Bias, Repeatability and Reproducibility of Liver T_1 Mapping With Variable Flip Angles

Sirisha Tadimalla, PhD,[1,2] Daniel J. Wilson, PhD,[3] David Shelley, BSc,[3] Gavin Bainbridge, BSc,[3] Margaret Saysell, BSc,[3] Iosif A. Mendichovszky, MD,[4] Martin J. Graves, PhD,[4] J. Ashley Guthrie, MB,[3] John C. Waterton, PhD,[5,6] Geoffrey J.M. Parker, PhD,[5,7] and Steven P. Sourbron, PhD[8]*

Background: Three-dimensional variable flip angle (VFA) methods are commonly used for T_1 mapping of the liver, but there is no data on the accuracy, repeatability, and reproducibility of this technique in this organ in a multivendor setting.
Purpose: To measure bias, repeatability, and reproducibility of VFA T_1 mapping in the liver.
Study Type: Prospective observational.
Population: Eight healthy volunteers, four women, with no known liver disease.
Field Strength/Sequence: 1.5-T and 3.0-T; three-dimensional steady-state spoiled gradient echo with VFAs; Look-Locker.
Assessment: Traveling volunteers were scanned twice each (30 minutes to 3 months apart) on six MRI scanners from three vendors (GE Healthcare, Philips Medical Systems, and Siemens Healthineers) at two field strengths. The maximum period between the first and last scans among all volunteers was 9 months. Volunteers were instructed to abstain from alcohol intake for at least 72 hours prior to each scan and avoid high cholesterol foods on the day of the scan.
Statistical Tests: Repeated measures ANOVA, Student *t*-test, Levene's test of variances, and 95% significance level. The percent error relative to literature liver T_1 in healthy volunteers was used to assess bias. The relative error (RE) due to intrascanner and interscanner variation in T_1 measurements was used to assess repeatability and reproducibility.
Results: The 95% confidence interval (CI) on the mean bias and mean repeatability RE of VFA T_1 in the healthy liver was 34 $\pm$ 6% and 10 $\pm$ 3%, respectively. The 95% CI on the mean reproducibility RE at 1.5 T and 3.0 T was 29 $\pm$ 7% and 25 $\pm$ 4%, respectively.
Data Conclusion: Bias, repeatability, and reproducibility of VFA T_1 mapping in the liver in a multivendor setting are similar to those reported for breast, prostate, and brain.
Level of Evidence: 1
Technical Efficacy Stage: 1

J. MAGN. RESON. IMAGING 2022.

Renewed interest in T_1 as a quantitative imaging biomarker (QIB) has sparked an increase in research and development of fast and accurate T_1 mapping methods across the body.[1–4] A wide range of traditional methods for T_1 mapping are available, which include fully relaxed methods such as inversion-recovery (IR), Look-Locker (LL), and steady-state

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.28127

Received Nov 24, 2021, Accepted for publication Feb 10, 2022.

*Address reprint requests to: S.P.S., POLARIS, University Imaging, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, C Floor, Royal Hallamshire Hospital, Glossop Road, Sheffield S10 2JF, Sheffield, UK. Email: s.sourbron@sheffield.ac.uk

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (IMI2JU) under grant agreement No. 116106. This Joint Undertaking received support from the European Union's Horizon 2020 research and innovation program and EFPIA. This publication reflects only the authors' view and IMI2JU is not responsible for any use that may be made of the information it contains. This research was also funded by the National Institute for Health Research (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

From the [1]Institute of Medical Physics, University of Sydney, Sydney, Australia; [2]Department of Biomedical Imaging Sciences, University of Leeds, Leeds, UK; [3]Leeds Teaching Hospitals Trust, Leeds, UK; [4]Department of Radiology, University of Cambridge, Cambridge, UK; [5]Bioxydyn Ltd, Manchester, UK; [6]Centre for Imaging Sciences, Division of Informatics Imaging and Data Sciences, School of Health Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK; [7]Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK; and [8]Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

Additional supporting information may be found in the online version of this article

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Magnetic Resonance Imaging* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

methods such as variable flip angle (VFA) or variable repetition time saturation recovery (VTR). IR methods are the most accurate[5] but are usually too slow for practical use, and VTR methods are not routinely used in clinical settings. LL type methods including modified LL imaging (MOLLI) are commonly used in abdomen and thorax for multislice, two-dimensional T_1 mapping.[6] These methods are valued for their high reproducibility[7] but are unsuitable for applications that require volumetric coverage, as in liver disease. In such applications, there is a rationale for VFA methods, which allow fast, three-dimensional T_1 mapping of large volumes. VFA is widely used, and the past few years in particular have seen an increase in the application of VFA techniques in organs such as brain, breast, and prostate.[8–10]

A downside of VFA methods is that they are more susceptible to bias caused by B_1^+ nonuniformities, imperfect spoiling, and magnetization transfer (MT) effects.[5,11] These errors depend on scanner hardware and sequence optimization and may vary spatially across the field of view, affecting accuracy, intrascanner repeatability, and interscanner reproducibility in multicenter clinical trials or diagnostic methods that require relaxation time measurements, for example, for the assessment of treatment response.

A multicenter phantom study across 10 scanners of three vendors and two field strengths using VFA T_1 mapping found that the combined effect of these errors can be substantial, producing a bias up to 32%, intrascanner relative error (RE) (Different definitions of repeatability and reproducibility metrics are in common use. Due to repeatability effects up to 26% and interscanner reproducibility RE of 22% at 1.5 T and 45% at 3.0 T.[12] (Different definitions of repeatability and reproducibility metrics are in common use. For the purposes of this paper, whenever literature values are cited, they are converted to the RE definition used in this study [see Methods section] to allow direct numerical comparison between results of this study and the literature.) While such phantom studies are a valuable and necessary contribution to characterizing the performance of quantitative measurements, their findings cannot be used directly to infer performance in vivo, due to subject- and organ-specific sources of variation. These include B_1^+ errors caused by nonuniform RF penetration and standing wave effects, and the impact of physiological motion including breathing and blood flow on the measurements.

Correction techniques for B_1^+ errors have been proposed,[13,14] and multisite studies in organs such as the brain[15] and breast[16,17] suggest this improves repeatability and reproducibility. In the brain, a multiparametric VFA protocol with corrections for B_1^+ and imperfect spoiling on six 3.0 T scanners from two vendors reported repeatability- and reproducibility RE for R_1 ($=1/T_1$) up to 16% and 20%, respectively.[8] In breast fibroglandular tissue, a single vendor study reported reproducibility RE across three sites of 14% in VFA T_1 after B_1^+ correction at 3.0 T.[9] Similar values were found in the prostate.[10] However, these studies do not provide a comprehensive coverage of the two main clinical field strengths and three main vendors. As a result, biases, repeatability, and reproducibility RE may be underestimated.

