Title
Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis

Permalink
https://escholarship.org/uc/item/5464j698

Journal
RADIOLOGY, 286(2)

ISSN
0033-8419

Authors
Yokoo, T
Serai, SD
Pirasteh, A
et al.

Publication Date
2018-02-01

DOI
10.1148/radiol.2017170550

Peer reviewed
Purpose: To determine the linearity, bias, and precision of hepatic proton density fat fraction (PDFF) measurements by using magnetic resonance (MR) imaging across different field strengths, imager manufacturers, and reconstruction methods.

Materials and Methods: This meta-analysis was performed in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. A systematic literature search identified studies that evaluated the linearity and/or bias of hepatic PDFF measurements by using MR imaging (hereafter, MR imaging–PDFF) against PDFF measurements by using colocalized MR spectroscopy (hereafter, MR spectroscopy–PDFF) or the precision of MR imaging–PDFF. The quality of each study was evaluated by using the Quality Assessment of Studies of Diagnostic Accuracy 2 tool. De-identified original data sets from the selected studies were pooled. Linearity was evaluated by using linear regression between MR imaging–PDFF and MR spectroscopy–PDFF measurements. Bias, defined as the mean difference between MR imaging–PDFF and MR spectroscopy–PDFF measurements, was evaluated by using Bland-Altman analysis. Precision, defined as the agreement between repeated MR imaging–PDFF measurements, was evaluated by using a linear mixed-effects model, with field strength, imager manufacturer, reconstruction method, and region of interest as random effects.

Results: Twenty-three studies (1679 participants) were selected for linearity and bias analyses and 11 studies (425 participants) were selected for precision analyses. MR imaging–PDFF was linear with MR spectroscopy–PDFF ($R^2 = 0.96$). Regression slope (0.97; $P < .001$) and mean Bland-Altman bias (−0.13%; 95% limits of agreement: −3.95%, 3.40%) indicated minimal underestimation by using MR imaging–PDFF. MR imaging–PDFF was precise at the region-of-interest level, with repeatability and reproducibility coefficients of 2.99% and 4.12%, respectively. Field strength, imager manufacturer, and reconstruction method each had minimal effects on reproducibility.

Conclusion: MR imaging–PDFF has excellent linearity, bias, and precision across different field strengths, imager manufacturers, and reconstruction methods.

Online supplemental material is available for this article.
Hepatic steatosis, or intracellular accumulation of triglycerides in hepatocytes, is a common histologic manifestation of many liver diseases. In particular, obesity-related steatosis, or nonalcoholic fatty liver disease, has become one of the leading causes of liver disease worldwide (1) paralleling the obesity pandemic. Nonalcoholic fatty liver disease can progress to cirrhosis and liver cancer, and it is the most rapidly growing indication for liver transplantation in the United States (2). Moreover, even in those patients with nonprogressive disease, nonalcoholic fatty liver disease is a risk factor for future development of diabetes, cardiovascular death, and other cancers (3–9). Control of nonalcoholic fatty liver disease and its complications has thus become a major public health priority.

Percutaneous liver biopsy has been the clinical reference standard for diagnosis and grading of hepatic steatosis (10). However, biopsy is costly, painful, and invasive with rare but serious complication risks including hemorrhage and death (11). In addition, inherently small tissue volume of biopsy (approximately 1/50,000th of the entire liver) is a concern for sampling variability and misclassification of disease severity (12,13). Because of the limitations of biopsy, especially for longitudinal disease monitoring (14,15), interest in noninvasive methods to diagnose and grade hepatic steatosis has increased.

Recently, proton density fat fraction (PDFF) has emerged as the leading noninvasive quantitative imaging biomarker (QIB) of hepatic steatosis (16,17). PDFF is a fundamental tissue property and an objective magnetic resonance (MR) imaging-based measure of tissue triglyceride concentration, calculated as the ratio of MR imaging-visible triglyceride protons to the sum of triglyceride and water protons. Spatially localized MR spectroscopy has been the accepted noninvasive reference standard for quantifying hepatic steatosis, used in epidemiologic studies and randomized controlled clinical trials to derive the highest level of evidence (18–34). Acquired and analyzed by using a standardized approach, MR spectroscopy can measure proton densities of triglyceride and water in a small volume of liver tissue in vivo, from which PDFF is calculated. However, PDFF measurements by using MR spectroscopy can be technically challenging for several reasons, including potential biases because of the selection of sampling volume in livers with nonuniform distribution of fat, difficulty in colocalizing measurement volumes across longitudinal time points, and a need for offline spectral analysis and data quality assessment. To address these technical challenges, advanced chemical shift-encoded MR imaging methods have been developed to automatically “map” hepatic PDFF values pixel by pixel throughout the entire liver. These specialized imaging methods are now commercially available with many 1.5-T and 3.0-T MR imaging systems, and the opportunity for widespread use of hepatic PDFF measurements as a QIB has become a reality.

