Repeat it without me: Crowdsourcing the T_1 mapping common ground via the ISMRM reproducibility challenge

Mathieu Boudreau[1,2] | Agah Karakuzu[1] | Julien Cohen-Adad[1,2,3,4,5] |
Ecem Bozkurt[6] | Madeline Carr[7,8] | Marco Castellaro[9] | Luis Concha[10] |
Mariya Doneva[11] | Seraina A. Dual[12] | Alex Ensworth[13,14] | Alexandru Foias[1] |
Véronique Fortier[15,16] | Refaat E. Gabr[17] | Guillaume Gilbert[18] |
Carri K. Glide-Hurst[19] | Matthew Grech-Sollars[20,21] | Siyuan Hu[22] |
Oscar Jalnefjord[23,24] | Jorge Jovicich[25] | Kübra Keskin[6] | Peter Koken[11] |
Anastasia Kolokotronis[13,26] | Simran Kukran[27,28] | Nam G. Lee[6] | Ives R. Levesque[13,29] |
Bochao Li[6] | Dan Ma[22] | Burkhard Mädler[30] | Nyasha G. Maforo[31,32] |
Jamie Near[33,34] | Erick Pasaye[10] | Alonso Ramirez-Manzanares[35] | Ben Statton[36] |
Christian Stehning[30] | Stefano Tambalo[25] | Ye Tian[6] | Chenyang Wang[37] |
Kilian Weiss[30] | Niloufar Zakariaei[38] | Shuo Zhang[30] | Ziwei Zhao[6] |
Nikola Stikov[1,2,39] | on behalf of the ISMRM Reproducible Research Study Group and
the ISMRM Quantitative MR Study Group

Correspondence
Nikola Stikov, Polytechnique Montréal,
500 Chem. De Polytechnique, Montréal,
QC H3T 1J4, Canada.
Email: nikola.stikov@polymtl.ca

Abstract

Purpose: T_1 mapping is a widely used quantitative MRI technique, but its tissue-specific values remain inconsistent across protocols, sites, and vendors. The ISMRM Reproducible Research and Quantitative MR study groups jointly launched a challenge to assess the reproducibility of a well-established inversion-recovery T_1 mapping technique, using acquisition details from a seminal T_1 mapping paper on a standardized phantom and in human brains.

Methods: The challenge used the acquisition protocol from Barral et al. (2010). Researchers collected T_1 mapping data on the ISMRM/NIST phantom and/or in human brains. Data submission, pipeline development, and analysis were conducted using open-source platforms. Intersubmission and intrasubmission comparisons were performed.

Results: Eighteen submissions (39 phantom and 56 human datasets) on scanners by three MRI vendors were collected at 3 T (except one, at 0.35 T). The mean coefficient of variation was 6.1% for intersubmission phantom measurements, and 2.9% for intrasubmission measurements. For humans, the intersubmission/intrasubmission coefficient of variation was 5.9/3.2% in the genu and 16/6.9% in the cortex. An interactive dashboard for data visualization was also developed: https://rrsg2020.dashboards.neurolibre.org.

Conclusion: The T_1 intersubmission variability was twice as high as the intrasubmission variability in both phantoms and human brains, indicating that the acquisition details in

Mathieu Boudreau and Agah Karakuzu contributed equally to this work.

For affiliations refer to page 1123

the original paper were insufficient to reproduce a quantitative MRI protocol. This study reports the inherent uncertainty in T_1 measures across independent research groups, bringing us one step closer to a practical clinical baseline of T_1 variations in vivo.

KEYWORDS

inversion recovery, open data, quantitative MRI, reproducibility, T_1 mapping

1 | INTRODUCTION

Significant challenges exist in the reproducibility of quantitative MRI.[1] Despite its promise of improving the specificity and reproducibility of MRI acquisitions, few quantitative MRI techniques have been integrated into clinical practice. Even the most fundamental MR parameters cannot be measured with sufficient reproducibility and precision across clinical scanners to pass the second of six stages of technical assessment for clinical biomarkers.[2–4] Half a century has passed since the first quantitative T_1 (spin–lattice relaxation time) measurements were first reported as a potential biomarker for tumors,[5] followed shortly thereafter by the first in vivo T_1 maps[6] of tumors, but there is still disagreement in reported values for this fundamental parameter across different sites, vendors, and measurement techniques.[7]

T_1 represents the time constant for recovery of the equilibrium longitudinal magnetization, and it is one of the fundamental MRI parameters.[8] T_1 values will vary depending on the molecular mobility and magnetic field strength.[9–11] Knowledge of the T_1 values for tissue is crucial for optimizing clinical MRI sequences for contrast and time efficiency[12–14] and to calibrate other quantitative MRI techniques.[15,16] Inversion recovery (IR)[17,18] is considered the gold standard for T_1 measurement due to its robustness against effects like B_1 inhomogeneity,[7] but its long acquisition times limit the clinical use of IR for T_1 mapping.[7] In practice, it is often used as a reference for validating other T_1 mapping techniques, such as variable flip-angle imaging (VFA),[19–21] Look-Locker,[22–24] and MP2RAGE.[25,26]

In ongoing efforts to standardize T_1 mapping methods, researchers have been actively developing quantitative MRI phantoms.[27] The International Society for Magnetic Resonance in Medicine (ISMRM) and the National Institute of Standards and Technology (NIST) collaborated on a standard system phantom,[28] which was subsequently commercialized (Premium System Phantom; CaliberMRI, Boulder, CO, USA). This phantom has since been used in large multicenter studies, such as Bane et al.,[29] which

concluded that acquisition protocols and field strength influence accuracy, repeatability, and interplatform reproducibility. Another NIST-led study[30] found no significant T_1 discrepancies among measurements using NIST protocols across 27 MRI systems from three vendors at two clinical field strengths.

The 2020 ISMRM reproducibility challenge[1] posed a slightly different question: Can an imaging protocol, independently implemented at multiple centers, consistently measure one of the fundamental MRI parameters (T_1)? To assess this, we proposed using IR on a standardized phantom (ISMRM/NIST system phantom) and the healthy human brain. Specifically, **this challenge explored whether the acquisition details provided in a seminal paper on T_1 mapping[31] is sufficient to ensure the reproducibility across independent research groups.** To evaluate reproducibility within the framework of this challenge, we explored whether the intersubmission variability in T_1 measurements is the same as intrasubmission variability.

2 | METHODS

2.1 | Phantom and human data

The challenge asked researchers with access to the ISMRM/NIST system phantom[28] (Premium System Phantom) to measure T_1 maps of the phantom's T_1 plate (Table 1). Researchers who participated in the challenge were instructed to record the temperature before and after scanning the phantom using the phantom's internal thermometer. Instructions for positioning and setting up the phantom were devised by NIST and were provided to researchers through the NIST website.[2] In brief, the instructions explained how to orient the phantom and how long the phantom should be in the scanner room before scanning to achieve thermal equilibrium.[3]

Researchers were also instructed to collect T_1 maps in healthy human brains and were asked to measure a single slice positioned parallel to the ante-

TABLE 1 Reference T_1 values of the $NiCl_2$ array of the standard system phantom (for both phantom versions) measured at 20°C and 3 T.