In the liver, VFA T_1 mapping has been proposed to assess conditions such as liver fibrosis and cirrhosis and for calibration of dynamic contrast enhanced (DCE) MRI.[18–27] A few studies of repeatability and/or reproducibility of liver T_1 have been reported but to our knowledge none employed VFA and therefore do not address the influence of varying B_1^+ fields on VFA-derived liver T_1.[28,29] It is not guaranteed that results in relatively static organs like the brain, breast, and prostate will translate to the liver, which exhibits significant deformable breathing motion and may be more susceptible to inhomogeneities due to its large size.

Thus, the aim of this study was therefore to determine the bias, repeatability, and reproducibility of VFA T_1 mapping in the liver, in real-world conditions at 1.5 T and 3.0 T and in scanners of three main vendors in widespread use today. These values might be used in uncertainty analysis and for estimating study power and will establish a baseline against which subsequent methodological developments can be benchmarked.

Materials and Methods

Subjects

Eight healthy volunteers (age = 23–58 years, mean 37 years; 4 women) with no known liver disease or MRI contraindications were scanned twice each (between 30 minutes to 3 months apart) on six MRI scanners. The maximum period between the first and last scans among all volunteers was 9 months. The study was approved by the institutional research ethics committee (University of Leeds, Faculty of Medicine and Health: MREC17-111), and written informed consent was obtained from all volunteers. Volunteers were instructed to abstain from alcohol intake for at least 72 hours prior to each scan and to avoid high cholesterol foods on the day of the scan.

Scanners

Details of the scanners and coils used are summarized in Table 1. A total of six scanners (two field strengths × three vendors [GE Healthcare, Philips Medical Systems, and Siemens Healthineers]) located across four different sites were used.

MRI Protocol

The MRI protocol was developed initially on the Siemens 3.0 T scanner and then transferred as closely as possible on the other scanners. Where an exact one-to-one correspondence of sequence parameters was not possible, the spatiotemporal geometries were aligned first (acquired voxel size, field of view (FOV), and acquisition time) to ensure a fair comparison of scanners in terms of signal to noise ratio (SNR) and to match breath-hold times, after which contrast

TABLE 1. Details of Scanners Used in This Study

Scanner	Vendor	Model	RF Coil
S1	Siemens Healthineers	3.0 T Prisma VE11C	18 channel array
S2	Siemens Healthineers	1.5 T Aera VE11A	18 channel array
P1	Philips Medical Systems	3.0 T Achieva	32 channel array
P2	Philips Medical Systems	1.5 T Ingenia	dStream Torso
G1	GE Healthcare	3.0 T Discovery MR750	32 channel array
G2	GE Healthcare	1.5 T Discovery MR450	32 channel array

and other parameters were optimized to match the reference protocol as closely as possible. Pilot data were acquired on standardized phantoms and volunteers to ensure sufficient image quality.

The protocol consisted of standard localizer/survey and calibration scans, followed by multislice two-dimensional anatomical reference T_2-weighted scans in the 1) coronal and 2) transverse planes with full liver coverage, 3) where available, a reference two-dimensional T_1 mapping sequence using an LL or MOLLI type sequence, 4) a three-dimensional coronal RF spoiled gradient echo (SPGR) breath-hold (BH, ~16 seconds) sequence with six flip angles (VFA BH), and 5) a fast (~2 seconds) three-dimensional coronal free-breathing (FB) RF spoiled SPGR sequence with six flip angles (VFA FB). The fast sequence was acquired continuously for up to a minute to average out breathing motion.

Two-dimensional LL T_1 mapping was implemented as a reference on the Siemens 3.0 T scanner using a nonselective IR magnetization preparation and gradient echo readout, with a simulated heart rate of 80 beats/minute to ensure sufficient sampling of the recovery curve for liver T_1. On the Philips 1.5 T and 3.0 T scanners, a dedicated MOLLI sequence was used with simulated electrocargiogram (ECG) using the same heart rate of 80 beats/minute. The two-dimensional MOLLI sequence was not available on the other three scanners.

Three-dimensional VFA T_1 mapping was implemented on the Siemens scanners using a three-dimensional Fast Low Angle Shot (FLASH) sequence and on the Philips scanners using a T_1-fast field echo sequence. Three-dimensional T_1-weighted images were acquired with a set of six flip angles, with receiver gains set using a preparation scan at 15°. Preparation scans were turned off for subsequent flip angles to ensure a constant receiver gain. On the GE scanners, it was not possible to automatically (without user-controlled manual prescan) prevent a change of the receiver gain between different flip angle acquisitions using the product three-dimensional Fast Spoiled Gradient Echo (FSPGR) sequence. It was also not possible to set up the VFA FB sequence with multiple measurements for each flip angle acquisition. Severe phase-wrap artifacts were also observed when using the FSPGR product sequence for coronal VFA acquisition. Therefore, a modified version of the FSPGR sequence, named the multiphase multiflip angle (MPMFA) sequence, was developed in-house. The MPMFA sequence allowed 1) single acquisition for all flip angles with a fixed receiver gain set at flip angle 15°, 2) multiple measurements within acquisition of each flip angle, and 3) a rotated slab excitation to reduce phase-wrap artifacts and sufficient anterior posterior coverage. The code for the MPMFA sequence will be made available upon request from sites with access to the GE research sharing database. B_1^+ mapping was not implemented on any scanner due to lack of product mapping sequences on all scanners.

Detailed imaging parameters for each scanner are given in Table S1.

Image Processing

Anonymized images were transferred from all scanners in DICOM format. Image processing was performed centrally by a single user (S.T.—9 years of experience) using the open-source software PMI (https://github.com/plaresmedima/PMI-0.4) customized for this purpose (compiled version freely available as supplementary material at https://doi.org/10.5281/zenodo.5589509). MOLLI T_1 maps were obtained by fitting to signal intensities as a function of inversion time as described by Messroghli et al.[30] VFA T_1 maps were obtained by fitting signal intensities at the six flip angles with the linearized steady-state SPGR equation.[5] Continuously acquired free-breathing three-dimensional coronal SPGR images were motion-corrected by using nonrigid registration with free form deformation between each image and magnitude averaged prior to VFA T_1 mapping.[31] As the study was performed on healthy volunteers with no known liver disease, liver fat and iron levels were assumed to be normal and no corrections were applied during T_1 calculations. Liver regions of interest (ROIs) were drawn on the T_1 maps as follows: 1) a central slice containing the portal vein at its largest was chosen; and 2) the entire liver within the slice was manually outlined and large blood vessels were removed from the ROI by user-defined thresholding. The user was not blinded to time points or subjects and compared segmentations from the same subject across scans to avoid intrareader segmentation differences impacting on the result.