According to the Radiological Society of North America Quantitative Imaging Biomarkers Alliance (QIBA), three key technical performance metrics of QIBs are linearity, bias, and precision (35). Linearity and bias together assess the degree to which a QIB (eg, PDFF measurements by using MR imaging [hereafter, MR imaging–PDFF]) provides an estimate of the true value (eg, a phantom with known triglyceride concentration) or of an accepted in vivo reference value (eg, PDFF measurements by using MR spectroscopy [hereafter, MR spectroscopy–PDFF]) over the entire range of expected values (eg,
PDFF measurements of approximately 0%-55%, or the range from normal lean liver to the most severe steatosis observed in patients). Precision assesses the agreement between repeated measurements of a QIB (eg, MR imaging–PDFF) and can be reported in two different ways: repeatability, or the agreement between repeated QIB measurements under identical or near-identical conditions (eg, scan-rescan repeatability of MR imaging–PDFF after interscan recalibration), and reproducibility, or the agreement between repeated measurements under different conditions (eg, MR imaging–PDFF by using equipment with different field strengths, imager manufacturers, and/or reconstruction methods).

Multiple previous studies, almost all single center, have shown MR imaging–PDFF to have high agreement with MR spectroscopy–PDFF in terms of the in vivo reference value (30–38), scan-rescan repeatability (39,59–61), cross-imager reproducibility (35,62,63), and cross-field strength reproducibility (48,55,62,63). Despite the excellent performance reported in these studies, there are limited comprehensive data available on the performance of MR imaging–PDFF in multicenter research or clinical settings in which participants may undergo MR imaging–PDFF by using equipment that varies in field strength, imager manufacturer, and/or reconstruction method. An understanding of the technical performance of MR imaging–PDFF in such settings is needed to inform, qualify, and support its use as a QIB for clinical trials and patient care.

Therefore, the purpose of this meta-analysis was to determine the linearity, bias, and precision of hepatic MR imaging–PDFF across different field strengths, imager manufacturers, and reconstruction methods.

### Materials and Methods

No industry funding was used to support this meta-analysis. None of the authors were industry employees. The lead author (T.Y.) had full control of the data and the information submitted for publication.

#### Definition and Criteria of PDFF

PDFF is defined as follows:

$$\text{PDFF} = \frac{\Sigma \text{PD}_{\text{WP}}}{(\Sigma \text{PD}_{\text{WP}} + \Sigma \text{PD}_{\text{FP}})}$$

where PD is MR imaging–visible proton density, or equivalently the spectral peak area, of the water molecules having a single resonance frequency of 4.7 ppm, or the triglyceride molecules having multiple frequencies as described elsewhere; WP is fat peak; and WP is water peak (64). Various MR imaging–based methods of PDFF measurement have been proposed, including chemical shift two- and three-dimensional spoiled gradient-recalled echo sequences at 1.5 T and 3.0 T by using different reconstruction methods (45,49,58,65), as detailed in Appendix E1 (online).

The QIBA PDFF Biomarker Committee currently adopts the following protocol design criteria for MR imaging– and MR spectroscopy–based methods to measure PDFF (detailed in Table E1 [online]). Briefly, three major confounders of PDFF must be either minimized or corrected (66,67): the T1 relaxation effect, the T2 or T2* relaxation effects, and multiple proton resonance frequencies of triglycerides (so-called spectral complexity). Various pulse sequence–specific confounders (eg, phase errors [68–72] and J coupling [73]) also need to be addressed either at acquisition or during reconstruction steps.

#### Literature Search

This study was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (74,75). Separate literature searches were performed for two pooled analyses for performance of MR imaging–PDFF: (a) linearity and bias and (b) precision. A systematic search of PubMed/MEDLINE, the Cochrane Library, and the Web of Science databases was performed on February 22, 2017, by a senior radiology resident (A.P.) to identify primary research studies satisfying inclusion criteria (or equivalent, depending on the syntax requirement of the specific search engine, listed in Tables E2 and E3 [online]) as follows: (a) linearity and bias analysis: liver “fat fraction” (imaging AND spectroscopy) “magnetic resonance” NOT review [publication type]; (b) precision analysis: liver “fat fraction” imaging (repeatability OR reproducibility OR precision OR agreement) “magnetic resonance” NOT review [publication type].

Only in vivo studies on humans that were either published in or translated to English were included.

#### Study Selection

Titles and abstracts, followed by the full text of these eligible studies, were screened by the same author performing the systematic search (A.P.) and then independently verified by another author (T.Y., with 10 years of experience in MR imaging fat quantification methods) using the following exclusion criteria: (a) For either analysis: secondary analysis of previously published data, unable to verify or did not meet the above criteria for PDFF; (b) linearity and bias analysis: PDFF measurements were not performed by using both MR imaging and MR spectroscopy; (c) precision analysis: multiple PDFF measurements were not performed per participant.