Sphere #	T_1 (ms)	
	Version 1	Version 2
1	1989 ± 1.0	1883.97 ± 30.32
2	1454 ± 2.5	1330.16 ± 20.41
3	984.1 ± 0.33	987.27 ± 14.22
4	706 ± 1.0	690.08 ± 10.12
5	496.7 ± 0.41	484.97 ± 7.06
6	351.5 ± 0.91	341.58 ± 4.97
7	247.13 ± 0.086	240.86 ± 3.51
8	175.3 ± 0.11	174.95 ± 2.48
9	125.9 ± 0.33	121.08 ± 1.75
10	89.0 ± 0.17	85.75 ± 1.24
11	62.7 ± 0.13	60.21 ± 0.87
12	44.53 ± 0.090	42.89 ± 0.44
13	30.84 ± 0.016	30.40 ± 0.62
14	21.719 ± 0.0054	21.44 ± 0.31

Note: Phantoms with serial numbers 0042 or less are referred to as "Version 1," and those 0043 or greater are "Version 2."

rior commissure–posterior commissure (AC-PC) line. Before imaging, the subjects consented[4] to share their de-identified data with the challenge organizers and on the Open Science Framework (OSF.io) website. As the submitted data were a single slice, the researchers were not instructed to de-face the data of their imaging subjects. Researchers submitting human data provided written confirmation to the organizers that their data were acquired in accordance with their institutional ethics committee (or equivalent regulatory body) and that the subjects had consented to data sharing as outlined in the challenge.

2.2 | MRI acquisition protocol

Researchers followed the IR T_1 mapping protocol optimized for the human brain as described in the paper published by Barral et al.,[31] which used $TR = 2550$ ms, $TI = 50, 400, 1100$ and 2500 ms, $TE = 14$ ms, 2-mm slice thickness, and 1×1 mm^2 in-plane resolution. Note that this protocol is not suitable for fitting models that assume $TR > 5\,T_1$. Instead, the more general Barral et al.[31] fitting model described in Section 2.4 can be used, and this model is compatible with both magnitude-only and complex data. Researchers were instructed to closely adhere to this protocol and report any deviations due to technical limitations.

2.3 | Data submissions

Data submissions for the challenge were handled through a GitHub repository (https://github.com/rrsg2020/data _submission), enabling a standardized and transparent process. All data sets were converted to the NIfTI format, and images for all TIs were concatenated into a single NIfTI file. Each submission included a YAML file to store additional information (submitter details, acquisition details, and phantom or human subject details). Submissions were reviewed[5], and following acceptance, the data sets were uploaded to OSF.io (osf.io/ywc9g/). A Jupyter Notebook[32,33] pipeline using qMRLab[34,35] was used to process the T_1 maps and to conduct quality control checks. MyBinder links to Jupyter Notebooks that reproduced each T_1 map were shared in each submission's GitHub issue to easily reproduce the results in web browsers while maintaining consistent computational environments. Eighteen submissions were included in the analysis, which resulted in 39 T_1 maps of the NIST/system phantom and 56 brain T_1 maps. Figure 1 illustrates all the submissions that acquired phantom data (Figure 1A) and human data (Figure 1B), the respective MRI scanner vendors, and the resulting T_1 mapping data sets. Some submissions included measurements in which both complex and magnitude-only data from the same acquisition were used to fit T_1 maps; thus, the total number of unique acquisitions is lower than the numbers reported previously (27 for phantom data and 44 for human data). The data sets were collected on systems from three MRI manufacturers (Siemens, GE, and Philips) and were acquired at 3T[6], except for one data set acquired at 0.35 T (the ViewRay Mridian MR-linac).

2.4 | Fitting model and pipeline

A reduced-dimension nonlinear least-squares approach was used to fit the complex general IR signal equation as follows:

$$S(TI) = a + be^{-\frac{TI}{T_1}} \tag{1}$$

where a and b are complex constants. This approach, developed by Barral et al.,[31] offers a model for the general T_1 signal equation without relying on the long-TR approximation. The a and b constants inherently factor TR in them, as well as other imaging parameters such as excitation flip angle, inversion-pulse flip angles, TR, TE, TI, and a constant that has contributions from T_2 and the receive coil sensitivity. Barral et al.[31] shared their *MATLAB* (Math-Works, Natick, MA, USA) code for the fitting algorithm used in their paper[7]. Magnitude-only data were fitted to a modified version of Eq. (1) (Eq. [15] of Barral et al.[31])
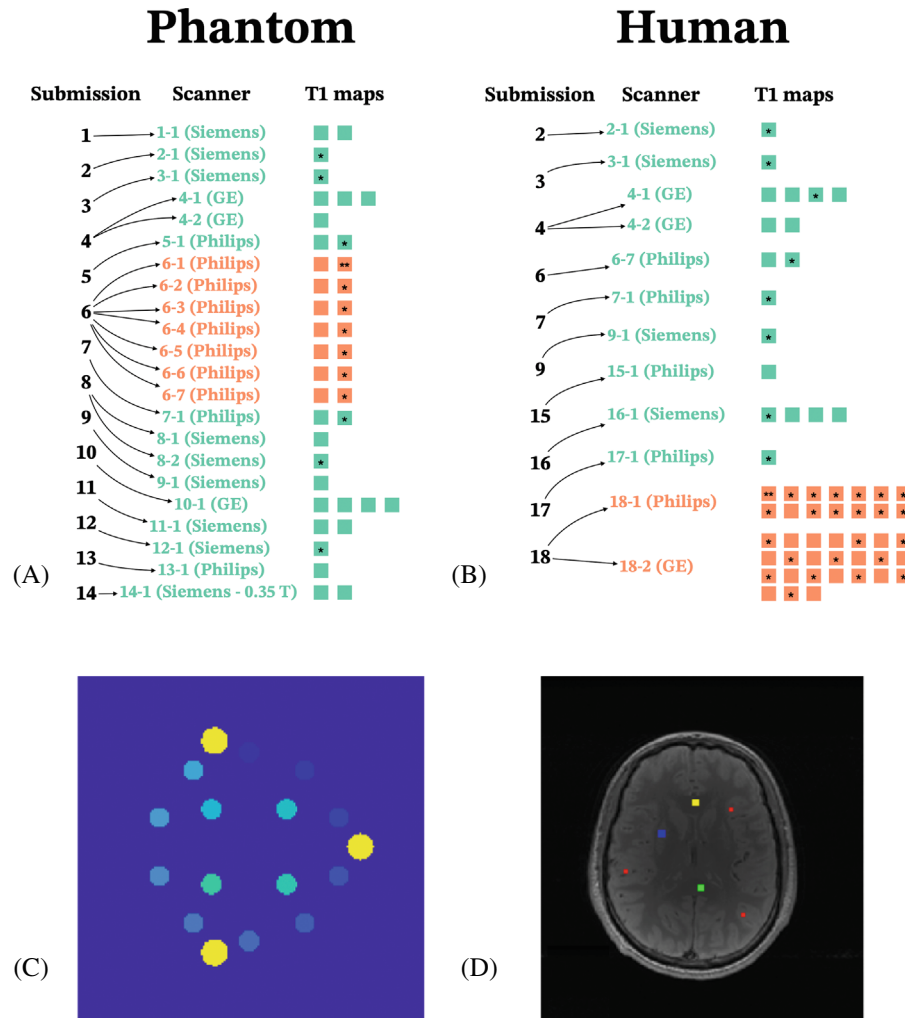
F I G U R E 1 List of the data sets submitted to the challenge. (A) Submissions that included phantom data. (B) Submissions that included human brain data. For the phantom (A), each submission acquired its data using a single phantom, but some researchers shared the same physical phantom with each other. Green indicates submissions used for intersubmission analyses, and orange indicates the sites used for intrasubmission analyses. T_1 maps used in the calculations of intersubmission (*green*) and intrasubmission (*orange*) coefficients of variation are indicated with asterisks. A more detailed figure can be found in Figure S1A. Images (C) and (D) illustrate the region-of-interest (ROI) choice in phantoms and humans.

