FULL PAPER

Magnetic Resonance in Medicine 00:00–00 (2017)

Accuracy, Repeatability, and Interplatform Reproducibility of T_1 Quantification Methods Used for DCE-MRI: Results From a Multicenter Phantom Study

Octavia Bane,[1,2] Stefanie J. Hectors,[1,2] Mathilde Wagner,[1,2] Lori L. Arlinghaus,[3] Madhava P. Aryal,[4] Yue Cao,[4] Thomas L. Chenevert,[5] Fiona Fennessy,[6] Wei Huang,[7] Nola M. Hylton,[8] Jayashree Kalpathy-Cramer,[9] Kathryn E. Keenan,[10] Dariya I. Malyarenko,[5] Robert V. Mulkern,[6] David C. Newitt,[8] Stephen E. Russek,[10] Karl F. Stupic,[10] Alina Tudorica,[11] Lisa J. Wilmes,[8] Thomas E. Yankeelov,[12] Yi-Fei Yen,[9] Michael A. Boss ,[10] and Bachir Taouli[1,2]*

Purpose: To determine the in vitro accuracy, test-retest repeatability, and interplatform reproducibility of T_1 quantification protocols used for dynamic contrast-enhanced MRI at 1.5 and 3 T.

Methods: A T_1 phantom with 14 samples was imaged at eight centers with a common inversion-recovery spin-echo (IR-SE) protocol and a variable flip angle (VFA) protocol using seven flip angles, as well as site-specific protocols (VFA with different flip angles, variable repetition time, proton density, and Look-Locker inversion recovery). Factors influencing the accuracy (deviation from reference NMR T_1 measurements) and repeatability were assessed using general linear mixed models. Interplatform reproducibility was assessed using coefficients of variation.

Results: For the common IR-SE protocol, accuracy (median error across platforms = 1.4–5.5%) was influenced predominantly by T_1 sample ($P < 10^{-6}$), whereas test-retest repeatability (median error = 0.2–8.3%) was influenced by the scanner ($P < 10^{-6}$). For the common VFA protocol, accuracy (median error = 5.7–32.2%) was influenced by field strength ($P = 0.006$), whereas repeatability (median error = 0.7–25.8%) was influenced by the scanner ($P < 0.0001$). Interplatform reproducibility with the common VFA was lower at 3 T than 1.5 T ($P = 0.004$), and lower than that of the common IR-SE protocol (coefficient of variation 1.5T: VFA/IR-SE = 11.13%/8.21%, $P = 0.028$; 3T: VFA/IR-SE = 22.87%/5.46%, $P = 0.001$). Among the site-specific protocols, Look-Locker inversion recovery and VFA (2–3 flip angles) protocols showed the best accuracy and repeatability (errors < 15%).

Conclusions: The VFA protocols with 2 to 3 flip angles optimized for different applications achieved acceptable balance of extensive spatial coverage, accuracy, and repeatability in T_1 quantification (errors < 15%). Further optimization in terms of flip-angle choice for each tissue application, and the use of B_1 correction, are needed to improve the robustness of VFA protocols for T_1 mapping. **Magn Reson Med 000:000–000, 2017.**

Key words: T_1 mapping; DCE-MRI; phantom; multicenter

[1]Translational and Molecular Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

[2]Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

[3]Vanderbilt University Medical Center, Nashville, Tennessee, USA.

[4]Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, USA.

[5]Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA.

[6]Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts, USA.

[7]Advanced Imaging Research Center, Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon, USA.

[8]Department of Radiology, University of California–San Francisco, San Francisco, California, USA.

[9]Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts, USA.

[10]Physical Measurement Laboratory, National Institute of Standards and Technology, Boulder, Colorado, USA.

[11]Department of Radiology, Oregon Health and Science University, Portland, Oregon, USA.

[12]Department of Radiology, University of Texas at Austin, Austin, Texas, USA.

*Correspondence to: Bachir Taouli, M.D., Department of Radiology and Translational and Molecular Imaging Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1234, New York, NY 10029-6574. E-mail: bachir.taouli@mountsinai.org.

This study was sponsored by the National Cancer Institute Quantitative Imaging Network (NCI QIN). The sites participating in this QIN Working Group project receive funding from the following National Institutes of Health National Cancer Institute (NIH NCI) grants: U01 CA172320, U01 CA142565, U01 CA183848, U01 CA166104, U01 CA151261, U01 CA154602, U01 CA151235, and U01 CA154601. The work of postdoctoral trainees was partly supported by the NCI training grant 5T32CA078207-15 (O.B.), and by Fondation ARC pour la Recherche sur le Cancer (France) SAE20140601302 (M.W.). The contribution of NIST is not subject to copyright in the United States. Certain commercial equipment, instruments, and software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Received 21 May 2017; revised 14 August 2017; accepted 16 August 2017

DOI 10.1002/mrm.26903

Published online 00 Month 2017 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Dynamic contrast-enhanced MRI (DCE-MRI) is often used to quantify flow and permeability in various tumors (1–3) and organs (brain, breast, liver, kidney, prostate) (4–10). Dynamic contrast-enhanced MRI captures the signal change in time with the intravenous injection of a gadolinium-based contrast agent by acquiring T_1-weighted images before, during, and after injection of contrast agent at a high temporal resolution. The signal-time curves are usually converted to gadolinium

concentration-time curves (5), reflecting the contrast agent uptake and washout of the tissue of interest. Pharmacokinetic modeling (11–13) can then be applied to the concentration-time curves to determine tissue biological parameters such as intravascular-extravascular transfer rate constant for the contrast agent, K^{trans}, and extravascular and extracellular volume fraction, v_e. The precision of pharmacokinetic parameters is highly dependent on the conversion of T_1-weighted signal to gadolinium concentration, and thus on the precontrast (native) T_1 value of the tissue of interest (8,14).

A simple approach is to assign a published T_1 value to the tissue of interest (15–17). However, because literature-derived T_1 values are often based on studies in healthy volunteers (18), this approach does not account for changes in tissue T_1 that occur with age (8) or disease (19,20). Further limitations of using a constant, literature-based baseline T_1 are that it does not account for interpatient variability and cellular heterogeneity in the tissue of interest (11,19,20). The need for patient and tissue-specific DCE-MRI parameters motivates the effort to develop accurate, widely available precontrast T_1 measurements.

The longitudinal relaxation time T_1 can be measured by a variety of methods (21,22). The multiple-delay inversion-recovery (IR) method that originated from historical NMR experiments (23,24) is considered the reference standard, but has limited spatial coverage and long acquisition times, which makes it impractical in clinical settings. The Look-Locker modified IR method (25–27) decreases the acquisition time by sampling the signal recovery curve multiple times per repetition time (TR) after application of several low flip-angle pulses during the acquisition. Despite shortened acquisition time, Look-Locker IR methods are still limited in spatial coverage. Another method that shares some of the same limitations as the IR method but with shorter overall scan times is the variable TR (VTR) method, in which T_1-weighted signal is acquired with multiple TR values (8). A frequently used alternative approach is to vary the radiofrequency tip angle while keeping the TR constant in a 3D spoiled gradient-echo acquisition, as in variable flip angle (VFA) methods (8,22,28,29). Variable flip angle measurements allow for voxel-based baseline (pre-DCE-MRI) T_1 mapping with the same spatial resolution and coverage as the DCE-MRI scan in a short amount of time. Similarly time-efficient, although less popular, is the proton density (PD) approach (30), in which T_1 is derived by comparing PD-weighted images with DCE baseline (precontrast) images.

Previous publications have reported interscanner and intersite variability in the T_1 values measured with different methods (31,32), including the IR method (22). In the brain, the Look-Locker modified IR method was found to consistently underestimate, and the VFA method to consistently overestimate, T_1 values in the white matter (22). The accuracy and test-retest repeatability of measured T_1 values are influenced by the same factors that affect the generation and acquisition of MR signal, such as temperature in the magnet bore (32), incomplete spoiling of transverse magnetization (33) and B_1 field inhomogeneity (8,22), the choice of the optimal sequence, and sequence parameters for the range of T_1 values to be measured. Additional factors contributing to variability include postprocessing, such as signal scaling by the image processing software (34) and the assumptions for the fitting function (22).

This study seeks to identify and systematically investigate some of the sources of variability in the T_1 measurement process using a dedicated phantom in a multicenter setting. The described research was conducted by the Data Acquisition Working Group of the National Cancer Institute–sponsored Quantitative Imaging Network (35,36), whose purpose is to improve the repeatability and reproducibility of quantitative imaging, and promote its adoption as an evaluation tool in oncology clinical trials.

The purpose of our study was to determine the (i) accuracy (with respect to reference standard NMR T_1 values) of T_1 measurements obtained with two predefined T_1 measurement protocols common among participating sites, and with the specific T_1 measurement protocol(s) used at each participating site for DCE-MRI studies; (ii) test-retest repeatability of T_1 measurements obtained with the common and site-specific T_1 measurement protocol(s); and (iii) Interplatform reproducibility of the common T_1 mapping protocols at 1.5 and 3 T.

