NeuroLibre
Reproducible Preprints

Paper is not enough: Crowdsourcing the T1 mapping common ground via the ISMRM reproducibility challenge

***Mathieu Boudreau**[1,2], ***Agah Karakuzu**[1,2], **Julien Cohen-Adad**[1,3,4,5], **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 Maforo**[31,32], **Jamie Near**[33,34], **Erick Pasaye**[35], **Alonso Ramirez-Manzanares**[35], **Ben Statton**[36], **Christian Stehning**[30], **Stefano Tambalo**[25], **Ye Tian**[6], **Chenyang Wang**[37], **Kilian Weis**[30], **Niloufar Zakariaei**[38], **Shuo Zhang**[30], **Ziwei Zhao**[6], **Nikola Stikov**[1,2,39]

* These authors contributed equally.

DOI: 10.55458/neurolibre.00023

Reproducible Preprint

- Jupyter Book

Code

- Technical Screening
- Submitted Repository

Reproducibility Assets

- Repository
- Dataset
- Jupyter Book
- Container

Moderator: Pierre Bellec
Screener(s):

- @pbellec

Submitted: 24 January 2024
Published: 01 February 2024

THIS PDF IS INTENDED FOR CONTENT REGISTRATION PURPOSES ONLY! FOR FULL ACCESS AND INTERACTIVE READING OF THIS PUBLICATION, PLEASE VISIT THE REPRODUCIBLE PREPRINT.

Summary

We present the results of the ISMRM 2020 joint Reproducible Research and Quantitative MR study groups reproducibility challenge on T1 mapping in phantom and human brain. T1 mapping, a widely used quantitative MRI technique, exhibits inconsistent tissue-specific values across protocols, sites, and vendors. The challenge aimed to assess the reproducibility of a well-established inversion recovery T1 mapping technique, with acquisition details published solely as a PDF, on a standardized phantom and in human brains. Participants acquired T1 mapping data from MRIs of three manufacturers at 3T, resulting in 39 phantom datasets and 56 datasets from healthy human subjects. The T1 inter-submission variability was twice as high as the intra-submission variability in both phantoms and human brains, indicating that the acquisition details in the selected paper were insufficient to reproduce a quantitative MRI protocol. This study reports the inherent uncertainty in T1 measures across independent research groups, bringing us one step closer to a practical clinical baseline of T1 variations in vivo. This challenge resulted in the creation of a comprehensive open database of T1 mapping acquisitions, accessible at osf.io/ywc9g/, and an interactive dashboard for wider community access and engagement.

Figures

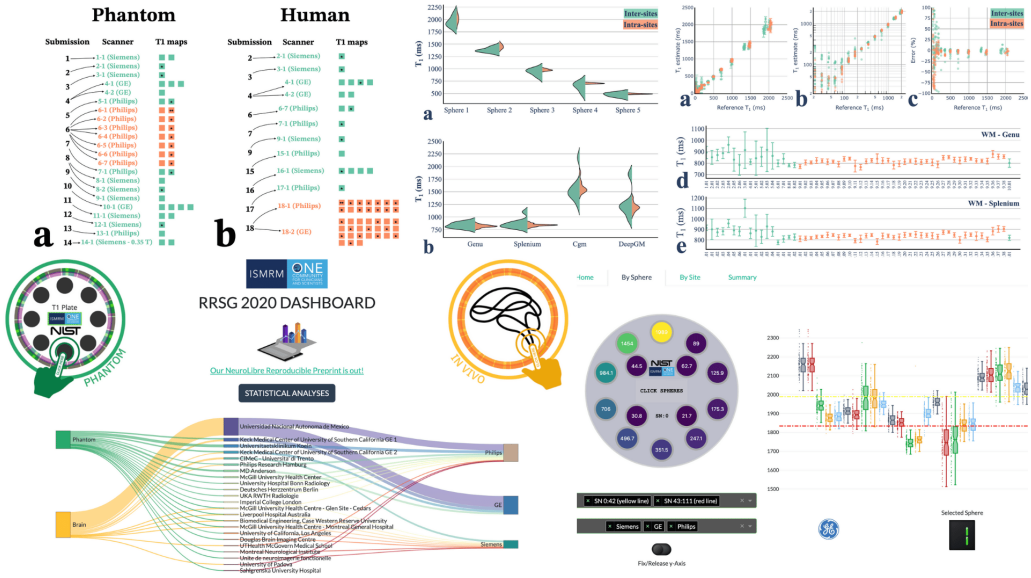

Figure 1: Dashboard. a) welcome page listing all the sites, the types of subject, and scanner, and the relationship between the three. Row b) shows two of the phantom dashboard tabs, and row c) shows two of the human data dashboard tabs Link: https://rrsg2020.db.neurolibre.org

