Radiology

An earlier incorrect version of this article appeared online. This article was corrected on October 2, 2017.

Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis[1]

Takeshi Yokoo, MD, PhD
Suraj D. Serai, PhD
Ali Pirasteh, MD
Mustafa R. Bashir, MD
Gavin Hamilton, PhD
Diego Hernando, PhD
Houchun H. Hu, PhD
Holger Hetterich, MD
Jens-Peter Kühn, MD
Guido M. Kukuk, MD
Rohit Loomba, MD
Michael S. Middleton, MD, PhD
Nancy A. Obuchowski, PhD
Ji Soo Song, MD, PhD
An Tang, MD, MSc
Xinhuai Wu, MD
Scott B. Reeder, MD, PhD
Claude B. Sirlin, MD
For the RSNA-QIBA PDFF Biomarker Committee[2,3]

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.

[1] From the Department of Radiology, University of Texas Southwestern Medical Center, 2201 Inwood Rd, NE2.210B, Dallas, TX 75390-9085 (T.Y., A.P.). Received March 9, 2017; revision requested April 21; revision received May 4; accepted June 5; final version accepted June 8. **Address correspondence to** T.Y. (e-mail: Takeshi.Yokoo@UTSouthwestern.edu).

Quantitative Imaging Biomarkers Alliance projects and activities have been funded in whole or in part with Federal funds from National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Department of Health and Human Services (HHSN268201000050C, HHSN268201300071C, HHSN268201500021C). T.Y supported in part by research grant from Radiological Society of North America (RSCH1625). S.B.R. supported in part by research grants from National Institutes of Health (R01DK083380, R01DK088925, R01DK100651, K24DK102595). A.T. supported in part by Fonds de recherche du Québec-Santé (Career Award #26993). C.B.S. supported in part by research grants from National Institutes of Health (R01DK106419, R01DK088925, R01DK110096, U01DK061734).

[2] The complete list of authors and affiliations is at the end of this article.

[3] The members of the RSNA-QIBA PDFF Biomarker Committee are listed in the Acknowledgments.

S.B.R. and C.B.S. contributed equally to this work.

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,

Advances in Knowledge

- Hepatic proton density fat fraction (PDFF) measured by using MR imaging (hereafter, MR imaging–PDFF) has excellent pooled linearity against the reference values measured by using MR spectroscopy (hereafter, MR spectroscopy-PDFF), with a linear model coefficient of determination of 0.96.
- The pooled linear regression analysis of data from MR imaging–PDFF and MR spectroscopy-PDFF demonstrated intercept of −0.07% and slope of 0.97, indicating minimal underestimation by using MR imaging–PDFF.
- The pooled bias, or the mean difference between MR imaging–PDFF and MR spectroscopy-PDFF, was −0.13%.
- Hepatic MR imaging–PDFF is highly precise at the region-of-interest level, with repeatability and reproducibility coefficients of 2.99% and 4.12%, respectively.

Implications for Patient Care

- As measured in published studies, hepatic MR imaging–PDFF has excellent technical performance as a quantitative imaging biomarker for widespread use in clinical trials and patient care.
- Hepatic MR imaging–PDFF can be measured with excellent linearity, negligible bias, and high precision by using different field strengths, imager manufacturers, and reconstruction methods.

https://doi.org/10.1148/radiol.2017170550

Content codes: GI MR

Radiology 2018; 286:486–498

Abbreviations:
CI = confidence interval
PDFF = proton density fat fraction
QIB = quantitative imaging biomarker
QIBA = Quantitative Imaging Biomarkers Alliance
ROI = region of interest

Author contributions:
Guarantors of integrity of entire study, T.Y., S.D.S., M.S.M., S.B.R.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, T.Y., S.D.S., A.P., M.R.B., G.H., D.H., H.H.H., H.H., J.P.K., G.M.K., R.L., M.S.M., X.W., S.B.R., C.B.S.; clinical studies, T.Y., S.D.S., H.H., J.P.K., G.M.K., A.T., X.W., S.B.R.; experimental studies, S.D.S., G.M.K., R.L., X.W.; statistical analysis, T.Y., S.D.S., M.R.B., J.P.K., N.A.O., S.B.R.; and manuscript editing, T.Y., S.D.S., A.P., M.R.B., G.H., D.H., H.H.H., H.H., J.P.K., G.M.K., R.L., M.S.M., N.A.O., J.S.S., A.T., S.B.R., C.B.S.

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

See also the editorial by Jackson in this issue.

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 (36–58), scan-rescan repeatability (53,56,59–61), cross-imager reproducibility (55,62,63), 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:

$$\text{PDFF} = \Sigma_{\text{all}}\,\text{PD}_{\text{FP}} \big/ \left(\text{PD}_{\text{WP}} + \Sigma_{\text{all}}\,\text{PD}_{\text{FP}}\right),$$

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; FP 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).

Table 1

Data Collected from Included Studies

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

Note.—ROI = region of interest.