METHODS

A preliminary survey identified eight Quantitative Imaging Network sites that measured the baseline-tissue T_1 for DCE-MRI quantification, rather than using a fixed T_1 based on the literature. The DCE-MRI with T_1 measurement is performed at 1.5 T (two sites, two scanners) and 3 T (seven sites, eight scanners), for different organs or neoplasms (Table 1). The methods and protocols used for T_1 measurement varied among sites, with VFA methods being the most common (Table 1) because of the achievable extensive spatial coverage in a short amount of time.

National Institute of Standards and Technology T_1 Phantom and NMR Reference T_1 Measurement

To assess accuracy, repeatability and reproducibility of T_1 measurements among platforms, sequences, and protocols, we used a phantom containing the T_1 elements (Fig. 1) from the National Institute of Standards and Technology (NIST) system phantom (31,32,37). The phantom consists of a 192-mm outer-diameter polycarbonate sphere filled with deionized water, containing a plastic plate with positioning markers and T_1 solution array. The T_1 array consists of 14 polycarbonate spheres (15-mm inner diameter, 20-mm outer diameter each), with 10 spherical samples equally spaced on a 50-mm diameter circle, and four inside the circle on a 40-mm grid. The anatomical directions on the central plate facilitate positioning of the phantom at scanner isocenter, with the central plate parallel to the coronal plane of the scanner. T_1 values in the range of 20 to 2000 ms were obtained by doping deionized water with $NiCl_2$. The $NiCl_2$ concentrations of the solutions used to create the samples were determined by traceable inductively coupled plasma mass spectrometry measurements. The

Table 1
Overview of Sites, Organs of Interest, Scanner Systems, and Site-Specific T_1 Mapping Protocols

Site	Organ	Scanner No.	Scanner	Site-Specific Protocols
1	Prostate	1	GE Discovery w750 (3 T)	VTR
2	Brain	2	Siemens Skyra (3 T)	VFA
		3	Tim Trio (3 T)	
3	Liver, prostate	4	Siemens Skyra (3 T)	VFA, Look-Locker
		9	Siemens Aera (1.5 T)	
4	Soft tissue sarcoma	5	Siemens Tim Trio (3 T)	PD
5	Breast	10	GE HDx (1.5 T)	VFA
6	Brain	6	Philips Ingenia (3 T)	VTR
7	Brain	7	Siemens Skyra (3 T)	VFA
8	Breast	8	Philips Achieva (3 T)	VFA

Note: Sites are listed in random order; scanners are ordered by decreasing field strength (3 T: scanners 1–8; 1.5 T: scanners 9 and 10) and by site number.

solutions are well-characterized and monitored by NIST for stability and accuracy (37).

The NMR spectroscopy reference measurements were performed at 1.5 and 3 T by NIST investigators. Aliquots of each of the 14 $NiCl_2$ solutions were sealed into 2-mm outer-diameter quartz NMR tubes, which were then flame-sealed using a methane/oxygen torch. A fiber-optic temperature probe was positioned with the sensor in the middle of the radiofrequency coil. Each sample was equilibrated to 293.15 K (20°C) for a minimum of 15 min. Samples were shimmed using the Berger-Braun shimming method before collecting NMR-IR relaxation-time data (24).

Image Acquisition

The Working Group developed a uniform T_1 phantom imaging protocol in collaboration with NIST. The T_1 phantom, along with a digital thermometer and scanning instructions, was shipped to the eight Quantitative Imaging Network centers, where it was scanned in duplicate (test-retest) sessions. The phantom was placed in the scanner room 8 h before scanning, to allow its temperature to stabilize. Temperature was measured in the bulk water of the phantom at the start and end of each experiment, with the provided digital thermometer. Sites were instructed to image the phantom using the same receive coils (head or body array) used for their DCE-MRI studies. Sites (5 and 8) performing breast DCE-MRI studies used head coils, as the phantom was too large to fit into breast coils. The phantom was positioned with the aid of a level to ensure that the T_1 array plate was aligned with the principal scanner directions (superior–inferior, anterior–posterior), and that the central NIST marker was located at scanner isocenter.

The phantom was scanned on 10 different platforms (Table 1) at two field strengths (two at 1.5 T and eight at 3 T) from three major vendors (Siemens Healthineers (Erlangen, Germany), GE Healthcare (Waukesha, WI), and Philips Healthcare (Best, the Netherlands)). All sites collected data with a common inversion-recovery spin-echo (IR-SE) protocol and a common VFA protocol with seven flip angles. Both protocols were optimized to measure the full T_1 range of the NIST phantom (20–2000 ms). A VFA sequence was chosen to build a common scanning protocol, as the VFA method was used by most sites (Table 1) for T_1 quantification in DCE-MRI. The scanning parameters of the common protocols, matched closely among the platforms, are summarized in Table 2.

Additionally, seven of the eight sites collected data using the site-specific T_1 measurement protocols. The organs of interest for DCE-MRI studies, the T_1 mapping method, and scanner(s) on which the data were acquired at each site are summarized in Table 1. The 14 site-

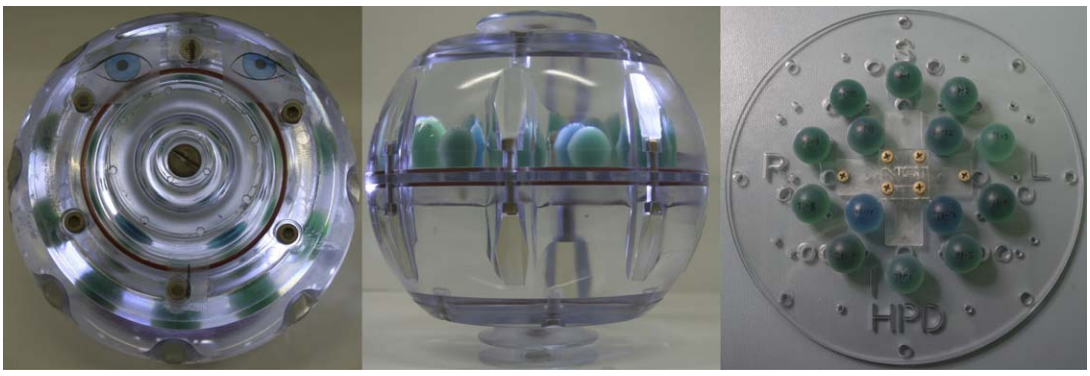

FIG. 1. (left) View of the NIST T_1 phantom from the top. Phantom positioning mimics the positioning of a patient lying head-first, supine on the MRI table (eye decals facing up, toward the scanner bore), so that the central plate is parallel to the coronal imaging plane. (center) View of the T_1 phantom from the side. (right) Central plate of the NIST T_1 phantom demonstrating positioning of the 14 samples (S1–S14) of $NiCl_2$ solution, with T_1 ranging from 22 to 2033 ms in descending order, determined by NMR spectroscopy at 1.5 and 3 T (Supporting Table S3).

Table 2
Standardized Acquisition Parameters of IR-SE and VFA Methods, Used to Image the NIST T_1 Phantom at All Sites

	IR-SE	VFA
Orientation	Coronal	Coronal
Flip angle (°)	180	2, 5, 10, 15, 20, 25, 30
Echo time (ms)	9	2
Repetition time (ms)	5000	12
Inversion time (ms)	24, 50, 75, 100, 125, 150, 250, 500, 750, 1000, 2000, 3000	—
Field of view (mm^2)	200 × 200	200 × 200
Number of slices	1	16
Slice thickness (mm)	5–6	5–6
Matrix	256 × 256	256 × 256
Echo train length	5–6	—
Number of averages	1	3
Acquisition time (min)	45	13

specific T_1 measurement protocols, their associated acquisition parameters, as well as the range of T_1 values they were optimized to measure, are summarized in order of scanner number in Supporting Tables S1 and S2. Briefly, site-specific protocols included four methods: (i) VFA (nine protocols: protocols 2 and 3 for brain; 4b and 9b for liver; 4c, 9c, and 9d for prostate; and 8 and 10 for breast), (ii) VTR (two protocols: protocol 1 for prostate; 6 for brain); (iii) Look-Locker modified IR (two protocols: 4a and 9a for liver); and (iv) PD (one protocol: protocol 5 for soft-tissue sarcoma). The site-specific VFA protocols had 2 to 10 flip angles, and choice of flip angles was optimized for the anatomical application.

Image Analysis

Centralized data analysis was performed at one site (Icahn School of Medicine at Mount Sinai, New York, NY). The acquired T_1 data were uploaded by participating sites in DICOM format. Data were analyzed by a single observer (O.B., a physicist with 4 years of postdoctoral experience in image analysis), who placed circular regions of interest (average size 1.0 cm^2) in a central slice of each sphere using OsiriX Lite (version 8.0.2, Pixmeo SARL, Bernex, Switzerland). For Philips scanners, the signal intensity (SI) values (arbitrary units) were modified by scaling factors and stored in "private" DICOM fields by the vendor (38), converted to floating point values (34), and then used for calculations. The mean region-of-interest SI was fitted according to the signal equation for each sequence (Supporting Information Eqs. [1]–[5]) to obtain T_1 values, using custom routines written in MATLAB R2015 (The MathWorks, Natick, MA).

