Linearity and Bias of Proton Density Fat Fraction as a Quantitative Imaging Biomarker: A Multicenter, Multiplatform, Multivendor Phantom Study

Houchun H. Hu, PhD* • Takeshi Yokoo, MD, PhD* • Mustafa R. Bashir, MD • Claude B. Sirlin, MD • Diego Hernando, PhD • Dariya Malyarenko, PhD • Thomas L. Chenevert, PhD • Mark A. Smith, MS • Suraj D. Serai, PhD • Michael S. Middleton, MD, PhD • Walter C. Henderson, BA • Gavin Hamilton, PhD • Jean Shaffer, RT • Yunhong Shu, PhD • Jean A. Tkach, MS, PhD • Andrew T. Trout, MD • Nancy Obuchowski, PhD • Jean H. Brittain, PhD • Edward F. Jackson, PhD • Scott B. Reeder, MD, PhD • for the RSNA Quantitative Imaging Biomarkers Alliance PDFF Biomarker Committee

From the Department of Radiology, Nationwide Children's Hospital, 700 Children's Dr, Columbus, OH 43235 (H.H.H., M.A.S.); Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Tex (T.Y.); Department of Radiology (M.R.B., J.S.), Department of Medicine, Division of Gastroenterology (M.R.B.), and Center for Advanced Magnetic Resonance Development (M.R.B., J.S.), Duke University Medical Center, Durham, NC; Liver Imaging Group, Department of Radiology, University of California San Diego, San Diego, Calif (C.B.S., M.S.M., W.C.H., G.H.); Departments of Radiology (D.H., J.H.B., S.B.R.), Medical Physics (D.H., E.F.J., S.B.R.), Biomedical Engineering (S.B.R.), Medicine (S.B.R.), and Emergency Medicine (S.B.R.), University of Wisconsin, Madison, Wis; Department of Radiology, University of Michigan, Ann Arbor, Mich (D.M., T.L.C.); Department of Radiology, Children's Hospital of Philadelphia, Philadelphia, Pa (S.D.S.); Department of Radiology, Mayo Clinic, Rochester, Minn (Y.S.); Department of Radiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio (J.A.T., A.T.T.); Department of Radiology, University of Cincinnati College of Medicine, Cincinnati, Ohio (J.A.T., A.T.T.); Department of Quantitative Health Science, Cleveland Clinic Foundation, Cleveland, Ohio (N.O.); and Calimetrix, LLC, Madison, Wis (J.H.B.). Received June 30, 2020; revision requested July 20; revision received October 11; accepted November 9. **Address correspondence to** H.H.H. (e-mail: *harryhhu@gmail.com*).

Supported by the National Institutes of Health (grants R41EB025729, R44EB025729, R01DK088925, K24DK102595, R01DK083380, R01CA23802301, and U24CA237683). Quantitative Imaging Biomarkers Alliance projects and activities have been funded in whole or in part with Federal funds from the National Institutes of Health.

* H.H.H. and T.Y. contributed equally to this work.

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

See also the editorial by Dyke in this issue.

Radiology 2021; 298:640–651 • https://doi.org/10.1148/radiol.2021202912 • Content codes: IN MR

Background: Proton density fat fraction (PDFF) estimated by using chemical shift–encoded (CSE) MRI is an accepted imaging biomarker of hepatic steatosis. This work aims to promote standardized use of CSE MRI to estimate PDFF.

Purpose: To assess the accuracy of CSE MRI methods for estimating PDFF by determining the linearity and range of bias observed in a phantom.

Materials and Methods: In this prospective study, a commercial phantom with 12 vials of known PDFF values were shipped across nine U.S. centers. The phantom underwent 160 independent MRI examinations on 27 1.5-T and 3.0-T systems from three vendors. Two three-dimensional CSE MRI protocols with minimal T1 bias were included: vendor and standardized. Each vendor's confounder-corrected complex or hybrid magnitude-complex based reconstruction algorithm was used to generate PDFF maps in both protocols. The Siemens reconstruction required a configuration change to correct for water-fat swaps in the phantom. The MRI PDFF values were compared with the known PDFF values by using linear regression with mixed-effects modeling. The 95% CIs were calculated for the regression slope (ie, proportional bias) and intercept (ie, constant bias) and compared with the null hypothesis (slope = 1, intercept = 0).

Results: Pooled regression slope for estimated PDFF values versus phantom-derived reference PDFF values was 0.97 (95% CI: 0.96, 0.98) in the biologically relevant 0%–47.5% PDFF range. The corresponding pooled intercept was −0.27% (95% CI: −0.50%, −0.05%). Across vendors, slope ranges were 0.86–1.02 (vendor protocols) and 0.97–1.0 (standardized protocol) at 1.5 T and 0.91–1.01 (vendor protocols) and 0.87–1.01 (standardized protocol) at 3.0 T. The intercept ranges (absolute PDFF percentage) were −0.65% to 0.18% (vendor protocols) and −0.69% to −0.17% (standardized protocol) at 1.5 T and −0.48% to 0.10% (vendor protocols) and −0.78% to −0.21% (standardized protocol) at 3.0 T.

Conclusion: Proton density fat fraction estimation derived from three-dimensional chemical shift–encoded MRI in a commercial phantom was accurate across vendors, imaging centers, and field strengths, with use of the vendors' product acquisition and reconstruction software.

©RSNA, 2021

Abnormal hepatic intracellular triglyceride accumulation (ie, hepatic steatosis) is the earliest and hallmark feature of nonalcoholic fatty liver disease (1). Nonalcoholic fatty liver disease is the leading cause of liver disease in the Western World and affects an estimated 1 billion people worldwide (2). Nonalcoholic steatohepatitis, its aggressive form, is widely expected to become the leading liver transplantation indication in the coming decade (2). Furthermore, hepatic steatosis increasingly is recognized as an independent risk factor for cardiovascular disease, cancer, metabolic disorders, and obesity (3–7). Accordingly, there is interest in developing noninvasive, accurate, and precise methods to quantify hepatic steatosis for diagnosis, quantitative staging,

Abbreviations

CSE = chemical shift encoded, PDFF = proton density fat fraction, QIBA = Quantitative Imaging Biomarkers Alliance

Summary

Proton density fat fraction estimated by using chemical shift–encoded MRI in a commercial phantom was accurate across three vendors, nine imaging centers, and two magnetic field strengths, with use of two protocols and complex or hybrid magnitude-complex based vendor reconstruction software.
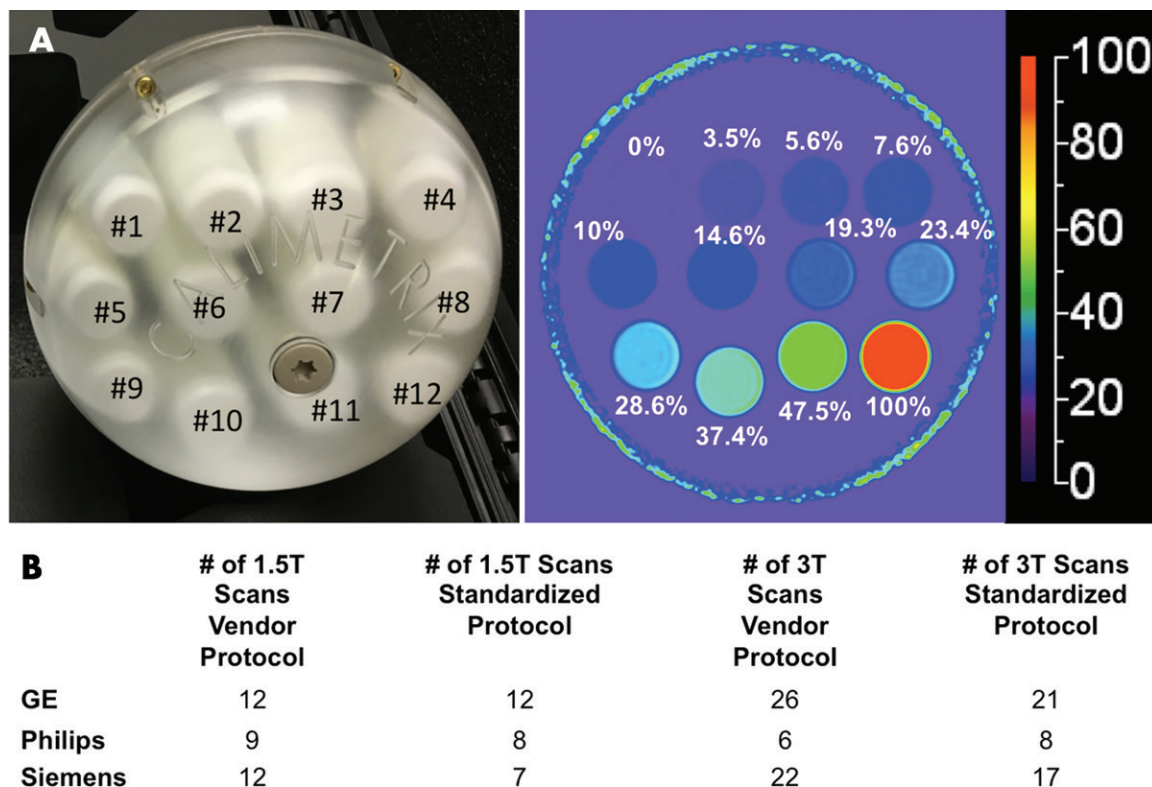
Key Results