#### Data Collection and Quality Assessment

After all articles satisfying the selection criteria were identified, the corresponding authors of these articles were invited to submit anonymized individual participant data from MR imaging–PDFF and MR spectroscopy–PDFF for a meta-analysis. The local investigational review board of each participating institution either approved or waived formal review for the transfer and central analysis of anonymized study data. For each participant’s MR imaging–PDFF and MR spectroscopy–PDFF measurements, data listed in Table 1 were recorded in a pooled database. Data quality of included studies (ie, the bias and applicability of each study) was assessed by one author (S.D.S.) using the Quality Assessment of Studies of Diagnostic Accuracy 2 tool (76,77). The possibility of publication bias across studies was assessed by using funnel plots and Egger test (78).
EVIDENCE-BASED PRACTICE: Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements

Yokoo et al

Statistical Analysis

The pooled data were analyzed by using R version 3.1.3 (R Foundation for Statistical Computing, Vienna, Austria) by one author (T.Y., with >10 years of experience in statistical computing) under supervision of an expert biostatistician (N.A.B., with >20 years of experience).

Linearity was initially evaluated by using a second-degree polynomial regression model of MR imaging–PDFF against MR spectroscopy–PDFF. A quadratic (second-order) term was first evaluated in the model; if the quadratic term did not reach statistical significance at the α level of .05 or if its relative effect size was two orders of magnitude smaller than the linear (first-order) term, then the model was reduced to linear regression and the first- and zeroth-order terms (ie, slope and intercept) were estimated. A linear mixed-effects model was used for linear regression to account for clustered measurements within the same participants. The coefficient of determination ($R^2$) was calculated as the strength metric of linearity (35).

Bias, defined as the average difference between MR imaging–PDFF and MR spectroscopy–PDFF measurements per participant (by using MR spectroscopy as the reference technique) was evaluated by using Bland-Altman analysis (79). The 95% limits of agreement were calculated. To determine the contributing factors to bias, a linear mixed-effects model was used with field strength, imager manufacturer, and reconstruction method as fixed effects and with participant, ROI (nested within participant), examination, and acquisition as random effects.

Precision, defined as the closeness of agreement between repeated MR imaging–PDFF measurements (35), was evaluated by using a linear mixed-effects model, with field strength, reconstruction, participant, ROI (nested within participant), examination, and acquisition as random effects. Precision was measured as the standard deviation of PDFF measurement associated with each random effect. Repeatability was assessed as the precision under varying circumstances (ie, different field strength, imager manufacturer, and reconstruction method), such as in a scan-rescan setting. Reproducibility was assessed as precision under varying circumstances (ie, different field strength, imager manufacturer, and/or reconstruction method), such as would be encountered in a multicenter setting. Both participant-level and ROI-level repeatability and reproducibility were assessed, because differences in ROI placement within the liver may contribute to PDFF variability due to biologic factors (rather than technical factors), especially in participants with nonuniform distribution of hepatic fat.

Heterogeneity of bias or precision across sites was not explicitly evaluated because differences in sites are modeled by (and expected to be strongly correlated with) the differences in the MR imaging equipment (ie, field strength, imager manufacturer) or reconstruction method. The 95% confidence intervals (CIs) and/or $P$ values were computed for each statistic when appropriate. $P$ values < .05 were considered to indicate statistical significance.

Results

Study Selection and Subjects

A flow diagram summarizing study selection for inclusion and exclusion criteria according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines is shown in Figure 1. Twenty-eight studies fulfilled the selection criteria. Anonymized data sets from all 28 selected studies were submitted for quantitative synthesis.

The pooled data included 16624 MR imaging–PDFF measurements in 1960 unique participants. For three studies (36,38,63) data were unavailable on the age and/or sex of 48, 50, and 10 participants, respectively. Mean participant age was 43 years (range, 8–89 years), with a man-to-woman ratio of 52.4. Mean MR imaging–PDFF value was 9.6% (range, −2.8% to 55.4%). A total of 1679 participants from 23 different studies were included in the linearity and bias analysis (36–58). 423 participants from 11 studies were included in the precision analysis (36,41,48,53,55,56,59–63), and 195 participants from six studies were included in both analyses.

Table 2 summarizes the characteristics of the included studies.

Quality Assessment and Publication Bias

Figure 2 summarizes quality assessment of the included studies by using the Quality Assessment of Studies of Diagnostic Accuracy 2 tool. No study met criteria for high risk of bias or applicability concern. Categories 2, 3, and 4 had greater than 90% compliance, suggesting that the studies were adequately blinded to