Statistical Analysis

Statistical analysis was performed using MATLAB R2015 and SAS 9.4 (SAS Institute, Cary, NC). Statistical significance was defined as $P < 0.05$. Agreement of IR-SE T_1 measurements with reference NMR T_1 measurements was tested by Lin's concordance correlation coefficient (39,40) and Bland-Altman statistics.

Accuracy was assessed with respect to NMR T_1 values as the reference. The accuracy error was computed as the percentage difference between T_1 measured with the common or site-specific protocols during the first (test) scanning session, and reference NMR T_1 (Eq. [1]). Smaller values represent higher accuracy.

$$\text{Accuracy Error } (\%) = 100 \cdot |T_{1\ \text{protocol}} - T_{1\ \text{NMR}}|/T_{1\ \text{NMR}} \quad [1]$$

Test-retest repeatability was assessed by the precision error, calculated as the percentage difference of T_1 values measured in duplicate relative to the mean of the two measured values (Eq. [2]). Smaller values represent higher repeatability.

$$\text{Precision Error } (\%) = 100 \cdot |T_{1\ \text{test}} - T_{1\ \text{retest}}| / \text{Mean } (T_{1\ \text{test}},\ T_{1\ \text{retest}}) \quad [2]$$

A general linear mixed model was used to compare the accuracy and precision errors of T_1 measurements across samples, field strengths, scanners, vendors, methods, and protocols (26). The effects of sample (solution), field, scanner (individual platform), and vendor on the accuracy and precision errors of the common IR-SE and VFA protocol were tested separately and in combination. For site-specific protocols data, the acquisition protocol and measurement method were tested as additional predictive variables. The "protocol" variable represented each site-specific combination of sequence implementation and acquisition parameters (14 site-specific protocols listed in Supporting Tables S1 and S2), whereas the "method" variable encoded each of the four T_1 measurement methods (VFA, Look-Locker modified IR, VTR, and PD) used to generate data submitted by the sites.

A stepwise model selection procedure was performed (41,42) to identify the best subset of one or more independent, uncorrelated predictors of accuracy or precision. Model results were reported as least-square means ± standard error, with smaller numbers representing better accuracy or precision. Comparisons of accuracy and precision errors were performed among the following factors: samples, field strengths, scanners, vendors, protocols (in the case of site-specific protocols for different anatomical applications), and methods (for site-specific protocol data). Type 3 P values were reported for the general linear mixed model (43), and Tukey adjusted P values (44) were reported for comparisons between factors.

Interplatform reproducibility for the T_1 measured in each sample with the common IR-SE and VFA protocols was assessed by intraclass correlation coefficient (ICC) and coefficient of variation (% CV = 100 x standard deviation/mean calculated between platforms, for each reference T_1 value). Paired sample Wilcoxon signed-rank tests were used to compare CVs between field strengths and common protocols.

RESULTS

T_1 measurements were obtained on 10 scanners at two field strengths (1.5 T: two scanners; 3 T: eight scanners),

in duplicate sessions, with two common protocols and 14 site-specific protocols. All sites were able to implement the common protocols, with two minor exceptions (lowest TI = 50 ms for the common IR-SE on the GE scanners 1 and 10, instead of 24 ms, and FA = 19° instead of 20° for the common VFA protocol at site 4, on scanner 5, a Siemens Trio machine). Fourteen phantom samples were analyzed for the scans with the common protocols, totaling 560 (14 × 10 × 2 × 2) T_1 measurements. Only the samples with T_1 values within the measurement range for which the site-specific protocols were optimized (8,19,20,28,30,45). T_1 ranges for each anatomical application, given in Supporting Tables S1 and S2, were analyzed, totaling 112 (56 × 2) T_1 measurements with the site-specific protocols. Goodness of fit was assessed by the coefficient of determination (R^2), which ranged between 0.85 and 0.99 for T_1 measurement with the common IR-SE protocol, between 0.88 and 0.99 for measurements with the common VFA protocol, and between 0.82 and 0.99 for measurements with site-specific protocols. Typical fit plots for the common protocols are shown in Supporting Figures S2 and S3.

Temperature measurements in the bulk water of the phantom ranged between 18.5 and 22.6°C at the start (mean 20.8 ± 1.15°C), and between 19.6 and 22.7°C at the end (mean 21.3 ± 1.1°C) of experiments, with a temperature change observed between the start and end of experiments of 0.4 ± 0.4°C.

Comparison of IR-SE and NMR Reference Measurements

The NMR reference measurements are listed in Supporting Table S3 for each solution sample. There was excellent concordance between IR-SE and NMR measurements (concordance correlation coefficient > 0.99; $P < 10^{-6}$), with some deviations from unity line observed at $T_1 = 500$ ms and $T_1 > 1500$ ms (Fig. 2a). Bland-Altman plots for all scanners (Fig. 2b) show near-zero bias, limits of agreement (−30%, 30%), and highlight discrepancies between IR-SE and NMR T_1 at reference $T_1 = 50$ ms (scanners 2–9) and $T_1 = 500$ ms (scanners 1 and 10), with these values outside the limits of agreement. The deviation at 500 ms is caused by underperformance of fit (Supporting Information Eq. [1]) in some data sets (scanners 1, 6, 8, and 10) (Supporting Fig. S1). When used without lower and upper bounds for the parameters, the fit performs better in some cases (Supporting Fig. S1), but in other cases returns nonphysical fitted parameters (e.g., negative noise, 1-cos (Inversion Angle) > 2). Because of these deviations, only the NMR measurements were used as the reference standard for the calculation of accuracy.

Accuracy Assessment

Common IR-SE Protocol

The results of the accuracy assessment of the common IR-SE protocol are displayed in Table 3 and Figure 3. For the full range of reference T_1 values of the phantom solutions (20–2000 ms), the range of median accuracy error among the 10 platforms (Table 3) was 1.4 to 5.5%. The distributions of accuracy errors for ranges of low (40–100 ms), intermediate (100–500 ms), and high (500–2000 ms) T_1 values are summarized in Supporting Table S4. Among the factors tested separately in a general linear mixed model, the sample (solution), or the actual T_1 value, was identified as a significant ($P < 10^{-6}$) independent predictor of accuracy (Fig. 3) for T_1 measurements with the common IR-SE protocol. T_1 measurements were significantly less accurate for solution sample S12 (reference T_1 ~45 ms) than for all other samples (Fig. 3; adjusted $P < 10^{-6}$).

Common VFA Protocol

For the full range of reference T_1 values of the phantom solutions (20–2000 ms), the range of median accuracy error among the 10 platforms (Table 3) was 5.7 to 32.2%. The distributions of accuracy errors for ranges of low (40–100 ms), intermediate (100–500 ms), and high (500–2000 ms) T_1 values are summarized in Supporting Table S4. Field strength was identified as a significant ($P = 0.006$) independent predictor (Fig. 3), and scanner (Fig 3; $P = 0.0003$) as a significant predictor of accuracy of T_1 measurements with the common VFA protocol.

T_1 measurements with the common VFA protocol were overall less accurate (adjusted $P = 0.006$) at 3 T than at 1.5 T (Fig. 3). However, of the eight 3T scanners, seven scanners did not reach a statistically different accuracy error than the two 1.5T scanners (Fig. 3). Significant differences in accuracy of the common VFA protocol among scanners are summarized in the Supporting Information.

Site-Specific Protocols

The results of the accuracy assessment of the site-specific protocols are shown in Figure 4. Accuracy errors ranged between 3.2 and 37.2% for the site-specific protocols. The VTR protocol 1 for the prostate, Look-Locker protocol 4a for the liver, VFA protocol 4b for the liver, VFA protocol 9b for the liver, and VFA 9c and 9d for the prostate had accuracy errors of less than 15%. Protocol ($P = 0.002$) and scanner ($P = 0.0014$) were identified as significant predictors of accuracy of T_1 measurements with the site-specific protocols. The stepwise model-selection procedure identified no independent predictors of accuracy.

The VTR protocol 1 for the prostate was significantly more accurate (adjusted $P = 0.03$) than protocol 2 (VFA for brain at 3 T with six flip angles; Supporting Table S2). Among VFA protocols (2, 3, 4b, 4c, 8, 9b, 9c, 9d; Supporting Tables S1 and S2), protocols 9c and 9d for the prostate (VFA with two flip angles at 1.5 T; Supporting Table S2) were significantly (adjusted $P = 0.03$) more accurate than protocol 2.