- Pooled regression slope for estimated proton density fat fraction (PDFF) values versus phantom-derived reference PDFF values was 0.97. Slope ranged from 0.86 to 1.02 (vendor protocols) and from 0.97 to 1.0 (standardized protocol) at 1.5 T and from 0.91 to 1.01 (vendor protocols) and from 0.87 to 1.01 (standardized protocol) at 3.0 T.
- Pooled regression intercept was −0.27% PDFF. Intercept (absolute PDFF percentage) ranged from −0.65% to 0.18% (vendor protocols) and −0.69% to −0.17% (standardized protocol) at 1.5 T and −0.48% to 0.10% (vendor protocols) and −0.78% to −0.21% (standardized protocol) at 3.0 T.

and treatment monitoring of nonalcoholic fatty liver disease. Confounder-corrected chemical shift–encoded (CSE) MRI proton density fat fraction (PDFF) is an accurate and precise biomarker of hepatic steatosis (8–11). CSE MRI methods are increasingly used for both research and clinical purposes in the liver and in other organs and are commercially available on multiple vendor platforms (12,13). Repeatability and reproducibility of CSE MRI–derived PDFF have been investigated in vitro and in vivo (14–22).

In 2015 the PDFF Biomarker Committee was formed within the Quantitative Imaging Biomarkers Alliance (QIBA) of the Radiological Society of North America. The committee was tasked with promoting standardized use and facilitating widespread adoption of MRI-based PDFF for clinical care and research and with developing technical standards on the basis of in vivo data from the literature. One of the QIBA missions is to characterize biomarker accuracy through assessment of biomarker linearity (ie, proportional) and bias (ie, constant) under expected hardware and software variations in clinical practice—for example, magnetic field strength and vendor-specific image reconstruction algorithms.

Although linearity and bias of MRI-derived PDFF have been validated in vivo against MR spectroscopy–derived PDFF as reference standard (23), there is a lack of definitive data validating linearity and bias across vendors, platforms, and magnetic field strength against an independent non–MRI-based reference standard. Therefore, the purpose of our work was to assess the accuracy of commercially available CSE MRI methods for estimating PDFF in a multicenter, multivendor, multiplatform round-robin phantom study at 1.5 T and 3.0 T at nine academic centers. The data from this study will inform future technical development of CSE MRI methods for PDFF estimation and establish

B	# of 1.5T Scans Vendor Protocol	# of 1.5T Scans Standardized Protocol	# of 3T Scans Vendor Protocol	# of 3T Scans Standardized Protocol
GE	12	12	26	21
Philips	9	8	6	8
Siemens	12	7	22	17

Figure 1: *A,* Photograph of phantom, which consisted of spherical acrylic housing containing 12 vials with known proton density fat fraction (PDFF) values, submerged in contrast medium–doped water bath. Representative PDFF parametric map with 0%–100% color scale. Phantom-derived reference PDFF values are shown based on volumes of ingredients used. *B,* Total number of examinations performed, categorized according to protocol, magnetic field strength, and vendor.

Table 1: MRI Platforms

Site	Manufacturer and System	Software Version	No. of Systems	Total No. of Examinations
University of Wisconsin–Madison				48
1.5 T	GE Signa HDx	HD23_V02	2	
1.5 T	GE Signa Artist	DV26_R02/3	1	
1.5 T	GE Optima MR450w	DV26_R02/3	1	
3.0 T	GE Discovery MR750	DV26_R01/3	1	
3.0 T	GE Signa PET/MR	MP26_R01	1	
3.0 T	GE Signa Premier	RX27_R01/2	1	
3.0 T	GE Signa Architect	DV26_R03	1	
University of Michigan, Ann Arbor				9
1.5 T	Philips Ingenia	R5.4.1.0	1	
3.0 T	Philips Ingenia	R5.3.1.1	1	
3.0 T	Philips Ingenia	R5.4.1.0	1	
Nationwide Children's Hospital				34
3.0 T	Siemens Magnetom Skyra	VE11C	1	
3.0 T	Siemens Magnetom Prisma	VE11C	1	
3.0 T	GE Discovery MR750	DV26_R02	1	
Children's Hospital of Philadelphia				13
1.5 T	Siemens Magnetom Avanto	VE11C	1	
3.0 T	Siemens Magnetom Skyra	VE11C	1	
University of Texas Southwestern Medical Center				8
1.5 T	Philips Ingenia	R5.3.0.3	1	
3.0 T	Philips Ingenia	R5.3.0.3	1	
University of California San Diego				6
3.0 T	GE Discovery MR750	DV26_R01	2	
Duke University				15
1.5 T	Siemens Magnetom Aera	VE11C	1	
3.0 T	Siemens Magnetom Skyra	VE11C	1	
Cincinnati Children's Hospital Medical Center				14
1.5 T	Philips Ingenia	R5.3.1.1	1	
3.0 T	Philips Ingenia	R5.3.1.1	1	
Mayo Clinic				13
1.5 T	Siemens Magnetom Aera	VE11C	1	
3.0 T	GE Discovery MR750w	DV26_R01	1	
3.0 T	Siemens Magnetom Skyra	VE11C	1	

appropriate boundaries and limits on bias. The data will provide confirmation of linearity and bias of MRI-derived PDFF, thus guiding QIBA and other technical specifications for PDFF as a quantitative imaging biomarker of fat content (24,25), including in organs and tissue beyond the liver (26–28).

Materials and Methods

This prospective study was conducted at nine participating centers affiliated with the QIBA PDFF Biomarker Committee between September 2018 and October 2019. The study included no human participants and was exempt from institutional board review and Health Insurance Portability and Accountability Act requirements.

Phantom

A single, commercially available PDFF phantom (Fat Fraction Phantom, Model 300; Calimetrix) was constructed from 12 vials (20 mL, 27 mm in diameter, 61 mm long). The volume of the emulsion ingredients defined the reference PDFF values in each vial, which were verified by standard Calimetrix quality procedures. A representative photograph and PDFF map of the phantom is shown in Figure 1. The PDFF values of these vials were selected to capture the relevant biologic range of PDFF in human liver (0% to ~50%) and the entire physical range (0%–100%). The vials were agar gel based, with the exception of the 100% PDFF vial, to avoid motion-related artifacts (ie, vibrations) and to minimize susceptibility artifacts caused by air bubbles. The vials were housed within a spherical acrylic casing (19-cm outer diameter) filled with doped water to optimize magnetic field homogeneity. The phantom was shipped between centers during the study period by next-day courier. When not in use, the phantom was kept in a foam-padded storage case at room temperature. At each center, staff was instructed to allow the phantom to equilibrate to room temperature in the MRI suite before imaging.