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field strength</td>
<td>1.5 T or 3.0 T</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>GE, Siemens, or Philips</td>
</tr>
<tr>
<td>Reconstruction method</td>
<td>Magnitude, complex, or hybrid</td>
</tr>
<tr>
<td>No. of examinations</td>
<td>Repeated image acquisitions (if any) in two or more examinations</td>
</tr>
<tr>
<td></td>
<td>by using independent instrument setup and/or calibration per examination</td>
</tr>
<tr>
<td>No. of acquisitions</td>
<td>Repeated image acquisitions (if any) during a single examination</td>
</tr>
<tr>
<td></td>
<td>by using identical instrument setup and/or calibration</td>
</tr>
<tr>
<td>ROI</td>
<td>Repeated ROI placement in different locations in the liver per methods described in each study</td>
</tr>
<tr>
<td>MR imaging–PDFF (%)</td>
<td>Average PDFF within an ROI</td>
</tr>
<tr>
<td>MR spectroscopy–PDFF (%)</td>
<td>PDFF measurement colocalized to MR imaging ROI, if available</td>
</tr>
</tbody>
</table>

Note.—ROI = region of interest.
the reference standard and variations in reference standard were minimal. Categories 1 and 5 had approximately 60% compliance because some studies used case-control selection based on age, sex, and/or status of the following risk factors: healthy, obesity, diabetes, and/or nonalcoholic steatohepatitis. Because these risk factors were not considered to be high risk for bias or applicability for this technical validation study, no specific subanalyses were performed. Funnel plots and Egger test P values are shown in Figure E2 (online). No statistically significant asymmetry was found to indicate publication bias (all P values > .5).

Assessment of Linearity
Figure 3a illustrates the relationship between the 3191 paired MR imaging–PDFF and MR spectroscopy–PDFF measurements. In the second-degree polynomial model, the quadratic term was smaller than two orders of magnitude than was the linear term (ratio of quadratic to linear terms, 0.002), so it was removed and the model was reduced to a first-degree linear model. Linearity was strong (R^2 = 0.96). The
Table 2