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, manufacturer, 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 of using identical equipment (same 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%:48%. 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), 425 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

Figure 1

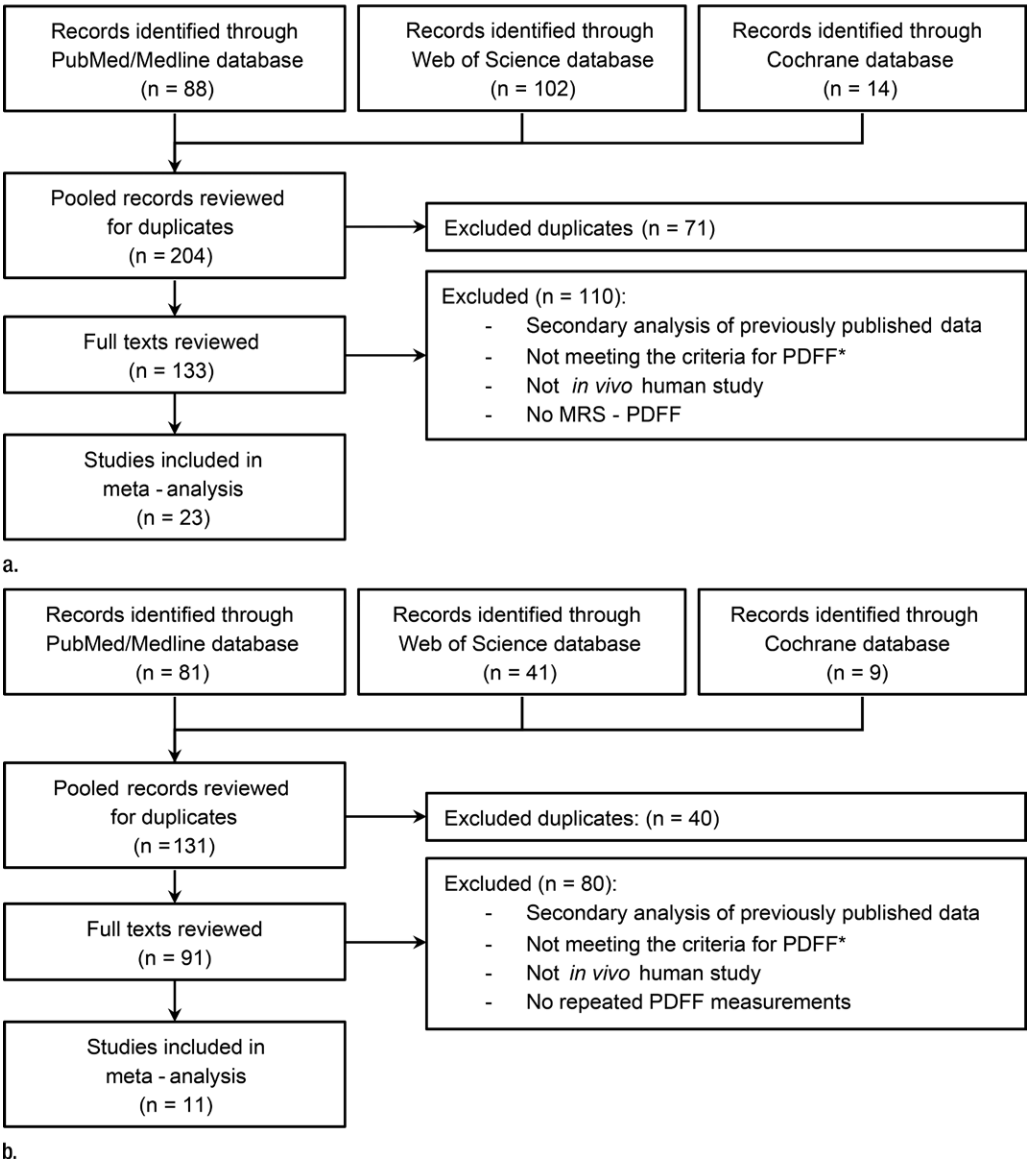


Figure 1: Flow diagrams show study selection according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for **(a)** linearity and bias analysis and **(b)** precision analysis. ∗ = Based on RSNA-QIBA PDFF Biomarker Committee for PDFF criteria as described in Materials and Methods section and Table E1 (online).

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

First Author	Year	Study Region	Data Collection	No. of Participants	Mean Age (y)	Manufacturer	Field Strength (T)	Reconstruction Method	PDFF Minimum (%)	Mean PDFF (%)	PDFF Maximum (%)	Analysis
Hines (56)	2011	North America	Prospective	42	51	GE	1.5	Hybrid	−2.8	5.0	39.5	Both
Kang (55)	2011	North America	Prospective	21	55	GE, Siemens	1.5, 3.0	Magnitude	−0.1	15.1	41.5	Both
Johnson (53)	2014	North America	Prospective	30	38	GE	3.0	Hybrid	0.5	12.1	28.7	Both
Artz (48)	2015	North America	Prospective	25	48	GE	1.5,3.0	Hybrid, magnitude	0.2	10.8	34.9	Both
Tyagi (41)	2015	North America	Prospective	29	24	GE	3.0	Hybrid, magnitude	0.1	13.0	37.3	Both
Motosugi (36)	2017	North America	Prospective	48	45	GE	1.5	Hybrid, magnitude	−0.2	6.2	27.7	Both
Meisamy (57)	2011	North America	Prospective	53	70	GE	1.5	Hybrid	0.5	6.6	25.6	Linearity and bias
Yokoo (58)	2011	North America	Prospective	163	39	GE	3.0	Magnitude	0.0	10.6	46.0	Linearity and bias
Motosugi (47)	2015	North America	Prospective	50	56	GE	3.0	Hybrid	0.0	6.6	36.7	Linearity and bias
Le (54)	2012	North America	Prospective	48	48	GE	3.0	Magnitude	1.9	18.4	55.4	Linearity and bias
Levin (52)	2014	North America	Prospective	84	36	GE	3.0	Magnitude	−0.8	10.8	32.9	Linearity and bias
Kühn (51)	2014	Europe	Prospective	50	57	Siemens	3.0	Hybrid	−0.2	13.1	43.5	Linearity and bias
Kühn (50)	2014	Europe	Prospective	33	57	Siemens	3.0	Hybrid	3.1	14.0	46.2	Linearity and bias
Bashir (49)	2015	North America	Prospective	42	53	Siemens	3.0	Hybrid	0.3	8.6	30.1	Linearity and bias
Rehm (44)	2015	North America	Prospective	132	13	GE	3.0	Hybrid	0.8	4.4	39.1	Linearity and bias
Zand (46)	2015	North America	Retrospective	286	14	GE	3.0	Magnitude	−0.1	12.2	41.2	Linearity and bias
Kukuk (45)	2015	Europe	Prospective	59	52	Philips	3.0	Complex	0.0	12.8	43.5	Linearity and bias
Kramer (38)	2016	North America	Prospective	50	56	GE	1.5	Hybrid	1.0	7.1	39.3	Linearity and bias
Kim (43)	2015	Asia	Retrospective	156	58	Siemens	3.0	Hybrid	0.3	5.2	36.0	Linearity and bias
Hetterich (42)	2016	Europe	Prospective	215	57	Siemens	3.0	Hybrid	0.6	9.2	37.5	Linearity and bias
Vu (40)	2016	North America	Prospective	35	60	Philips	3.0	Magnitude	1.0	14.3	38.6	Linearity and bias
Cui (39)	2016	North America	Prospective	50	54	GE	3.0	Magnitude	5.1	16.8	33.3	Linearity and bias
Park (37)	2016	North America	Prospective	31	53	GE	3.0	Magnitude	0.5	7.6	32.3	Linearity and bias
Wu (59)	2016	Asia	Prospective	15	33	GE	3.0	Hybrid	−1.8	3.5	10.7	Precision
Mashhood (63)	2013	North America	Prospective	10	50	GE, Siemens	1.5, 3.0	Magnitude	−0.1	8.1	39.2	Precision
Negrete (61)	2014	North America	Prospective	29	24	GE	3.0	Magnitude	−0.1	11.8	40.5	Precision
Sofue (60)	2015	North America	Prospective	150	57	Siemens	3.0	Hybrid	0.5	6.9	39.3	Precision
Serai (62)	2017	North America	Prospective	24	39	GE, Philips	1.5, 3.0	Complex, hybrid	0.3	5.8	26.0	Precision