Test-Retest Repeatability Assessment

Common IR-SE Protocol

The median precision error range across scanners was 0.2 to 8.3% for the full range of reference T_1 values in the phantom (Table 3). The distributions of precision errors for ranges of low, intermediate, and high reference T_1 values are summarized in Supporting Table S5. Among the factors tested, scanner was identified as a significant ($P < 10^{-6}$) independent predictor of test-retest repeatability for T_1 measurements with the IR-SE

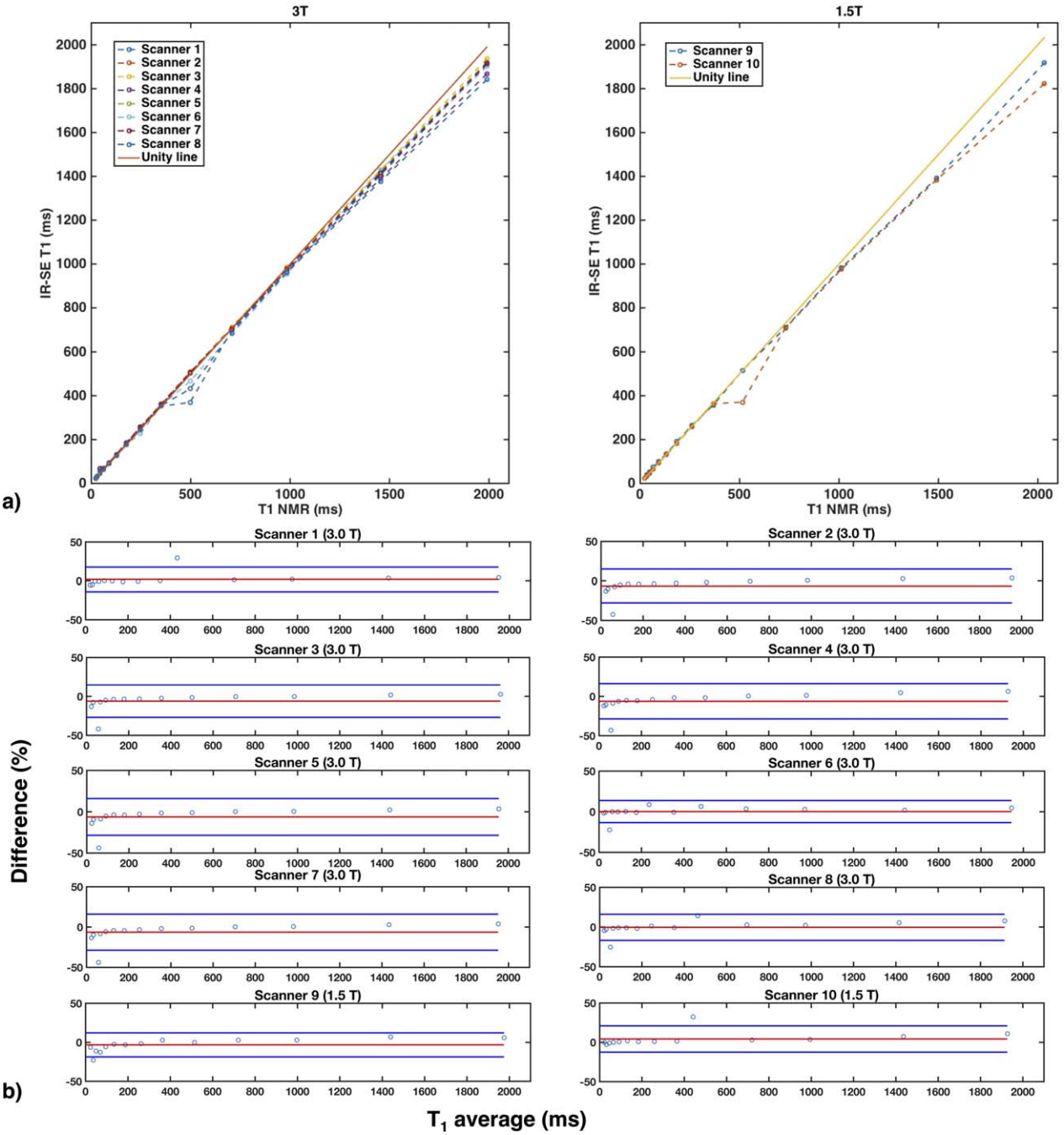

FIG. 2. **a**: Correlation plots between IR-SE and reference NMR T_1 values at 3 T (left) and 1.5 T (right). Values were highly correlated (Lin's concordance correlation coefficient > 0.9). However, systematic deviation of IR-SE T_1 values from reference NMR values was observed at long T_1. **b**: Bland-Altman plots showing differences (%) between IR-SE and reference NMR T_1 values for all scanners. There is high agreement between values, with bias near zero and limits of agreement (−30%, 30%), with measurements at reference $T_1 = 50$ ms and $T_1 = 500$ ms outside the limits of agreement.

common protocol (Fig. 5). The IR-SE T_1 measurements were significantly less repeatable on scanner 7 than on all of the other scanners (Fig. 5; adjusted $P < 10^{-6}$).

Common VFA Protocol

The median precision error range across scanners was 0.7 to 25.8% for the full range of reference T_1 values in the phantom solutions (Table 3). The distributions of precision errors for ranges of low, intermediate, and high reference T_1 values are summarized in Supporting Table S5. Field ($P < 0.0001$), scanner ($P < 10^{-6}$), and vendor ($P = 0.0012$) were identified as significant predictors of test-retest repeatability of T_1 measurements with the common VFA protocol, when tested separately in a general linear mixed model. The model-selection procedure identified scanner as the independent predictor of test-retest repeatability (Fig. 5). Among the variable subsets

Table 3
Accuracy and Precision Errors for Each Scanner, With the Common IR-SE and VFA Protocols, for Reference T_1 Values (22.9–2033 ms at 1.5 T; 21.7–1989 ms at 3 T) in the Phantom Solution Samples

		Accuracy Error (%)		Precision Error (%)	
Scanner No.	Field (T)	IR-SE	VFA	IR-SE	VFA
1	3	1.4 (3.1)	10.4 (34.0)	0.3 (0.7)	5.1 (5.7)
2	3	4.2 (5.1)	21.3 (21.5)	0.2 (0.5)	0.7 (1.4)
3	3	3.3 (5.7)	24.1 (11.7)	0.2 (0.2)	15.3 (3.5)
4	3	5.5 (7.2)	8.1 (12.4)	0.4 (0.2)	25.8 (2.3)
5	3	3.6 (7.6)	10.4 (12.7)	0.4 (0.2)	19.0 (1.2)
6	3	1.7 (3.7)	32.2 (13.6)	1.4 (1.3)	22.9 (5.8)
7	3	4.1 (6.8)	5.7 (7.2)	8.3 (12.3)	3.4 (1.5)
8	3	2.4 (4.1)	6.9 (10.0)	0.4 (0.4)	1.4 (1.3)
9	1.5	4.5 (4.0)	8.0 (5.1)	0.2 (0.2)	2.1 (1.0)
10	1.5	1.6 (2.8)	6.1 (7.8)	0.7 (0.9)	1.9 (0.9)

Note: Values are given as median (interquartile range).

not containing scanner, field was the independent predictor of test-retest repeatability. In subsets excluding scanner or field, vendor was the independent predictor of test-retest repeatability. Overall, T_1 measurements were less repeatable at 3 T than at 1.5 T (adjusted $P < 0.0001$) (Fig. 5). Similar to the accuracy results, five of the eight 3T scanners had substantially worse performance for repeatability of measurements than the two 1.5T scanners (Fig. 5; see also Supporting Information).

Site-Specific Protocols

The results of repeatability assessment of the site-specific protocols are shown in Figure 4. Precision errors ranged between 0.25 and 40% for the site-specific protocols. Look-Locker protocols 4a and 9a for the liver, VFA protocols 2 and 3 for the brain, VFA protocol 8 for the breast, VFA protocol 9b for the liver, and VTR protocol 6 for the brain had precision errors of less than 15%. Complete statistical comparison results for the repeatability of site-specific protocols are provided in the Supporting Information.

Protocol ($P < 0.0001$), scanner ($P = 0.0001$), vendor ($P = 0.0001$), and method ($P = 0.035$) were identified as significant predictors of test-retest repeatability of T_1 measurements with the site-specific protocols, when tested separately in a general linear-mixed model. The model-selection procedure identified protocol as an independent predictor of repeatability.