with signal-polarity restoration by finding the signal minima, fitting the IR curve for two cases (data points for $TI < TI_{minimum}$ flipped, and data points for $TI \leq TI_{minimum}$ flipped), and selecting the case that resulted in the best fit based on minimizing the residual between the model and the measurements[8]. This code is available as part of the *qMRLab* open-source software,[34,35] which provides a standardized application program interface to call the fitting in *MATLAB/Octave* scripts.

A data-processing pipeline was written using *MATLAB/Octave* in a Jupyter Notebook. This pipeline downloads every data set from OSF (osf.io/ywc9g/), loads its configuration file, fits the T_1 maps, and then saves them to NifTI and PNG formats. The code is available on GitHub (https://github.com/rrsg2020/t1_fitting_pipeline, filename: RRSG_T1_fitting.ipynb). Finally, T_1 maps were manually uploaded to OSF (https://osf.io/ywc9g/).

2.5 | Image labeling and registration

The T_1 plate (NiCl$_2$ array) of the phantom has 14 spheres that were labeled as the regions of interest (ROIs) using

a numerical mask template created in *MATLAB*, provided by NIST researchers (Figure 1C). To avoid potential edge effects in the T_1 maps, the ROI labels were reduced to 60% of the expected sphere diameter. A registration pipeline in *Python* using the Advanced Normalization Tools (ANTs)[36] was developed and shared in the analysis repository of our GitHub organization (https://github.com/rrsg2020/analysis, filename: register_t1maps_nist.py, commit ID: 8d38644). Briefly, a label-based registration was first applied to obtain a coarse alignment, followed by an affine registration (gradientStep: 0.1, metric: cross correlation, number of steps: 3, iterations: 100/100/100, smoothness: 0/0/0, subsampling: 4/2/1) and a BsplineSyN registration (gradientStep:0.5, meshSizeAtBaseLevel:3, number of steps: 3, iterations: 50/50/10, smoothness: 0/0/0, subsampling: 4/2/1). The ROI label template was nonlinearly registered to each T_1 map uploaded to OSF.

For the human data, manual ROIs were segmented by a single researcher (M.B., 12+ years of neuroimaging experience) using FSLeyes[37] in four regions (Figure 1D), located in the genu, splenium, deep gray matter, and cortical gray matter. Automatic segmentation was not used because

the data were single-slice and there was inconsistent slice positioning between datasets.

2.6 | Analysis and statistics

Analysis code and scripts were developed and shared in a version-controlled public GitHub repository[9]. The T_1 fitting and data analysis were performed by M.B., one of the challenge organizers. Computational environment requirements were containerized in *Docker*[38,39] to create an executable environment that allows for analysis reproduction in a web browser via *MyBinder*[10],[40] Backend *Python* files handled reference data, database operations, ROI masking, and general analysis tools. Configuration files handled data-set information, and the data sets were downloaded and pooled using a script (*make_pooled_datasets.py*). The databases were created using a publicly available Jupyter Notebook script and subsequently saved in the repository.

The mean T_1 values of the ISMRM/NIST phantom data for each ROI were compared with temperature-corrected reference values and visualized in three different types of plots (linear axes, log–log axes, and error relative to the reference value). Temperature correction involved nonlinear interpolation[11] of a NIST reference table of T_1 values for temperatures ranging from 16°C to 26°C (2°C intervals) as specified in the phantom's technical specifications. For the human data sets, the mean and SDs for each tissue ROI were calculated from all submissions across all sites. Two of the submissions (one of phantom data [Submission 6 in Figure 1A] and one of human data [Submission 18 in Figure 1B]) were much larger than the others, because they included multiple acquisitions. Submission 6 consisted of data from one traveling phantom acquired at seven Philips 3T imaging sites, and Submission 18 was a large cohort of volunteers who were imaged on two 3T scanners, one GE, and one Philips. These data sets (identified in orange in Figures 1, 3, and 4) were used to calculate intrasubmission coefficients of variation (CoVs) (one per scanner/volunteer, identified by asterisks in Figure 1A,B), and intersubmission CoVs were calculated using one T_1 map from each of these (orange) along with one from all other submissions[12] (identified as green in Figures 1, 3, and 4; the T_1 maps used in those CoV calculations are also indicated with asterisks in Figure 1A,B). All quality assurance and analysis plot images were stored in the repository. Additionally, the database files of ROI values and acquisition details for all submissions were also stored in the repository.

2.7 | Dashboard

To widely disseminate the challenge results, a web-based dashboard was developed (Figure 2, https://rrsg2020.dashboards.neurolibre.org). The landing page (Figure 2A) showcases the relationship between the phantom and brain data sets acquired at different sites/vendors. Selecting the icons labeled as "phantom" or "in vivo" and then clicking a ROI will display whisker plots for that region. Additional sections of the dashboard allow for displaying statistical summaries for both sets of data: a magnitude versus complex data fitting comparison, and hierarchical shift function analyses.

3 | RESULTS

Figure 3 presents a comprehensive overview of the challenge results through violin plots, depicting intersubmission and intrasubmission comparisons in both phantoms (A) and human (B) data sets. For the phantom (Figure 3A), the average intersubmission CoV for the T_1 values in the human brain (Spheres 1–5, approximately 500 to 2000 ms) was 6.1%. By addressing outliers from two sites associated with specific challenges for Sphere 4 (signal null near a TI), the mean intersubmission CoV was reduced to 4.1%. One participant (Submission 6, Figure 1) measured T_1 maps using a consistent protocol at seven different sites, and the mean intrasubmission CoV across the first five spheres for this submission was calculated to be 2.9%.