Figure 2

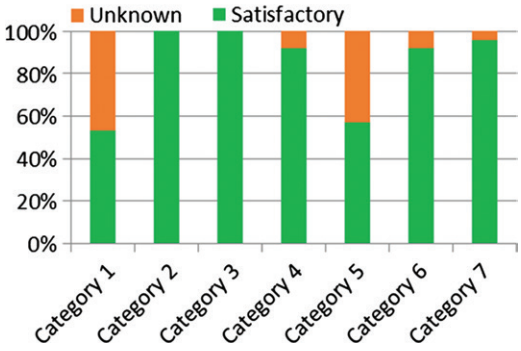


Figure 2: Bar chart shows estimates of percentage compliance by using Quality Assessment of Studies of Diagnostic Accuracy 2 tool. All 28 studies had moderate to high scores and low risk of bias in all seven categories; all fulfilled five or more of the seven quality categories. Summary of scores for each category were as follows: Category 1, Risk of bias–Patient selection, 57% (16 of 28); Category 2, Risk of bias–Index test, 100% (28 of 28); Category 3, Risk of bias–Reference standard, 100% (28 of 28); Category 4, Risk of bias–Flow and timing, 93% (26 of 28); Category 5, Applicability concerns–Patient selection, 61% (17 of 28); Category 6, Applicability concerns–Index test, 93% (26 of 28); and Category 7, Applicability concerns–Reference standard, 96% (27 of 28).

estimated intercept of −0.07 was not significantly different from zero (P = .70; 95% CI: −0.50, 0.32). The estimated slope of the regression line was significantly below unity at 0.97 ($P < .001$; 95% CI: 0.96, 0.98), indicating 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 manufacturer (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 either GE or Siemens systems.

Assessment of Precision

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

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

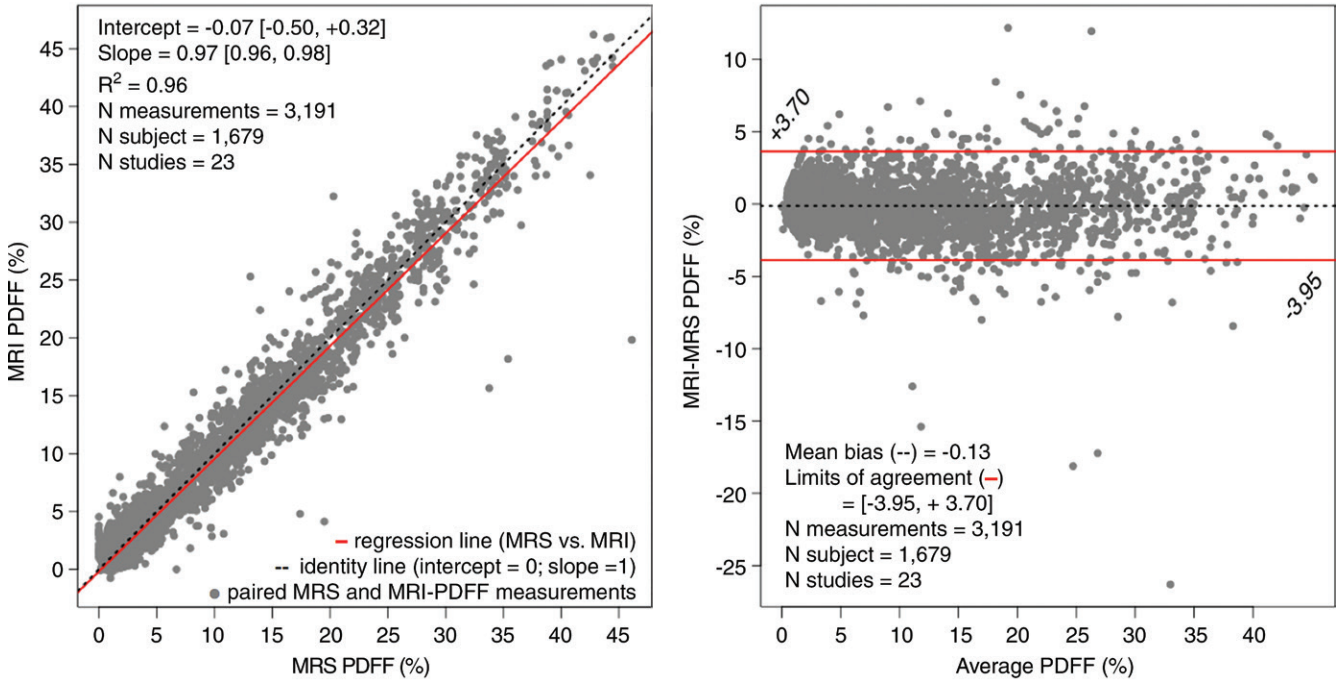
The estimated variance components from linear mixed-effects models (Fig 4) indicated that the main contributors to MR imaging–PDFF measurement variability were ROI locations (standard deviation, 1.30%) and random measurement error because of repeated acquisitions (standard deviation, 1.07%). Other technical 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 primary research studies from geographically 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 criteria 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 of emphasis in different subpopulations of various age range, sex, and steatosis risk factors (eg, obesity, diabetes), the potential impact of the variability is thought to be small in this meta-analysis of technical validation.