Statistical Analysis

Median T_1 values within the ROIs were extracted from each T_1 map. Bias estimate, repeatability, reproducibility, and spatial inhomogeneity were calculated for each volunteer as described below, and averages over the volunteers were reported along with their standard error.

$$\text{Bias estimate}(\%) = 100 \times \frac{T_1(\text{averaged over scans 1 and 2}) - \text{reference } T_1}{\text{reference } T_1}$$

$$\text{Repeatability RE}(\%) = 100 \times 1.96 \times \frac{\text{Standard deviation (scan1, scan2)}}{\text{Average (scan1, scan2)}}$$

$$\text{Reproducibility RE}\,(\%) = 100 \times 1.96 \times \frac{\text{Standard deviation (all vendors)}}{\text{Average (all vendors)}}$$

$$\text{Spatial heterogeneity}(\%) = 100 \times \frac{\text{Inter quartile range } T_1}{\text{Median } T_1}, \text{averaged over scans 1 and 2}$$

The repeatability RE measures the relative random error (half of the 95% confidence interval, CI) on median T_1 when all measurements are done on the same machine; the reproducibility RE measures the relative random error (half of the 95% CI) on median T_1 when measurements are performed on machines from different vendors. Reference T_1 values in the liver were obtained from literature for healthy liver (752 msec at 3.0 T, 602 msec at 1.5 T).[32] Separate pairwise reproducibility REs were also calculated for the three possible pairs of vendors.

Between-subject variation in T_1, estimated for each scanner and method, was calculated as,

$$\text{Between-subject variation}\,(\%) = 100 \times 1.96 \times \frac{\text{Standard deviation (all subjects)}}{\text{Average (all subjects)}}$$

Comparisons of means across subjects, sequences, and scanners were performed using repeated measures analysis of variance (rANOVA). When the *P*-value was less than 0.05, post hoc pairwise *t*-tests were performed. Levene's test was used to compare between-subject measurement variances across sequences and scanners.

Incidental Findings

Scans were also read by a radiologist so that any incidental findings could be followed up confidentially: any such findings were not disclosed to the investigators and so are not reported in this paper.

Results

All volunteers completed the study except one who withdrew after completing scans on three of the six scanners. All acquired data were included in the analysis.

Summary Parameters

Table 2 presents the 95% CI on the overall bias estimate, repeatability, reproducibility, and spatial heterogeneity of MOLLI and VFA T_1 values in the liver. The table shows that VFA overestimates median liver T_1 by 30% on average, relative to literature estimates. The RE is $\pm 10\%$ if all scans are performed on a single scanner and $\pm 27\%$ if scanners from different vendors and field strengths are used. Bias estimate, repeatability, and reproducibility of MOLLI T_1 measurements are consistent with those reported in the literature using IR/LL T_1 mapping methods in the liver.[29,32] More detailed performance metrics for each sequence, field strength, and vendor are provided in Table S2. When compared with MOLLI, VFA T_1 values are 15 times more biased, four times less repeatable, two times less reproducible, and two times less homogeneous. Table 3 compares the repeatability and reproducibility in liver T_1 as measured in this study against other results in breast, brain, and prostate after correcting for different definitions of repeatability and reproducibility. Corrections applied to convert literature definitions of repeatability and reproducibility to the definitions used in this study are given in Table S3.

Effect of Field Strength

Figure 1 shows the bias, repeatability, reproducibility, and spatial heterogeneity of VFA T_1 measurements for the two clinical field strengths separately. Results of the corresponding rANOVA and Levine's tests are given in Table 4. Unlike the VFA FB sequence, the T_1 bias was not significantly different between field strengths for the VFA BH sequence, despite increased B_1^+ nonuniformity expected at 3.0 T. However, the variance in the T_1 bias was larger at 3.0 T for the VFA FB sequence, in line with a higher spatial heterogeneity for this sequence at 3.0 T. While field strength had no effect on the repeatability, both the mean and the variance of the relative reproducibility error were lower at 3.0 T for VFA BH.

Effect of Vendor

Figure 2 summarizes the bias, repeatability, reproducibility, and spatial heterogeneity of VFA T_1 measurements across the three vendors (S, P, and G), and for the possible pairs of two vendors for both field strengths. Results of the corresponding rANOVA and Levene's tests are given in Table 5. Vendor has no effect on repeatability of liver VFA T_1 values; however, variance in bias on vendor S scanners is significantly higher than on vendors G and P. This manifests also as a significant improvement in reproducibility when vendor S is removed.

TABLE 2. The 95% CI on Overall Mean Bias, Repeatability, Reproducibility, and Spatial Heterogeneity of T_1 Values in the Normal Liver

	MOLLI	VFA BH	VFA FB
Bias (%)	1.7 ± 3.7**	31 ± 8.0	29 ± 8.1
Relative repeatability Error (%)	2.4 ± 0.7**	11 ± 3.1	9.4 ± 2.0
Relative reproducibility error (%) at 1.5 T		34 ± 9.2[a]	24 ± 8.1
Relative reproducibility error (%) at 3.0 T	14 ± 5.8*	22 ± 4.0[a]	29 ± 7.1
Spatial heterogeneity (%)	11 ± 1.5**	25 ± 1.9[a]	18 ± 2.3

MOLLI, modified Look-Locker imaging; rANOVA, repeated measures ANOVA; VFA, variable flip angle.
[a]Paired *t*-test (VFA BH vs. VFA FB) *P*-value < 0.0001.
**rANOVA *P*-value < 0.0001; *rANOVA *P*-value < 0.01.

TABLE 3. Repeatability and Reproducibility of VFA T_1 Values in the Liver, Compared to Literature Values in Phantoms and Other Organs

Application Area	Corrections	Scanners	Relative Repeatability Error[a] (%)	Relative Reproducibility Error[a] (%)
Liver (from this study)	None	1.5 T: 1 × S, 1 × P, 1 × G 3.0 T: 1 × G, 1 × P, 1 × S	10%	28%
Breast (reference 8)	B_1^+ correction	3.0 T: 3 × S	11%	14%
Brain (R_1) (reference 7)	B_1^+ correction; imperfect spoiling	3.0 T: 4 × S, 2 × P	10%–16%	10%–20%
Prostate (transitional zone) (reference 9)	With and without B_1^+ correction	3.0 T: 2 × S	12% (14% without correction)	14% (18% without correction)

VFA, variable flip angle.
[a]Literature values were converted to the definition used in this study.

Vendor choice also affects spatial heterogeneity, as seen by the lower mean and variance in heterogeneity in VFA FB T_1 values on vendor G.

Effect of Sequence Optimization

No significant differences in bias or repeatability were found between the VFA BH and VFA FB methods ($P = 0.22$ and 0.29, respectively). However, reproducibility errors were significantly different at both the field strengths. The use of very low spatial resolution during imaging with VFA FB resulted in significantly lower spatial heterogeneity than the BH approach.

Comparison of T_1 Values

T_1 values in the normal liver in all volunteers on all six scanners are provided as supplementary material (https://doi.org/10.5281/zenodo.5589509) and presented in Fig. 3. The 95% CI of the mean T_1 values at 1.5 T and 3.0 T are summarized in Table 6. The data confirm that in general the average VFA T_1 is overestimated relative to the MOLLI reference (as seen from the bias estimates). One notable exception is the VFA T_1 measurements on vendor S at 1.5 T which are close to the corresponding MOLLI T_1 measurements.