Acknowledgements

The conception of this collaborative reproducibility challenge originated from discussions with experts, including Paul Tofts, Joëlle Barral, and Ilana Leppert, who provided valuable insights. Additionally, Kathryn Keenan, Zydrunas Gimbutas, and Andrew Dienstfrey from NIST provided their code to generate the ROI template for the ISMRM/NIST phantom. Dylan Roskams-Edris and Gabriel Pelletier from the Tanenbaum Open Science Institute (TOSI) offered valuable insights and guidance related to data ethics and data sharing in the context of this international multi-center conference challenge. The 2020 RRSG study group committee members who launched the challenge, Martin Uecker, Florian Knoll, Nikola Stikov, Maria Eugenia Caligiuri, and Daniel Gallichan, as well as the 2020 qMRSG committee members, Kathryn Keenan, Diego Hernando, Xavier Golay, Annie Yuxin Zhang, and Jeff Gunter, also played an essential role in making this challenge possible. We would also like to thank the Canadian Open Neuroscience Platform (CONP), the Quebec Bioimaging Network (QBIN), and the Montreal Heart Institute Foundation for their support in creating the NeuroLibre preprint. Finally, we extend our thanks to all the volunteers and individuals who helped with the scanning at each imaging site.

The authors thank the ISMRM Reproducible Research Study Group for conducting a code review of the code (Version 1) supplied in the Data Availability Statement. The scope of the code review covered only the code's ease of download, quality of documentation, and ability to run, but did not consider scientific accuracy or code efficiency.

Lastly, we acknowledge use of ChatGPT (v3), a generative language model, for accelerating manuscript preparation. The co-first authors employed ChatGPT in the initial draft for transforming bullet point sentences into paragraphs, proofreading for typos, and refining the academic tone. ChatGPT served exclusively as a writing aid, and was not used to create or interpret results.

NOTE

The following section in this document repeats the narrative content exactly as found in the corresponding NeuroLibre Reproducible Preprint (NRP). The content was automatically incorporated into this PDF using the NeuroLibre publication workflow (Karakuzu, DuPre, et al., 2022) to credit the referenced resources. The submitting author of the preprint has verified and approved the inclusion of this section through a GitHub pull request made to the source repository from which this document was built. Please note that the figures and tables have been excluded from this (static) document. **To interactively explore such outputs and re-generate them, please visit the corresponding NRP.** For more information on integrated research objects (e.g., NRPs) that bundle narrative and executable content for reproducible and transparent publications, please refer to (DuPre et al., 2022). NeuroLibre is sponsored by the Canadian Open Neuroscience Platform (CONP) (Harding et al., 2023).

1 | INTRODUCTION

Significant challenges exist in the reproducibility of quantitative MRI (qMRI) (Keenan et al., 2019). Despite its promise of improving the specificity and reproducibility of MRI acquisitions, few qMRI 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 (Fryback & Thornbury, 1991; Schweitzer, 2016; Seiberlich et al., 2020). Half a century has passed since the first quantitative T1 (spin-lattice relaxation time) measurements were first reported as a potential biomarker for tumors (Damadian, 1971), followed shortly thereafter by the first in vivo T1 maps (Pykett & Mansfield, 1978) of tumors, but there is still disagreement in reported values for this fundamental parameter across different sites, vendors, and measurement techniques (Stikov et al., 2015).

Among fundamental MRI parameters, T1 holds significant importance (Boudreau et al., 2020). T1 represents the time constant for recovery of equilibrium longitudinal magnetization. T1 values will vary depending on the molecular mobility and magnetic field strength (Bottomley et al., 1984; Dieringer et al., 2014; Wansapura et al., 1999). Knowledge of the T1 values for tissue is crucial for optimizing clinical MRI sequences for contrast and time efficiency (Ernst & Anderson, 1966; Redpath & Smith, 1994; Tofts, 1997) and to calibrate other quantitative MRI techniques [Sled & Pike (2001);Yuan2012-xh]. Inversion recovery (IR) (Drain, 1949; Hahn, 1949) is considered the gold standard for T1 measurement due to its robustness against effects like B1 inhomogeneity (Stikov et al., 2015), but its long acquisition times limit the clinical use of IR for T1 mapping (Stikov et al., 2015). In practice, IR is often used as a reference for validating other T1 mapping techniques, such as variable flip angle imaging (VFA) (Cheng & Wright, 2006; Deoni et al., 2003; Fram et al., 1987), Look-Locker (Look & Locker, 1970; Messroghli et al., 2004; Piechnik et al., 2010), and MP2RAGE (Marques et al., 2010; Marques & Gruetter, 2013).

In ongoing efforts to standardize T1 mapping methods, researchers have been actively developing quantitative MRI phantoms (Keenan et al., 2018). The International Society for Magnetic Resonance in Medicine (ISMRM) and the National Institute of Standards and Technology (NIST) collaborated on a standard system phantom (Stupic et al., 2021), which was subsequently commercialized (Premium System Phantom, CaliberMRI, Boulder, Colorado). This phantom has since been used in large multicenter studies, such as Bane et al. (Bane et al., 2018) which concluded that acquisition protocols and field strength influence accuracy, repeatability, and interplatform reproducibility. Another NIST-led study (Keenan et al., 2021) found no significant T1 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 (T1)? To assess this, we proposed using inversion recovery 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 T1 mapping (Barral et al., 2010) is sufficient to ensure the reproducibility across independent research groups.**

2 | METHODS

2.1 | Phantom and human data

2 | METHODS

2.1 | Phantom and human data

The challenge asked researchers with access to the ISMRM/NIST system phantom (Stupic et al., 2021) (Premium System Phantom, CaliberMRI, Boulder, Colorado) to measure T1 maps of the phantom's T1 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 prior to scanning to achieve thermal equilibrium.