For the human data sets (Figure 3B), intersubmission CoVs for independently implemented imaging protocols were 5.9% for genu, 10.6% for splenium, 16% for cortical gray matter (GM), and 22% for deep GM. One participant (Submission 18, Figure 1) measured a large data set (13 individuals) on three scanners and two vendors, and the intrasubmission CoVs for this submission were 3.2% for genu, 3.1% for splenium, 6.9% for cortical GM, and 7.1% for deep GM. The binomial appearance for the splenium, deep GM, and cortical GM for the sites used in the intersubmission analyses (green) can be explained by an outlier measurement, which can be seen in Figure 4E-G (Submission 3.001).

A scatterplot of the T_1 data for all submissions and their ROIs is shown in Figure 4 (phantom [A–C] and human brains [D–G]). The NIST phantom T_1 measurements are presented in each plot for different axes types (linear, log, and error) to better visualize the results. Figure 4A shows good agreement for this data set in comparison with the temperature-corrected reference T_1 values. However, this trend did not persist for low T_1 values ($T_1 < 100$–200 ms), as seen in the log–log plot (Figure 4B),

FIGURE 2 Dashboard. (A) Welcome page listing all the sites, the scan type (phantom/brain), the scanner vendor, and the corresponding site. (B) Phantom tab for a selected region of interest (ROI). (C) In vivo tab for a selected ROI. Link: https://rrsg2020.dashboards.neurolibre.org. GM, gray matter.

which was expected because the imaging protocol is optimized for human-brain T_1 values ($T_1 > 500$ ms). Higher variability is seen for long T_1 values ($T_1 \sim 2000$ ms) in Figure 4A. Errors exceeding 10% are observed in the phantom spheres with T_1 values below 300 ms (Figure 4C), and 3–4 measurements with outlier values exceeding 10%

error were observed in the human brain tissue range ($\sim$500–2000 ms).

Figure 4D–F displays the scatter plot data for human data sets submitted to this challenge, showing mean and SD T_1 values for the white matter (WM; genu and splenium) and GM (cerebral cortex and deep GM) ROIs. Mean

FIGURE 3 Summary of results of the challenge as violin plots displaying the intersubmission and intrasubmission comparisons for phantoms (A) and human brains (B). Green indicates submissions used for intersubmission analyses, and orange indicates the sites used for intrasubmission analyses. Interactive figure available at: https://preprint .neurolibre.org/10.55458/neurolibre .00023/. cGM, cortical gray matter; GM, gray matter.
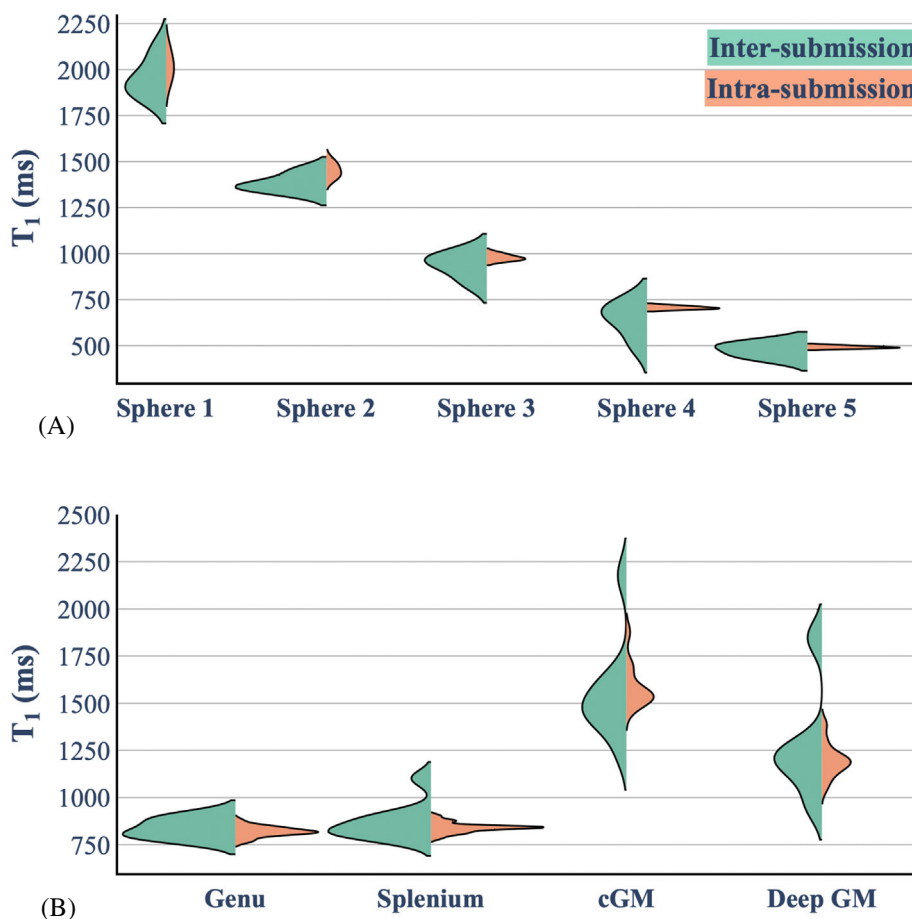
WM T_1 values across all submissions were 828 ± 38 ms in the genu and 852 ± 49 ms in the splenium, and mean GM T_1 values were 1548 ± 156 ms in the cortex and 1188 ± 133 ms in the deep GM, with less variations overall in WM compared with GM, possibly due to better ROI placement and less partial voluming in WM. The lower SDs for the ROIs of human database ID site 9 (Submission 18 in Figure 1, and seen in orange in Figure 4D–G) are due to good slice positioning, cutting through the AC-PC line and the genu for proper ROI placement, particularly for the corpus callosum and deep GM.

4 | DISCUSSION

This challenge explored whether different research groups could reproduce T_1 maps based on the protocol information reported in a seminal publication.[31] Eighteen submissions independently implemented the IR T_1 mapping acquisition protocol as outlined in Barral et al.,[31] and reported T_1 mapping data in a standard quantitative MRI phantom and/or human brains at 27 MRI sites, using systems from three different vendors (GE, Philips, and Siemens). The collaborative effort produced

an open-source database of 95 T_1 mapping data sets, including 39 ISMRM/NIST phantom and 56 human-brain data sets. The intersubmission variability was twice as high as the intrasubmission variability in both phantom and human-brain T_1 measurements, **demonstrating that acquisition details communicated via a paper are not sufficient for reproducing quantitative MRI measurements**. This study reports the inherent uncertainty in T_1 measures across independent research groups, which brings us one step closer to producing a practical baseline of variations for this metric.

