Repeatability of MR Elastography of Liver: A Meta-Analysis

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.

© RSNA, 2017
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 as for 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.

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.

Content codes:
- 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
- CI = confidence interval
- ROI = region of interest
- RC = repeatability coefficient
- ROE = 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:

\[ \text{Percentage } RC = 1.96 \times \sqrt{2 \times \text{wCV}}^2, \]

where %wCV is the wCV expressed as a percentage of the mean stiffness value (ie, wSD/\(F \cdot 100)\) 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 calculated the design effect, def, 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:

\[ 2.77 \times \sqrt{\frac{M \times \text{wCV}^2}{k_{0.1}^2}}, \]

where \(k_{0.1}\) is the 11th percentile of the \(\chi^2\) distribution with \(M\) degrees of freedom. For the lower bound, \(\alpha\) is .975, and for the upper bound, \(\alpha\) 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 \(RC\) is the percentage RC from the \(i\)th article and \(M_i\) is the effective sample size from the \(i\)th study (20):

\[ \hat{\theta} = \sqrt{\frac{\sum_{i=1}^{n} RC_i \cdot M_i}{\sum_{i=1}^{n} M_i}}. \]

A bootstrap 95% percentile CI was constructed for each \(\theta\). Let \(N_i\) denote the number of studies in the \(i\)th subgroup. For each of the \(n\) bootstrap samples, \(N_y\) articles were randomly selected with replacement from the original \(N_i\) articles and \(\hat{\theta}_i\) was calculated. A total of 10000 bootstrap samples were constructed. The 250th and 9750th largest values of \(\hat{\theta}_i\) 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
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 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 fell near the summary value of 22% at the top of the plot, and studies with smaller sample sizes fall fairly symmetrically on either side of the bottom of the funnel (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...
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 Bolte et al (21), they reported an intrapatient 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 contribute 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 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 a postdoctoral fellow with 2 years of experience, 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