Table 2: Representative Imaging Parameters for Vendor Three-dimensional Acquisitions

| Parameter	GE Healthcare IDEAL IQ	Philips Healthcare mDIXON Quant	Siemens Healthineers LiverLab Package	QIBA Standardized Phantom
1.5 T				
FOV (cm)	40–44	40	36–40	As small as possible
Matrix size	130 × 130	192 × 192	160 × 160	≥128 × 128
Section thickness (mm)	10	5–10	5	5–10
TR (msec)	11.79–12.67	5.32	13.3	Minimum TR allowed or 7–12
First TE (msec)	0.87–1.02	0.92	2.0–2.3	2.0–2.5
TE spacing (msec)	1.46–1.98	0.66	2.0–2.3	2.0–2.5
No. of echo trains/no. of echoes	1/6	1/6	1/6	Single interleaf preferred, multiple interleaves allowed/6
BW (approximate kHz/pixel)	1	1.8–2	1	As high as possible
FA (degrees)	5	5	5	3–5 for three-dimensional; 10 for two-dimensional
Readout type	Monopolar	Bipolar	Bipolar	Monopolar or bipolar
3.0 T				
FOV (cm)	40–44	40	36–40	As small as possible
Matrix size	160 × 160	192 × 192	160 × 160	≥128 × 128
Section thickness (mm)	10	10	5	5–10
TR (msec)	5.82–7.32	5.55–5.65	9–10	Minimum TR allowed or 7–12 preferred for three-dimensional; 150 for two-dimensional
First TE (msec)	0.87–0.99	0.69–0.97	1.21–1.25	0.8–1.5
TE spacing (msec)	0.69–0.83	0.71–0.97	1.23–1.25	1.0–1.5
No. of echo trains/no. of echoes	2/6	1/6	1/6	Single interleaf preferred, multiple interleaves allowed/6
BW (approximate kHz/pixel)	1.8–2	1.8–2	1	As high as possible
FA (degrees)	3	3	3	3–5 for TR of 7–12 for three-dimensional; 10 for TR of 150 for two-dimensional
Readout type	Monopolar	Bipolar	Bipolar	Monopolar or bipolar

Note.—Base pulse sequence was either a two-dimensional or a three-dimensional multiecho spoiled gradient-echo acquisition. All sites in this study, however, implemented the three-dimensional sequence for the standardized phantom protocol. No parallel imaging was used; all echoes were fully sampled and acquired with one signal average. No phase wrap, reduced-phase field of view, flow compensation, distortion correction, or postprocessing data filters were used. BW = bandwidth, FA = flip angle, FOV = field of view, QIBA = Quantitative Imaging Biomarkers Alliance, TE = echo time, TR = repetition time.

Imaging Experiments

Table 1 summarizes the details of the nine imaging centers and their MRI systems. Each center was asked to acquire data by using two protocols. The first protocol was a vendor default CSE MRI PDFF protocol (LiverLab package, Siemens Healthineers; mDIXON Quant, Philips Healthcare; IDEAL IQ, GE Healthcare). Each center ran this protocol with default system parameters for the acquisition. The second protocol was a standardized protocol developed for phantom imaging by the QIBA PDFF Biomarker Committee for this study. Each center was requested to strictly follow parameters for the standardized protocol. On Philips platforms, adherence to the standardized protocol required the local clinical scientist to remove parameter limitations (ie, allowed range) to achieve the requested echo times. Hereafter, these will be referred to as vendor and standardized protocols, respectively. Table 2 summarizes the typical imaging parameters used for each protocol and MRI system.

For both protocols, three-dimensional volumetric CSE MRI image data sets were acquired in the axial orientation, centered on the phantom, capturing the cross sections of the 12 vials as shown in Figure 1. Vendor-supplied multichannel head or torso array coils were used. Online software from each vendor was used to reconstruct quantitative parametric PDFF maps, accounting for multipeak fat spectral modeling and correcting for T2* (1/R2*) decay (29–31), with no modifications to PDFF reconstruction or postprocessing parameters. Water-fat swaps were observed at all Siemens Healthineers sites in the majority of acquisitions. When water-fat swaps occurred in the phantom, they were commonly observed across the entire PDFF map, regardless of protocol. In such cases, the on-line reconstruction was repeated by utilizing a built-in option of the product reconstruction algorithm to force unswapping of the water-fat misassignment, by using a script provided by the manufacturer. The script makes a change in the configuration file accessed by the product reconstruction algorithm. Water-fat swaps in the phantom were not observed on GE Healthcare and Philips Healthcare systems used in this study. To assess temporal phantom stability, one site (University of
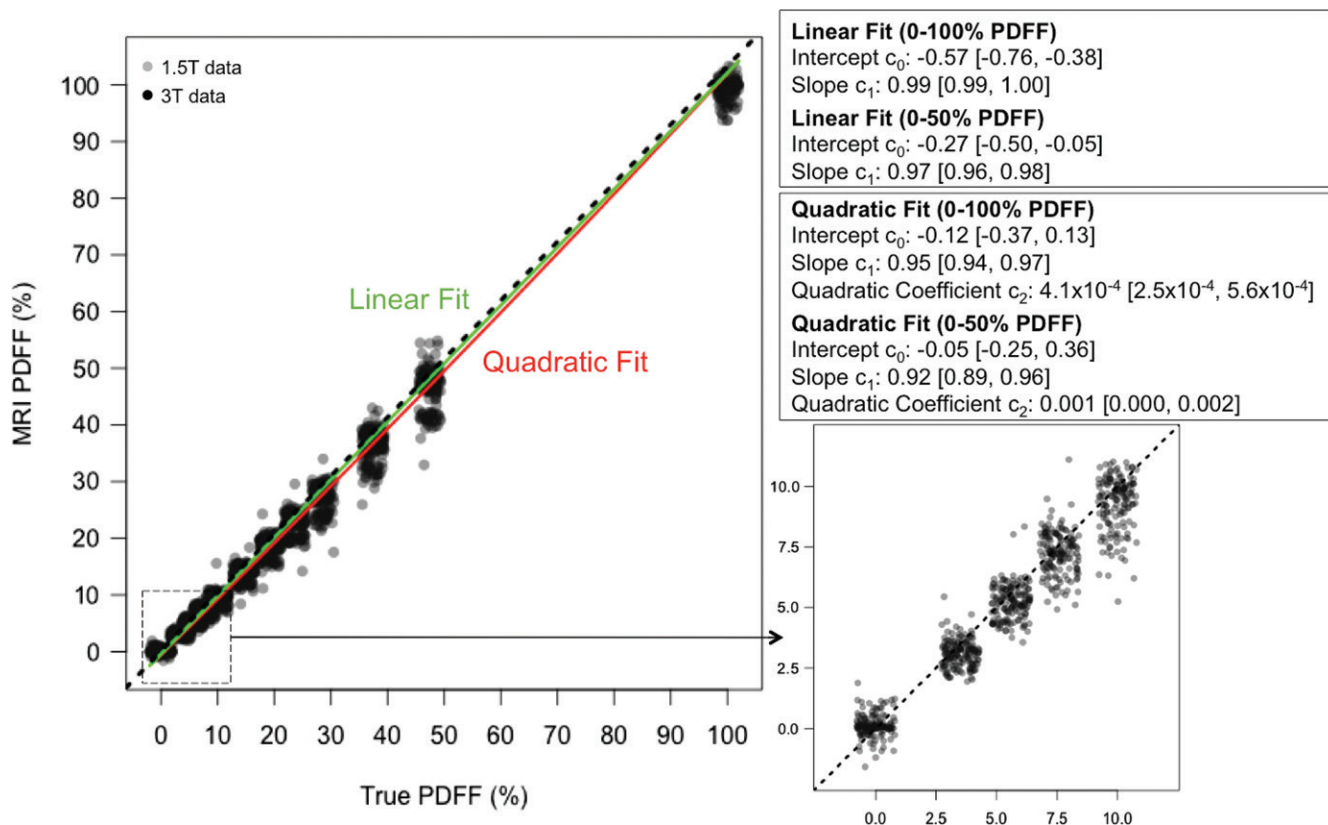
Figure 2: Plots show chemical shift–encoded MRI-based proton density fat fraction (PDFF) versus phantom's reference PDFF values, summarizing linear (green) and quadratic (red) regression fits. Data pooled from all acquisitions, independent of protocol, magnetic field strength, and vendor. Dashed line represents identity. Corresponding table shows linear and quadratic regression fit parameters for entire physical range (0%–100% PDFF). For comparison, linear and quadratic regression fit parameters for biologically relevant PDFF range (0%–47.5%) were also computed. Enlarged inset further illustrates data from 0%–10% PDFF, showing strong linearity of measurements in low-PDFF range.

Wisconsin–Madison) imaged the phantom with both protocols on three separate occasions across all of their imaging systems (September 2018, April 2019, and October 2019).

Image Processing

The PDFF maps were submitted to a central data processing center not affiliated with Calimetrix, any MRI vendor, or the University of Wisconsin–Madison. A single analyst (H.H.H., with 20 years of MRI experience), unblinded, reviewed each center's data to ensure adherence to acquisition protocol parameters and to confirm absence of gross artifacts. The analyst subsequently made all PDFF estimations with software (OsiriX MD 10; Pixmeo). A single, large, circular region of interest (~3.5–4 cm^2) was drawn on the cross section of each vial.

Statistical Analysis

A separate team member (T.Y., with 20 years of experience) performed statistical analysis under supervision of a faculty biostatistician (N.O., with 30 years of experience). These investigators were not affiliated with Calimetrix, the University of Wisconsin–Madison, or any MRI vendor. Statistical analyses were performed with software (R version 3.6.3; R Foundation for Statistical Computing).