Results of our meta-analysis demonstrated that MR imaging–PDFF has excellent linearity and negligible bias with respect to the reference standard of MR spectroscopy-PDFF measurements over the entire range of observed steatosis severity. A slight deviation of the MR imaging versus MR spectroscopy regression from perfect agreement was statistically significant, 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 (repeatability and reproducibility) were inherent 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

Figure 3



a. b.

Figure 3: Scatterplots show linearity and bias of MR imaging–PDFF. **(a)** Linearity of MR imaging–PDFF against MR spectroscopy-PDFF, both measured in colocalized ROIs in the liver. Red line represents linear regression fit (after correcting for within-participant correlations between replicated measurements, if any) with coefficient of determination indicating very strong linear fit (R^2 = 0.96). Estimated intercept (−0.07%) was not significantly different from zero; estimated slope was close to but less than unity (0.97), indicating very slight underestimation of MR imaging–PDFF values at larger MR spectroscopy-PDFF values. **(b)** Bland-Altman plot of MR imaging–PDFF relative to MR spectroscopy-PDFF as reference technique demonstrates very small mean bias (−0.13%) and limits of agreement within ± 4%. Estimated size of various bias components are presented in Table 3.

as field strength, imager manufacturer, and reconstruction method. Based on these observations, we conclude that MR imaging–PDFF has excellent technical performance characteristics as a QIB for widespread use in clinical trials and patient care.

Our meta-analysis findings are consistent with previous single-center studies demonstrating high technical accuracy and precision of MR imaging–PDFF as a QIB of hepatic steatosis. These earlier data suggested that PDFF should be well suited for multicenter trials and longitudinal assessment of hepatic steatosis. Our meta-analysis provides higher-level evidence supporting this claim. Several observational studies and therapeutic trials of nonalcoholic fatty liver disease demonstrated the feasibility of using PDFF as a study end point (44,46,80), as well as a patient-level diagnostic and therapeutic response biomarker for clinical practice (42,81–84). In addition, PDFF methods are used in large population-based cohort trials (including studies with more than 1000 participants) (85–87). U.S. Food and Drug

Table 3

Bias Components of MR Imaging–PDFF

Bias Components	Estimated Change PDFF (%)	*P* Value
Intercept	0.42 (0.21, 0.63)	<.0001
Manufacturer		
GE	Reference	…
Siemens	−0.01 (−0.19, 0.17)	.933
Philips	2.24 (1.51, 2.96)	<.0001
Reconstruction method		
Magnitude	Reference	…
Complex	−1.36 (−2.22, −0.50)	.002
Hybrid	0.92 (0.80, 1.05)	<.0001
Field strength		
1.5 T	Reference	…
3.0 T	−1.20 (−1.40, −1.00)	<.0001

Note.—Data in parentheses are 95% CIs.