Researchers were also instructed to collect T1 maps in healthy human brains, and were asked to measure a single slice positioned parallel to the anterior commissure - posterior commissure (AC-PC) line. Prior to imaging, the imaging subjects consented [3] to share their de-identified data with the challenge organizers and on the Open Science Framework (OSF.io) website. As the submitted data was 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 was 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 inversion recovery T1 mapping protocol optimized for the human brain as described in the paper published by Barral et al. (Barral et al., 2010), which used: TR = 2550 ms, TIs = 50, 400, 1100, 2500 ms, TE = 14 ms, 2 mm slice thickness and 1×1 mm2 in-plane resolution. Note that this protocol is not suitable for fitting models that assume TR > 5T1. Instead, the more general Barral et al. (Barral et al., 2010) 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 datasets

[1] ISMRM blog post announcingn the RRRSG challenge

[2] The website provided to the researchers has since been removed from the NIST website.

[3] This website was provided as a resource to the participants for best practices to obtain informed consent for data sharing.

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 [4], and following acceptance the datasets were uploaded to OSF.io (osf.io/ywc9g/). A Jupyter Notebook (Beg et al., 2021; Kluyver et al., 2016) pipeline using qMRLab (Cabana et al., 2015; Karakuzu et al., 2020) was used to process the T1 maps and to conduct quality-control checks. MyBinder links to Jupyter notebooks that reproduced each T1 map were shared in each respective submission 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 T1 maps of the NIST/system phantom, and 56 brain T1 maps. Figure 1 illustrates all the submissions that acquired phantom data (Figure 1-a) and human data (Figure 1-b), the MRI scanner vendors, and the resulting T1 mapping datasets. Some submissions included measurements where both complex and magnitude-only data from the same acquisition were used to fit T1 maps, thus the total number of unique acquisitions is lower than the numbers reported above (27 for phantom data and 44 for human data). The datasets were collected on systems from three MRI manufacturers (Siemens, GE, Philips) and were acquired at 3T [5], except for one dataset acquired at 0.35T (the ViewRay MRidian MR-linac).

Figure 1 A snapshot of the figures (top row) included in the reproducible preprint (https://preprint.neurolibre.org/10.55458/neurolibre.00023) and the dashboard (bottom row, https://rrsg2020.db.neurolibre.org).

2.4 | Data Processing

A reduced-dimension non-linear least squares (RD-NLS) approach was used to fit the complex general inversion recovery signal equation:

where a and b are complex constants. This approach, developed by Barral et al. (Barral et al., 2010), offers a model for the general T1 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 pulse angle, inversion pulse flip angles, TR, TE, TI, and a constant that has contributions from T2 and the receive coil sensitivity. Barral et al. [31] shared their MATLAB (MathWorks, Natick, MA) code for the fitting algorithm used in their paper [6]. Magnitude-only data were fitted to a modified version of Eq. 1 (Eq. 15 of Barral et al. 2010) with signal-polarity restoration by finding the signal minima, fitting the inversion recovery curve for two cases (data points for TI $<$ TIminimum flipped, and data points for TI $>$ TIminimum flipped), and selecting the case that resulted in the best fit based on minimizing the residual between the model and the measurements [7]. This code is available as part of the open-source software qMRLab (Cabana et al., 2015; Karakuzu et al., 2020), which provides a standardized application program interface (API) 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 dataset from OSF.io (osf.io/ywc9g/), loads its configuration file,

[4] 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. Lastly, the submitted data was passed to the T1 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 datasets were found and if possible, corrected (e.g., scaling issues between inversion time data points).

[5] Strictly speaking, not all manufacturers operate at 3.0 T. Even though this is the field strength advertised by the system manufacturers, there is some deviation in actual field strength between vendors. The actual center frequencies are typically reported in the DICOM files, and these were shared for most datasets and are available in our OSF.io repository (https://osf.io/ywc9g/). From these datasets, the center frequencies imply participants that used GE and Philips scanners were at 3.0T (~127.7 MHz), whereas participants that used Siemens scanners were at 2.89T (~123.2 MHz). For simplicity, we will always refer to the field strength in this article as 3T.

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

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

fits the T1 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, T1 maps were manually uploaded to OSF (osf.io/ywc9g/).

2.5 | Image Labeling & Registration

The T1 plate (NiCl2 array) of the phantom has 14 spheres that were labeled as the regions-of-interest (ROI) using a numerical mask template created in MATLAB, provided by NIST researchers (Figure 1-c). To avoid potential edge effects in the T1 maps, the ROI labels were reduced to 60% of the expected sphere diameter. A registration pipeline in Python using the Advanced Normalization Tools (ANTs) {cite}`Avants2009-cw` 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, sub-sampling: 4/2/1) and a BSplineSyN registration (gradientStep:0.5, meshSizeAtBaseLevel:3, number of steps: 3, iterations: 50/50/10, smoothness: 0/0/0, sub-sampling: 4/2/1). The ROI labels template was nonlinearly registered to each T1 map uploaded to OSF.