Linearity was assessed over the biologically relevant range for human liver fat (ie, 0% to ~50% PDFF) and over the entire physical range of 0%–100% PDFF. This was done by fitting the MRI data collected on the 0%–47.5% vials (ie, biologically relevant range) and the 0%–100% vials (ie, full physical PDFF range), respectively, at all imaging centers and by using all vendors, magnetic field strengths, and protocols. Linearity, or lack of nonlinearity, is generally assessed by comparing a linear fit with alternative nonlinear (ie, quadratic, cubic, and higher order polynomial) fits. Per typical QIBA convention, a linear model ($y = c_0 + c_1*x$), a quadratic model ($y = c_0 + c_1*x + c_2*x^2$), and a cubic model ($y = c_0 + c_1*x + c_2*x^2 + c_3*x^3$) were postulated herein, where phantom-derived reference PDFF value was the independent variable and estimated PDFF was the dependent variable. The second-degree quadratic coefficient c_2 and its 95% CI were compared with the maximum acceptable value of c_2 less than 0.01, above which the quadratic effect of a quantitative imaging biomarker cannot be ignored in clinical use (32).

Estimation bias of MRI PDFF was assessed against phantom-derived reference PDFF values by using linear regression with mixed-effects modeling. The regression intercept (ie, constant bias) and the slope (ie, proportional bias) were compared with the null hypotheses of intercept of 0 and slope of 1, respectively, and 95% CIs were predominantly used. Repeated examinations in the same system per vial were modeled as a random effect. Because scanners from different vendors at different field strengths were unequally
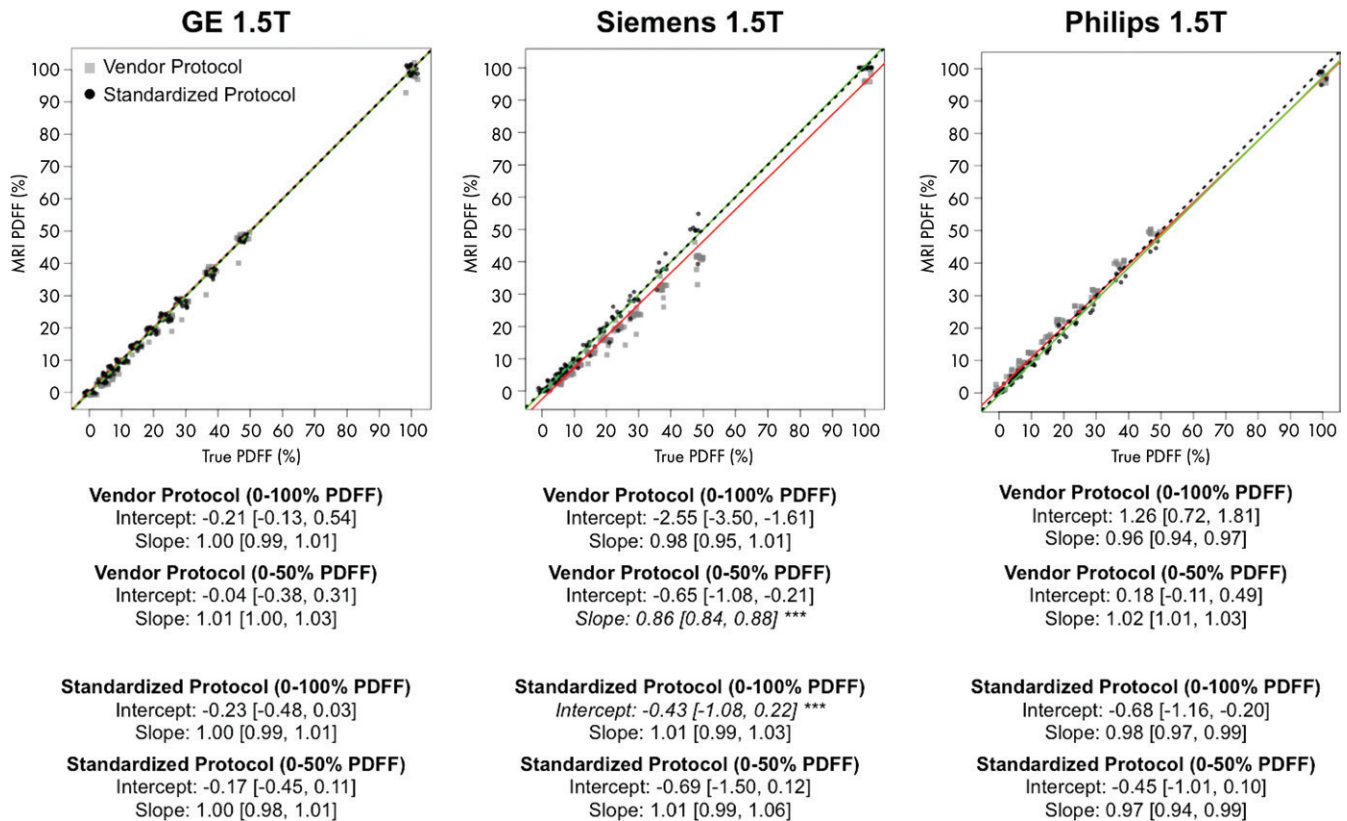
Figure 3: Plots show chemical shift–encoded MRI-based proton density fat fraction (PDFF) versus phantom-derived reference PDFF values, categorized according to vendor (columns), at 1.5 T. Data have been slightly jittered for visualization purposes. Red line represents linear regression fit of vendor protocol data, and green line represents regression fit of standardized phantom protocol data. The two lines nearly overlap in some instances. Regression parameters are summarized below each plot for 0%–47.5% PDFF and for 0%–100% PDFF ranges. Asterisks (***) and italicized text under each subplot indicate statistically smaller bias (ie, intercept and/or slope) compared with other protocol for given vendor, magnetic field strength, and PDFF range considered. Dashed lines represent identity. Data in brackets are 95% CIs.

represented in the sample (Table 1), inverse probability sampling weights were used in the fitting procedure, where appropriate, to correct for the imbalanced population of MRI systems. For point estimates, 95% CIs outside of the null hypothesis (ie, $P < .05$) were considered to indicate a statistically significant difference.

Next, the PDFF estimation bias assessment was stratified for each combination of vendor and field strength, graphically by using scatter plots and statistically by using mixed-model regression, with repeated examinations modeled as a random effect. Regression intercept and slope estimates and their 95% CIs were compared with the null hypotheses of intercept of 0 and slope of 1, respectively. Bias was also compared between the two protocols by calculating the magnitude of biases in estimated intercept and slope (between 0 and 1, respectively) and by testing the statistical significance of their differences by using bootstrap 95% CIs. Bias in the regression intercept and slope was also assessed on a per-scanner basis.

Last, potential temporal drift in the phantom-derived reference PDFF values was assessed by comparing PDFF values between paired points (September 2018 vs April 2019, April 2019 vs October 2019). Analysis was performed by using a mixed-effects linear model, with repeated estimations on the same vial in the same imaging session modeled as a random effect.

A priori sample-size calculation was not performed. Instead, we required at least three different MRI systems for each combination of vendor and magnetic field strength to calculate the mean and standard deviation of each point estimate.

Results

At all centers, data from both protocols were successfully acquired. MRI PDFF linearity was assessed by comparing the linear and alternative quadratic model fits (Fig 2). The second-degree coefficient (c_2) was significant but very close to 0, regardless of whether it was calculated from the 0%–50% or 0%–100% PDFF data. The 95% CI for c_2 did not exceed the 0.01 threshold for clinically significant nonlinearity (32). A third-degree cubic fit yielded $c_0 = -0.001$ (95% CI: -0.38, 0.38), $c_1 = 0.94$ (95% CI: 0.86, 1.02), $c_2 = 0.00$ (95% CI: -0.01, 0.01), and $c_3 = -4.3$ (95% CI: -4.3, 7.1) $\times 10^{-5}$, further confirming the small contribution from nonlinear terms. Therefore, linear modeling was used for all subsequent analyses. Pooling across all vendors, magnetic field strengths, protocols, and systems, the intercept and slope of the linear model were lower than the identity line, with absolute PDFF percentage of the intercept equal to -0.27% (95% CI: -0.50, -0.05) and the slope equal to 0.97 (95% CI: 0.96, 0.98), but the differences from

GE 3T

Vendor Protocol (0–100% PDFF)
Intercept: 0.37 [-0.48, 0.03]
Slope: 1.00 [0.99, 1.01]

Vendor Protocol (0–50% PDFF)
Intercept: 0.10 [-0.18, 0.38]
Slope: 1.01 [1.00, 1.02]

Standardized Protocol (0–100% PDFF)
Intercept: -0.09 [-0.29, 0.12]
Slope: 1.01 [1.00, 1.01]