Overall, our approach did show improvement in the reproducibility of T_1 measurements in vivo compared with researchers implementing T_1 mapping protocols completely independently (i.e., with no central guidance), as literature T_1 values in vivo vary more than reported here (e.g., Bojorquez et al.[41] reports that reported T_1 values in WM vary between 699 and 1735 ms in published literature). We were aware that coordination was essential for a quantitative MRI challenge, which is why the protocol specifications we provided to researchers were more detailed than any public guidelines for quantitative MRI that were available at the time. Yet, even in combination with the same T_1 mapping processing tools, this level of
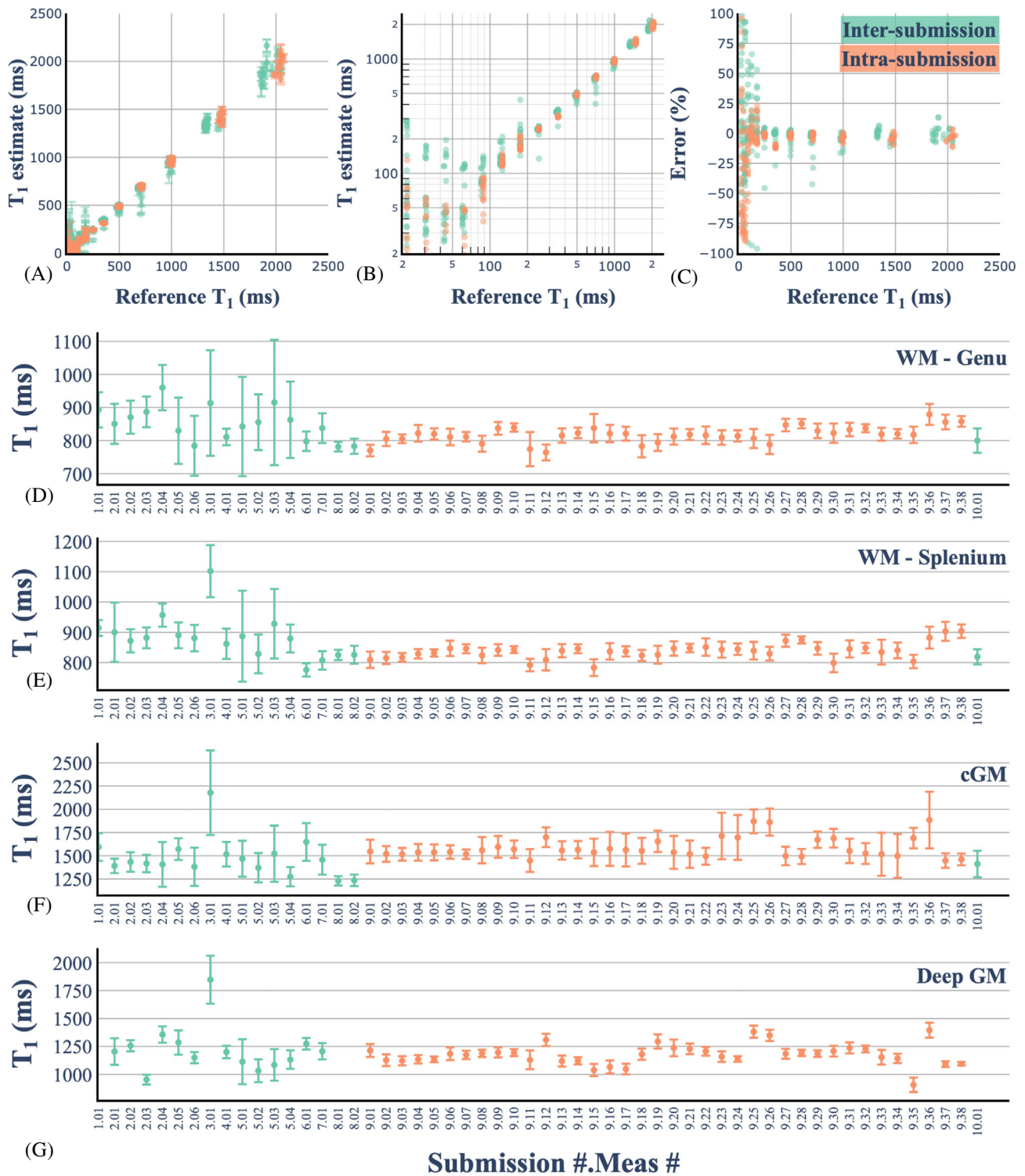
FIGURE 4 Measured mean T_1 values versus temperature-corrected NIST reference values of the phantom spheres are presented as linear plots (A), log–log plots (B), and plots of the error relative to the reference T_1 value (C). Green indicates submissions used for intersubmission analyses, and orange indicates the sites used for intrasubmission analyses. The dashed lines in (C) represent a ± 10% error. Mean T_1 values in two sets of regions of interest (ROIs), white matter (one 5 × 5 voxel ROI for genu, one 5 × 5 voxel ROI for splenium) and gray matter (GM; three 3 × 3 voxel ROIs for cortex, one 5 × 5 voxel ROI for deep GM). (G) The missing datapoints for deep GM for Submissions 1, 8, and 10 were due to the slice positioning of the acquisition not containing deep GM. Interactive figure available at: https://preprint.neurolibre.org/10.55458/neurolibre.00023/. cGM, cortical gray matter.

description (a paper + post-processing tools) leaves something to be desired.

This analysis highlights that more information is needed to unify all the aspects of a pulse sequence across sites, beyond what is routinely reported in a scientific publication. However, in a vendor-specific setting, this is a major challenge, given the disparities between proprietary development libraries.[42] Vendor-neutral pulse sequence design platforms[43–45] have emerged as a powerful solution to standardize sequence components at the implementation level (e.g., RF pulse shape, gradients). Vendor neutrality has been shown to significantly reduce the variability of T_1 maps acquired using VFA across vendors.[45] In the absence of a vendor-neutral framework, a vendor-specific alternative is the implementation of a strategy to control the saturation of magnetization transfer across TRs.[46] Nevertheless, this approach can still benefit from a vendor-neutral protocol to enhance accessibility and unify implementations. This is because vendor-specific constraints are known to impose limitations on the adaptability of sequences, resulting in significant variability even when implementations are closely aligned within their respective vendor-specific development environments.[47]

After reflecting on our reproducibility challenge design, we believe there are some improvements that would give additional insights if the challenge was to be repeated in the future. One major addition would be to distribute (1) a full T_1 mapping protocol file that can be imported on the scanners (matched as closely as possible for each vendor) and (2) a vendor-neutral sequence file (e.g., using Pulseq,[43] Gammastar,[44] or RTHawk[45]), assuming sufficient sites would have the setup to use it. It would also be important to standardize the image reconstruction and postprocessing of the acquired data; this could be done using open tools such as Gadgetron[48] or BART.[49] However, this would require the authors to submit raw k-space data, which would substantially increase the dataset sizes and complicate the transfer and storage of the submissions. These two additions (matched full protocols and vendor-neutral sequences) would provide further information on how much each component of the scanner-to-T_1 map pipeline contributes to the variation across independent sites. Another change to the challenge framework could be to substitute (or supplement) the IR T_1 mapping protocol with a technique that is used more widely in practice (i.e., a rapid and 3D technique), such as MP2RAGE[25] or VFA/DESPOT1.[19–21] However, these protocols have greater B_1 sensitivity,[26,50] requiring an additional B_1 mapping protocol to be established and distributed to the researchers.