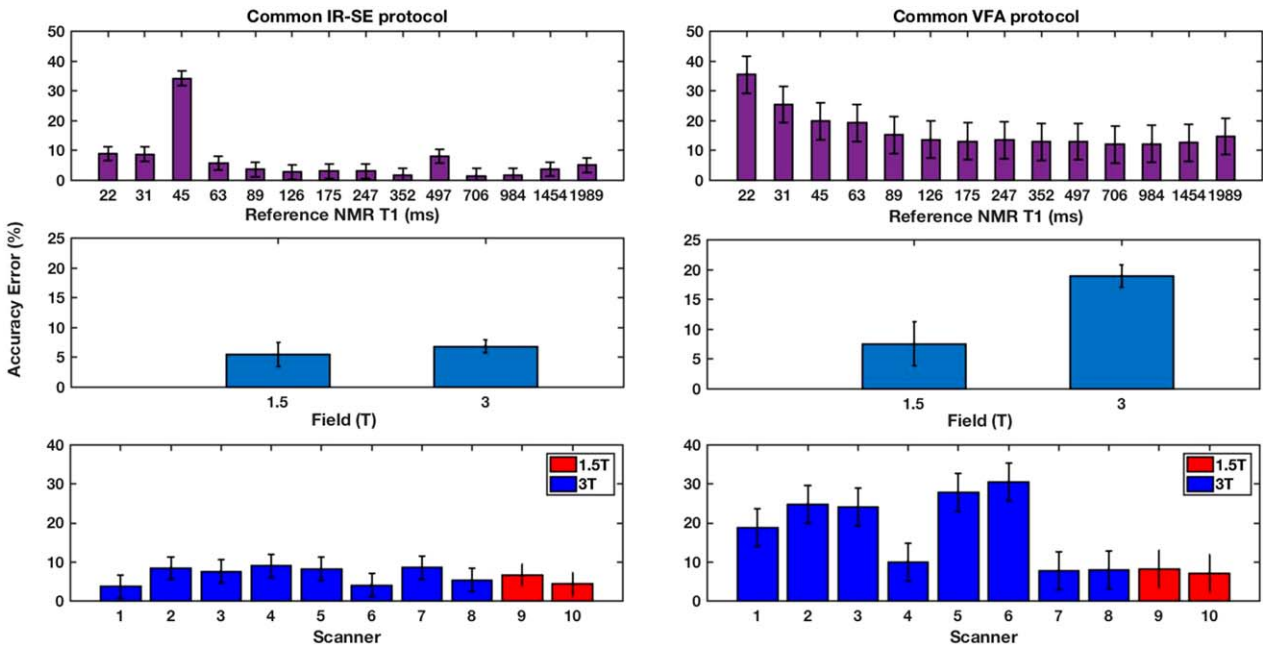

FIG. 3. Accuracy errors of the IR-SE and VFA common protocols for each solution sample with reference T_1 value (top), field strength (middle), and scanner (bottom), using NMR measurement as the reference. The accuracy errors were calculated for the T_1 measurements in the test scanning session, using NMR measurements as the reference. Bar graphs represent general linear model least square means ± standard error. Smaller numbers represent better accuracy. Sample was the significant ($P < 10^{-6}$) independent predictor of accuracy of IR-SE T_1 measurements; field was the significant ($P = 0.006$) independent predictor; and scanner was a significant ($P = 0.0003$) predictor of accuracy of T_1 measurements with the VFA common protocol.

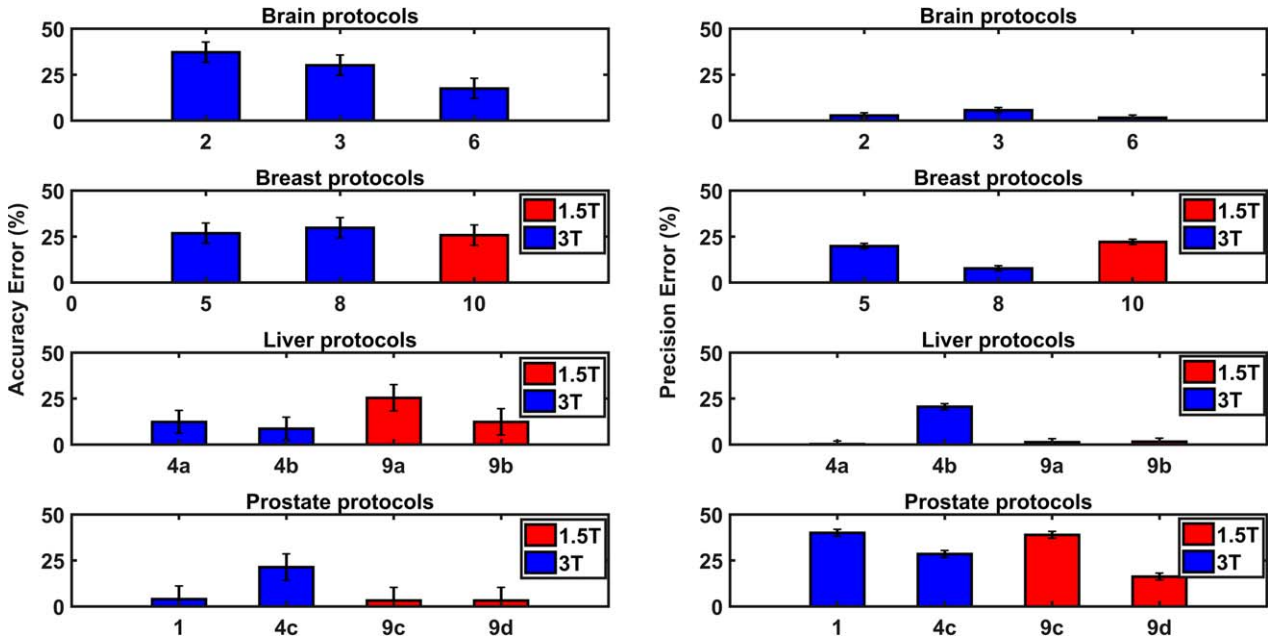

FIG. 4. Significant predictors of accuracy (left) and test-retest precision (right) errors of T_1 measurements with site-specific protocols. The accuracy errors were calculated for the T_1 measurements in the test scanning session, using NMR measurements as the reference. The precision error was calculated as the relative mean percentage difference between T_1 values measured in the test and re-test scanning sessions. The results of general linear mixed models are presented as least square means ± standard error. Smaller numbers represent better accuracy/ test-retest repeatability. Site-specific protocols, by scanner number, are as follows: 1, prostate VTR; 2, brain VFA I; 3, brain VFA II; 4a, liver Look-Locker 3 T; 4b, liver VFA 3 T; 4c, prostate VFA 3 T; 5, soft-tissue sarcoma PD; 6, brain VTR; 8, breast VFA 3 T; 9a, liver Look-Locker 1.5 T; 9b, liver VFA 1.5 T; 9c, prostate VFA I 1.5 T; 9d, prostate VFA II 1.5 T; 10, breast VFA 1.5 T. Site/scanner 7 did not provide site-specific data. (left) Protocol was a significant predictor of accuracy ($P = 0.002$). (right) Protocol was a significant predictor of repeatability ($P < 0.0001$).

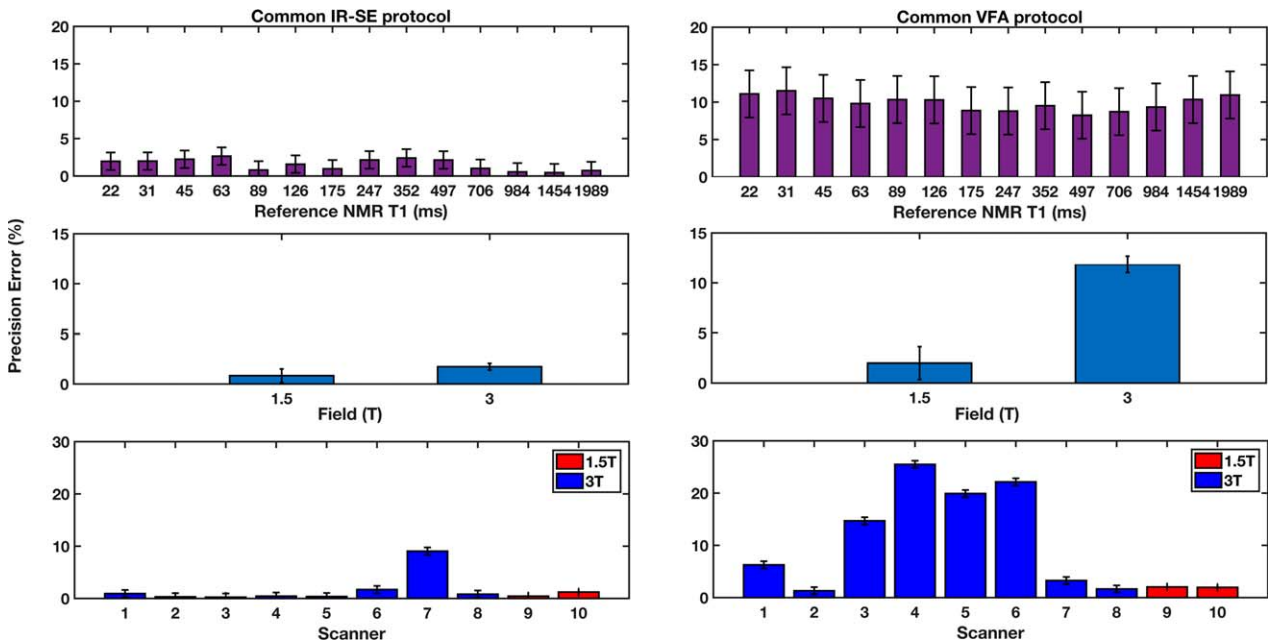

FIG. 5. Precision errors of the IR-SE and VFA common protocols for each solution sample with reference T_1 value (top), field strength (middle), and scanner (bottom). The precision error was calculated as the relative mean percentage difference between T_1 values measured in the test and re-test scanning sessions. Bar graphs represent general linear model least square means ± standard error. Smaller numbers represent better repeatability. Field was a significant ($P < 0.0001$) predictor of precision errors with the common VFA protocol. Scanner was the significant, independent predictor of precision error of T_1 measurements with the common IR-SE protocol ($P < 10^{-6}$) and the common VFA protocol ($P < 10^{-6}$).