For human data, manual ROIs were segmented by a single researcher (M.B., 11+ years of neuroimaging experience) using FSLeyes {cite}`McCarthy2019-qd` in four regions (Figure 1-d): 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 [8]. The T1 fitting and data analysis were performed by M.B., one of the challenge organizers. Computational environment requirements were containerized in Docker (Boettiger, 2015; Merkel, 2014) to create an executable environment that allows for analysis reproduction in a web browser via MyBinder [9] (Project Jupyter et al., 2018). Backend Python files handled reference data, database operations, ROI masking, and general analysis tools. Configuration files handled dataset information, and the datasets were downloaded and pooled using a script (`make_pooled_datasets.py`). The databases were created using a reproducible Jupyter Notebook script and subsequently saved in the repository.

The mean T1 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 [10] of a NIST reference table of T1 values for temperatures ranging from 16 °C to 26 °C (2 °C intervals) as specified in the phantom's technical specifications. For the human datasets, the mean and standard deviations for each tissue ROI were calculated from all submissions across all sites. Two submissions (one of phantom data – submission 6 in Figure 1-a, and one of human data – submission 18 in Figure 1-b) were received that measured large T1 mapping datasets. Submission 6 consisted of data from one traveling phantom acquired at seven Philips imaging sites, and submission 18 was a large cohort of volunteers that were imaged on two 3T scanners, one GE and one Philips. These datasets (identified in orange in Figures 1, 3, and 4) were used to calculate intra-submission coefficients of variation (COV) (one per scanner/volunteer, identified by asterisks in Figure 1-a and 1-b), and inter-submission COVs were calculated using one T1 map from each of these (orange) along with one from

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

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

[10] The T1 values vs 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, and a Jupyter Notebook used in temperature interpolation development is here.

all other submissions [11] (identified as green in Figures 1, 3, and 4, and the T1 maps used in those COV calculations are also indicated with asterisks in Figure 1-a and 1-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 2-a) showcases the relationship between the phantom and brain datasets acquired at different sites/vendors. Selecting the Phantom or In Vivo icons 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 vs 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 inter- and intra- submission comparisons in both phantoms (a) and human (b) datasets. For the phantom (Figure 3-a), the average inter-submission COV for the first five spheres, representing the expected T1 value range in the human brain (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 inter-submission COV was reduced to 4.1%. One participant (submission 6, Figure 1) measured T1 maps using a consistent protocol at 7 different sites, and the mean intra-submission COV across the first five spheres for this submission was calculated to be 2.9%.

For the human datasets (Figure 3-b), inter-submission COVs for independently-implemented imaging protocols were 5.9% for genu, 10.6 % for splenium, 16 % for cortical GM, and 22% for deep GM. One participant (submission 18, Figure 1) measured a large dataset (13 individuals) on three scanners and two vendors, and the intra-submission 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 inter-site analyses (green) can be explained by an outlier measurement, which can be seen in (Figure 4 e-f, site 3.001).

A scatterplot of the T1 data for all submissions and their ROIs is shown in Figure 4 (phantom a-c, and human brains d-f). The NIST phantom T1 measurements are presented in each plot for different axes types (linear, log, and error) to better visualize the results. Figure 4-a shows good agreement for this dataset in comparison with the temperature-corrected reference T1 values. However, this trend did not persist for low T1 values (T1 $<$ 100-200 ms), as seen in the log-log plot (Figure 4-b), which was expected because the imaging protocol is optimized for human brain T1 values (T1 $>$ 500 ms). Higher variability is seen at long T1 values (T1 $\sim$ 2000 ms) in Figure 4-a. Errors exceeding 10% are observed in the phantom spheres with T1 values below 300 ms (Figure 4-c), and 3-4 measurements with outlier values exceeding 10% error were observed in the human brain tissue range (~500-2000 ms).

Figure 4 d-f displays the scatter plot data for human datasets submitted to this challenge, showing mean and standard deviation T1 values for the WM (genu and splenium) and GM (cerebral cortex and deep GM) ROIs. Mean WM T1 values across all submissions were 828 $\pm$ 38 ms in the genu and 852 $\pm$ 49 ms in the splenium, and mean GM T1 values were 1548 $\pm$ 156 ms in the cortex and 1188 $\pm$ 133 ms in the deep GM, with less variations overall in WM compared to GM, possibly due to better ROI placement and less partial voluming in WM. The

[11] Only T1 maps measured using phantom version 1 were included in this inter-submission COV, as including both sets would have increased the COV due to the differences in reference T1 values. There were seven research groups that used version 1, and six that used version 2.