Characteristics of 28 Selected Studies

<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Study Region</th>
<th>Data Collection</th>
<th>No. of Participants</th>
<th>Mean Age (y)</th>
<th>Manufacturer</th>
<th>Field Strength (T)</th>
<th>Reconstruction Method</th>
<th>PDFF Minimum (%)</th>
<th>Mean PDFF (%)</th>
<th>PDFF Maximum (%)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hines (56)</td>
<td>2011</td>
<td>North America</td>
<td>Prospective</td>
<td>42</td>
<td>51</td>
<td>GE</td>
<td>1.5</td>
<td>Hybrid</td>
<td>-2.8</td>
<td>5.0</td>
<td>39.5</td>
<td>Both</td>
</tr>
<tr>
<td>Kang (55)</td>
<td>2011</td>
<td>North America</td>
<td>Prospective</td>
<td>21</td>
<td>55</td>
<td>GE, Siemens</td>
<td>1.5, 3.0</td>
<td>Magnitude</td>
<td>-0.1</td>
<td>15.1</td>
<td>41.5</td>
<td>Both</td>
</tr>
<tr>
<td>Johnson (53)</td>
<td>2014</td>
<td>North America</td>
<td>Prospective</td>
<td>30</td>
<td>38</td>
<td>GE</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.5</td>
<td>12.1</td>
<td>28.7</td>
<td>Both</td>
</tr>
<tr>
<td>Artz (48)</td>
<td>2015</td>
<td>North America</td>
<td>Prospective</td>
<td>25</td>
<td>48</td>
<td>GE</td>
<td>1.5, 3.0</td>
<td>Hybrid, magnitude</td>
<td>0.2</td>
<td>10.8</td>
<td>34.9</td>
<td>Both</td>
</tr>
<tr>
<td>Trajci (41)</td>
<td>2015</td>
<td>North America</td>
<td>Prospective</td>
<td>29</td>
<td>24</td>
<td>GE</td>
<td>3.0</td>
<td>Hybrid, magnitude</td>
<td>0.1</td>
<td>13.0</td>
<td>37.3</td>
<td>Both</td>
</tr>
<tr>
<td>Matsuzuki (36)</td>
<td>2017</td>
<td>North America</td>
<td>Prospective</td>
<td>46</td>
<td>45</td>
<td>GE</td>
<td>1.5</td>
<td>Hybrid, magnitude</td>
<td>-0.2</td>
<td>6.2</td>
<td>27.7</td>
<td>Both</td>
</tr>
<tr>
<td>Meisamy (57)</td>
<td>2011</td>
<td>North America</td>
<td>Prospective</td>
<td>53</td>
<td>70</td>
<td>GE</td>
<td>1.5</td>
<td>Hybrid</td>
<td>0.5</td>
<td>6.6</td>
<td>25.6</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Yokoo (58)</td>
<td>2011</td>
<td>North America</td>
<td>Prospective</td>
<td>163</td>
<td>39</td>
<td>GE</td>
<td>3.0</td>
<td>Magnitude</td>
<td>0.0</td>
<td>10.6</td>
<td>46.0</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Matsuzuki (47)</td>
<td>2015</td>
<td>North America</td>
<td>Prospective</td>
<td>50</td>
<td>56</td>
<td>GE</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.0</td>
<td>6.6</td>
<td>36.7</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Le (54)</td>
<td>2012</td>
<td>North America</td>
<td>Prospective</td>
<td>48</td>
<td>48</td>
<td>GE</td>
<td>3.0</td>
<td>Magnitude</td>
<td>1.9</td>
<td>18.4</td>
<td>55.4</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Levin (52)</td>
<td>2014</td>
<td>North America</td>
<td>Prospective</td>
<td>84</td>
<td>36</td>
<td>GE</td>
<td>3.0</td>
<td>Magnitude</td>
<td>-0.8</td>
<td>10.8</td>
<td>32.9</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Kühn (51)</td>
<td>2014</td>
<td>Europe</td>
<td>Prospective</td>
<td>50</td>
<td>57</td>
<td>Siemens</td>
<td>3.0</td>
<td>Hybrid</td>
<td>-0.2</td>
<td>13.1</td>
<td>43.5</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Kühn (50)</td>
<td>2014</td>
<td>Europe</td>
<td>Prospective</td>
<td>33</td>
<td>57</td>
<td>Siemens</td>
<td>3.0</td>
<td>Hybrid</td>
<td>3.1</td>
<td>14.0</td>
<td>46.2</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Bashir (49)</td>
<td>2015</td>
<td>North America</td>
<td>Prospective</td>
<td>42</td>
<td>53</td>
<td>Siemens</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.3</td>
<td>8.6</td>
<td>30.1</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Rehm (44)</td>
<td>2015</td>
<td>North America</td>
<td>Prospective</td>
<td>132</td>
<td>13</td>
<td>GE</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.8</td>
<td>4.4</td>
<td>39.1</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Zand (46)</td>
<td>2015</td>
<td>North America</td>
<td>Retrospective</td>
<td>286</td>
<td>14</td>
<td>GE</td>
<td>3.0</td>
<td>Magnitude</td>
<td>-0.1</td>
<td>12.2</td>
<td>41.2</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Kukuk (45)</td>
<td>2015</td>
<td>Europe</td>
<td>Prospective</td>
<td>59</td>
<td>52</td>
<td>Philips</td>
<td>3.0</td>
<td>Complex</td>
<td>0.0</td>
<td>12.8</td>
<td>43.5</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Kramer (38)</td>
<td>2016</td>
<td>North America</td>
<td>Prospective</td>
<td>50</td>
<td>56</td>
<td>GE</td>
<td>1.5</td>
<td>Hybrid</td>
<td>1.0</td>
<td>7.1</td>
<td>39.3</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Kim (43)</td>
<td>2015</td>
<td>Asia</td>
<td>Retrospective</td>
<td>156</td>
<td>58</td>
<td>Siemens</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.3</td>
<td>5.2</td>
<td>36.0</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Hettinger (42)</td>
<td>2016</td>
<td>Europe</td>
<td>Prospective</td>
<td>215</td>
<td>57</td>
<td>Siemens</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.6</td>
<td>9.2</td>
<td>37.5</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Wu (40)</td>
<td>2016</td>
<td>North America</td>
<td>Prospective</td>
<td>35</td>
<td>60</td>
<td>Philips</td>
<td>3.0</td>
<td>Magnitude</td>
<td>1.0</td>
<td>14.3</td>
<td>38.6</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Cui (39)</td>
<td>2016</td>
<td>North America</td>
<td>Prospective</td>
<td>50</td>
<td>54</td>
<td>GE</td>
<td>3.0</td>
<td>Magnitude</td>
<td>5.1</td>
<td>16.8</td>
<td>33.3</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Park (37)</td>
<td>2016</td>
<td>North America</td>
<td>Prospective</td>
<td>31</td>
<td>53</td>
<td>GE</td>
<td>3.0</td>
<td>Magnitude</td>
<td>0.5</td>
<td>7.6</td>
<td>32.3</td>
<td>Linearity and bias</td>
</tr>
<tr>
<td>Wu (59)</td>
<td>2016</td>
<td>Asia</td>
<td>Prospective</td>
<td>15</td>
<td>33</td>
<td>GE</td>
<td>3.0</td>
<td>Hybrid</td>
<td>-1.8</td>
<td>3.5</td>
<td>10.7</td>
<td>Precision</td>
</tr>
<tr>
<td>Mashood (63)</td>
<td>2013</td>
<td>North America</td>
<td>Prospective</td>
<td>10</td>
<td>50</td>
<td>GE, Siemens</td>
<td>1.5, 3.0</td>
<td>Magnitude</td>
<td>-0.1</td>
<td>8.1</td>
<td>39.2</td>
<td>Precision</td>
</tr>
<tr>
<td>Negrete (61)</td>
<td>2014</td>
<td>North America</td>
<td>Prospective</td>
<td>29</td>
<td>24</td>
<td>GE</td>
<td>3.0</td>
<td>Hybrid</td>
<td>-0.1</td>
<td>11.8</td>
<td>40.5</td>
<td>Precision</td>
</tr>
<tr>
<td>Sokul (60)</td>
<td>2015</td>
<td>North America</td>
<td>Prospective</td>
<td>150</td>
<td>57</td>
<td>Siemens</td>
<td>3.0</td>
<td>Hybrid</td>
<td>0.5</td>
<td>6.9</td>
<td>39.3</td>
<td>Precision</td>
</tr>
<tr>
<td>Serai (62)</td>
<td>2017</td>
<td>North America</td>
<td>Prospective</td>
<td>24</td>
<td>39</td>
<td>GE, Philips</td>
<td>1.5, 3.0</td>
<td>Complex, hybrid</td>
<td>0.3</td>
<td>5.8</td>
<td>26.0</td>
<td>Precision</td>
</tr>
</tbody>
</table>
compared with those obtained with ei
approximately 2% higher PDFF values
The estimated intercept of \(-0.07\) was not
significantly different from zero \((P =
0.70; 95\% CI: -0.50, 0.32)\). The esti-
mated slope of the regression line was
significantly below unity at 0.97
\((P < .001; 95\% CI: 0.96, 0.98)\), in-
dicating underestimation by using MR
imaging–PDFF compared with MR
spectroscopy-PDFF (corresponds to
about 1.5% underestimation at 50%
MR spectroscopy-PDFF).