Between-Subject Variation

Between-subject variation measured in T_1 values within the population of volunteers in this study for each scanner and method are shown in Fig. 3. Overall, between-subject differences in measured T_1 values are significant for both MOLLI and VFA methods. The MOLLI method estimates a between-subject variation of approximately 15%, irrespective

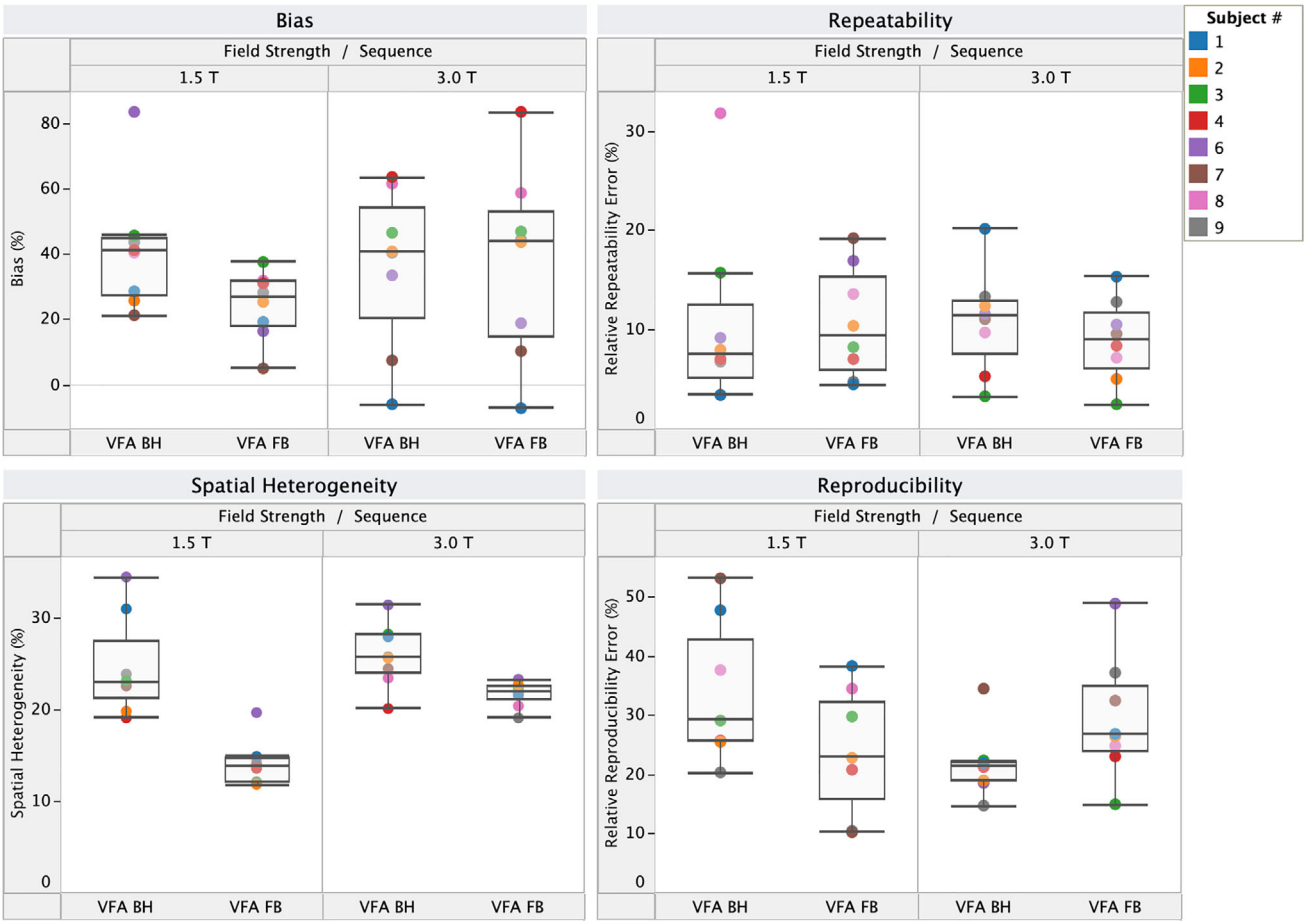


FIGURE 1: Bias, repeatability, reproducibility, and spatial heterogeneity of variable flip angle (VFA) T_1 measurements across field strengths. The box plots indicate the median and interquartile range (IQR), and the whiskers encompass all points within 1.5 times the IQR. Each color-coded datapoint represents a single volunteer.

TABLE 4. Results (*P*-values) of Repeated Measures ANOVA (Levene's Tests in Parentheses) on the Effect of Field Strength on Bias, Repeatability, Reproducibility, and Spatial Heterogeneity

Sequence	Bias	Repeatability	Reproducibility	Spatial Heterogeneity
MOLLI	**<0.001** (0.70)	0.29 (0.36)	-	**<0.0001** (0.30)
VFA BH	0.16 (0.30)	0.99 (0.77)	**0.01 (0.03)**	0.30 (0.26)
VFA FB	**0.004 (0.006)**	0.69 (0.75)	0.68 (0.17)	**<0.0001** (0.31)

Significant results are highlighted in bold font.
MOLLI, modified Look-Locker imaging; VFA, variable flip angle.

of vendor and field strength. With the VFA methods, subject-wise differences are larger, the degree depending on both field strength as well as vendor. The scanner from vendor S at 3.0 T generates the largest between-subject variation with both VFA methods, at 56% on average.

Discussion

In this work, bias, repeatability, reproducibility, and spatial heterogeneity of VFA T_1 values were measured in the liver of healthy volunteers, on a representative set of six scanners at two field strengths from three vendors.

Bias in VFA T_1 values are rarely reported in vivo, but the values reported in this study were smaller than those reported in a standardized phantom.[12,33] Bias was calculated relative to reference liver T_1 measurements at the two field strengths obtained from literature; the literature values used as reference were close to the MOLLI T_1 measurements. Root-mean-squared deviation of MOLLI, VFA BH, and VFA