The 2020 Reproducibility Challenge, jointly organized by the Reproducible Research and Quantitative MR ISMRM study groups, led to the creation of a large open database of standard quantitative MR phantom and human-brain IR T_1 maps. These maps were measured using independently implemented imaging protocols on MRI scanners from three different manufacturers. All collected data, processing pipeline code, computational environment files, and analysis scripts were shared with the goal of promoting reproducible research practices, and an interactive dashboard was developed to broaden the accessibility and engagement of the resulting data sets (https://rrsg2020.dashboards.neurolibre.org). The differences in stability between independently implemented (intersubmission) and centrally shared (intrasubmission) protocols observed both in phantoms and in vivo could help inform future meta-analyses of quantitative MRI metrics[51,52] and better guide multicenter collaborations.

By providing access and analysis tools for this multicenter T_1 mapping data set, we aim to provide a benchmark for future T_1 mapping approaches. We also hope that this data set will inspire new acquisition, analysis, and standardization techniques that address non-physiological sources of variability in T_1 mapping. This could lead to more robust and reproducible quantitative MRI and ultimately better patient care.

AFFILIATIONS

[1]NeuroPoly Lab, Polytechnique Montréal, Montréal, Quebec, Canada

[2]Montreal Heart Institute, Montréal, Quebec, Canada

[3]Unité de Neuroimagerie Fonctionnelle, Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Quebec, Canada

[4]Mila–Quebec AI Institute, Montréal, Québec, Canada

[5]Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada

[6]Magnetic Resonance Engineering Laboratory, University of Southern California, Los Angeles, California, USA

[7]Medical Physics, Ingham Institute for Applied Medical Research, Liverpool, Australia

[8]Department of Medical Physics, Liverpool and Macarthur Cancer Therapy Centers, Liverpool, Australia

[9]Department of Information Engineering, University of Padova, Padova, Italy

[10]Institute of Neurobiology, Universidad Nacional Autónoma de México Campus Juriquilla, Querétaro, Mexico

[11]Philips Research Hamburg, Hamburg, Germany

[12]Department of Radiology, Stanford University, Stanford, California, USA

[13]Medical Physics Unit, McGill University, Montréal, Québec, Canada

[14]University of British Columbia, Vancouver, British Columbia, Canada

[15]Department of Medical Imaging, McGill University Health Center, Montréal, Québec, Canada

[16]Department of Radiology, McGill University, Montréal, Québec, Canada

[17]Department of Diagnostic and Interventional Imaging, University of Texas Health Science Center at Houston, McGovern Medical School, Houston, Texas, USA

[18]MR Clinical Science, Philips Canada, Mississauga, Ontario, Canada

[19]Department of Human Oncology, University of Wisconsin–Madison, Madison, Wisconsin, USA

[20]Center for Medical Image Computing, Department of Computer Science, University College London, London, UK

[21]Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, London, UK

[22]Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio, USA

[23]Department of Medical Radiation Sciences, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[24]Department of Medical Physics and Biomedical Engineering, Sahlgrenska University Hospital, Gothenburg, Sweden

[25]Center for Mind/Brain Sciences, University of Trento, Trento, Italy

[26]Hopital Maisonneuve-Rosemont, Montréal, Québec, Canada

[27]Bioengineering, Imperial College London, London, UK

[28]Radiotherapy and Imaging, Institute of Cancer Research, Imperial College London, London, UK

[29]Research Institute of the McGill University Health Center, Montréal, Québec, Canada

[30]Clinical Science, Philips Healthcare, Hamburg, Germany

[31]Department of Radiological Sciences, University of California Los Angeles, Los Angeles, California, USA

[32]Physics and Biology in Medicine IDP, University of California Los Angeles, Los Angeles, California, USA

[33]Douglas Brain Imaging Center, Montréal, Québec, Canada

[34]Sunnybrook Research Institute, Toronto, Ontario, Canada

[35]Computer Science Department, Centro de Investigación en Matemáticas, A.C., Guanajuato, Mexico

[36]Medical Research Council, London Institute of Medical Sciences, Imperial College London, London, UK

[37]Department of Radiation Oncology–CNS Service, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[38]Department of Biomedical Engineering, University of British Columbia, Vancouver, British Columbia, Canada

[39]Center for Advanced Interdisciplinary Research, Ss. Cyril and Methodius University, Skopje, North Macedonia

ACKNOWLEDGMENTS

CONFLICT OF INTEREST
The following authors are employees of Phillips: Mariya Doneva, Guillaume Gilbert, Peter Koken, Burkhard Mädler, Christian Stehning, Kilian Weiss, and Shuo Zhang.

DATA AVAILABILITY STATEMENT
An interactive NeuroLibre preprint of this manuscript is available at https://preprint.neurolibre.org/10.55458/neurolibre.00023/. All imaging data submitted to the challenge, data-set details, registered ROI maps, and processed T_1 maps are hosted on OSF: https://osf.io/ywc9g/. The data-set submissions and quality assurance were handled through GitHub issues in this repository: https://github.com/rrsg2020/data_submission (commit: 9d7eff1). Note that accepted submissions are closed issues, and the GitHub branches associated with the issue numbers contain the Dockerfile and Jupyter Notebook scripts that reproduce these preliminary quality assurance results and can be run in a browser using *MyBinder*. The ROI registration scripts for the phantoms and the T_1 fitting pipeline to process all data sets are hosted in this GitHub repository: https://github.com/rrsg2020/t1_fitting_pipeline (commit: 3497a4e). All the analyses of the data sets were done using Jupyter Notebooks and are

available in this repository: https://github.com/rrsg2020/analysis (commit: 8d38644), which also contains a Dockerfile to reproduce the environment using a tool like *MyBinder*. A dashboard was developed to explore the data set's information and results in a browser, which is accessible here: https://rrsg2020.dashboards.neurolibre.org; the code is also available on GitHub: https://github.com/rrsg2020/rrsg2020-dashboard (commit: 6ee9321).

ENDNOTES

[1] https://blog.ismrm.org/2019/12/12/reproducibility-challenge-2020-join-the-reproducible-research-and-quantitative-mr-study-groups-in-their-efforts-to-standardize-t1-mapping/.