Figure 4

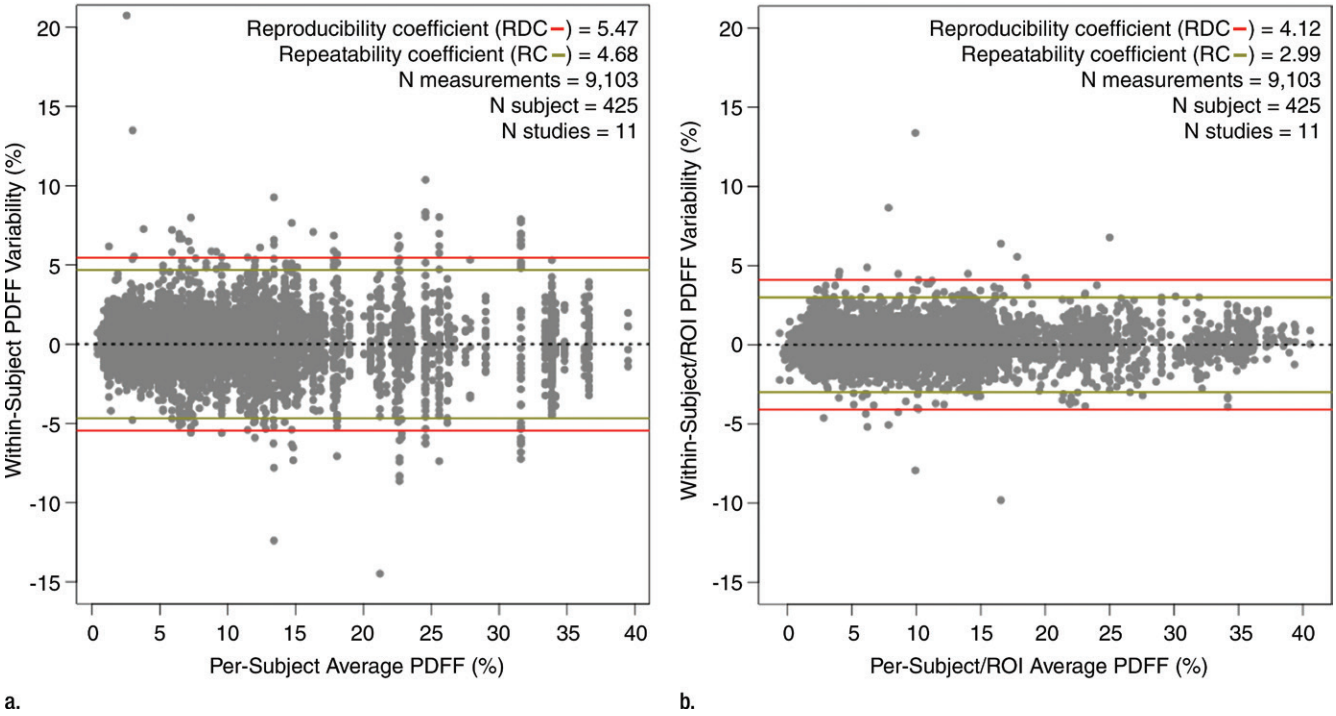


Figure 4: **(a, b)** Scatterplots illustrate participant-level and ROI-level precision of MR imaging–PDFF measurements, respectively, because of differences in field strength, imager manufacturer, reconstruction method, equipment setup, and random noise (effect sizes are detailed in Table 4). Repeatability refers to precision under an identical experimental condition (ie, scan-rescan repeatability by using fixed hardware and software). Reproducibility refers to precision under variable experimental condition (ie, by using different hardware or software). At ROI level **(b)**, MR imaging–PDFF is highly precise with repeatability coefficients (RCs) and reproducibility coefficients (RDCs) indicating the 95th percentile of precision to be approximately ± 3% and ± 4%, respectively. At participant level **(a)**, repeatability and reproducibility are approximately ± 5% and ± 5.5%. Subject-level precision is lower than ROI-level precision because of differences in ROI placement, that is, heterogeneity of underlying steatosis. In both cases, RCs and RDCs are similar to each other, indicating small impact of technical factors (field strength, imager manufacturer, reconstruction method) on the precision of the measurement.

Administration–approved MR imaging pulse sequences for hepatic PDFF measurement are now available from major imager manufacturers worldwide including GE, Siemens, and Philips.

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 statistically powerful 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

Table 4

Within-Subject Variance Components of MR Imaging–PDFF

Component	Variable ROI and Variable Technique (%)	Variable ROI and Fixed Technique (%)	Fixed ROI and Variable Technique (%)	Fixed ROI and Fixed Technique (%)
Nonuniform steatosis (different ROI location)	1.30	1.30	…	…
Manufacturer (GE, Siemens, or Philips)	0.81	…	0.81	…
Field strength (1.5 T or 3.0 T)	0.51	…	0.51	…
Reconstruction method (magnitude, complex, or hybrid)	0.35	…	0.35	…
Interexamination (equipment setup)	0.16	0.16	0.16	0.16
Interacqusition (random error)	1.07	1.07	1.07	1.07
Sum of components	1.98	1.69	1.49	1.08

Note.—Data are standard deviations. Data are from 11 precision studies including 425 participants, with two or three examinations per participant. See Figure 4 for repeatability and reproducibility coefficients.

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.); Deptartment 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.R.B.); 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.); 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, Jeonju, 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, PhD, MD, Takeda Pharmaceuticals; Cathy Elsinger, PhD, NordicNeuroLab; Manuela França, MD, Hospital de San Antonio; Alexander Guimaraes, MD, PhD, Oregon Health & Science University; Gavin Hamilton, PhD, University of California, San Diego (UCSD); Diego Hernando, PhD, University of Wisconsin-Madison; Harry H. Hu, PhD, Phoenix Children's Hospital; Edward F. Jackson, PhD, University of Wisconsin, School of Medicine and Public Health; Tommy Johansson, Advanced MR Analytics (MR) (Oslo, Norway); So Yeon Kim, MD, Adan Medical Center, Korea; Sonja Kinner, MD, University of Duisburg-Essen (Duisburg, Germany); Hendrik Laue, 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 Middleton, MD, University of California, San Diego (UCSD); Carlo E. Neumaier, MD, IRCCS-Azienda Ospedaliera Universitaria San Martino-IST-National Cancer Institute, Genoa; Nancy Obuchowski, 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 Serai, PhD, Cincinnati Children's Hospital; Claude Sirlin, MD, University of California, San Diego (UCSD); Judd Storrs, PhD, University of Mississippi Medical Center; Satoshi Sugiura, PhD, Toshiba; Shuji Yamamoto, 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.R.B.** 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, TaiwanJ 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. **H.H.** disclosed no relevant relationships. **J.P.K.** disclosed no relevant relationships. **G.M.K.** disclosed no relevant relationships. **R.L.** Activities related to the present

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 contracted work through university for Alexion, AstraZeneca, Bioclinica, Biomedical Systems, Bristol-Myers Squibb, Galmed, GE, Gilead, ICON, Intercept, Isis, Janssen, NuSirt, Pfizer, Profil, Sanofi Genzyme, Shire, Siemens, Takeda, 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. **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 Elucid 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.B.R.** 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, LCC; has stock ownership in Cellectar Biosciences; is a shareholder in Elucent Medical; is a consultant for Parexel International; institution receives research support from Bracco Diagnostics and GE. Other relationships: disclosed no relevant relationships. **C.B.S.** disclosed no relevant relationships.