lower standard deviations for the ROIs of human database ID site 9 (by submission 18, 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 focused on exploring if different research groups could reproduce T1 maps based on the protocol information reported in a seminal publication (Barral et al., 2010). Eighteen submissions independently implemented the inversion recovery T1 mapping acquisition protocol as outlined in Barral et al. (Barral et al., 2010), and reported T1 mapping data in a standard quantitative MRI phantom and/or human brains at 27 MRI sites, using systems from three different vendors (GE, Philips, Siemens). The collaborative effort produced an open-source database of 94 T1 mapping datasets, including 38 ISMRM/NIST phantom and 56 human brain datasets. The inter-submission variability was twice as high as the intra-submission variability in both phantom and human brain T1 measurements, **demonstrating that written instructions communicated via a PDF are not enough for reproducibility in quantitative MRI**. This study reports the inherent uncertainty in T1 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 T1 measurements in vivo compared to researchers implementing T1 mapping protocols completely independently (i.e. with no central guidance), as literature T1 values in vivo vary more than reported here (e.g., Bojorquez et al. (Bojorquez et al., 2017) reports that reported T1 values in WM vary between 699 to 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 guidelines for quantitative MR imaging that were available at the time. Yet, even in combination with the same T1 mapping processing tools, this level of description (a PDF + 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 (Gracien et al., 2020). Vendor-neutral pulse sequence design platforms (Cordes et al., 2020; Karakuzu, Biswas, et al., 2022; Layton et al., 2017) have emerged as a powerful solution to standardize sequence components at the implementation level (e.g., RF pulse shape, gradients, etc.). Vendor neutrality has been shown to significantly reduce the variability of T1 maps acquired using VFA across vendors (Karakuzu, Biswas, et al., 2022). In the absence of a vendor-neutral framework, a vendor-specific alternative is the implementation of a strategy to control the saturation of MT across TRs (A G Teixeira et al., 2020). Nevertheless, this approach can still benefit from a vendor-neutral protocol to enhance accessibility and unify implementations. This is because vendor-specific constraints are recognized 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 (Lee et al., 2019).

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 inversion recovery T1 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 datasets (https://rrsg2020.dashboards.neurolibre.org). The differences in stability between independently implemented (inter-submission) and centrally shared (intra-submission) protocols observed both in phantoms and in vivo could help inform

future meta-analyses of quantitative MRI metrics (Lazari & Lipp, 2021; Mancini et al., 2020) and better guide multi-center collaborations.

By providing access and analysis tools for this multi-center T1 mapping dataset, we aim to provide a benchmark for future T1 mapping approaches. We also hope that this dataset will inspire new acquisition, analysis, and standardization techniques that address non-physiological sources of variability in T1 mapping. This could lead to more robust and reproducible quantitative MRI and ultimately better patient care.

References

A G Teixeira, R. P., Neji, R., Wood, T. C., Baburamani, A. A., Malik, S. J., & Hajnal, J. V. (2020). Controlled saturation magnetization transfer for reproducible multivendor variable flip angle T1 and T2 mapping. *Magn. Reson. Med.*, *84*(1), 221–236. https://doi.org/10.1002/mrm.28109

Bane, O., Hectors, S. J., Wagner, M., Arlinghaus, L. L., Aryal, M. P., Cao, Y., Chenevert, T. L., Fennessy, F., Huang, W., Hylton, N. M., Kalpathy-Cramer, J., Keenan, K. E., Malyarenko, D. I., Mulkern, R. V., Newitt, D. C., Russek, S. E., Stupic, K. F., Tudorica, A., Wilmes, L. J., … Taouli, B. (2018). Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. *Magn. Reson. Med.*, *79*(5), 2564–2575. https://doi.org/10.1002/mrm.26903

Barral, J. K., Gudmundson, E., Stikov, N., Etezadi-Amoli, M., Stoica, P., & Nishimura, D. G. (2010). A robust methodology for in vivo T1 mapping. *Magn. Reson. Med.*, *64*(4), 1057–1067. https://doi.org/10.1002/mrm.22497

Beg, Taka, Kluyver, Konovalov, Ragan-Kelley, Thiery, & Fangohr. (2021). Using jupyter for reproducible scientific workflows. *Https://Www.computer.org › Csdl › Magazine › 2021/02https://Www.computer.org › Csdl › Magazine › 2021/02*, *23*, 36–46. https://doi.org/10.1109/MCSE.2021.3052101

Boettiger, C. (2015). An introduction to docker for reproducible research. *Oper. Syst. Rev.*, *49*(1), 71–79. https://doi.org/10.1145/2723872.2723882

Bojorquez, J. Z., Bricq, S., Acquitter, C., Brunotte, F., Walker, P. M., & Lalande, A. (2017). What are normal relaxation times of tissues at 3 t? *Magn. Reson. Imaging*, *35*, 69–80. https://doi.org/10.1016/j.mri.2016.08.021

Bottomley, P. A., Foster, T. H., Argersinger, R. E., & Pfeifer, L. M. (1984). 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.*, *11*(4), 425–448. https://doi.org/10.1118/1.595535