[2] The website provided to the researchers, https://collaborate.nist.gov/mriphantoms/bin/view/MriPhantoms/SimpleImagingInstructions, has since been removed from the NIST website.

[3] Source: https://qmri.com/cmri-product-resources/#premium-system-resources.

[4] This website was provided as a resource to the researchers for best practices to obtain informed consent for data sharing: https://www.uu.nl/en/research/research-data-management/guides/informed-consent-for-data-sharing.

[5] Submissions were reviewed by MB and AK. Submission guidelines (https://github.com/rrsg2020/data_submission/blob/master/README.md) and a GitHub issue checklist (https://github.com/rrsg2020/data_submission/blob/master/.github/ISSUE_TEMPLATE/data-submission-request.md) were checked. The submitted data were passed to the T_1 processing pipeline and verified for quality and expected values. Feedback was sent to the authors if their submission did not adhere to the requested guidelines or if issues with the submitted data sets were found, and if possible, corrected (e.g., scaling issues between TI data points).

[6] Strictly speaking, not all manufacturers operate at 3 T. Even though this is the field strength advertised by the system manufacturers, there is some deviation in actual field strength among vendors. The actual center frequencies are typically reported in the DICOM files, and these were shared for most data sets and are available in our OSF.io repository (https://osf.io/ywc9g/). From these data sets, the center frequencies imply that researchers who used GE and Philips scanners operated at 3 T ($\sim$127.7 MHz), whereas researchers who used Siemens scanners operated at 2.89 T ($\sim$123.2 MHz). For simplicity, we always refer to the field strength in this article as 3 T.

[7] http://www-mrsrl.stanford.edu/~jbarral/t1map.html.

[8] https://github.com/qMRLab/qMRLab/blob/master/src/Models_Functions/IRfun/rdNlsPr.m#L118-L129.

[9] https://github.com/rrsg2020/analysis.

[10] https://mybinder.org/v2/gh/rrsg2020/analysis/master?filepath=analysis.

[11] The T_1 values-versus-temperature tables reported by the phantom manufacturer did not always exhibit a linear relationship. We explored the use of spline fitting on the original data and quadratic fitting on the log–log representation of the data. Both methods yielded good results, and we opted to use the latter in our analyses. The code is found here: https://github.com/rrsg2020/analysis/blob/master/src/nist.py. A Jupyter Notebook used in temperature interpolation development is found here: https://github.com/rrsg2020/analysis/blob/master/temperature_correction.ipynb.

[12] Only T_1 maps measured using phantom Version 1 were included in this intersubmission CoV, as including both sets would have increased the CoV due to the differences in reference T_1 values. There were seven research groups that used Version 1 and six that used Version 2.

ORCID
Mathieu Boudreau https://orcid.org/0000-0002-7726-4456
Agah Karakuzu https://orcid.org/0000-0001-7283-271X
Julien Cohen-Adad https://orcid.org/0000-0003-3662-9532
Madeline Carr https://orcid.org/0000-0002-4915-5076
Luis Concha https://orcid.org/0000-0002-7842-3869
Refaat E. Gabr https://orcid.org/0000-0002-8802-3201
Matthew Grech-Sollars https://orcid.org/0000-0003-3881-4870
Siyuan Hu https://orcid.org/0000-0003-3137-0605
Oscar Jalnefjord https://orcid.org/0000-0003-2741-5890
Nam G. Lee https://orcid.org/0000-0001-5462-1492
Bochao Li https://orcid.org/0000-0002-5267-9129
Jamie Near https://orcid.org/0000-0003-3516-936X
Alonso Ramirez-Manzanares https://orcid.org/0000-0001-6645-9162
Ben Statton https://orcid.org/0000-0001-5118-7977
Christian Stehning https://orcid.org/0000-0002-0660-840X
Ye Tian https://orcid.org/0000-0002-8559-4404
Shuo Zhang https://orcid.org/0000-0002-1057-7255
Ziwei Zhao https://orcid.org/0000-0003-0281-1141
Nikola Stikov https://orcid.org/0000-0002-8480-5230

TWITTER
Nikola Stikov Stikov

REFERENCES

1. Keenan KE, Biller JR, Delfino JG, et al. Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. *J Magn Reson Imaging*. 2019;49:e26-e39.
2. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Mak*. 1991;11:88-94.
3. Schweitzer M. Stages of technical efficacy: journal of magnetic resonance imaging style. *J Magn Reson Imaging*. 2016;44:781-782.
4. Seiberlich N, Gulani V, Campbell A, et al. *Quantitative Magnetic Resonance Imaging*. Academic Press; 2020.
5. Damadian R. Tumor detection by nuclear magnetic resonance. *Science*. 1971;171:1151-1153.
6. Pykett IL, Mansfield P. A line scan image study of a tumorous rat leg by NMR. *Phys Med Biol*. 1978;23:961-967.
7. 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.

8. Boudreau M, Keenan KE, Stikov N. Quantitative T_1 and T_1rho mapping. *Quantitative Magnetic Resonance Imaging*. Elsevier; 2020:19-45.

9. Bottomley PA, Foster TH, Argersinger RE, Pfeifer LM. A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age. *Med Phys.* 1984;11:425-448.

10. Wansapura JP, Holland SK, Dunn RS, Ball WS Jr. NMR relaxation times in the human brain at 3.0 Tesla. *J Magn Reson Imaging.* 1999;9:531-538.

11. Dieringer MA, Deimling M, Santoro D, et al. Rapid parametric mapping of the longitudinal relaxation time T_1 using two-dimensional variable flip angle magnetic resonance imaging at 1.5 Tesla, 3 Tesla, and 7 Tesla. *PLoS One.* 2014;9:e91318.

12. Ernst RR, Anderson WA. Application of Fourier transform spectroscopy to magnetic resonance. *Rev Sci Instrum.* 1966;37:93-102.

13. Redpath TW, Smith FW. Technical note: use of a double inversion recovery pulse sequence to image selectively grey or white brain matter. *Br J Radiol.* 1994;67:1258-1263.

14. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imaging.* 1997;7:91-101.

15. Sled JG, Pike GB. Quantitative imaging of magnetization transfer exchange and relaxation properties in vivo using MRI. *Magn Reson Med.* 2001;46:923-931.

16. Yuan J, Chow SKK, Yeung DKW, Ahuja AT, King AD. Quantitative evaluation of dual-flip-angle T_1 mapping on DCE-MRI kinetic parameter estimation in head and neck. *Quant Imaging Med Surg.* 2012;2:245-253.

17. Drain LE. A direct method of measuring nuclear spin-lattice relaxation times. *Proc Phys Soc A.* 1949;62:301-306.

18. Hahn EL. An accurate nuclear magnetic resonance method for measuring spin-lattice relaxation times. *Phys Rev.* 1949;76:145-146. doi:10.1103/physrev.76.145

19. Fram EK, Herfkens RJ, Johnson GA, et al. Rapid calculation of T_1 using variable flip angle gradient refocused imaging. *Magn Reson Imaging.* 1987;5:201-208.