References

1. Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 2013;10(11):686–690.
2. Wong RJ, Cheung R, Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. Hepatology 2014;59(6):2188–2195.
3. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology 1999;116(6):1413–1419.
4. Ekstedt M, Franzén LE, Mathiesen UL, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. Hepatology 2006;44(4):865–873.
5. Rubinstein E, Lavine JE, Schwimmer JB. Hepatic, cardiovascular, and endocrine outcomes of the histological subphenotypes of nonalcoholic fatty liver disease. Semin Liver Dis 2008;28(4):380–385.
6. Guzman G, Brunt EM, Petrovic LM, Chejfec G, Layden TJ, Cotler SJ. Does nonalcoholic fatty liver disease predispose patients to hepatocellular carcinoma in the absence of cirrhosis? Arch Pathol Lab Med 2008;132(11):1761–1766.
7. Sanyal AJ, Banas C, Sargeant C, et al. Similarities and differences in outcomes of cirrhosis due to nonalcoholic steatohepatitis and hepatitis C. Hepatology 2006;43(4):682–689.
8. Adams LA, Lymp JF, St Sauver J, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. Gastroenterology 2005;129(1):113–121.
9. Adams LA, Sanderson S, Lindor KD, Angulo P. The histological course of nonalcoholic fatty liver disease: a longitudinal study of 103 patients with sequential liver biopsies. J Hepatol 2005;42(1):132–138.
10. Angulo P. Nonalcoholic fatty liver disease. N Engl J Med 2002;346(16):1221–1231.
11. Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD; American Association for the Study of Liver Diseases. Liver biopsy. Hepatology 2009;49(3):1017–1044.
12. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128(7):1898–1906.
13. Bravo AA, Sheth SG, Chopra S. Liver biopsy. N Engl J Med 2001;344(7):495–500.
14. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 2012;55(6):2005–2023.
15. Castera L, Vilgrain V, Angulo P. Noninvasive evaluation of NAFLD. Nat Rev Gastroenterol Hepatol 2013;10(11):666–675.
16. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. J Magn Reson Imaging 2012;36(5):1011–1014.
17. Hu HH, Börnert P, Hernando D, et al. ISMRM workshop on fat-water separation: insights, applications and progress in MRI. Magn Reson Med 2012;68(2):378–388.
18. Wei JL, Leung JC, Loong TC, et al. Prevalence and severity of nonalcoholic fatty liver disease in non-obese patients: a population study using proton-magnetic resonance spectroscopy. Am J Gastroenterol 2015;110(9):1306–1314; quiz 1315.
19. Szczepaniak LS, Nurenberg P, Leonard D, et al. Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. Am J Physiol Endocrinol Metab 2005;288(2):E462–E468.
20. Longo R, Pollesello P, Ricci C, et al. Proton MR spectroscopy in quantitative in vivo determination of fat content in human liver steatosis. J Magn Reson Imaging 1995;5(3):281–285.
21. Rajeev SP, Sprung VS, Roberts C, et al. Compensatory changes in energy balance during dapagliflozin treatment in type 2 diabetes mellitus: a randomised double-blind, placebo-controlled, cross-over trial (ENERGIZE)-study protocol. BMJ Open 2017;7(1):e013539.
22. Smits MM, Tonneijck L, Muskiet MH, et al. Twelve week liraglutide or sitagliptin does not affect hepatic fat in type 2 diabetes: a randomised placebo-controlled trial. Diabetologia 2016;59(12):2588–2593.
23. Zhang HJ, He J, Pan LL, et al. Effects of moderate and vigorous exercise on nonalcoholic fatty liver disease: a randomized clinical trial. JAMA Intern Med 2016;176(8):1074–1082.
24. Papamiltiadous ES, Roberts SK, Nicoll AJ, et al. A randomised controlled trial of a Mediterranean Dietary Intervention for Adults with Non Alcoholic Fatty Liver Disease (MEDINA): study protocol. BMC Gastroenterol 2016;16:14.
25. Bhatia L, Scorletti E, Curzen N, Clough GF, Calder PC, Byrne CD. Improvement in non-alcoholic fatty liver disease severity is associated with a reduction in carotid intima-media thickness progression. Atherosclerosis 2016;246:13–20.
26. Hallsworth K, Thoma C, Hollingsworth KG, et al. Modified high-intensity interval training reduces liver fat and improves cardiac function in non-alcoholic fatty liver disease: a randomized controlled trial. Clin Sci (Lond) 2015;129(12):1097–1105.
27. Scorletti E, Bhatia L, McCormick KG, et al. Effects of purified eicosapentaenoic and docosahexaenoic acids in nonalcoholic fatty liver disease: results from the Welcome* study. Hepatology 2014;60(4):1211–1221.
28. Koopman KE, Caan MW, Nederveen AJ, et al. Hypercaloric diets with increased meal frequency, but not meal size, increase intrahepatic triglycerides: a randomized controlled trial. Hepatology 2014;60(2):545–553.
29. Wong VW, Wong GL, Yeung DK, et al. Fatty pancreas, insulin resistance, and β-cell function: a population study using fat-water magnetic resonance imaging. Am J Gastroenterol 2014;109(4):589–597.
30. de Mutsert R, den Heijer M, Rabelink TJ, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. Eur J Epidemiol 2013;28(6):513–523.
31. Sullivan S, Kirk EP, Mittendorfer B, Patterson BW, Klein S. Randomized trial of exercise effect on intrahepatic triglyceride con-

tent and lipid kinetics in nonalcoholic fatty liver disease. Hepatology 2012;55(6):1738–1745.