<table>
<thead>
<tr>
<th>Study No.</th>
<th>Study Design</th>
<th>Sample Size</th>
<th>Age Range (y)</th>
<th>Male Subjects (%)</th>
<th>MR Imaging Unit(s)</th>
<th>Field Strength (T)</th>
<th>Frequency (Hz)</th>
<th>Property Measured* Subjects</th>
<th>No. of Readers</th>
<th>Time Interval</th>
<th>CV (%)</th>
<th>RC (%)</th>
<th>RC 95% CI (%)</th>
<th>Fasting vs Feeding</th>
<th>ROI Drawn by</th>
<th>ROI Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Prospective</td>
<td>5</td>
<td>23–63</td>
<td>44</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>2</td>
<td>2 Weeks</td>
<td>9–12</td>
<td>23</td>
<td>14.3, 56.4</td>
<td>No information</td>
<td>Two radiologists</td>
<td>Three round or oval</td>
</tr>
<tr>
<td>2</td>
<td>Prospective</td>
<td>41</td>
<td>20–57</td>
<td>44</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>2</td>
<td>4–6 Weeks</td>
<td>8.4</td>
<td>4.6</td>
<td>15.3, 31.0</td>
<td>4–6 Hours fasting</td>
<td>One experienced reader (5 y of experience)</td>
<td>Two or three round or oval ROIs</td>
</tr>
<tr>
<td>3</td>
<td>Prospective</td>
<td>9</td>
<td>20–57</td>
<td>44</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>Five healthy subjects, four patients</td>
<td>3</td>
<td>1–2 Weeks</td>
<td>6–11</td>
<td>17</td>
<td>12.2, 28.0</td>
<td>8 Hours fasting</td>
<td>Software</td>
<td>One large</td>
</tr>
<tr>
<td>4</td>
<td>Prospective</td>
<td>10</td>
<td>27–63</td>
<td>90</td>
<td>GE 750 W</td>
<td>3.0</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>1</td>
<td>1 Week</td>
<td>NA</td>
<td>14</td>
<td>7.0, 17.5</td>
<td>No information</td>
<td>Two radiologists</td>
<td>One large</td>
</tr>
<tr>
<td>5</td>
<td>Prospective</td>
<td>15</td>
<td>57 (Mean)</td>
<td>44</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>G* Patients</td>
<td>2</td>
<td>2 Weeks</td>
<td>NA</td>
<td>14</td>
<td>10.3, 21.7</td>
<td>No information</td>
<td>One experienced reader (3 y of experience)</td>
<td>Three round ROIs</td>
</tr>
<tr>
<td>6</td>
<td>Prospective</td>
<td>22</td>
<td>18–56</td>
<td>41</td>
<td>GD HD</td>
<td>3.0</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>2</td>
<td>1 Week (short term); 27–30 weeks (long term)</td>
<td>5.75</td>
<td>15.9</td>
<td>12.7, 21.4</td>
<td>8 Hours fasting</td>
<td>Two radiologists</td>
<td>One large</td>
</tr>
<tr>
<td>7</td>
<td>Retro</td>
<td>47</td>
<td>27–82</td>
<td>6.8</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>G* Patients</td>
<td>2</td>
<td>8–10 Minutes</td>
<td>13</td>
<td>25.3</td>
<td>21.0, 31.7</td>
<td>No information</td>
<td>Two radiologists</td>
<td>Three round or oval</td>
</tr>
<tr>
<td>8</td>
<td>Prospective</td>
<td>30</td>
<td>55.8 (Mean)7.7</td>
<td>11 Healthy subjects, 19 patients</td>
<td>GE 750</td>
<td>3.0</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>2</td>
<td>20 Minutes</td>
<td>3.8</td>
<td>10.5</td>
<td>8.6, 13.4</td>
<td>6 Hours fasting and then repeated after feeding</td>
<td>One radiologist (10 y of experience)</td>
<td>One large</td>
</tr>
<tr>
<td>9</td>
<td>Prospective</td>
<td>30</td>
<td>21–68</td>
<td>53</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>G* 20 Volunteers, 10 patients</td>
<td>2</td>
<td>2–4 Weeks</td>
<td>17.4</td>
<td>37</td>
<td>30.4, 47.3</td>
<td>No information</td>
<td>Operator</td>
<td>One large</td>
</tr>
<tr>
<td>10</td>
<td>Prospective</td>
<td>11</td>
<td>23–39</td>
<td>75</td>
<td>GE HDx</td>
<td>1.5</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>1</td>
<td>5 Weeks</td>
<td>8.5</td>
<td>23.5</td>
<td>16.6, 39.9</td>
<td>Feeding between examinations</td>
<td>Operator</td>
<td>One large</td>
</tr>
<tr>
<td>11</td>
<td>Prospective</td>
<td>30</td>
<td>19–59</td>
<td>60</td>
<td>Philips</td>
<td>3.0</td>
<td>50</td>
<td>G* 16 Volunteers, 14 patients</td>
<td>1</td>
<td>1–4 Weeks</td>
<td>10.1</td>
<td>22.2</td>
<td>17.7, 29.7</td>
<td>No information</td>
<td>One radiologist</td>
<td>One large</td>
</tr>
<tr>
<td>12</td>
<td>Prospective</td>
<td>24</td>
<td>22–55</td>
<td>21</td>
<td>GE and Philips</td>
<td>1.5 And 3.0</td>
<td>60</td>
<td>G* Healthy subjects</td>
<td>1</td>
<td>Same day</td>
<td>10.7</td>
<td>16.6</td>
<td>13.0, 23.1</td>
<td>6–8 Hours fasting</td>
<td>One reviewer (6 y of experience)</td>
<td>One large</td>
</tr>
</tbody>
</table>

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”).
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 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 measurable 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 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...
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.R.S. disclosed no relevant relationships. E.H.M. disclosed no relevant relationships. E.A. Activities related to the present article: disclosed no relevant relationships. F.E.C. 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