Standardized Protocol (0–50% PDFF)
Intercept: -0.21 [-0.46, 0.04]
Slope: 1.01 [1.00, 1.02]

Siemens 3T

Vendor Protocol (0–100% PDFF)
Intercept: -1.81 [-2.50, -1.12]
Slope: 0.99 [0.97, 1.01]

Vendor Protocol (0–50% PDFF)
Intercept: -0.48 [-1.04, 0.09]
*Slope: 0.91 [0.88, 0.93] ****

Standardized Protocol (0–100% PDFF)
*Intercept: -0.72 [-1.04, -0.40] ****
Slope: 1.00 [0.99, 1.01]

Standardized Protocol (0–50% PDFF)
Intercept: -0.40 [-0.76, -0.04]
Slope: 0.98 [0.96, 1.00]

Philips 3T

Vendor Protocol (0–100% PDFF)
*Intercept: -0.26 [-0.55, -0.04] ****
Slope: 0.98 [0.97, 0.99]

Vendor Protocol (0–50% PDFF)
Intercept: -0.27 [-0.62, -0.08]
Slope: 0.98 [0.97, 1.00]

Standardized Protocol (0–100% PDFF)
Intercept: -2.82 [-3.76, -1.89]
*Slope: 1.00 [0.97, 1.02] ****

Standardized Protocol (0–50% PDFF)
Intercept: -0.78 [-1.35, -0.20]
*Slope: 0.87 [0.84, 0.89] ****

Figure 4: Plots show chemical shift–encoded MRI-based proton density fat fraction (PDFF) versus phantom-derived reference PDFF values, categorized according to vendor (columns), at 3.0 T. Data have been slightly jittered for visualization purposes. Red line represents linear regression fit of vendor protocol data, and green line represents regression fit of standardized phantom protocol data. The two lines nearly overlap in some instances. Regression parameters are summarized below each plot for 0%–47.5% PDFF and for 0%–100% PDFF ranges. Asterisks (***) and italicized text under each subplot indicate statistically smaller bias (ie, intercept and/or slope) compared with other protocol for given vendor, magnetic field strength, and PDFF range considered. Dashed lines represent identity. Data in brackets are 95% CIs.

Table 3: Bias Comparison of Vendor Default versus Standardized Phantom Protocols

	GE Healthcare		Philips Healthcare		Siemens Healthineers	
Parameter	Intercept	Slope	Intercept	Slope	Intercept	Slope
1.5 T	−0.13 (−0.39, 0.02)	0.01 (−0.01, 0.03)	0.31 (−0.70, 0.051)	−0.02 (−0.04, 0.02)	−0.05 (−0.62, 0.75)	0.11 (0.07, 0.16)*
3.0 T	−0.15 (−0.37, 0.05)	−0.01 (−0.02, 0.01)	−0.51 (−1.07, −0.02)*	−0.11 (−0.13, −0.08)*	0.09 (−0.20, 0.46)	0.06 (−0.09, 0.06)

Note.—Data in parentheses are 95% CIs. All intercept and associated CI values are expressed as absolute proton density fat fraction percentage, and results are rounded to two decimal digits. Positive values of difference in bias indicate greater bias with the vendor's default protocol, whereas negative values indicate greater bias with the standardized phantom protocol.
* Statistically significant differences in bias (ie, difference in bias ≠ 0) between the two protocols for the given vendor and magnetic field strength at α = .05 (ie, 95% CIs that do not include 0).

ideal intercept of 0 and slope of 1 (ie, null hypothesis) were small (Fig 2).

The results of the stratified regression analyses per vendor and magnetic field strength are shown in Figures 3 and 4. Again, separate regression coefficients for 0%–50% PDFF and 0%–100% PDFF are presented. The differences in bias because of protocol (vendor default vs standardized protocol) are further summarized in Table 3. Figure 5 illustrates the data as residual plots, highlighting differences between vendors, magnetic field strength, and protocol. On GE Healthcare systems with IDEAL IQ (Fig 5), no deviation from the identity line was observed in the intercepts or slopes regardless of the protocols, magnetic field strengths,

or PDFF range of regression (0%–50% vs 0%–100%), with the 95% CIs of the intercept and slope estimates inclusive of the values 0 and 1, respectively.

On Siemens Healthineers systems with the LiverLab package (Fig 5), no bias in intercept was found regardless of the protocol or magnetic field strength with 95% CIs including 0. However, the slope was biased away from 1 for the vendor protocol, with values of 0.86 (95% CI: 0.85, 0.88) at 1.5 T and 0.91 (95% CI: 0.89, 0.94) at 3.0 T. However, no bias was found in the slope at either magnetic field strength with the standardized protocol, with 95% CIs inclusive of the ideal value 1. The standardized protocol was less biased than the vendor protocol in slope (differences in
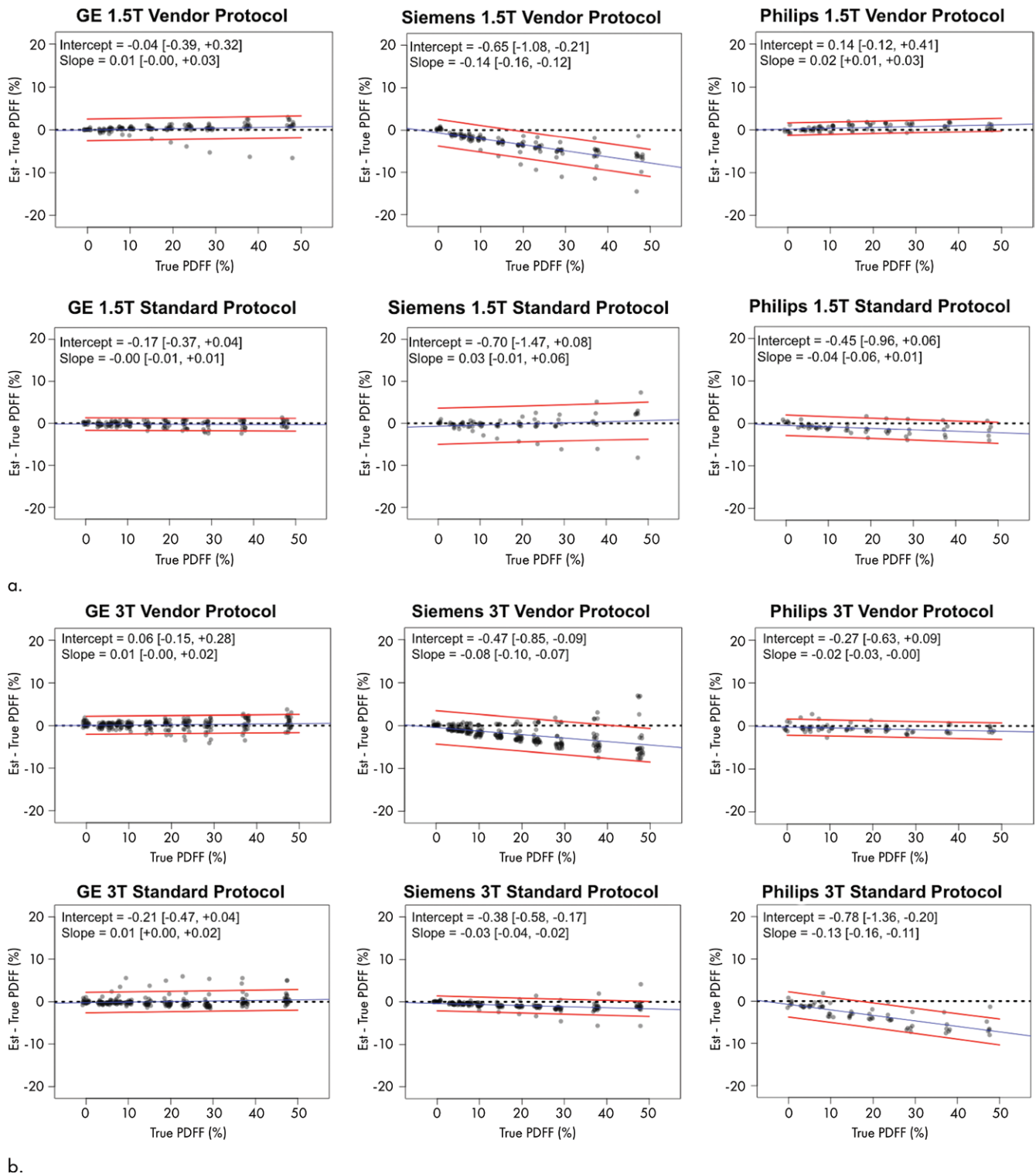
Figure 5: Plots show chemical shift–encoded MRI-based proton density fat fraction (PDFF) versus phantom-derived reference PDFF values, categorized according to vendor and protocol, shown for **(a)** 1.5 T and **(b)** 3.0 T. Difference between estimated (est.) PDFF versus reference (ie, true) PDFF is plotted on y-axis. Dashed black line denotes zero error. Blue line is linear regression fit of data. Two accompanying red lines are 95% prediction interval bounds. Data in brackets are 95% CIs.

bias 95% CI: 0.07, 0.16; exclusive of 0) at 1.5 T, but not at 3.0 T (Table 3).