32. Wong VW, Chu WC, Wong GL, et al. Prevalence of non-alcoholic fatty liver disease and advanced fibrosis in Hong Kong Chinese: a population study using proton-magnetic resonance spectroscopy and transient elastography. Gut 2012;61(3):409–415.

33. Sathyanarayana P, Jogi M, Muthupillai R, Krishnamurthy R, Samson SL, Bajaj M. Effects of combined exenatide and pioglitazone therapy on hepatic fat content in type 2 diabetes. Obesity (Silver Spring) 2011;19(12):2310–2315.

34. Guerrero R, Vega GL, Grundy SM, Browning JD. Ethnic differences in hepatic steatosis: an insulin resistance paradox? Hepatology 2009;49(3):791–801.

35. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27–67.

36. Motosugi U, Hernando D, Wiens C, Bannas P, Reeder SB. High SNR acquisitions improve the repeatability of liver fat quantification using confounder-corrected chemical shift-encoded MR imaging. Magn Reson Med Sci 2017 Feb 13. [Epub ahead of print]

37. Park CC, Hamilton G, Desai A, et al. Effect of intravenous gadoxetate disodium and flip angle on hepatic proton density fat fraction estimation with six-echo, gradient-recalled-echo, magnitude-based MR imaging at 3T. Abdom Radiol (NY) 2017;42(4):1189–1198.

38. Kramer H, Pickhardt PJ, Kliewer MA, et al. Accuracy of liver fat quantification with advanced CT, MRI, and ultrasound techniques: prospective comparison with MR spectroscopy. AJR Am J Roentgenol 2017;208(1):92–100.

39. Cui J, Philo L, Nguyen P, et al. Sitagliptin vs. placebo for non-alcoholic fatty liver disease: a randomized controlled trial. J Hepatol 2016;65(2):369–376.

40. Vu KN, Gilbert G, Chalut M, Chagnon M, Chartrand G, Tang A. MRI-determined liver proton density fat fraction, with MRS validation: comparison of regions of interest sampling methods in patients with type 2 diabetes. J Magn Reson Imaging 2016;43(5):1090–1099.

41. Tyagi A, Yeganeh O, Levin Y, et al. Intra- and inter-examination repeatability of magnetic resonance spectroscopy, magnitude-based MRI, and complex-based MRI for estimation of hepatic proton density fat fraction in overweight and obese children and adults. Abdom Imaging 2015;40(8):3070–3077.

42. Hetterich H, Bayerl C, Peters A, et al. Feasibility of a three-step magnetic resonance imaging approach for the assessment of hepatic steatosis in an asymptomatic study population. Eur Radiol 2016;26(6):1895–1904.

43. Kim KY, Song JS, Kannengiesser S, Han YM. Hepatic fat quantification using the proton density fat fraction (PDFF): utility of free-drawn-PDFF with a large coverage area. Radiol Med (Torino) 2015;120(12):1083–1093.

44. Rehm JL, Wolfgram PM, Hernando D, Eickhoff JC, Allen DB, Reeder SB. Proton density fat-fraction is an accurate biomarker of hepatic steatosis in adolescent girls and young women. Eur Radiol 2015;25(10):2921–2930.

45. Kukuk GM, Hittatiya K, Sprinkart AM, et al. Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. Eur Radiol 2015;25(10):2869–2879.

46. Zand KA, Shah A, Heba E, et al. Accuracy of multiecho magnitude-based MRI (M-MRI) for estimation of hepatic proton density fat fraction (PDFF) in children. J Magn Reson Imaging 2015;42(5):1223–1232.

47. Motosugi U, Hernando D, Bannas P, et al. Quantification of liver fat with respiratory-gated quantitative chemical shift encoded MRI. J Magn Reson Imaging 2015;42(5):1241–1248.

48. Artz NS, Haufe WM, Hooker CA, et al. Reproducibility of MR-based liver fat quantification across field strength: Same-day comparison between 1.5T and 3T in obese participants. J Magn Reson Imaging 2015;42(3):811–817.

49. Bashir MR, Zhong X, Nickel MD, et al. Quantification of hepatic steatosis with a multistep adaptive fitting MRI approach: prospective validation against MR spectroscopy. AJR Am J Roentgenol 2015;204(2):297–306.

50. Kühn JP, Jahn C, Hernando D, et al. T1 bias in chemical shift-encoded liver fat-fraction: role of the flip angle. J Magn Reson Imaging 2014;40(4):875–883.

51. Kühn JP, Hernando D, Mensel B, et al. Quantitative chemical shift-encoded MRI is an accurate method to quantify hepatic steatosis. J Magn Reson Imaging 2014;39(6):1494–1501.

52. Levin YS, Yokoo T, Wolfson T, et al. Effect of echo-sampling strategy on the accuracy of out-of-phase and in-phase multiecho gradient-echo MRI hepatic fat fraction estimation. J Magn Reson Imaging 2014;39(3):567–575.

53. Johnson BL, Schroeder ME, Wolfson T, et al. Effect of flip angle on the accuracy and repeatability of hepatic proton density fat fraction estimation by complex data-based, T1-independent, T2*-corrected, spectrum-modeled MRI. J Magn Reson Imaging 2014;39(2):440–447.