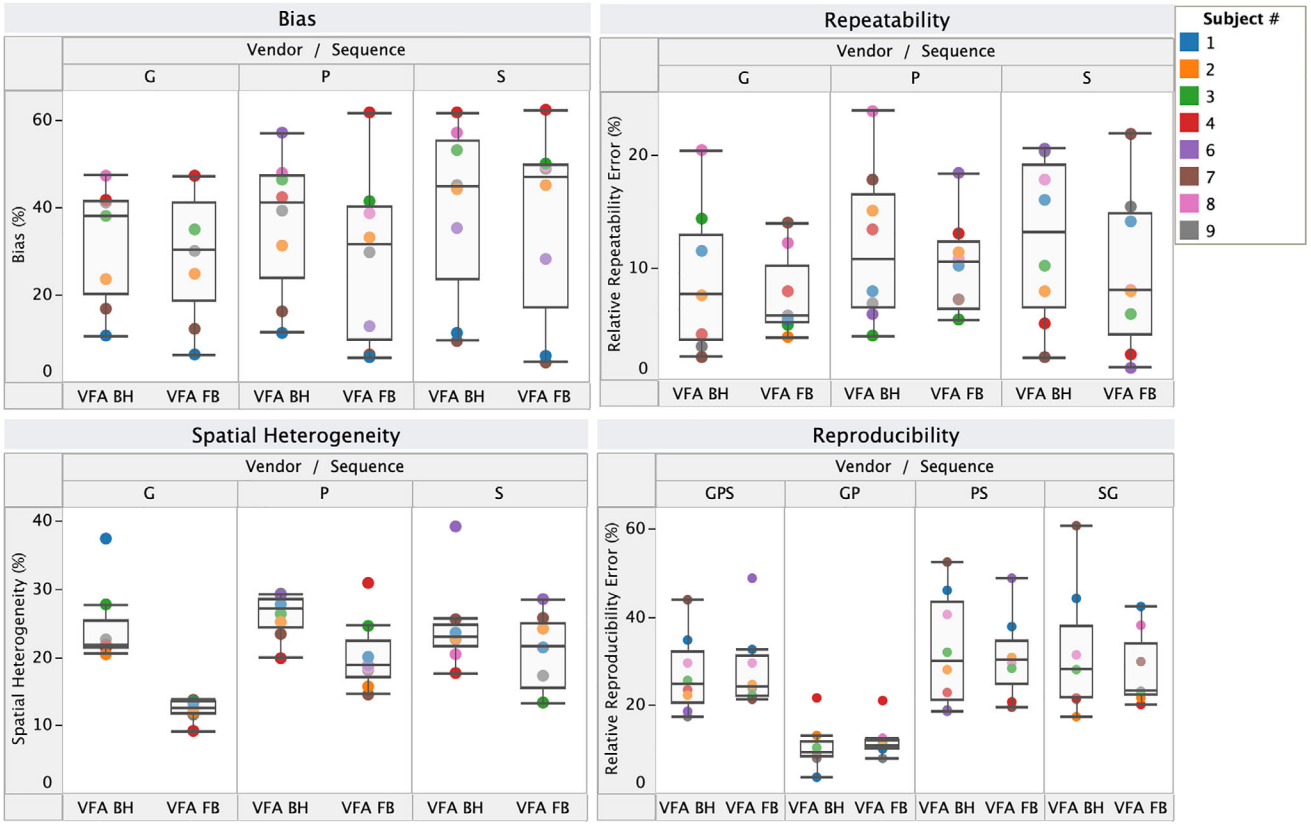


FIGURE 2: Bias, repeatability, reproducibility, and spatial heterogeneity of variable flip angle (VFA) T_1 measurements across vendors. The box plots indicate the median and interquartile range (IQR), and the whiskers encompass all points within 1.5 times the IQR. Each color-coded datapoint represents a single volunteer.

TABLE 5. Results (*P*-Values) of Repeated Measures ANOVA (Levene's Tests in Parentheses) on the Effect of Vendor on Bias, Repeatability, and Spatial Heterogeneity

Sequence	Bias	Repeatability	Reproducibility	Spatial Heterogeneity
MOLLI[a]	**0.003** (0.30)	0.96 (0.88)	-	**0.002** (0.58)
VFA BH	0.08 (**0.005**)	0.69 (0.80)	**<0.0001** (0.05)	0.31 (0.65)
VFA FB	0.69 (**0.0003**)	0.40 (0.13)	**<0.0001** (0.22)	**0.007 (0.003)**

Significant results are highlighted in bold font.
MOLLI, modified LL imaging; VFA, variable flip angle.
[a]Vendors P and S only.

FB T_1 measurements in this study from literature values are 3.4%, 25%, and 23%, respectively.[32]

The between-subject variation in liver T_1 estimated by the MOLLI method is also consistent with the literature, while the variation in VFA T_1 values between subjects is considerably larger.[32] This is consistent with the effect of known errors in VFA-based T_1 measurement, in particular the impact of B_1 effects, which are known to cause nonuniform excitations across the liver in a manner that depends on body size. The hypothesis is also supported by the observed trends in the spatial heterogeneity measurements. At 1.5 T, the VFA FB shows a comparable spatial heterogeneity to the MOLLI, consistent with the observation that the between-subject variability at 1.5 T is similar between MOLLI and VFA FB. At 3.0 T, the spatial heterogeneity of the VFA FB is substantially larger than MOLLI for vendors P and S, but not for vendor G, again in agreement with the observed T_1-variability in the population. The data also indicate that these errors are to some extent reproducible. For example, low liver T_1 values are recorded for subjects 1 and 7, and high liver T_1