Table 4
Interplatform Reproducibility of T_1 Measurements Expressed as CV (%) Using Common IR-SE and VFA Protocols for Reference T_1 Values (22.9–2033 ms at 1.5 T; 21.7–1989 ms at 3 T) in the Phantom Solution Samples

	RMS CV (%)	
Field strength	IR-SE	VFA
1.5 T	8.21[a]	11.13[b]
3 T	5.46[c]	22.87

Note: Results are summarized as the RMS of within-sample CVs of the IR-SE and VFA measurements at each field strength. *P* values are given for paired-sample Wilcoxon signed-rank tests used to compare the CV values. IR-SE 1.5 T versus 3 T: $P = 0.379$.
[a]1.5 T: IR-SE versus VFA, $P = 0.028$.
[b]VFA 1.5 T versus 3 T, $P = 0.004$.
[c]3 T: IR-SE versus VFA, $P = 0.001$.

Protocols with more prescribed flip angles (protocols 2 and 3 for the brain at 3 T with six and seven flip angles, and protocol 8 for the breast at 3 T with 10 flip angles) had significantly better test-retest repeatability than VFA protocols with two or three flip angles (protocol 4b for the liver, 9c and 9d for the prostate; adjusted $P < 10^{-4}$). The VFA protocols for the same anatomical application were more repeatable at 1.5 T than at 3 T (e.g., liver protocol 9b at 1.5 T was more repeatable than liver protocol 4b at 3 T (adjusted $P < 10^{-4}$), and prostate protocol 9c at 1.5 T was more repeatable than prostate protocol 4c at 3 T (adjusted $P = 0.0022$)), with one exception: Breast protocol 8 at 3 T was more repeatable than breast protocol 10 at 1.5 T (adjusted $P < 10^{-4}$).

Unlike the VFA approach, the Look-Locker IR method did not show significantly different repeatability between protocols at different field strengths (e.g., protocols 4a and 9a for the liver did not have significantly different repeatability). Look-Locker IR protocols 4a (at 3 T) and 9a (at 1.5 T) for the liver were significantly (adjusted $P < 10^{-4}$) more repeatable than the VFA protocol 4b for the liver at 3 T. The superior repeatability of Look-Locker IR protocols versus VFA is also apparent from the comparison of precision errors between the methods. Look-Locker IR had significantly higher test-retest repeatability (precision error: $0.68 \pm 4.64\%$) than VFA (precision error: $14.7 \pm 2.04\%$, $P = 0.038$) and PD (precision error: $19.8 \pm 5.49\%$, $P = 0.049$), and borderline significantly higher repeatability than VTR ($15.98 \pm 4.34\%$, $P = 0.088$). There were no significant differences in repeatability among the other methods (*P* range 0.8–0.99).

Interplatform Reproducibility

Excellent agreement was found in T_1 measurements between platforms, for both the IR-SE (1.5 T: ICC = 0.997, $P < 10^{-6}$; 3 T: ICC = 0.999, $P < 10^{-6}$) and VFA (1.5 T: ICC = 0.993, $P < 10^{-6}$; 3 T: ICC = 0.947, $P < 10^{-6}$) common protocols. Interplatform CVs of common IR-SE and VFA T_1 values for 1.5 T (two scanners) and 3 T (eight scanners) are displayed in Table 4 and Figure 6. Interplatform reproducibility was higher at 1.5 T (root mean square (RMS) CV 11.13%, range: 0.1–18.8%) than at 3 T (RMS CV 22.87%, range: 16.6–45.5%) for the common VFA protocol ($P = 0.004$; Table 4). For the common IR-SE protocol, there was no significant difference in reproducibility at different field strengths (1.5 T: RMS CV 8.21%, range: 0.11–23%; 3 T: RMS CV 5.46%, range 0.99–14.6%; $P = 0.379$) (Table 4 and Fig. 6). As expected, at both field strengths the overall reproducibility was higher for the common IR-SE protocol ($P = 0.028$) than for the common VFA protocol ($P = 0.001$). Interplatform CV was less than 10% for the IR-SE sequence for most phantom samples, with the exception of samples with reference NMR T_1s of 33, 45, and 500 ms (Fig. 6).

DISCUSSION AND CONCLUSIONS

Accurate, repeatable, and reproducible quantification of baseline-tissue T_1 is highly desired for patient-specific,

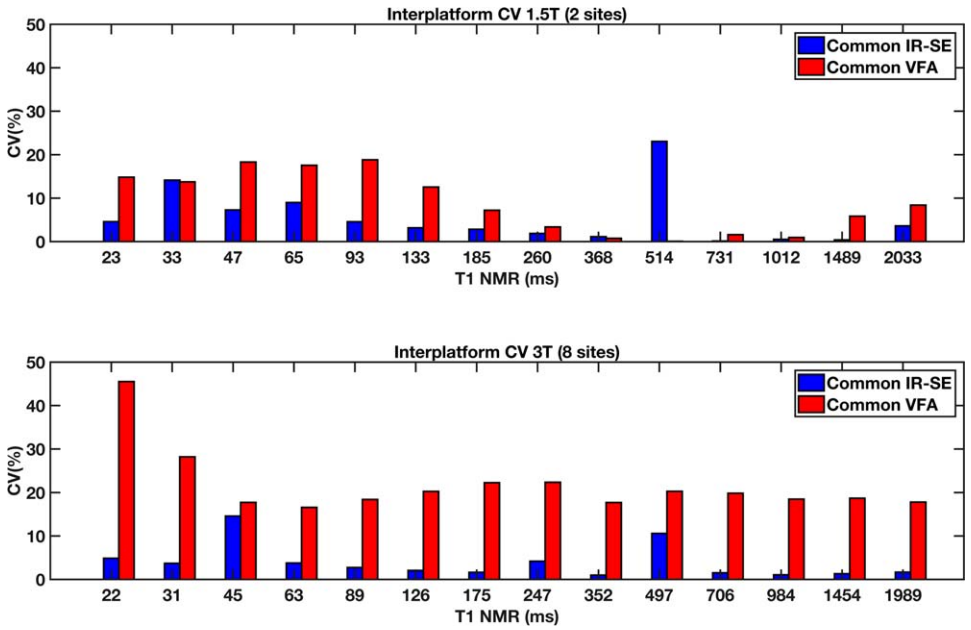

FIG. 6. Interplatform CVs at 1.5 T (top) and 3 T (bottom) for each sample with known NMR T_1. Interplatform CV was less than 10% for the IR-SE sequence for all phantom samples, with the exception of the ones with reference NMR T_1 values of 33 ms at 1.5 T, 45 ms at 3 T, and approximately 500 ms at both field strengths, as a result of fit underperformance at these values.

robust perfusion quantification from DCE-MRI studies. The present study expands the scope of previous multicenter studies of the variability of quantitative MR relaxometry metrics (31,32,46). In addition to investigating sources of error with the same T_1 mapping protocols across scanner platforms, our study also compared the performance of 14 site-specific T_1 mapping protocols used for different oncological applications. Our study performed a centralized analysis of site data, to control for sources of error such as the data-fitting method and software packages used for analysis. In particular, different results may have been obtained for the VFA protocols by fitting the VFA data with the linearized version of the signal equation, rather than the full signal equation. The knowledge gained on the magnitude of accuracy and repeatability errors, as well as of interplatform variability of T_1 measurements, obtained with different protocols potentially allows future studies on the propagation of T_1 measurement errors to pharmacokinetic parameters obtained from DCE-MRI studies. Our work showed the field strength and scanner dependence of accuracy and test-retest repeatability of T_1 measurements. We observed significantly lower interplatform reproducibility at 3 T compared with 1.5 T for the common VFA protocol, and significantly lower reproducibility of the common VFA protocol compared with the IR-SE protocol at both field strengths.

Our study also confirmed the expected high accuracy, repeatability, and interplatform reproducibility of T_1 measurements with a common IR-SE protocol. However, because of the small number of 1.5T systems, the reproducibility findings for both common protocols would need to be validated with a larger 1.5T scanner pool. For site-specific T_1 quantification protocols for different DCE-MRI applications, accuracy and repeatability of measurements were influenced primarily by the type of protocol and by scanner. Our study identified several VFA and Look-Locker IR protocols that achieved clinically acceptable accuracy and repeatability errors of less than 15%.

We observed high concordance between IR-SE and the reference NMR T_1 values, with deviations from unity line at the solution samples with 500 ms and higher T_1 values (1500–2000 ms). Bland-Altman plots showed discrepancy between IR-SE and reference NMR T_1 values at low T_1 values (< 50 ms) and 500 ms, but not at high T_1 values. Greater interplatform CV was also observed for the IR-SE sequence at 500 ms at both field strengths. The deviation from the reference standard in samples with high T_1 values can be explained by not fully recovered longitudinal magnetization (22,46) at the relatively short TR (5000 ms) chosen for manageable scanning time. The variability of IR-SE T_1 measurements of the 500-ms sample is caused by underperformance of the fitting function with magnitude IR-SE signal data, which can be remedied in future studies by fitting complex signal data (22,46).