Boudreau, M., Keenan, K. E., & Stikov, N. (2020). Quantitative T1 and T1r mapping. In *Quantitative magnetic resonance imaging* (pp. 19–45). https://doi.org/10.1016/b978-0-12-817057-1.00004-4

Cabana, J.-F., Gu, Y., Boudreau, M., Levesque, I. R., Atchia, Y., Sled, J. G., Narayanan, S., Arnold, D. L., Pike, G. B., Cohen-Adad, J., Duval, T., Vuong, M.-T., & Stikov, N. (2015). Quantitative magnetization transfer imaging*made*easy with *qMTLab*: Software for data simulation, analysis, and visualization. *Concepts Magn. Reson. Part A Bridg. Educ. Res.*, *44A*(5), 263–277. https://doi.org/10.1002/cmr.a.21357

Cheng, H.-L. M., & Wright, G. A. (2006). Rapid high-resolutionT1 mapping by variable flip angles: Accurate and precise measurements in the presence of radiofrequency field inhomogeneity. In *Magnetic Resonance in Medicine* (No. 3; Vol. 55, pp. 566–574). https://doi.org/10.1002/mrm.20791

Cordes, C., Konstandin, S., Porter, D., & Günther, M. (2020). Portable and platform-independent MR pulse sequence programs. *Magn. Reson. Med.*, *83*(4), 1277–1290. https://doi.org/10.1002/mrm.28020

Damadian, R. (1971). Tumor detection by nuclear magnetic resonance. *Science*, *171*(3976), 1151–1153. https://doi.org/10.1126/science.171.3976.1151

Deoni, S. C. L., Rutt, B. K., & Peters, T. M. (2003). Rapid combinedT1 andT2 mapping using gradient recalled acquisition in the steady state. In *Magnetic Resonance in Medicine* (No. 3; Vol. 49, pp. 515–526). https://doi.org/10.1002/mrm.10407

Dieringer, M. A., Deimling, M., Santoro, D., Wuerfel, J., Madai, V. I., Sobesky, J., Knobelsdorff-Brenkenhoff, F. von, Schulz-Menger, J., & Niendorf, T. (2014). Rapid parametric mapping of the longitudinal relaxation time T1 using two-dimensional variable flip angle magnetic resonance imaging at 1.5 tesla, 3 tesla, and 7 tesla. *PLoS One*, *9*(3), e91318. https://doi.org/10.1371/journal.pone.0091318

Drain, L. E. (1949). A direct method of measuring nuclear Spin-Lattice relaxation times. *Proc. Phys. Soc. A*, *62*(5), 301. https://doi.org/10.1088/0370-1298/62/5/306

DuPre, E., Holdgraf, C., Karakuzu, A., Tetrel, L., Bellec, P., Stikov, N., & Poline, J.-B. (2022). Beyond advertising: New infrastructures for publishing integrated research objects. *PLOS Computational Biology*, *18*(1), e1009651. https://doi.org/10.1371/journal.pcbi.1009651

Ernst, R. R., & Anderson, W. A. (1966). Application of fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instrum.*, *37*(1), 93–102. https://doi.org/10.1063/1.1719961

Fram, E. K., Herfkens, R. J., Johnson, G. A., Glover, G. H., Karis, J. P., Shimakawa, A., Perkins, T. G., & Pelc, N. J. (1987). Rapid calculation of T1 using variable flip angle gradient refocused imaging. *Magn. Reson. Imaging*, *5*(3), 201–208. https://doi.org/10.1016/0730-725X(87)90021-X

Fryback, D. G., & Thornbury, J. R. (1991). The efficacy of diagnostic imaging. *Med. Decis. Making*, *11*(2), 88–94. https://doi.org/10.1177/0272989X9101100203

Gracien, Maiworm, Brüche, Shrestha, & others. (2020). How stable is quantitative MRI?–Assessment of intra-and inter-scanner-model reproducibility using identical acquisition sequences and data analysis …. *Neuroimage*. https://doi.org/10.1016/j.neuroimage.2019.116364

Hahn, E. L. (1949). An accurate nuclear magnetic resonance method for measuring Spin-Lattice relaxation times. In *Physical Review* (No. 1; Vol. 76, pp. 145–146). https://doi.org/10.1103/PhysRev.76.145

Harding, R. J., Bermudez, P., Bernier, A., Beauvais, M., Bellec, P., Hill, S., Karakuzu, A., Knoppers, B. M., Pavlidis, P., Poline, J.-B., Roskams, J., Stikov, N., Stone, J., Strother, S., Consortium, C., & Evans, A. C. (2023). The Canadian Open Neuroscience Platform—An open science framework for the neuroscience community. *PLOS Computational Biology*, *19*(7), 1–14. https://doi.org/10.1371/journal.pcbi.1011230

Karakuzu, A., Biswas, L., Cohen-Adad, J., & Stikov, N. (2022). Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. *Magn. Reson. Med.*, *88*(3), 1212–1228. https://doi.org/10.1002/mrm.29292

Karakuzu, A., Boudreau, M., Duval, T., Boshkovski, T., Leppert, I., Cabana, J.-F., Gagnon, I., Beliveau, P., Pike, G., Cohen-Adad, J., & Stikov, N. (2020). qMRLab: Quantitative MRI analysis, under one umbrella. *J. Open Source Softw.*, *5*(53), 2343. https://doi.org/10.21105/joss.02343