Assessment of Bias

Figure 3b illustrates the MR imaging–
PDFF measurement bias. The mean
bias was small \((-0.13\%; 95\% limits
of agreement: -3.95\%, 3.70\%)\). In the
analysis of individual bias components
(Table 3), all effects except manufac-
turer (between GE and Siemens) had
statistically significant effects on bias,
but most bias components had effects
smaller than 1.5% in absolute PDFF.
Bias because of manufacturer for the
Philips system was the exception, with
approximately 2% higher PDFF values
compared with those obtained with ei-
ther GE or Siemens systems.

Assessment of Precision

Figure 4 illustrates the precision pro-
files, represented as the difference
between the repeated measurements as
a function of the per-participant (Fig 4a)
and per-ROI (Fig 4b) mean MR imag-
ing–PDFF values based on 9103 mea-
surements in 425 participants.

Under repeatability conditions
(same field strength, imager manu-
facturer, and reconstruction method),
within-participant standard deviation
was 1.69% and within-ROI standard
deviation was 1.08% in absolute PDFF
value. Under reproducibility conditions
(different field strength, imager manu-
facturer, or reconstruction method),
within-participant standard deviation
was 1.97% and within-ROI standard
deviation was 1.48% in absolute PDFF
value.

The estimated variance compo-
nents from linear mixed-effects models
(Fig 4) indicated that the main contrib-
utors to MR imaging–PDFF measure-
ment variability were ROI locations
(standard deviation, 1.30%) and ran-
dom measurement error because of
repeated acquisitions (standard devi-
ation, 1.07%). Other technical factors
(field strength, imager manufacturer,
and reconstruction method) invariably
had smaller standard deviations of less
than 0.8% in absolute PDFF value.

Discussion

Our meta-analysis included nearly 2000
participants in 28 independent pri-
mary research studies from geographi-
cally diverse sites that varied in field
strength, imager manufacturer, and
reconstruction method. These studies
were generally considered satisfactory
in quality per Quality Assessment of
Studies of Diagnostic Accuracy 2 crite-
ria with minimal bias in index test (MR
imaging–PDFF) and reference standard
(MR spectroscopy-PDFF for linearity
and bias analyses and repeated MR
imaging–PDFF for precision analyses).
Although patient selection criteria in
some studies were not uniform because
definition of emphasis in different subpopu-
lations of various age range, sex, and steato-
sis risk factors (eg, obesity, diabetes),
the potential impact of the variability is
thought to be small in this meta-analy-
sis of technical validation.

Results of our meta-analysis demon-
strated that MR imaging–PDFF has
excellent linearity and negligible bias
with respect to the reference stan-
dard of MR spectroscopy-PDFF mea-
surements over the entire range of
observed steatosis severity. A slight
deviation of the MR imaging versus
MR spectroscopy regression from per-
fact agreement was statistically sig-
nificant, but the effect size was small
(up to 1.5% absolute PDFF value in
very high PDFF range) and is unlikely
to be meaningful either clinically or
in research studies. In addition, our
results demonstrated that the largest
contributors to the variability in MR
imaging–PDFF measurements (repeat-
ability and reproducibility) were inher-
ent heterogeneity of steatosis across
the liver (ie, because of different ROI
locations; standard deviation, 1.3% in
PDFF) and the random measurement
error (1%), with smaller contributions
(<0.8%) from technical factors such
EVIDENCE-BASED PRACTICE: Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements

Yokoo et al

population-based cohort trials (including studies with more than 1000 participants) (85–87). U.S. Food and Drug...
EVIDENCE-BASED PRACTICE: Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements

Yokoo et al

Our work was motivated by the activities of QIBA, whose mission is to improve the practical value of QIBs by promoting standardization to reducing variability between hardware and software, thereby facilitating the use of QIBs in clinical trials and patient care. These goals are achieved by preparing a QIBA profile document that is intended to span various commonly encountered potential confounders, such as field strength, imager manufacturer, and reconstruction methods, so that expected variability is included to the maximum extent possible. This profile document is intended for a broad audience, including imager 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 linearity, bias, and precision claimed in the profile to be realistic and reasonably achievable across imaging centers and readers spanning a relevant range of technical variations. Our meta-analysis aims to provide generalizable technical performance standards to inform future QIBA profile statements.