20. Deoni SCL, Rutt BK, Peters TM. Rapid combined T_1 and T_2 mapping using gradient recalled acquisition in the steady state. *Magn Reson Med.* 2003;49:515-526. doi:10.1002/mrm.10407

21. Cheng H-LM, Wright GA. Rapid high-resolution T_1 mapping by variable flip angles: accurate and precise measurements in the presence of radiofrequency field inhomogeneity. *Magn Reson Med.* 2006;55:566-574. doi:10.1002/mrm.20791

22. Look DC, Locker DR. Time saving in measurement of NMR and EPR relaxation times. *Rev Sci Instrum.* 1970;41:250-251.

23. 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.

24. Piechnik SK, Ferreira VM, Dall'Armellina E, et al. Shortened modified Look-Locker inversion recovery (ShMOLLI) for clinical myocardial T_1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. *J Cardiovasc Magn Reson.* 2010;12:69.

25. Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T_1-mapping at high field. *NeuroImage.* 2010;49:1271-1281. doi:10.1016/j.neuroimage.2009.10.002

26. Marques JP, Gruetter R. New developments and applications of the MP2RAGE sequence—focusing the contrast and high spatial resolution R1 mapping. *PLoS One.* 2013;8:e69294.

27. Keenan KE, Ainslie M, Barker AJ, et al. Quantitative magnetic resonance imaging phantoms: a review and the need for a system phantom. *Magn Reson Med.* 2018;79:48-61.

28. Stupic KF, Ainslie M, Boss MA, et al. A standard system phantom for magnetic resonance imaging. *Magn Reson Med.* 2021;86:1194-1211.

29. Bane O, Hectors SJ, Wagner M, et al. 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.

30. 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.

31. Barral JK, Gudmundson E, Stikov N, Etezadi-Amoli M, Stoica P, Nishimura DG. A robust methodology for in vivo T_1 mapping. *Magn Reson Med.* 2010;64:1057-1067.

32. Kluyver T, Ragan-Kelley B, Granger B, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press; 2016:87-90.

33. Beg T, Kluyver K, Ragan-Kelley T, et al. Using Jupyter for reproducible scientific workflows. *Comput Sci Eng.* 2021;23:36-46.

34. Karakuzu A, Boudreau M, Duval T, et al. qMRLab: quantitative MRI analysis, under one umbrella. *J Open Source Softw.* 2020;5:2343.

35. Cabana J-F, Gu Y, Boudreau M, et al. Quantitative magnetization transfer imaging*made*easy with *qMTLab*: software for data simulation, analysis, and visualization. *Concepts Magn Reson Part A Bridg Educ Res.* 2015;44A:263-277.

36. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight J.* 2009;2:1-35.

37. McCarthy P. FSLeyes. 2019. doi:10.5281/zenodo.3403671

38. Merkel D. Docker: lightweight Linux containers for consistent development and deployment. 2014. Available at: https://www.seltzer.com/margo/teaching/CS508.19/papers/merkel14.pdf Accessed February 14, 2023.

39. Boettiger C. An introduction to Docker for reproducible research. *Oper Syst Rev.* 2015;49:71-79.

40. Bussonnier M, Forde J, Freeman J, et al. Binder 2.0—reproducible, interactive, sharable environments for science at scale. *In: Proceedings of the Python in Science Conference*. scipy.org; 2018. doi:10.25080/majora-4af1f417-011

41. Bojorquez JZ, Bricq S, Acquitter C, Brunotte F, Walker PM, Lalande A. What are normal relaxation times of tissues at 3 T? *Magn Reson Imaging.* 2017;35:69-80.

42. Gracien RM, Maiworm M, Brüche N, et al. How stable is quantitative MRI? –Assessment of intra- and inter-scanner-model reproducibility using identical acquisition sequences and data analysis. *NeuroImage.* 2020;207:116364.

43. Layton KJ, Kroboth S, Jia F, et al. Pulseq: a rapid and hardware-independent pulse sequence prototyping framework. *Magn Reson Med.* 2017;77:1544-1552.

44. Cordes C, Konstandin S, Porter D, Günther M. Portable and platform-independent MR pulse sequence programs. *Magn Reson Med.* 2020;83:1277-1290.

45. Karakuzu A, Biswas L, Cohen-Adad J, Stikov N. Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. *Magn Reson Med.* 2022;88:1212-1228.

46. Teixeira RPAG, Neji R, Wood TC, Baburamani AA, Malik SJ, Hajnal JV. Controlled saturation magnetization transfer for reproducible multivendor variable flip angle T_1 and T_2 mapping. *Magn Reson Med.* 2020;84:221-236.

47. Lee Y, Callaghan MF, Acosta-Cabronero J, Lutti A, Nagy Z. Establishing intra- and inter-vendor reproducibility of T_1 relaxation time measurements with 3T MRI. *Magn Reson Med.* 2019;81:454-465.

48. Hansen MS, Sørensen TS. Gadgetron: an open source framework for medical image reconstruction. *Magn Reson Med.* 2013;69:1768-1776.

49. Uecker M, Tamir JI, Ong F, Lustig M. The BART toolbox for computational magnetic resonance imaging. *Proceedings of the 24th Annual Meeting of ISMRM, Singapore.* ISMRM; 2016.

50. Boudreau M, Tardif CL, Stikov N, Sled JG, Lee W, Pike GB. 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.

51. Mancini M, Karakuzu A, Cohen-Adad J, Cercignani M, Nichols TE, Stikov N. An interactive meta-analysis of MRI biomarkers of myelin. *Elife.* 2020;9:e61523. doi:10.7554/eLife.61523

52. Lazari A, Lipp I. Can MRI measure myelin? Systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology. *NeuroImage.* 2021;230:117744.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. Complete list of the data sets submitted to the challenge. (A) Submissions that included phantom data. (B) Submissions that included human brain data. Submissions were assigned numbers to keep track of which submissions included both phantom and human data. Some submissions included data sets acquired on multiple scanners. For the phantom (A), each submission acquired all their data using a single phantom; however some researchers shared the same physical phantom with each other (*same color*). Some additional details about the data sets are included in the T_1 maps column, if relevant. Note that for complex data sets in the magnitude/phase format, T_1 maps were calculated both using magnitude-only data and complex-data, but these were from the same measurement (*branching off arrow*).

How to cite this article: Boudreau M, Karakuzu A, Cohen-Adad J, et al. Repeat it without me: Crowdsourcing the T_1 mapping common ground via the ISMRM reproducibility challenge. *Magn Reson Med*. 2024;92:1115-1127. doi: 10.1002/mrm.30111