Karakuzu, A., DuPre, E., Tetrel, L., Bermudez, P., Boudreau, M., Chin, M., Poline, J.-B., Das, S., Bellec, P., & Stikov, N. (2022). *NeuroLibre : A preprint server for full-fledged*

reproducible neuroscience. OSF Preprints. https://doi.org/10.31219/osf.io/h89js

Keenan, K. E., Ainslie, M., Barker, A. J., Boss, M. A., Cecil, K. M., Charles, C., Chenevert, T. L., Clarke, L., Evelhoch, J. L., Finn, P., Gembris, D., Gunter, J. L., Hill, D. L. G., Jack, C. R., Jr, Jackson, E. F., Liu, G., Russek, S. E., Sharma, S. D., Steckner, M., … Zheng, J. (2018). Quantitative magnetic resonance imaging phantoms: A review and the need for a system phantom. *Magn. Reson. Med.*, *79*(1), 48–61. https://doi.org/10.1002/mrm.26982

Keenan, K. E., Biller, J. R., Delfino, J. G., Boss, M. A., Does, M. D., Evelhoch, J. L., Griswold, M. A., Gunter, J. L., Hinks, R. S., Hoffman, S. W., Kim, G., Lattanzi, R., Li, X., Marinelli, L., Metzger, G. J., Mukherjee, P., Nordstrom, R. J., Peskin, A. P., Perez, E., … Sullivan, D. C. (2019). Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. *J. Magn. Reson. Imaging*, *49*(7), e26–e39. https://doi.org/10.1002/jmri.26598

Keenan, K. E., Gimbutas, Z., Dienstfrey, A., Stupic, K. F., Boss, M. A., Russek, S. E., Chenevert, T. L., Prasad, P. V., Guo, J., Reddick, W. E., Cecil, K. M., Shukla-Dave, A., Aramburu Nunez, D., Shridhar Konar, A., Liu, M. Z., Jambawalikar, S. R., Schwartz, L. H., Zheng, J., Hu, P., & Jackson, E. F. (2021). Multi-site, multi-platform comparison of MRI T1 measurement using the system phantom. *PLoS One*, *16*(6), e0252966. https://doi.org/10.1371/journal.pone.0252966

Kluyver, T., Ragan-Kelley, B., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Abdalla, S., & Willing, C. (2016). Jupyter notebooks – a publishing format for reproducible computational workflows. In *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Layton, K. J., Kroboth, S., Jia, F., Littin, S., Yu, H., Leupold, J., Nielsen, J.-F., Stöcker, T., & Zaitsev, M. (2017). Pulseq: A rapid and hardware-independent pulse sequence prototyping framework. *Magn. Reson. Med.*, *77*(4), 1544–1552. https://doi.org/10.1002/mrm.26235

Lazari, A., & Lipp, I. (2021). Can MRI measure myelin? Systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology. *Neuroimage*, *230*, 117744. https://doi.org/10.1016/j.neuroimage.2021.117744

Lee, Y., Callaghan, M. F., Acosta-Cabronero, J., Lutti, A., & Nagy, Z. (2019). Establishing intra- and inter-vendor reproducibility of T1 relaxation time measurements with 3T MRI. *Magn. Reson. Med.*, *81*(1), 454–465.

Look, D. C., & Locker, D. R. (1970). Time saving in measurement of NMR and EPR relaxation times. *Rev. Sci. Instrum.*, *41*(2), 250–251. https://doi.org/10.1063/1.1684482

Mancini, M., Karakuzu, A., Cohen-Adad, J., Cercignani, M., Nichols, T. E., & Stikov, N. (2020). An interactive meta-analysis of MRI biomarkers of myelin. *Elife*, *9*. https://doi.org/10.55458/neurolibre.00004

Marques, J. P., & Gruetter, R. (2013). New developments and applications of the MP2RAGE sequence–focusing the contrast and high spatial resolution R1 mapping. *PLoS One*, *8*(7), e69294. https://doi.org/10.1371/journal.pone.0069294

Marques, J. P., Kober, T., Krueger, G., Zwaag, W. van der, Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. In *NeuroImage* (No. 2; Vol. 49, pp. 1271–1281). https://doi.org/10.1016/j.neuroimage.2009.10.002

Merkel, D. (2014). *Docker: Lightweight linux containers for consistent development and deployment*. https://www.seltzer.com/margo/teaching/CS508.19/papers/merkel14.pdf.