On Philips Healthcare systems with mDIXON Quant (Fig 5), no bias in intercept or slope was found for the vendor protocol at either magnetic field strength, with 95%

CIs including 0 and 1, respectively. By using the standardized protocol, however, the slope was 0.97 (95% CI: 0.94, 0.99) at 1.5 T, and at 3.0 T the slope was 0.87 (95% CI: 0.84, 0.89). Both slopes were biased from identity. This may be because the echo times of the standardized protocol

Table 4: Intercepts and Slopes per System

System	Vendor Default Protocol		Standardized Phantom Protocol	
	Intercept	Slope	Intercept	Slope
GE 1.5 T				
UW	−0.17(−0.78, 0.45)	0.97 (0.94, 1.00)	−0.05 (−0.36, 0.28)	1.01 (1.00, 1.02)
UW	0.11 (−0.51, 0.72)	1.03 (1.00, 1.06)	−0.45 (−0.71, −0.13)	0.97 (0.96, 0.99)
UW	0.04 (−0.57, 0.65)	1.03 (1.01, 1.06)	−0.24 (−0.55, 0.09)	1.00 (0.99, 1.01)
UW	−0.13 (−0.74, 0.48)	1.02 (1.00, 1.05)	0.07 (−0.26, 0.39)	1.01 (0.99, 1.02)
Summary	−0.04 (−0.38, 0.31)	1.01 (1.00, 1.03)	−0.17 (−0.48, 0.11)	1.00 (0.98, 1.01)
GE 3.0 T				
Mayo Clinic	0.62 (0.16, 1.08)	1.01 (0.99, 1.03)	−0.46 (−1.55, 0.62)	1.01 (0.96, 1.05)
Mayo Clinic	−0.17 (−0.82, 0.49)	0.97 (0.94, 1.00)	−0.57 (−1.66, 0.51)	1.01 (0.96, 1.06)
NCH	−0.61 (−0.93, −0.28)	0.97 (0.95, 0.98)	0.03 (−0.51, 0.58)	1.00 (0.98, 1.02)
UCSD	0.28 (−0.38, 0.93)	1.02 (0.99, 1.05)	−0.08 (−1.16, 1.01)	1.00 (0.96, 1.06)
UCSD	0.02 (−0.44, 0.49)	1.04 (1.01, 1.06)	−0.26 (−1.03, 0.51)	0.99 (0.96, 1.03)
UW	−0.07 (−0.45, 0.31)	1.01 (0.99, 1.03)	0.28 (−0.34, 0.91)	1.04 (1.01, 1.07)
UW	0.43 (0.048, 0.81)	1.04 (1.01, 1.05)	−0.82 (−1.45, −0.20)	1.02 (1.00, 1.05)
UW	0.03 (−0.35, 0.41)	1.02 (1.01, 1.04)	−0.19 (−0.82, 0.44)	0.99 (0.97, 1.02)
UW	0.42 (0.043, 0.80)	1.01 (0.99, 1.02)	−0.25 (−0.88, 0.38)	1.03 (1.01, 1.06)
Summary	0.10 (−0.18, 0.38)	1.01 (1.00, 1.01)	−0.21 (−0.46, 0.04)	1.01 (1.00, 1.02)
Siemens 1.5 T				
CHOP	−0.61 (−1.36, 0.14)	0.88* (0.84, 0.91*)	−0.65 (−1.79, 0.49)	0.94 (0.89, 0.99)
Duke	−0.67 (−1.54, 0.20)	0.84* (0.80, 0.88*)	−0.60 (−1.74, 0.54)	1.09 (1.04, 1.14)
Mayo Clinic	−0.66 (−1.33, 0.01)	0.86* (0.83, 0.89*)	−0.79 (−1.72, 0.14)	1.04 (1.00, 1.08)
Summary	−0.65 (−1.08, −0.21)	0.86* (0.84, 0.88*)	−0.69 (−1.50, 0.12)	1.01 (0.99, 1.06)
Siemens 3.0 T				
CHOP	−0.59 (−1.39, 0.20)	0.89* (0.85, 0.92*)	−0.52 (−0.91, −0.13)	0.99 (0.97, 1.01)
Duke	−0.57 (−1.22, 0.08)	0.89* (0.86, 0.92*)	−0.52 (−1.00, −0.04)	1.03 (1.01, 1.05)
NCH	−0.43 (−0.96, 0.10)	0.98 (0.95, 1.00)	−0.26 (−0.65, 0.13)	0.94 (0.92, 0.96)
NCH	−0.32 (−1.03, 0.39)	0.87* (0.84, 0.90*)	−0.34 (−0.56, −0.11)	0.97 (0.96, 0.98)
Summary	−0.48 (−1.04, 0.09)	0.91* (0.88, 0.93*)	−0.40 (−0.76, −0.04)	0.98 (0.96, 0.93)
Philips 1.5 T				
CCHMC	−0.03 (−0.34, 0.28)	1.02 (1.00, 1.03)	−0.40 (−1.01, 0.21)	1.00 (0.97, 1.03)
UM	0.84 (0.22, 1.45)	1.00 (0.98, 1.03)	−0.53 (−1.39, 0.33)	0.98 (0.94, 1.02)
UTSW T	0.15 (−0.29, 0.59)	1.04 (1.02, 1.06)	−0.46 (−1.07, 0.15)	0.92 (0.90, 0.95)
Summary	0.18 (−0.11, 0.49)	1.02 (1.01, 1.03)	−0.45 (−1.01, 0.10)	0.97 (0.94, 0.99)
Philips 3.0 T				
CCHMC	0.14 (−0.47, 0.75)	0.97 (0.95, 1.00)	−0.31 (−1.20, 0.57)	0.90* (0.86, 0.94*)
UM	−1.10 (−1.95, −0.24)	1.00 (0.96, 1.04)	−0.88 (−2.13, 0.37)	0.84* (0.79, 0.90*)
UM	−0.62 (−1.48, 0.24)	0.98 (0.94, 1.02)	−0.81 (−2.06, 0.44)	0.83* (0.77, 0.88*)
UTSW	−0.09 (−0.70, 0.51)	0.98 (0.95, 1.01)	−1.18 (−2.06, −0.21)	0.88* (0.84, 0.92*)
Summary	−0.27 (−0.62, −0.08)	0.98 (0.97, 1.00)	−0.78 (−1.35, −0.20)	0.87* (0.84, 0.89*)

Note.—Data in parentheses are 95% CIs. Each row represents a unique MRI system. Intercept and associated CI values are expressed as absolute PDFF percentage; results are rounded to two decimal digits. The intercepts and slopes per system were on the basis of linear regression for the physiologic range of 0%–50% proton density fat fraction (11 vials). CCHMC = Cincinnati Children's Hospital Medical Center, CHOP = Children's Hospital of Philadelphia, Duke = Duke University, NCH = Nationwide Children's Hospital, PDFF = proton density fat fraction, UCSD = University of California, San Diego, UM = University of Michigan, UTSW = University of Texas Southwestern Medical Center, UW = University of Wisconsin.
* Values outside of the equivalence criteria of intercept (−1.5%, 1.5%) and/or slope (0.9, 1.1).

were outside the permissible range for the default Philips mDIXON Quant inversion algorithm. Although the standardized protocol parameters were achievable after the local clinical scientist removed such parameter constraints in the software, the resultant PDFF could have consequentially become more biased.

Figure 5 illustrates the data as residual plots, highlighting differences between vendors, magnetic field strength, and protocol. Individual scanner accuracy and bias assessments are shown in Table 4.