This study had several strengths and limitations. Major strengths of this meta-analysis included a large and geographically diverse cohort, multiple studies conducted by independent research groups, and variability in hardware and software as expected in real-world applications. An important and unique strength of this work was the close collaboration among the authors of the selected articles that enabled pooled analyses directly on the collective MR imaging–PDFF and MR spectroscopy-PDFF individual participant data, rather than the summary statistics extracted from the published articles as is done in a traditional meta-analysis. The main limitation of this study was that we used MR spectroscopy as the reference standard for linearity and bias analyses, rather than a tissue-based reference standard. However, MR spectroscopy has been widely accepted as the reference standard for hepatic fat quantification.
in research studies to derive highest levels of evidence, including randomized control trials and epidemiologic studies (18–34). Considering the large number of studies available comparing MR imaging and MR spectroscopy, MR spectroscopy is probably the most widely available reference, particularly in the light of high variability of biopsy to quantify hepatic steatosis (12,88).

In summary, our meta-analysis of pooled data collected from 28 published studies demonstrated excellent linearity, negligible bias, and high precision of MR imaging–PDFF across different field strengths, imager manufacturers, and reconstruction methods. We conclude that MR imaging–PDFF has excellent technical performance characteristics for widespread use in clinical trials and patient care.

Author affiliations: Advanced Imaging Research Center, University of Texas Southwestern Medical Center, Dallas, Tex (T.Y.); Department of Radiology, Cincinnati Children’s Hospital and Medical Center, Cincinnati, Ohio (S.D.S.); Department of Radiology and Center for Advanced Magnetic Resonance Development, Duke University Medical Center, Durham, NC (M.B.R.); Department of Radiology and Liver Imaging Group, University of California at San Diego, San Diego, Calif (G.H., M.S.M., C.B.S.); Division of Gastroenterology, Department of Medicine, University of California, San Diego, La Jolla, Calif (R.L.); Division of Epidemiology, Department of Family Medicine and Preventive Medicine, University of California, San Diego, La Jolla, Calif (R.L.); Departments of Radiology and Medical Physics, University of Wisconsin, Madison, Wis (D.H., S.B.R.); Biomedical Engineering, Medicine and Emergency Medicine, University of Wisconsin, Madison, Wis (S.B.R.); Department of Radiology, Phoenix Children’s Hospital, Phoenix, Ariz (H.H.H.); Institute of Clinical Radiology, Hospital of the Ludwig-Maximilian University, Munich, Germany (H.H.H.); Institute of Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany (J.P.K.); Department of Radiology, University Hospital, Carl Gustav Carus University, Dresden, Germany (J.P.K.); Department of Radiology, Rheinische Friedrich-Wilhelms Universität, Bonn, Germany (G.M.K.); Quantitative Health Sciences, The Cleveland Clinic Foundation, Cleveland, Ohio (N.A.O.); Department of Radiology, Chonbuk National University Medical School and Hospital, Jeongu, Chonbuk, Korea (J.S.S.); Department of Radiology, Université de Montréal, Montréal, QC, Canada (A.T.); and Department of Radiology, Beijing MILitary General Hospital, Beijing, China (X.W.).

Acknowledgments: The authors also thank GE, Siemens, and Philips who provide research support to many of the institutions included in this meta-analysis.