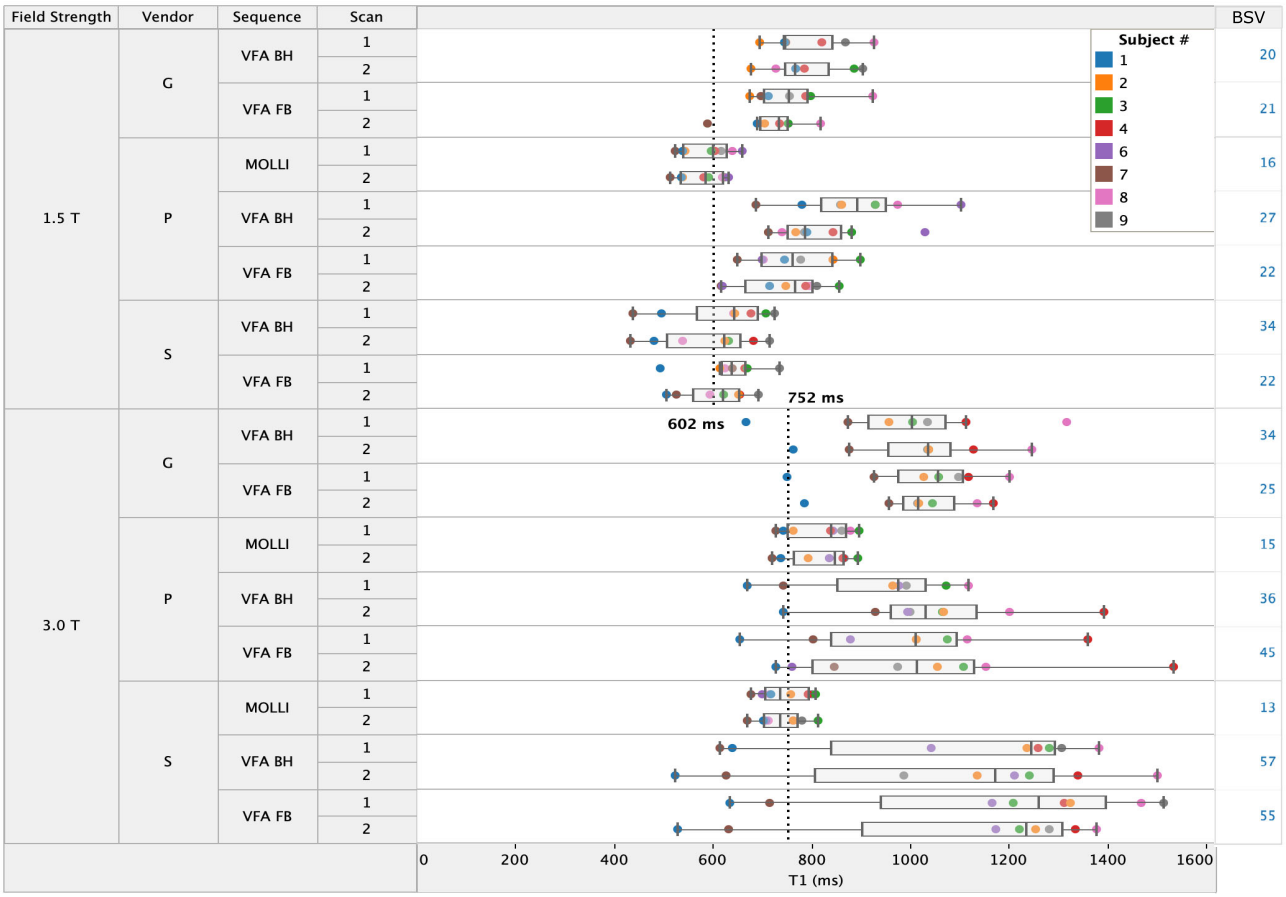


FIGURE 3: T_1 values in the normal liver at 1.5 T and 3.0 T. Each color-coded datapoint represents a single volunteer. Between-subject variation (BSV, expressed as a percentage) is given for each scanner and method on the right.

TABLE 6. The 95% CI of the mean T_1 values in the liver at 1.5 T and 3.0 T

	Literature reference	MOLLI	VFA BH	VFA FB
T_1 at 1.5 T (msec)	602	586 ± 33	780 ± 91	705 ± 39
T_1 at 3.0 T (msec)	752	782 ± 36	1036 ± 148	1053 ± 147

MOLLI, modified Look-Locker imaging; VFA, variable flip angle.

values are recorded for subject 4 consistently across all scanners and methods.

Repeatability and reproducibility of VFA T_1 in the liver as measured in this study are comparable to other static organs.[8–10] The repeatability RE in liver (10% ± 2%) is in fact at the lower end of published results in brain, breast, and prostate (10%–16%). The reproducibility RE in the liver (29% ± 7% at 1.5 T and 25% ± 4% at 3.0 T) is higher than previous studies (7%–20% at 3.0 T). These comparisons should be interpreted with caution as published studies are positively biased, using a narrower range of scanners (one field strength and no more than two vendors). Indeed, restricting the liver T_1 reproducibility to pairs of vendors improves the reproducibility RE to 10%–12% for the best aligned pair of vendors (G–P), well within the range of previous studies in other organs.

Unlike the repeatability and reproducibility studies on VFA T_1 mapping in the other organs, liver VFA T_1 values obtained in this study were acquired without corrections for B_1^+ effects, imperfect spoiling, MT effects, or other confounders. Correcting for these effects may improve the reproducibility, but evidence for this is limited, especially in view of the indication above that B_1 errors themselves may be reproducible. While previous multivendor studies in brain, breast, and prostate with B_1^+ corrections showed improved

reproducibility relative to an earlier phantom study, these studies also used a narrower sample of vendors, scanners, or field strengths. This potentially produced an optimistic assessment of repeatability and reproducibility. Only one multivendor study, in the prostate, assessed the impact of B_1^+ correction directly and found the improvement to be modest, improving the reproducibility RE from 18% to 14%.[10] This is consistent with our observation that results in the absence of B_1^+ correction are in the range of B_1^+ corrected results in other organs. Hence, the room for improvement in reproducibility using standard B_1^+ correction methods may be limited. However, the data in this study suggest that B_1^+ correction may have a significant impact on the overall bias and accuracy of the measurements on single-subject level. In the liver, B_1^+ inhomogeneities of 0.4–1.3 (ratio of actual to prescribed flip angle) have been reported at 3.0 T.[34] For the literature, T_1 value of 752 msec at 3.0 T, assuming a TR = 3.5 msec, bias in VFA T_1 measurements between −84% and 70% can be expected, which are much larger than the bias estimates obtained in this study.

A separate issue is that fast and validated B_1^+ correction methods are not routinely available on all clinical scanners and are therefore of limited use in clinical trials or clinical practice today. Indeed, in its recent profile revision, the QIBA DCE-MRI Biomarker Committee has specifically not included B_1^+ mapping as a requirement for VFA T_1 mapping due to "the dearth of literature and lack of access to vendor-specific B_1^+ mapping sequences" (QIBA DCE-MRI BC Call Summary, 21 Dec 2020, [http://qibawiki.rsna.org/images/d/d7/2020_12-21_QIBA_DCE-MRI_BC_Call_Summary-FINAL.pdf]). The committee also highlighted the lack of test–retest data on the effects of B_1^+ corrections on T_1 measurements in routine VFA T_1 mapping. This is crucial because B_1^+ corrections may themselves be subject to measurement error.[35] Indeed, there have been reports of exacerbation of B_1^+ nonuniformity in some vendor-provided maps[36] and increase in bias in T_1 values after inline B_1^+ corrections.[37] These observations indicate the importance of robust and accurate B_1^+ correction, which may come at a substantial cost in acquisition time.

Comparison with MOLLI indicates that there is significant room to improve on the accuracy of VFA and supports the common assumption that the faster acquisition afforded by VFA comes at a cost of accuracy, repeatability, and reproducibility. It may be likely that the differences between VFA and MOLLI as reported in this study are overestimated; as the MOLLI sequence was only available on three of the six scanners, its reproducibility is likely to be lower in a more representative population of scanners. However, repeatability of the MOLLI T_1 in this study is very consistent with literature using LL methods, while the repeatability RE of liver VFA T_1 is substantially higher.[32]

On the effect of field strength, the only significant differences between 1.5 T and 3.0 T are an improved reproducibility at 3.0 T for the BH VFA sequence, but a larger between-subject variability and spatial heterogeneity for the VFA FB. Hence, it appears the optimal field strength in terms of reproducibility is sequence specific, with 3.0 T preferred for the BH sequence and 1.5 T preferred for the FB sequence. Considering the results of individual vendors separately indicates that the optimal field strength is also vendor specific. In all three vendors, the between-subject variation increased at 3.0 T in line with subject-specific errors caused by B1-effects. In vendor S, the mean bias is larger at 3.0 T, whereas for vendors P and G it is comparable. On the whole, this indicates a preference for 1.5 T when using uncorrected VFA in view of the smaller between-subject variability and spatial heterogeneity.