The temporal stability data of the PDFF phantom are shown in Figure 6 and Table 5. The greatest change was a

mere −0.8% (95% CI: −2.0, 0.3), observed in the 47.5% vial. No temporal drift of 2.0% or greater in absolute PDFF occurred in any of the 0%–47.5% vials.
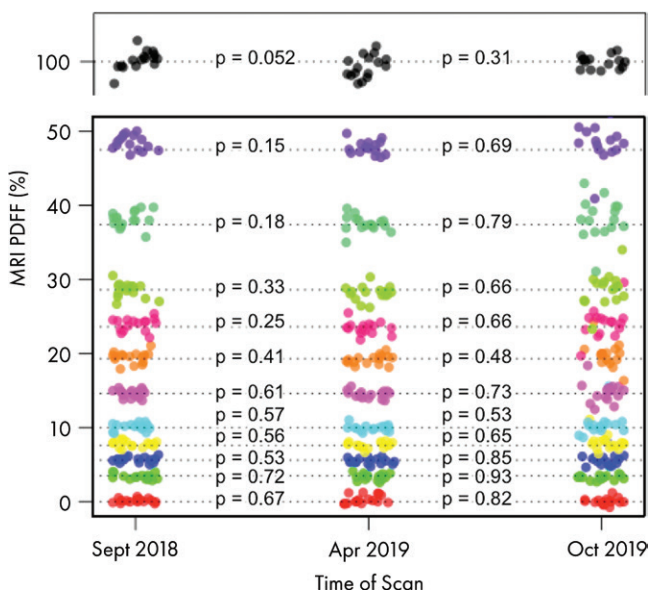
Finally, Table 6 summarizes distributions of PDFF across magnetic field strength and protocols by pooling data from all centers and across the three vendors.

Discussion

Proton density fat fraction (PDFF) estimated by using chemical shift–encoded (CSE) MRI is an accepted imaging biomarker of hepatic steatosis. We performed a phantom study involving nine academic centers and multiple vendor platforms to

Figure 6: Image shows results of repeated assessments of phantom during span of 13 months. Different colors denote each of 12 vials, and each point represents one acquisition. *P* values across three points for each vial are shown and reveal that no vial-specific changes in proton density fat fraction (PDFF) values were detected. Apr = April, Oct = October, Sept = September.

evaluate the accuracy of CSE MRI to quantify PDFF by using vendor default and Quantitative Imaging Biomarkers Alliance (QIBA)–standardized protocols and vendor-specific reconstruction algorithms. By using a commercially available phantom, we demonstrated high linearity and low bias of CSE MRI across the entire physiologic range of PDFF values. We found that PDFF bias can depend on vendor, magnetic field strength, or protocol, but it can be controlled within the intercept range (−1.5% to 1.5%) and the slope range (0.9 to −1.1), regardless of field strength or vendor, by informed selection of the appropriate protocol (either vendor default or QIBA standardized). Specifically, GE Healthcare systems regardless of vendor or standardized protocol, Siemens systems that use standardized protocol, and Philips systems that use vendor protocol are within these bias margins. These small constant and proportional biases are unlikely to be consequential in most human clinical use cases for liver imaging and assessment. Therefore, we propose these ranges to define the acceptance bias margins for PDFF derived from commercially available quantitative CSE MRI pulse sequences.

Previously, Hernando et al (19) performed a similar study to evaluate complex-based CSE MRI methods in estimating PDFF across multiple vendors by using a centralized reconstruction pipeline. That study relied on investigational reconstruction software to reconstruct data from participating centers. More recently, Jang et al (21) performed a study evaluating three commercial CSE MRI methods on a limited number of systems by using a fat fraction phantom. That study demonstrated excellent linearity and reproducibility of the evaluated commercial CSE MRI methods. However, it also relied on investigational reconstruction software to reconstruct data from participating centers. Decoupling contributions to deviations related specifically to acquisition (hardware, pulse sequence) and PDFF reconstruction method would require reconstruction of raw data by using a common, centralized reconstruction algorithm. Use of a common approach to spectral modeling of PDFF and T2* (1/R2*) correction, for example, may reduce

Table 5: Proton Density Fat Fraction Stability Measurements Assessed at Beginning, Middle, and End of Study

Vial	Reference PDFF	Baseline PDFF September 2018	Change in PDFF April 2019	Change in PDFF October 2019
1	0.0	0.1 (−0.4, 0.7)	0.1 (−0.6, 1.0)	−0.1 (−0.9, 0.7)
2	3.5	3.4 (2.8, 4.0)	−0.2 (−1.3, 0.9)	0.1 (−1.1, 1.2)
3	5.6	5.7 (5.1, 6.3)	−0.4 (−1.5, 0.7)	0.1 (−1.0, 1.2)
4	7.6	7.8 (7.2, 8.3)	−0.3 (−1.5, 0.8)	0.3 (−0.9, 1.4)
5	10.0	10.1 (9.5, 10.6)	−0.3 (−1.5, 0.8)	0.4 (−0.7, 1.5)
6	14.6	14.6 (14.0, 15.2)	−0.3 (−1.4, 0.8)	0.2 (−0.9, 1.3)
7	19.3	19.5 (18.9, 20.0)	−0.5 (−1.6, 0.6)	0.4 (−0.7, 1.5)
8	23.4	23.9 (23.3, 24.5)	−0.7 (−1.8, 0.4)	0.3 (−0.9, 1.4)
9	28.6	28.5 (28.0, 29.1)	−0.6 (−1.7, 0.5)	0.2 (−0.9, 1.3)
10	37.4	38.1 (37.6, 38.7)	−0.8 (−1.9, 0.3)	0.2 (−1.0, 1.3)
11	47.5	48.5 (47.9, 49.0)	−0.8 (−2.0, 0.3)	0.2 (−0.9, 1.6)
12	100.0	100.2 (99.6, 100.8)	−1.1 (−2.3, 0.0)	−0.6 (−1.7, 0.5)

Note.— All values are absolute PDFF percentages; data in parentheses are 95% CIs. Beginning, middle, and end refer to September 2018, April 2019, and October 2019, respectively. PDFF = proton density fat fraction.

Table 6: Proton Density Fat Fraction Information

Vial	Reference PDFF	Vendor Default Protocol, 1.5 T	Vendor Default Protocol, 3.0 T	Standardized Phantom Protocol, 1.5 T	Standardized Phantom Protocol, 3.0 T
1	0.0	0.1 ± 0.4 (−0.1, −0.0)	0.2 ± 0.6 (0.1, 0.4)	0.2 ± 0.4 (0.0, 0.3)	0.1 ± 0.8 (−0.1, 0.3)
2	3.5	3.0 ± 0.5 (2.8, 3.2)	3.1 ± 0.7 (3.0, 3.3)	3.2 ± 0.5 (3.0, 3.4)	3.1 ± 0.5 (2.9, 3.2)
3	5.6	5.1 ± 0.8 (4.8, 5.4)	5.2 ± 0.8 (5.0, 5.5)	5.3 ± 0.7 (5.1, 5.6)	5.2 ± 0.5 (5.1, 5.4)
4	7.6	7.1 ± 1.1 (6.7, 7.5)	7.2 ± 0.9 (7.0, 7.5)	7.4 ± 1.0 (7.0, 7.8)	7.3 ± 0.9 (7.1, 7.6)
5	10.0	9.4 ± 1.4 (8.9, 9.9)	9.2 ± 1.1 (8.9, 9.5)	9.6 ± 0.8 (9.3, 9.9)	9.3 ± 1.4 (8.9, 9.8)
6	14.6	13.8 ± 1.9 (13.1, 14.4)	13.6 ± 1.4 (13.3, 14.0)	14.0 ± 1.0 (13.6, 14.4)	13.6 ± 1.3 (13.2, 14.0)
7	19.3	18.2 ± 2.5 (17.4, 19.1)	18.1 ± 1.7 (17.7, 18.6)	18.9 ± 1.4 (18.3, 19.4)	18.2 ± 1.5 (17.8, 18.7)
8	23.4	22.3 ± 2.7 (21.4, 23.3)	22.2 ± 2.0 (21.7, 22.8)	23.2 ± 1.6 (22.6, 23.9)	22.5 ± 1.8 (22.0, 23.1)
9	28.6	26.8 ± 3.2 (25.7, 28.0)	26.6 ± 2.4 (26.0, 27.3)	27.7 ± 1.8 (27.0, 28.4)	26.9 ± 2.2 (26.2, 27.4)
10	37.4	35.7 ± 3.7 (34.4, 37.1)	35.9 ± 2.8 (35.1, 36.6)	37.0 ± 2.0 (36.2, 37.8)	36.1 ± 2.7 (35.3, 36.9)
11	47.5	45.4 ± 4.3 (43.9, 46.9)	45.9 ± 3.9 (45.9, 47.0)	47.5 ± 2.7 (46.5, 48.7)	46.5 ± 2.9 (45.6, 47.4)
12	100	98.1 ± 2.6 (97.2 99.0)	99.4 ± 1.6 (99.0, 99.8)	99.0 ± 2.2 (98.2, 99.9)	100.1 ± 0.5 (100.0, 100.3)

Note.—Mean data are ± standard deviation; data in parentheses are 95% CIs. PDFF = proton density fat fraction.

variability. However, use of a centralized reconstruction algorithm may not address vendor-specific hardware constraints or imperfections that are mitigated by vendor-specific reconstruction algorithms, and thus could introduce additional sources of error. In our study, we intentionally included all sources of variability and error likely to be encountered in clinical scenarios in which default vendor-supplied hardware and software are used to produce PDFF estimations.