54. Le TA, Chen J, Changchien C, et al. Effect of colesevelam on liver fat quantified by magnetic resonance in nonalcoholic steatohepatitis: a randomized controlled trial. Hepatology 2012;56(3):922–932.

55. Kang GH, Cruite I, Shiehmorteza M, et al. Reproducibility of MRI-determined proton density fat fraction across two different MR scanner platforms. J Magn Reson Imaging 2011;34(4):928–934.

56. Hines CD, Frydrychowicz A, Hamilton G, et al. T(1) independent, T(2) (*) corrected chemical shift based fat-water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. J Magn Reson Imaging 2011;33(4):873–881.

57. Meisamy S, Hines CD, Hamilton G, et al. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology 2011;258(3):767–775.

58. Yokoo T, Shiehmorteza M, Hamilton G, et al. Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology 2011;258(3):749–759.

59. Wu B, Han W, Li Z, et al. Reproducibility of intra- and inter-scanner measurements of liver fat using complex confounder-corrected chemical shift encoded MRI at 3.0 Tesla. Sci Rep 2016;6:19339.

60. Sofue K, Mileto A, Dale BM, Zhong X, Bashir MR. Interexamination repeatability and spatial heterogeneity of liver iron and fat quantification using MRI-based multistep adaptive fitting algorithm. J Magn Reson Imaging 2015;42(5):1281–1290.

61. Negrete LM, Middleton MS, Clark L, et al. Inter-examination precision of magnitude-based MRI for estimation of segmental hepatic proton density fat fraction in obese participants. J Magn Reson Imaging 2014;39(5):1265–1271.

62. Serai SD, Dillman JR, Trout AT. Proton density fat fraction measurements at 1.5- and 3-T hepatic MR imaging: same-day agreement among readers and across two imager manufacturers. Radiology 2017;284(1):244–254.

63. Mashhood A, Railkar R, Yokoo T, et al. Reproducibility of hepatic fat fraction measurement by magnetic resonance imaging. J Magn Reson Imaging 2013;37(6):1359–1370.

64. Hamilton G, Yokoo T, Bydder M, et al. In vivo characterization of the liver fat ^{1}H MR spectrum. NMR Biomed 2011;24(7):784–790.

65. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magn Reson Med 2008;60(5):1122–1134.

66. Reeder SB, Cruite I, Hamilton G, Sirlin CB. Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. J Magn Reson Imaging 2011;34(4):729–749.

67. Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. Magn Reson Imaging Clin N Am 2010;18(3):337–357, ix.

68. Ruschke S, Eggers H, Kooijman H, et al. Correction of phase errors in quantitative water-fat imaging using a monopolar time-interleaved multi-echo gradient echo sequence. Magn Reson Med 2016 Oct 31. [Epub ahead of print]

69. Colgan TJ, Hernando D, Sharma SD, Reeder SB. The effects of concomitant gradients on chemical shift encoded MRI. Magn Reson Med 2016 Sep 21. [Epub ahead of print]

70. Peterson P, Månsson S. Fat quantification using multiecho sequences with bipolar gradients: investigation of accuracy and noise performance. Magn Reson Med 2014;71(1):219–229.

71. Hernando D, Hines CD, Yu H, Reeder SB. Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. Magn Reson Med 2012;67(3):638–644.

72. Yu H, Shimakawa A, Hines CD, et al. Combination of complex-based and magnitude-based multiecho water-fat separation for accurate quantification of fat-fraction. Magn Reson Med 2011;66(1):199–206.

73. Hamilton G, Middleton MS, Bydder M, et al. Effect of PRESS and STEAM sequences on magnetic resonance spectroscopic liver fat quantification. J Magn Reson Imaging 2009;30(1):145–152.

74. Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. J Craniomaxillofac Surg 2011;39(2):91–92.

75. 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.

76. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–536.

77. 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.

78. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629–634.

79. 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.

80. Di Martino M, Pacifico L, Bezzi M, et al. Comparison of magnetic resonance spectroscopy, proton density fat fraction and histological analysis in the quantification of liver steatosis in children and adolescents. World J Gastroenterol 2016;22(39):8812–8819.

81. Patel J, Bettencourt R, Cui J, et al. Association of noninvasive quantitative decline in liver fat content on MRI with histologic response in nonalcoholic steatohepatitis. Therap Adv Gastroenterol 2016;9(5):692–701.

82. Vogt LJ, Steveling A, Meffert PJ, et al. Magnetic resonance imaging of changes in abdominal compartments in obese diabetics during a low-calorie weight-loss program. PLoS One 2016;11(4):e0153595.

83. Arulanandan A, Ang B, Bettencourt R, et al. Association between quantity of liver fat and cardiovascular risk in patients with nonalcoholic fatty liver disease independent of nonalcoholic steatohepatitis. Clin Gastroenterol Hepatol 2015;13(8):1513–1520.e1.

84. Tang A, Chen J, Le TA, et al. Cross-sectional and longitudinal evaluation of liver volume and total liver fat burden in adults with nonalcoholic steatohepatitis. Abdom Imaging 2015;40(1):26–37.

85. Bamberg F, Hetterich H, Rospleszcz S, et al. Subclinical disease burden as assessed by whole-body MRI in participants with prediabetes, participants with diabetes, and normal control participants from the general population: the KORA-MRI study. Diabetes 2017;66(1):158–169.

86. Bamberg F, Kauczor HU, Weckbach S, et al. Whole-body MR imaging in the German National Cohort: rationale, design, and technical background. Radiology 2015;277(1):206–220.

87. Völzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011;40(2):294–307.

88. Bedossa P, Dargère D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 2003;38(6):1449–1457.