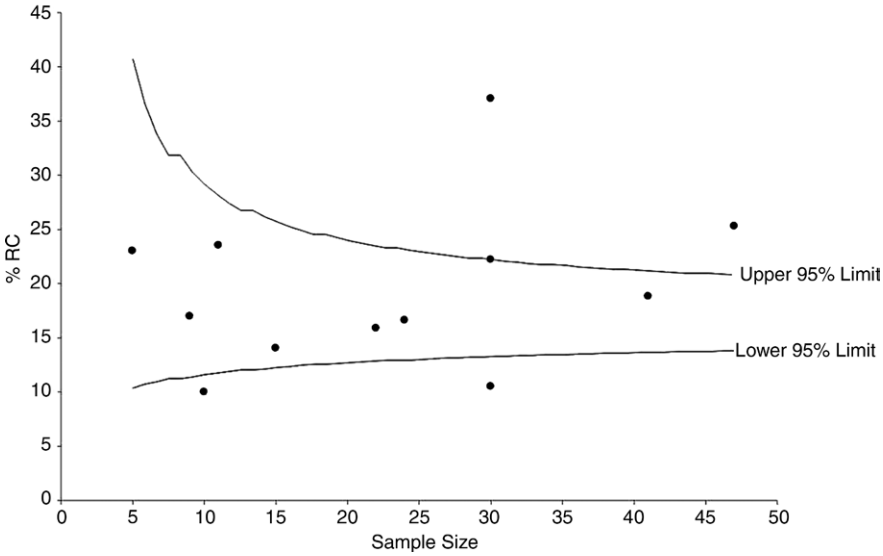

Figure 4: Funnel plot of the RC estimates from each study (y-axis) versus the effective sample size (x-axis) shows that studies with large sample sizes fall near the summary value of 22% and smaller studies fall fairly symmetrically on either side toward the bottom of the plot.

and model the effects of MR imaging unit field strength, ROI, and time interval between examinations on the RC; thus, we provide results based on subgroups of studies. The RSNA QIBA MR elastography profile is expected to serve as guidelines to standardize the acquisition, postprocessing (including selecting and providing training data sets for the selection of ROIs) and interpretation of MR elastography.

In conclusion, our study has shown that MR elastography is a repeatable, noninvasive method for the detection and staging of hepatic fibrosis. Our estimated meta-analysis summary RC for MR elastography was 22% (95% CI: 16.1%, 28.2%). Assuming no change in MR elastography hardware and software, a change in a subject's measured MR elastography over time of 22% or greater can be considered a true change. Ongoing efforts to further standardize MR elastography examination protocols and drawing contours to obtain liver stiffness values may provide further increases in performance. Future investigations should also assess the reproducibility of MR elastography measurements across different system and vendor platforms.

Disclosures of Conflicts of Interest: S.D.S. disclosed no relevant relationships. **N.A.O.** disclosed no relevant relationships. **S.K.V.** disclosed no relevant relationships. **C.B.S.** disclosed no relevant relationships. **F.H.M.** disclosed no relevant relationships. **E.A.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is an employee of Virtualscopics. Other relationships: disclosed no relevant relationships. **P.E.C.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is an employee of Takeda Pharmaceuticals. Other relationships: disclosed no relevant relationships. **R.L.E.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is the president and CEO of and holds equity in Resoundant; has patents related to MR elastography for which the licensee is Resoundant. Other relationships: disclosed no relevant relationships.

References

1. Pariente D, Franchi-Abella S. Paediatric chronic liver diseases: how to investigate and follow up? role of imaging in the diagnosis of fibrosis. Pediatr Radiol 2010;40(6):906–919.
2. Towbin AJ, Serai SD, Podberesky DJ. Magnetic resonance imaging of the pediatric liver: imaging of steatosis, iron deposition, and fibrosis. Magn Reson Imaging Clin N Am 2013;21(4):669–680.
3. Yin M, Talwalkar JA, Glaser KJ, et al. Assessment of hepatic fibrosis with magnetic resonance elastography. Clin Gastroenterol Hepatol 2007;5(10):1207–1213.e2.
4. Xanthakos SA, Podberesky DJ, Serai SD, et al. Use of magnetic resonance elastography to assess hepatic fibrosis in children with chronic liver disease. J Pediatr 2014;164(1):186–188.
5. Venkatesh SK, Yin M, Ehman RL. Magnetic resonance elastography of liver: technique, analysis, and clinical applications. J Magn Reson Imaging 2013;37(3):544–555.
6. Kessler LG, Barnhart HX, Buckler AJ, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 2015;24(1):9–26.
7. Hines CD, Bley TA, Lindstrom MJ, Reeder SB. Repeatability of magnetic resonance elastography for quantification of hepatic stiffness. J Magn Reson Imaging 2010;31(3):725–731.
8. Shi Y, Guo Q, Xia F, Sun J, Gao Y. Short- and midterm repeatability of magnetic resonance elastography in healthy volunteers at 3.0 T. Magn Reson Imaging 2014;32(6):665–670.
9. Yasar TK, Wagner M, Bane O, et al. Interplatform reproducibility of liver and spleen stiffness measured with MR elastography. J Magn Reson Imaging 2016;43(5):1064–1072.
10. Zhang J, Arena C, Pednekar A, et al. Short-term repeatability of magnetic resonance elastography at 3.0T: effects of motion-encoding gradient direction, slice position, and meal ingestion. J Magn Reson Imaging 2016;43(3):704–712.
11. Venkatesh SK, Wang G, Teo LL, Ang BW. Magnetic resonance elastography of liver in healthy Asians: normal liver stiffness quantification and reproducibility assessment. J Magn Reson Imaging 2014;39(1):1–8.
12. Trout AT, Dillman JR, Xanthakos S, et al. Prospective assessment of correlation between US acoustic radiation force impulse and MR elastography in a pediatric population: dispersion of US shear-wave speed measurement matters. Radiology 2016;281(2):544–552.
13. Wallihan DB, Podberesky DJ, Marino BS, Sticka JS, Serai S. Relationship of MR elastography determined liver stiffness with cardiac function after Fontan palliation. J Magn Reson Imaging 2014;40(6):1328–1335.