On the effect of vendor, including vendor S in a study increases the reproducibility RE substantially relative to studies that include vendors G and P only. This observation remains valid when field strengths are considered separately. Between the other vendors G and P, results are comparable, the only distinguishing feature being a lower spatial heterogeneity for vendor G at 3.0 T for the VFA FB. On the other hand, dedicated sequence development was needed to enable this study on vendor G, unlike the other vendors where product sequences were available. A different picture emerges when considering the field strengths separately. Unlike at 3.0 T, vendor S has the lowest bias at 1.5 T, showing a systematic error that is substantially smaller than vendors G and P. The repeatability RE for vendor S at 1.5 T is also lower than that of vendors G and P, though the differences are smaller. This indicates that vendor S reduces the reproducibility at 1.5 T only because it has a substantially lower bias—illustrating the limitation of using reproducibility measures alone to characterize an imaging biomarker assay.[38]

To test the effect of sequence optimization on the performance of VFA T_1 mapping, we included two sequences in the study that were optimized in different ways: a BH sequence at high spatial and low temporal resolution, and an FB sequence at low spatial and high temporal resolution. Reproducibility RE of the FB sequence was lower at 1.5 T and higher at 3.0 T compared to the BH sequence. And other than an improved spatial heterogeneity of the FB sequence, all remaining parameters were comparable between the sequences, indicating that the details of sequence optimization do not fundamentally impact on the accuracy of the measurement. Hence, the choice of sequence settings can be based on other criteria, such as the need for high spatial detail or a desire to avoid breath holds in frail patient populations.

Study Limitations

Only one scanner of each vendor and each field strength was available for this multivendor study, and therefore, we are not

able to test the effect of the variability induced by using two different scanners of the same make and model. Naturally, the small sample size of eight healthy volunteers is a limitation and has reduced the study's power to detect more subtle differences in means. Finally, bias in VFA T_1 in the liver could not be determined due to the lack of a true reference measurement in the study. Hence, we provided an estimated bias using a literature value. This was close to the measured MOLLI values in our population, which provides some confidence that the literature value is close to the ground truth.

Liver T_1 measurements are known to be affected by fat, iron, and glycogen content.[39,40] In this healthy volunteer study, liver fat and iron levels were assumed to be normal. While this may cause bias in subjects where these assumptions are invalid, both fat and iron levels can be assumed to remain stable throughout the study period. Their effect on T_1 repeatability and reproducibility should, therefore, be minimal, which was the main focus of this study. Participants in this study were also not instructed to attend scans in consistent fasted or fed states, and variation in liver glycogen levels between scans can impact T_1 repeatability and reproducibility. However, variations in T_1 in healthy volunteers between fed and fasted states[40] have been reported to be within the same-day test–retest T_1 repeatability ranges reported in volunteers in a fasting state.[29] Therefore, the impact of variation in meal intake between scans is not expected to have a major impact on the repeatability and reproducibility REs obtained in this study.

Another limitation of this study is that the optimization of sequences on each scanner was not independent. Some sequence parameters such as FOV, spatial resolution, and FA were kept fixed; however, it was not possible to directly match parameters such as parallel imaging acceleration factors or phase oversampling across scanners. For example, differences in acquisition parameters such as TR are known to affect the sensitivity of VFA T_1 measurements. For a given T_1, the Ernst angle increases with increasing TR.[41] However, the range of TR values used in this study was 3.19–6.04 msec. For the literature values of liver T_1, the range of e^{-TR/T_1} is 0.99–0.996. The corresponding shift in the Ernst angle is 3°; therefore, the impact of the mismatch in TR to VFA T_1 reproducibility RE is expected to be negligible. While an effort was made to minimize differences in sequence implementation on the scanners, any remaining differences could have contributed to the vendor and field strength effects.

In this study, the MOLLI measurements were performed in the transverse orientation, while the VFA T_1 images were acquired coronally. VFA T_1 mapping in the liver is often used for the calibration of DCE-MRI signals. In such a liver MRI protocol, the VFA acquisition is required to match the DCE-MRI acquisition. While the current clinical norm is to acquire liver DCE-MRI in the transverse orientation, coronal acquisitions are preferred to avoid inflow effects in arterial input function measurements for pharmacokinetic modeling and to simplify motion correction. Therefore, in this study, the VFA T_1 acquisitions were performed coronally. On the other hand, the MOLLI is performed transverse as a standard for liver T_1 studies. In retrospect, coronal MOLLI acquisition could have allowed direct comparisons with VFA. However, no differences in renal T_1 values between coronal and axial acquisitions were found in other studies.[42] Therefore, the impact of different acquisition orientations of VFA and MOLLI T_1 comparisons is not expected to be large.

On vendor G, the sequence was additionally modified at a sequence programming level in order to enable scans of all flip angles to run consecutively, without phase-wrap artifacts in the coronal acquisition with a large field of view, and no change in receiver gain between flip angles, whereas in vendors S and P only the sequence parameters were optimized. This may also have created a bias in favor of vendor G. Finally, the sequences were first set up and tested on a single vendor and a single field strength and subsequently translated to others. It is plausible that this has created a bias in favor of the reference scanner (vendor S, 3.0 T).

Conclusion

Bias, repeatability, and reproducibility of VFA T_1 mapping in the liver in a multivendor setting are similar to those reported in breast, prostate, and brain. The numerical values reported in this study can serve as benchmarks against which any future improvements of VFA T_1 mapping in the liver can be qualified.

References

1. Sun H, Cleary JO, Glarin R, Kolbe SC. Extracting more for less: Multi-echo MP2RAGE for simultaneous T_1-weighted imaging, T_1 mapping, mapping, SWI, and QSM from a single acquisition. Magn Reason Med 2020;83:1178-1191.
2. Li Z, Fu Z, Keerthivasan M, Bilgin A, Johnson K. Rapid high-resolution volumetric T_1 mapping using a highly accelerated stack-of-stars look locker technique. Magn Reson Imaging 2021;79:28-37.
3. Huang L, Neji R, Nazir MS, Whitaker J. FASt single-breathhold 2D multislice myocardial T_1 mapping ($FAST_1$) at 1.5T for full left ventricular coverage in three breathholds. J Magn Reson 2020;51:492-504.
4. Schmaranzer F, Afacan O, Lerch TD, Kim YJ. Magnetization-prepared 2 rapid gradient-Echo MRI for B1 insensitive 3D T_1 mapping of hip cartilage: An experimental and clinical validation. Radiology 2021;299: 150-158.
5. Stikov N, Boudreau M, Levesque IR, Tardif CL, Barral JK, Pike GB. On the accuracy of T_1 mapping: Searching for common ground. Magn Reson Med 2015;73:514-522.
6. Tirkes T, Zhao X, Lin C, et al. Evaluation of variable flip angle, MOLLI, SASHA, and IR-SNAPSHOT pulse sequences for T_1 relaxometry and extracellular volume imaging of the pancreas and liver. Magn Reson Mater Phys Biol Med 2019;32:559-566.
7. Roujol S, Weingärtner S, Foppa M, et al. Accuracy, precision, and reproducibility of four T_1 mapping sequences: A head-to-head comparison of MOLLI, ShMOLLI, SASHA, and SAPPHIRE. Radiology 2014;272: 683-689.