Our findings show that the common IR-SE protocol had the highest accuracy in the range of 60 to 2000 ms (accuracy error $< 11\%$). Thus, for applications in which accuracy is important and longer acquisition times are possible without motion artifacts, an IR-SE protocol can be used as calibration for a faster T_1 measurement protocol (22). Accuracy with the common IR-SE protocol was also shown to depend on the $NiCl_2$ solution T_1, with lower accuracy for samples with very low reference T_1 (23–50 ms). These samples also showed interplatform CV greater than 10%. Our implemented IR-SE protocol had up to two measurements with short inversion times (24–50ms), which makes it difficult to resolve low T_1.

Inversion-recovery methods (IR-SE, Look-Locker modified IR) provided repeatable and reproducible T_1 measurement. Lower interplatform CV was observed for the common IR-SE protocol compared with the VFA common protocol. The IR-SE precision error was less than 5% for most platforms, and a significantly lower test-retest precision error/higher repeatability was observed for the Look-Locker IR than for the VFA and PD methods in the comparison by general linear model. These results are in agreement with the work of Stikov et al (22), which found greater stability of T_1 measurements in vitro and in white matter with Look-Locker IR than with VFA. Of note, the observed dependence of repeatability of the common IR-SE protocol on the scanner platform was likely driven by outlier test-retest precision errors at one of the scanners. This could be caused by periodic signal variation on the outlier scanner, as observed in other studies (46). In our study, the Look-Locker IR protocol optimized for the liver did not have greater accuracy than VFA protocols for the liver, but had greater test-retest repeatability. Because superior repeatability with the Look-Locker IR method was observed on only two platforms (1.5 and 3 T) from the same vendor, this finding should be confirmed in a future multiplatform, multivendor study.

For T_1 measurements with the common VFA protocol, we observed field strength dependence of accuracy and scanner dependence of test-retest repeatability, as well as significantly higher interplatform variability at 3 T. The deviations from the reference NMR measurements seen with the common VFA and site-specific protocols (as high as 30–40%) cannot be attributed only to differences in temperature (accounting for less than 2% error in T_1 for the recorded temperature range (32)). These are likely caused by B_1 field inhomogeneity, which is more pronounced at 3 T compared with 1.5 T (8,22), and leads to variability in T_1 values among scanners, even for a common VFA protocol with the same prescribed flip angles. Vendor was also a significant predictor of repeatability of T_1 measurements with the common VFA protocol, which suggests differences in implementation across vendors. Although the participating sites did their best to implement a uniform VFA protocol, the B_1 pulse profiles were unknown and could vary among scanners and vendors and change across the range of flip angles (32). Future implementation of a standardized VFA protocol would involve outreach to vendors to ensure that the B_1 correction methods used are appropriate for the pulse profiles used by the vendors in their implementations of VFA sequences. Accounting for incomplete spoiling of the spoiled gradient-echo signal and using the vendor-specific radiofrequency phase-difference increment can further be used to eliminate T_2^* bias in T_1 quantification (33). Our results with the common VFA protocol can

help facilitate the effort of optimizing VFA protocols (47,48) by identifying combinations of flip angles that minimize accuracy error with respect to the reference NMR-measured T_1 values.

Our study showed that although the VTR protocol and VFA protocols with fewer flip angles were more accurate (accuracy error $< 5\%$) than a VFA protocol with multiple flip angles (e.g., VTR prostate protocol 1, and prostate VFA protocols 9c and 9d versus protocol 2 for the brain), these protocols were significantly less repeatable (precision errors of 20–40%). For clinical applications in which both accuracy and repeatability are important, we would recommend protocols with accuracy and repeatability errors of less than 15%, such as the Look-Locker protocol 4a for the liver, VFA protocol 9b for the liver, or VFA protocol 9d for the prostate at 1.5 T.

Our study had some limitations. First, we did not collect complex signal data for any of the T_1 mapping methods, to simulate the clinical setting, in which only magnitude data are typically collected. Second, we did not perform B_1 mapping to measure the applied flip angles for the VFA measurements. B_1 mapping methods (49–53) are usually time-consuming, not standardized among platforms and vendors, and may be specific absorption rate–prohibitive (51), which limits their clinical applicability. Furthermore, B_1 may not be accurately measured as a result of B_0 inhomogeneities and tissue conductivity (22,54). Third, because participation in the study was determined by voluntary response to a survey, the study design was not statistically balanced for field strength (two 1.5T scanners versus eight 3T scanners) and vendors (six Siemens scanners, two GE scanners, two Phillips scanners), which may explain why vendor was one of the significant predictors of repeatability, but not of accuracy. The number of site-specific protocols was also not balanced between sites. Fourth, the effects of small variations in phantom positioning away from isocenter were not studied.

In conclusion, our findings show that accuracy, repeatability, and interplatform reproducibility of T_1 measurements depend on the T_1 measurement sequence and protocol used, the field strength, and the range of reference T_1 values. Among the site-specific protocols tested, VFA protocols with two to three flip angles optimized for different applications achieved an acceptable balance of accuracy and repeatability in T_1 quantification (errors $< 15\%$). The VFA protocols with two to three flip angles may be of interest to investigators designing translational DCE-MRI studies, as they have the advantage of higher spatial coverage achieved in the acquisition time of one breath-hold. Further optimization in terms of flip-angle choice for each tissue application, and the use of B_1 correction standardized among vendors, is needed to improve the robustness of VFA protocols for T_1 mapping in DCE-MRI for the purpose of multicenter studies.

ACKNOWLEDGMENTS

The authors thank James Babb, Ph.D., for his useful suggestions on statistical methods.

REFERENCES

1. Abdullah SS, Pialat JB, Wiart M, Duboeuf F, Mabrut JY, Bancel B, Rode A, Ducerf C, Baulieux J, Berthezene Y. Characterization of hepatocellular carcinoma and colorectal liver metastasis by means of perfusion MRI. J Magn Reson Imaging 2008;28:390–395.
2. Evelhoch JL. Key factors in the acquisition of contrast kinetic data for oncology. J Magn Reson Imaging 1999;10:254–259.
3. Ferl GZ, Port RE. Quantification of antiangiogenic and antivascular drug activity by kinetic analysis of DCE-MRI data. Clin Pharmacol Ther 2012;92:118–124.
4. Annet L, Materne R, Danse E, Jamart J, Horsmans Y, Van Beers BE. Hepatic flow parameters measured with MR imaging and Doppler US: correlations with degree of cirrhosis and portal hypertension. Radiology 2003;229:409–414.
5. Aronhime S, Calcagno C, Jajamovich GH, et al. DCE-MRI of the liver: effect of linear and nonlinear conversions on hepatic perfusion quantification and reproducibility. J Magn Reson Imaging 2014;40:90–98.
6. Bokacheva L, Rusinek H, Chen Q, Oesingmann N, Prince C, Kaur M, Kramer E, Lee VS. Quantitative determination of Gd-DTPA concentration in T1-weighted MR renography studies. Magn Reson Med 2007; 57:1012–1018.
7. Buckley DL, Shurrab AE, Cheung CM, Jones AP, Mamtora H, Kalra PA. Measurement of single kidney function using dynamic contrast-enhanced MRI: comparison of two models in human subjects. J Magn Reson Imaging 2006;24:1117–1123.
8. Fennessy FM, Fedorov A, Gupta SN, Schmidt EJ, Tempany CM, Mulkern RV. Practical considerations in T1 mapping of prostate for dynamic contrast enhancement pharmacokinetic analyses. Magn Reson Imaging 2012;30:1224–1233.
9. Hagiwara M, Rusinek H, Lee VS, Losada M, Bannan MA, Krinsky GA, Taouli B. Advanced liver fibrosis: diagnosis with 3D whole-liver perfusion MR imaging—initial experience. Radiology 2008;246:926–934.
10. Patel J, Sigmund EE, Rusinek H, Oei M, Babb JS, Taouli B. Diagnosis of cirrhosis with intravoxel incoherent motion diffusion MRI and dynamic contrast-enhanced MRI alone and in combination: preliminary experience. J Magn Reson Imaging 2010;31:589–600.
11. Sourbron SP, Buckley DL. Classic models for dynamic contrast-enhanced MRI. NMR Biomed 2013;26:1004–1027.
12. Materne R, Smith AM, Peeters F, Dehoux JP, Keyeux A, Horsmans Y, Van Beers BE. Assessment of hepatic perfusion parameters with dynamic MRI. Magn Reson Med 2002;47:135–142.
13. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999;10:223–232.
14. Dale BM, Jesberger JA, Lewin JS, Hillenbrand CM, Duerk JL. Determining and optimizing the precision of quantitative measurements of perfusion from dynamic contrast enhanced MRI. J Magn Reson Imaging 2003;18:575–584.
15. Bane O, Wagner M, Zhang JL, Dyvorne HA, Orton M, Rusinek H, Taouli B. Assessment of renal function using intravoxel incoherent motion diffusion-weighted imaging and dynamic contrast-enhanced MRI. J Magn Reson Imaging 2016;44:317–326.
16. Oto A, Yang C, Kayhan A, Tretiakova M, Antic T, Schmid-Tannwald C, Eggener S, Karczmar GS, Stadler WM. Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis. AJR Am J Roentgenol 2011;197:1382–1390.
17. Fedorov A, Fluckiger J, Ayers GD, Li X, Gupta SN, Tempany C, Mulkern R, Yankeelov TE, Fennessy FM. A comparison of two methods for estimating DCE-MRI parameters via individual and cohort based AIFs in prostate cancer: a step towards practical implementation. Magn Reson Imaging 2014;32:321–329.
18. de Bazelaire CM, Duhamel GD, Rofsky NM, Alsop DC. MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T: preliminary results. Radiology 2004;230:652–659.
19. Besa C, Bane O, Jajamovich G, Marchione J, Taouli B. 3D T1 relaxometry pre and post gadoxetic acid injection for the assessment of liver cirrhosis and liver function. Magn Reson Imaging 2015;33:1075–1082.
20. Kim KA, Park MS, Kim IS, Kiefer B, Chung WS, Kim MJ, Kim KW. Quantitative evaluation of liver cirrhosis using T1 relaxation time with 3 tesla MRI before and after oxygen inhalation. J Magn Reson Imaging 2012;36:405–410.