Compared with prior works, our reference standard was a commercial PDFF phantom constructed with a wide range of PDFF values. Repeated acquisitions at one center (University of Wisconsin–Madison) demonstrated temporal phantom stability, providing important evidence that the phantom-derived reference PDFF values were stable during our study. Among the advantages of using commercial PDFF phantoms are the use of materials and geometries optimized during multiple development cycles and the use of quality controls during manufacturing and assembly. In addition, the same type of phantom can be used for comparison studies in the future.

Our study had limitations. First, although an evaluation of all commercially available CSE MRI methods was performed on the majority of available platforms, a comprehensive evaluation of all possible combinations of hardware platforms and software versions was not feasible. Second, the phantom did not modulate T2* (1/R2*) over the full physiologic range of R2* values, in particular not reflecting values seen in patients with moderate to severe iron overload. All of the CSE MRI methods assessed in our study used R2* correction (29–31) to correct for the presence of iron in vivo in patients with mild to moderate iron overload (33,34). Third, all acquisitions were performed at room temperature, and temperature can affect relative chemical shifts across water and fat resonance peaks (35,36). Differences in temperature at which imaging was performed could explain, in part, differences between estimated and reference PDFF values. The degree of the temperature dependency may also vary across vendors, imaging parameters, and reconstruction algorithms.

Fourth, in the vendor protocols, we did not make any changes to parameters such as flip angle, repetition time, echo time, echo spacing, and bandwidth. Although default values for each vendor's software were used, additional adjustments may improve acquisition time and signal-to-noise ratio, minimize T1 bias, and increase accuracy of PDFF values. Related to this, we did not explicitly measure the T1 and T2 properties of the phantom. Finally, we limited the scope of this work to complex and hybrid magnitude-complex based CSE protocols. The study can be repeated with magnitude-based CSE methods that use in-phase and opposed-phase data.

In conclusion, through this multicenter, multiplatform, multivendor round-robin phantom study across nine centers, the Quantitative Imaging Biomarkers Alliance Proton Density Fat Fraction (PDFF) Committee successfully demonstrated linearity of PDFF estimations from complex and hybrid magnitude-complex based chemical shift–encoded (CSE) MRI in a commercial phantom, and further determined the range of PDFF bias across existing commercially available CSE MRI methods at both 1.5 T and 3.0 T.

Author contributions: Guarantors of integrity of entire study, H.H.H., 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, H.H.H., T.Y., M.R.B., D.H., S.D.S., M.S.M., S.B.R.; clinical studies, H.H.H., S.D.S., J.S., J.A.T.; experimental studies, H.H.H., T.Y., M.R.B., D.H., D.M., T.L.C., S.D.S., W.C.H., G.H., J.S., Y.S., J.A.T., A.T.T., S.B.R.; statistical analysis, H.H.H., T.Y., W.C.H., N.O., S.B.R.; and manuscript editing, H.H.H., T.Y., M.R.B., C.B.S., D.H., D.M., S.D.S., M.S.M., Y.S., J.A.T., A.T.T., N.O., J.H.B., S.B.R.

References

1. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. JAMA 2015;313(22):2263–2273.
2. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology 2018;67(1):123–133.
3. Kumar R, Priyadarshi RN, Anand U. Non-alcoholic fatty liver disease: growing burden, adverse outcomes and associations. J Clin Transl Hepatol 2020;8(1):76–86.
4. Ahmed A, Wong RJ, Harrison SA. Nonalcoholic Fatty Liver Disease Review: Diagnosis, Treatment, and Outcomes. Clin Gastroenterol Hepatol 2015;13(12):2062–2070.
5. Schindhelm RK, Diamant M, Heine RJ. Nonalcoholic fatty liver disease and cardiovascular disease risk. Curr Diab Rep 2007;7(3):181–187.
6. Ismaiel A, Dumitraşcu DL. Cardiovascular Risk in Fatty Liver Disease: The Liver-Heart Axis-Literature Review. Front Med (Lausanne) 2019;6:202.
7. Targher G, Bertolini L, Padovani R, et al. Prevalence of nonalcoholic fatty liver disease and its association with cardiovascular disease among type 2 diabetic patients. Diabetes Care 2007;30(5):1212–1218.
8. 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.
9. Zhong X, Nickel MD, Kannengiesser SA, Dale BM, Kiefer B, Bashir MR. Liver fat quantification using a multi-step adaptive fitting approach with multi-echo GRE imaging. Magn Reson Med 2014;72(5):1353–1365.
10. 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.
11. Yokoo T, Serai SD, Pirasteh A, et al. Linearity, bias, and precision of hepatic proton density fat fraction measurements by using MR imaging: a meta-analysis. Radiology 2018;286(2):486–498.
12. Gu J, Liu S, Du S, et al. Diagnostic value of MRI-PDFF for hepatic steatosis in patients with non-alcoholic fatty liver disease: a meta-analysis. Eur Radiol 2019;29(7):3564–3573.
13. Hu HH, Branca RT, Hernando D, et al. Magnetic resonance imaging of obesity and metabolic disorders: summary from the 2019 ISMRM Workshop. Magn Reson Med 2020;83(5):1565–1576.
14. 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):928–934https://doi.org/10.1002/jmri.22701.
15. 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.
16. 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.
17. 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 subjects. J Magn Reson Imaging 2015;42(3):811–817.
18. 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.
19. Hernando D, Sharma SD, Aliyari Ghasabeh M, et al. Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat-water phantom. Magn Reson Med 2017;77(4):1516–1524.
20. 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.
21. Jang JK, Lee SS, Kim B, et al. Agreement and reproducibility of proton density fat fraction measurements using commercial MR sequences across different platforms: a multivendor, multi-institutional phantom experiment. Invest Radiol 2019;54(8):517–523.
22. Kim HJ, Cho HJ, Kim B, et al. Accuracy and precision of proton density fat fraction measurement across field strengths and scan intervals: a phantom and human study. J Magn Reson Imaging 2019;50(1):305–314.
23. 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.
24. Hansen KH, Schroeder ME, Hamilton G, Sirlin CB, Bydder M. Robustness of fat quantification using chemical shift imaging. Magn Reson Imaging 2012;30(2):151–157.
25. Keenan KE, Biller JR, Delfino JG, et al. Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. J Magn Reson Imaging 2019;49(7):e26–e39.
26. Schlaffke L, Rehmann R, Rohm M, et al. Multi-center evaluation of stability and reproducibility of quantitative MRI measures in healthy calf muscles. NMR Biomed 2019;32(9):e4119.
27. Schmeel FC, Vomweg T, Träber F, et al. Proton density fat fraction MRI of vertebral bone marrow: accuracy, repeatability, and reproducibility among readers, field strengths, and imaging platforms. J Magn Reson Imaging 2019;50(6):1762–1772.
28. Bainbridge A, Bray TJP, Sengupta R, Hall-Craggs MA. Practical approaches to bone marrow fat fraction quantification across magnetic resonance imaging platforms. J Magn Reson Imaging 2020;52(1):298–306.
29. Yu H, McKenzie CA, Shimakawa A, et al. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. J Magn Reson Imaging 2007;26(4):1153–1161.
30. 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.
31. Bydder M, Yokoo T, Hamilton G, et al. Relaxation effects in the quantification of fat using gradient echo imaging. Magn Reson Imaging 2008;26(3):347–359.
32. Obuchowski NA, Bullen J. Quantitative imaging biomarkers: effect of sample size and bias on confidence interval coverage. Stat Methods Med Res 2018;27(10):3139–3150.
33. Sofue K, Zhong X, Nickel MD, Dale BM, Bashir MR. Stability of liver proton density fat fraction and changes in R 2* measurements induced by administering gadoxetic acid at 3T MRI. Abdom Radiol (NY) 2016;41(8):1555–1564.
34. Meisamy S, Hines CDG, 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.
35. Kuroda K, Oshio K, Chung AH, Hynynen K, Jolesz FA. Temperature mapping using the water proton chemical shift: a chemical shift selective phase mapping method. Magn Reson Med 1997;38(5):845–851.
36. Hernando D, Sharma SD, Kramer H, Reeder SB. On the confounding effect of temperature on chemical shift-encoded fat quantification. Magn Reson Med 2014;72(2):464–470.