8. Leutritz T, Seif M, Helms G, et al. Multiparameter mapping of relaxation (R1, R2*), proton density and magnetization transfer saturation at 3 T: A multicenter dual-vendor reproducibility and repeatability study. Hum Brain Mapp 2020;41:4232-4247.
9. Sorace AG, Wu C, Barnes SL, et al. Repeatability, reproducibility, and accuracy of quantitative mri of the breast in the community radiology setting. J Magn Reson Imaging 2018;48:695-707.
10. Zhong X, Shakeri S, Liu D, Sayre J. Repeatability and reproducibility of variable flip angle T_1 quantification in the prostate at 3 T. J Magn Reson 2019;49:1730-1735.
11. AG Teixeira RP, Malik SJ, Hajnal JV. Fast quantitative MRI using controlled saturation magnetization transfer. Magn Reson Med 2019;81: 907-920.
12. Bane O, Hectors SJ, Wagner M. Accuracy, repeatability, and interplatform reproducibility of T_1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magn Reson Med 2018;79:2564-2575.
13. Yoon JH, Lee JM, Kim E, Okuaki T, Han JK. Quantitative liver function analysis: Volumetric T_1 mapping with fast multisection B_1 inhomogeneity correction in hepatocyte-specific contrast-enhanced liver MR imaging. Radiology 2017;282:408-417.
14. Theilig D, Elkilany A, Schmelzle M, et al. Consistency of hepatocellular gadoxetic acid uptake in serial MRI examinations for evaluation of liver function. Abdom Radiol 2019;44:2759-2768.
15. Boudreau M, Tardif CL, Stikov N. B_1 mapping for bias-correction in quantitative T_1 imaging of the brain at 3T using standard pulse sequences. J Magn Reson Imaging 2017;46:1673-1682.
16. Sung K, Daniel BL, Hargreaves BA. Transmit B1+ field inhomogeneity and T1 estimation errors in breast DCE-MRI at 3 tesla: B1+ and T1 estimation in breast DCE-MRI at 3T. J Magn Reson Imaging 2013;38: 454-459.
17. 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.
18. Xu X, Zhu H, Li R, et al. Whole-liver histogram and texture analysis on T1 maps improves the risk stratification of advanced fibrosis in NAFLD. Eur Radiol 2020;31:1748-1759.
19. Li Z, Sun J, Hu X, et al. Assessment of liver fibrosis by variable flip angle T1 mapping at 3.0T. J Magn Reson Imaging 2016;43:698-703.
20. 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.
21. Georgiou L, Penny J, Nicholls G, et al. Quantitative Assessment of liver function using gadoxetate-enhanced magnetic resonance imaging: Monitoring transporter-mediated processes in healthy volunteers. Invest Radiol 2017;52:111-119.
22. Kamimura K, Fukukura Y, Yoneyama T, et al. Quantitative evaluation of liver function with T_1 relaxation time index on Gd-EOB-DTPA-enhanced MRI: Comparison with signal intensity-based indices. J Magn Reson Imaging 2014;40:884-889.
23. Kim KA, Park M-S, Kim I-S, et al. Quantitative evaluation of liver cirrhosis using T_1 relaxation time with 3 tesla MRI before and after oxygen inhalation. J Magn Reson Imaging 2012;36:405-410.
24. Faller TL, Trotier AJ, Miraux S, Ribot EJ. Radial MP2RAGE sequence for rapid 3D T_1 mapping of mouse abdomen: Application to hepatic metastases. Eur Radiol 2019;29:5844-5851.
25. Steudel A, Träber F, Krahe T, Schiffmann O, Harder T. Qualitätskontrolle der quantitativen MR-Tomographie: In-vitro-und In-vivo-Überprüfung von Relaxationszeitmessungen. RöFo - Fortschritte Auf Dem Gebiet der Röntgenstrahlen Und der Nuklearmedizin 1990; 152:673-676.
26. Banerjee R, Pavlides M, Tunnicliffe EM, et al. Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J Hepatol 2014;60:69-77.
27. Bradley CR, Cox EF, Scott RA, et al. Multi-organ assessment of compensated cirrhosis patients using quantitative magnetic resonance imaging. J Hepatol 2018;69:1015-1024.
28. Li J, Liu H, Zhang C, et al. Native T_1 mapping compared to ultrasound elastography for staging and monitoring liver fibrosis: An animal study of repeatability, reproducibility, and accuracy. Eur Radiol 2020;30: 337-345.
29. Bachtiar V, Kelly MD, Wilman HR, et al. Repeatability and reproducibility of multiparametric magnetic resonance imaging of the liver. PLoS One 2019;14:e0214921.
30. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP. Modified look-locker inversion recovery (MOLLI) for high-resolution T_1 mapping of the heart. Magn Reson Med 2004;52:141-146.
31. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans Med Imaging 1999;18:712-721.
32. Waterton JC. Survey of water proton longitudinal relaxation in liver in vivo. Magma 2021;34:779-789.
33. Keenan KE, Gimbutas Z, Dienstfrey A, et al. Multi-site, multi-platform comparison of MRI T_1 measurement using the system phantom. PLoS One 2021;16:e0252966.
34. Roberts NT, Hinshaw LA, Colgan TJ, Ii T, Hernando D, Reeder SB. B0 and B1 inhomogeneities in the liver at 1.5 T and 3.0 T. Magn Reson Med 2020;85:2212-2220.
35. Bliesener Y, Zhong X, Guo Y, et al. Radiofrequency transmit calibration: A multi-center evaluation of vendor-provided radiofrequency transmit mapping methods. Med Phys 2019;46:2629-2637.
36. Schwartz DL, Tagge I, Powers K, et al. Multisite reliability and repeatability of an advanced brain MRI protocol. J Magn Reson Imaging 2019;50:878-888.
37. Wang Y, Tadimalla S, Rai R, Goodwin J. Quantitative MRI: Defining repeatability, reproducibility and accuracy for prostate cancer imaging biomarker development. Magn Reson Imaging 2021;77:169-179.
38. Hockings P, Saeed N, Simms R, et al. MRI biomarkers. In: Seiberlich N, Gulani V, Calamante F, et al., editors. *Advances in magnetic resonance technology and applications*, Vol 1: Academic Press; 2020. p liii-lxxxvi. https://doi.org/10.1016/B978-0-12-817057-1.00002-0
39. Hoad CL, Palaniyappan N, Kaye P, et al. A study of T_1 relaxation time as a measure of liver fibrosis and the influence of confounding histological factors. NMR Biomed 2015;28:706-714.
40. Mózes FE, Valkovič L, Pavlides M, Robson MD, Tunnicliffe EM. Hydration and glycogen affect T_1 relaxation times of liver tissue. NMR Biomed 2021;34:e4530.
41. Deoni SCL, Rutt BK, Peters TM. Rapid combined T_1 and T_2 mapping using gradient recalled acquisition in the steady state. Magn Res Med 2003;49:515-526.
42. Graham-Brown MP, Singh A, Wormleighton J, et al. Association between native T_1 mapping of the kidney and renal fibrosis in patients with IgA nephropathy. BMC Nephrol 2019;20:1-7.