The members of the RSNA-QIBA PDFF Biomarker Committee are Ángel Alberich-Bayarri, PhD, Quantitative Imaging Biomarkers in Medicine (QUIBIM); Mustafa Bashir, MD, Duke University; David Bennett, PhD, PAREXEL; Thomas L. Chenevert, PhD, University of Michigan Health System; Patricia E. Cole, MD, PhD, Take-da Pharmaceuticals; Cathy Elsinger, PhD, NordicNeuroLab; Manuela França, MD, Hospital de San Antonio; Alexander Guimaraes, PhD, PhD, Oregon Health & Science University; Gavin Hamilton, PhD, University of California, San Diego (UCSD); Diego Hernando, PhD, University of Wisconsin-Madison; Harry H. Hn, PhD, Phoenix Children’s Hospital; Edward F. Jackson, PhD, University of Wisconsin, School of Medicine and Public Health; Tommy Johansson, Advanced MR Analytics (MR) (Osl, Norway); So Yeon Kim, MD, Adan Medical Center, Korea; Sonja Kinner, MD, University of Duisburg-Essen (Duisburg, Germany); Hendrik Lane, PhD, Fraunhofer MEVIS, Institute for Medical Image Computing (Germany); Olof Dahlqvist Leinhard, PhD, Advanced MR Analytics (MR) (Oslo, Norway); Dariya Malyarenko, PhD, University of Michigan; Luis Marti-Bonmati, MD, PhD, Quirón Hospital (Valencia, Spain); Michael Middelton, MD, University of California, San Diego (UCSD); Carlo E. Neuwaller, MD, IRCCS-Azienda Ospedaliera Universitaria Sant’ Martino-Istituto Nazionale Cancer Institute, Genoa; Nancy Olochowski, PhD, Cleveland Clinic Foundation; Ali Pirasteh, MD, University of Texas Southwestern Medical Center; Balu Rajagopalan, PhD, John Muir Health; Chelsea Ranger, PA, MHSc, Advanced MR Analytics (MR) (Oslo, Norway); Scott B. Reeder, MD, PhD, University of Wisconsin-Madison; Mark Rosen, MD, PhD, University of Pennsylvania; Nozaki Seiji, PhD, Toshiba; Suraj Sirsi, PhD, Cincinnati Children’s Hospital; Claude Sirlin, MD, University of California, San Diego (UCSD); Judd Storrs, PhD, University of Mississippi Medical Center; Satoshi Sugiiura, PhD, Toshiba; Suiji Yamanoto, PhD, National Cancer Center (Japan); Takeshi Yokoo, MD, PhD, University of Texas, Southwestern Medical Center; Hui-Jing Yu, PhD, BioClinica; Gudrun Zahlmann, PhD, Institute of Electrical and Electronics Engineers (IEEE).

Disclosures of Conflicts of Interest: T.Y. disclosed no relevant relationships. S.D.S. disclosed no relevant relationships. A.P. disclosed no relevant relationships. M.B.R. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author reports grants from Siemens, GE, Guerbet, NGM Biopharmaceuticals, Taiwan Pharmaceuticals, and Madrigal Pharmaceuticals; personal fees from RadMD. Other relationships: disclosed no relevant relationships. G.H. disclosed no relevant relationships. D.H. disclosed no relevant relationships. H.H.H. disclosed no relevant relationships. J.P.K. disclosed no relevant relationships. M.R.B. disclosed no relevant relationships. R.L. Activities related to the present article.
article: author reports grant from Siemens. Activities not related to the present article: author reports grants from Siemens and GE. Other relationships: disclosed no relevant relationships.

M.S.M. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author reports contract work through university for Alexion, AstraZeneca, Bioinclusia, Biomedical Systems, Bristol-Myers Squibb, Galmed, GE, Gilead, ICON, Intercept, Isis, Janssen, NuSirt, Pfizer, Profi, Sanofi Genzyme, Shire, Siemens, Takeda, and Viralnterconcepts; discussions about contract work for Celgene, Genentech, Medien, Perspectum, and Zydus; grants from GE, Gilead, and Guebert; stockholder in GE and Pfizer; two patents pending. Other relationships: disclosed no relevant relationships. N.A.O. Activities related to the present article: author reports grant from National Institute of Biomedical Imaging and Bioengineering. Activities not related to the present article: disclosed consultancies with Siemens, QTUS, and Elfucid on breast cancer studies with the Cleveland Clinic Foundation. Other relationships: disclosed no relevant relationships. J.S.S. disclosed no relevant relationships. A.T. disclosed no relevant relationships. X.W. disclosed no relevant relationships. S.R.B. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author disclosed that he is a founder of Calimetrix, LLC; has stock ownership in Celllectar Biosciences; is a shareholder in Elucnt Medical; is a consultant for Parexel and Virtualscopics; discussions about contracted work for Celgene, Genentech, Median, Perspectum, and Zydus; grants from GE, Gilead, and Guebert; stockholder in GE and Pfizer; two patents pending. Other relationships: disclosed no relevant relationships. A.T. disclosed no relevant relationships. X.W. disclosed no relevant relationships. S.R.B. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author disclosed that he is a founder of Calimetrix, LLC; has stock ownership in Celllectar Biosciences; is a shareholder in Elucnt Medical; is a consultant for Parexel and Virtualscopics; discussions about contracted work for Celgene, Genentech, Median, Perspectum, and Zydus; grants from GE, Gilead, and Guebert; stockholder in GE and Pfizer; two patents pending. Other relationships: disclosed no relevant relationships. C.B.S. disclosed no relevant relationships.

References

6. Guzman G, Brun ET, Petrovic LM, Chejgc J, Layden TJ, Cotler SJ. Does non-alcoholic fatty liver disease predispose pa-


31. Sullivan S, Kirk EP, Mittendorf B, Patterson BW, Klein S. Randomized trial of exercise effect on intrahepatic triglyceride con-