14. Serai SD, Towbin AJ, Podberesky DJ. Pediatric liver MR elastography. Dig Dis Sci 2012;57(10):2713–2719.

15. Serai SD, Wallihan DB, Venkatesh SK, et al. Magnetic resonance elastography of the liver in patients status-post Fontan procedure: feasibility and preliminary results. Congenit Heart Dis 2014;9(1):7–14.

16. Trout AT, Serai S, Mahley AD, et al. Liver stiffness measurements with MR elastography: agreement and repeatability across imaging systems, field strengths, and pulse sequences. Radiology 2016;281(3):793–804.

17. Obuchowski NA, Reeves AP, Huang EP, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. Stat Methods Med Res 2015;24(1):68–106.

18. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336–341.

19. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.

20. Huang EP, Wang XF, Choudhury KR, et al. Meta-analysis of the technical performance of an imaging procedure: guidelines and statistical methodology. Stat Methods Med Res 2015;24(1):141–174.

21. Bohte AE, Garteiser P, De Niet A, et al. MR elastography of the liver: defining thresholds for detecting viscoelastic changes. Radiology 2013;269(3):768–776.

22. Hines CD, Lindstrom MJ, Varma AK, Reeder SB. Effects of postprandial state and mesenteric blood flow on the repeatability of MR elastography in asymptomatic subjects. J Magn Reson Imaging 2011;33(1):239–244.

23. Jajamovich GH, Dyvorne H, Donnerhack C, Taouli B. Quantitative liver MRI combining phase contrast imaging, elastography, and DWI: assessment of reproducibility and postprandial effect at 3.0 T. PLoS One 2014;9(5):e97355.

24. Lee YJ, Lee JM, Lee JE, et al. MR elastography for noninvasive assessment of hepatic fibrosis: reproducibility of the examination and reproducibility and repeatability of the liver stiffness value measurement. J Magn Reson Imaging 2014;39(2):326–331.

25. Shinagawa Y, Mitsufuji T, Morimoto S, et al. Optimization of scanning parameters for MR elastography at 3.0 T clinical unit: volunteer study. Jpn J Radiol 2014;32(7):441–446.

26. Shire NJ, Yin M, Chen J, et al. Test-retest repeatability of MR elastography for noninvasive liver fibrosis assessment in hepatitis C. J Magn Reson Imaging 2011;34(4):947–955.

27. Wang Y, Ganger DR, Levitsky J, et al. Assessment of chronic hepatitis and fibrosis: comparison of MR elastography and diffusion-weighted imaging. AJR Am J Roentgenol 2011;196(3):553–561.

28. Shin SU, Lee JM, Yu MH, et al. Prediction of esophageal varices in patients with cirrhosis: usefulness of three-dimensional MR elastography with echo-planar imaging technique. Radiology 2014;272(1):143–153.

29. Serai SD, Yin M, Wang H, Ehman RL, Podberesky DJ. Cross-vendor validation of liver magnetic resonance elastography. Abdom Imaging 2015;40(4):789–794.

30. Singh S, Venkatesh SK, Wang Z, et al. Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: a systematic review and meta-analysis of individual participant data. Clin Gastroenterol Hepatol 2015;13(3):440–451.e6.

31. Tang A, Cloutier G, Szeverenyi NM, Sirlin CB. Ultrasound elastography and MR elastography for assessing liver fibrosis. II. diagnostic performance, confounders, and future directions. AJR Am J Roentgenol 2015;205(1):33–40.

32. Yin M, Talwalkar JA, Glaser KJ, et al. Dynamic postprandial hepatic stiffness augmentation assessed with MR elastography in patients with chronic liver disease. AJR Am J Roentgenol 2011;197(1):64–70.

33. Yin M, Talwalkar JA, Glaser KJ, et al. Assessment of hepatic fibrosis with magnetic resonance elastography. Clin Gastroenterol Hepatol 2007;5(10):1207–1213.e2.

34. Venkatesh SK, Ehman RL. Magnetic resonance elastography of liver. Magn Reson Imaging Clin N Am 2014;22(3):433–446.

35. Dzyubak B, Glaser K, Yin M, et al. Automated liver stiffness measurements with magnetic resonance elastography. J Magn Reson Imaging 2013;38(2):371–379.

36. Dzyubak B, Venkatesh SK, Manduca A, Glaser KJ, Ehman RL. Automated liver elasticity calculation for MR elastography. J Magn Reson Imaging 2016;43(5):1055–1063.

37. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC; Group. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. Radiology 2011;258(3):906–914.