21. Kingsley P. Methods of measuring spin-lattice (t1) relaxation times: an annotated bibliography. Concepts Magn Reson 1999;11:243–276.
22. Stikov N, Boudreau M, Levesque IR, Tardif CL, Barral JK, Pike GB. On the accuracy of T1 mapping: searching for common ground. Magn Reson Med 2015;73:514–522.
23. Brown RW, Cheng Y-CN, Haacke EM, Thompson MR, Venkatesan R. Magnetic resonance imaging: physical principles and sequence design. New York: Wiley-Blackwell; 2014.
24. Berger S, Braun S. 200 and more NMR experiments: a practical course. Weinheim, Germany: Wiley-VCH; 2004. p. 854.
25. Look DC, Locker D. Time saving in measurement of NMR and EPR relaxation times. Rev Sci Instrum 1970;41:250–251.
26. Raman FS, Kawel-Boehm N, Gai N, Freed M, Han J, Liu CY, Lima JA, Bluemke DA, Liu S. Modified look-locker inversion recovery T1 mapping indices: assessment of accuracy and reproducibility between magnetic resonance scanners. J Cardiovasc Magn Reson 2013;15:64.
27. Roujol S, Weingartner S, Foppa M, Chow K, Kawaji K, Ngo LH, Kellman P, Manning WJ, Thompson RB, Nezafat R. Accuracy, precision, and reproducibility of four T1 mapping sequences: a head-to-head comparison of MOLLI, ShMOLLI, SASHA, and SAPPHIRE. Radiology 2014;272:683–689.
28. Aryal MP, Chenevert TL, Cao Y. Impact of uncertainty in longitudinal T1 measurements on quantification of dynamic contrast-enhanced MRI. NMR Biomed 2016;29:411–419.
29. Schabel MC, Parker DL. Uncertainty and bias in contrast concentration measurements using spoiled gradient echo pulse sequences. Phys Med Biol 2008;53:2345–2373.
30. Huang W, Wang Y, Panicek DM, Schwartz LH, Koutcher JA. Feasibility of using limited-population-based average R10 for pharmacokinetic modeling of osteosarcoma dynamic contrast-enhanced magnetic resonance imaging data. Magn Reson Imaging 2009;27:852–858.
31. Keenan KE, Boss M, Jackson EF, Kown S-j, Jennings D, Russek S. NIST/ISMRM MRI system phantom T1 measurements on multiple MRI systems. In Proceedings of the 21st Annual Meeting of ISMRM, Salt Lake City, Utah, USA, 2013. p. 4338.
32. Keenan KE, Stupic K, Boss M, et al. Multi-site, multi-vendor comparison of T1 measurement using NIST/ISMRM system phantom. In Proceedings of the 24th Annual Meeting of ISMRM, Singapore, 2016. p. 3290.
33. Heule R, Ganter C, Bieri O. Variable flip angle T1 mapping in the human brain with reduced T2 sensitivity using fast radiofrequency-spoiled gradient echo imaging. Magn Reson Med 2016;75:1413–1422.
34. Chenevert TL, Malyarenko DI, Newitt D, et al. Errors in quantitative image analysis due to platform-dependent image scaling. Transl Oncol 2014;7:65–71.
35. National Cancer Institute: Cancer Imaging Program. QIN Network Organization website. https://imaging.cancer.gov/programs_resources/specialized_initiatives/qin.htm Published 2012.Updated 10/28/16.Accessed 05/18/2017.
36. Kurland BF, Gerstner ER, Mountz JM, et al. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magn Reson Imaging 2012;30:1301–1312.
37. Russek S. NIST Phantom TWIKI. http://collaborate.nist.gov/mriphantoms/bin/view/MriPhantoms/PhantomOverview Published 2014. Updated August 19, 2014. Accessed May 18, 2017.
38. Clunie DA. DICOM structured reporting and cancer clinical trials results. Cancer Inform 2007;4:33–56.
39. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–268.
40. Lin LI-K. A note on the concordance correlation coefficient biometrics. Stata J 2000;56:324–325
41. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19:716–723.
42. Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A comparison of recent approaches to the analysis of repeated measurements. Br J Math Stat Psychol 1999;52:63–78.
43. Brunner E, Domhof S, Langer F. Nonparametric analysis of longitudinal data in factorial experiments. New York: John Wiley & Sons; 2002.
44. Tukey JW. The problem of multiple comparisons. In: Braun HI, editor. New York: Chapman & Hall; 1994.
45. Keenan KE, Wilmes LJ, Aliu SO, Newitt DC, Jones EF, Boss MA, Stupic KF, Russek SE, Hylton NM. Design of a breast phantom for quantitative MRI. J Magn Reson Imaging 2016;44:610–619.
46. Jerome NP, Papoutsaki MV, Orton MR, Parkes HG, Winfield JM, Boss MA, Leach MO, deSouza NM, Collins DJ. Development of a temperature-controlled phantom for magnetic resonance quality assurance of diffusion, dynamic, and relaxometry measurements. Med Phys 2016;43:2998.
47. Deoni SC, Peters TM, Rutt BK. Determination of optimal angles for variable nutation proton magnetic spin-lattice, T1, and spin-spin, T2, relaxation times measurement. Magn Reson Med 2004;51:194–199.
48. Deoni SC, Rutt BK, Peters TM. Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. Magn Reson Med 2003;49:515–526.
49. Sacolick LI, Wiesinger F, Hancu I, Vogel MW. B1 mapping by Bloch-Siegert shift. Magn Reson Med 2010;63:1315–1322.
50. Yarnykh VL. Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magn Reson Med 2007;57:192–200.
51. Fleysher R, Fleysher L, Liu S, Gonen O. TriTone: a radiofrequency field (B1)-insensitive T1 estimator for MRI at high magnetic fields. Magn Reson Imaging 2008;26:781–789.
52. Pohmann R, Scheffler K. A theoretical and experimental comparison of different techniques for B(1) mapping at very high fields. NMR Biomed 2013;26:265–275.
53. Whisenant JG, Dortch RD, Grissom W, Kang H, Arlinghaus LR, Yankeelov TE. Bloch-Siegert B1-mapping improves accuracy and precision of longitudinal relaxation measurements in the breast at 3 T. Tomography 2016;2:250–259.
54. Wang J, Qiu M, Kim H, Constable RT. T1 measurements incorporating flip angle calibration and correction in vivo. J Magn Reson 2006; 182:283–292.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Fig. S1. Sample fitting the IR-SE signal in phantom sample S5, with NMR T_1 approximately 500 ms. When fitting Equation [1] to the magnitude signal data with a Levenberg-Marquardt algorithm with constraints on parameters $T_1 = [0, Inf]$, $M0 = [0, Inf]$, $invF = [0,2]$ and $noise = [0, Inf]$, the fit (in black) underperforms and T_1 is underestimated ($T_1 = 370.6 \pm 73.6$ ms). When parameters are fitted without constraints, the fit to the data is improved (red) ($T_1 = 509.7 \pm 0.91$ ms).

Fig. S2. Typical model fit for the IR-SE signal obtained with the common IR-SE protocol on scanner 4 (3 T) in phantom samples S1 to S14.

Fig. S3. Typical model fit for the VFA signal obtained with the common VFA protocol on scanner 4 (3 T) in phantom samples S1 to S14.

Table S1. Acquisition Parameters for the Site-Specific T_1 Measurement Protocols at 3 T

Table S2. Acquisition Parameters for the Site-Specific T_1 Measurement Protocols at 1.5 T

Table S3. T_1 (ms) Values From NMR Spectroscopy Measurements Performed at NIST at 1.5 and 3 T Used as the Reference Standard

Table S4. Accuracy Error (%) for Each Scanner, With the Common IR-SE and VFA Protocols, for Ranges of Reference T_1 Values in the Phantom Solution Samples

Table S5. Precision Error (%) for Each Scanner, With the Common IR-SE and VFA Protocols, for Ranges of Reference T_1 Values in the Phantom Solution Samples