Messroghli, D. R., Radjenovic, A., Kozerke, S., Higgins, D. M., Sivananthan, M. U., & Ridgway, J. P. (2004). Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1

mapping of the heart. *Magn. Reson. Med.*, *52*(1), 141–146. https://doi.org/10.1002/mrm.20110

Piechnik, S. K., Ferreira, V. M., Dall'Armellina, E., Cochlin, L. E., Greiser, A., Neubauer, S., & Robson, M. D. (2010). Shortened modified Look-Locker inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. *J. Cardiovasc. Magn. Reson.*, *12*(1), 69. https://doi.org/10.1186/1532-429X-12-69

Project Jupyter, Bussonnier, M., Forde, J., Freeman, J., Granger, B., Head, T., Holdgraf, C., Kelley, K., Nalvarte, G., Osheroff, A., Pacer, M., Panda, Y., Perez, F., Ragan-Kelley, B., & Willing, C. (2018). Binder 2.0 - reproducible, interactive, sharable environments for science at scale. *Proceedings of the Python in Science Conference*. https://doi.org/10.25080/Majora-4af1f417-011

Pykett, I. L., & Mansfield, P. (1978). A line scan image study of a tumorous rat leg by NMR. *Phys. Med. Biol.*, *23*(5), 961–967. https://doi.org/10.1097/00004728-197904000-00056

Redpath, T. W., & Smith, F. W. (1994). Technical note: Use of a double inversion recovery pulse sequence to image selectively grey or white brain matter. *Br. J. Radiol.*, *67*(804), 1258–1263. https://doi.org/10.1259/0007-1285-67-804-1258

Schweitzer, M. (2016). Stages of technical efficacy: Journal of magnetic resonance imaging style. *J. Magn. Reson. Imaging*, *44*(4), 781–782. https://doi.org/10.1002/jmri.25417

Seiberlich, N., Gulani, V., Campbell, A., Sourbron, S., Doneva, M. I., Calamante, F., & Hu, H. H. (2020). *Quantitative magnetic resonance imaging*. Academic Press.

Sled, J. G., & Pike, G. B. (2001). Quantitative imaging of magnetization transfer exchange and relaxation properties in vivo using MRI. *Magn. Reson. Med.*, *46*(5), 923–931. https://doi.org/10.1002/mrm.1278

Stikov, N., Boudreau, M., Levesque, I. R., Tardif, C. L., Barral, J. K., & Pike, G. B. (2015). On the accuracy of T1 mapping: Searching for common ground. *Magn. Reson. Med.*, *73*(2), 514–522. https://doi.org/10.1002/mrm.25135

Stupic, K. F., Ainslie, M., Boss, M. A., Charles, C., Dienstfrey, A. M., Evelhoch, J. L., Finn, P., Gimbutas, Z., Gunter, J. L., Hill, D. L. G., Jack, C. R., Jackson, E. F., Karaulanov, T., Keenan, K. E., Liu, G., Martin, M. N., Prasad, P. V., Rentz, N. S., Yuan, C., & Russek, S. E. (2021). A standard system phantom for magnetic resonance imaging. *Magn. Reson. Med.*, *86*(3), 1194–1211. https://doi.org/10.1002/mrm.28779

Tofts, P. S. (1997). Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J. Magn. Reson. Imaging*, *7*(1), 91–101. https://doi.org/10.1002/jmri.1880070113

Wansapura, J. P., Holland, S. K., Dunn, R. S., & Ball, W. S., Jr. (1999). NMR relaxation times in the human brain at 3.0 tesla. *J. Magn. Reson. Imaging*, *9*(4), 531–538. https://doi.org/10.1002/(SICI)1522-2586(199904)9:4%3C531::AID-JMRI4%3E3.0.CO;2-L

Affiliations

1 NeuroPoly Lab, Polytechnique Montréal, Montreal, Quebec, Canada **2** Montreal Heart Institute, Montreal, Quebec, Canada **3** Unité de Neuroimagerie Fonctionnelle (UNF), Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montreal, Quebec, Canada **4** Mila - Quebec AI Institute, Montreal, QC, Canad **5** Centre de recherche du CHU Sainte-Justine, Université de Montréal, Montreal, QC, Canada **6** Magnetic Resonance Engineering Laboratory (MREL), 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 Centres, Liverpool, Australia **9** Department of Information Engineering, University of Padova, Padova, Italy **10** Institute of Neurobiology, Universidad Nacional Autónoma de México Campus Juriquilla, Querétaro, México **11** Philips Research Hamburg, Germany **12** Department of Radiology, Stanford University,

Stanford, California, United States **13** Medical Physics Unit, McGill University, Montreal, Canada **14** University of British Columbia, Vancouver, Canada **15** Department of Medical Imaging, McGill University Health Centre, Montrea, Quebec, Canada **16** Department of Radiology, McGill University, Montreal, Quebec, 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** Centre 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** Biomedical Engineering, Sahlgrenska University Hospital, Gothenburg, Sweden **25** Center for Mind/Brain Sciences, University of Trento, Italy **26** Hopital Maisonneuve-Rosemont, Montreal, Canada **27** Bioengineering, Imperial College London, UK **28** Radiotherapy and Imaging, Insitute of Cancer Research, Imperial College London, UK **29** Research Institute of the McGill University Health Centre, Montreal, Canada **30** Clinical Science, Philips Healthcare, Germany **31** Department of Radiological Sciences, University of California Los Angeles, Los Angeles, CA, USA **32** Physics and Biology in Medicine IDP, University of California Los Angeles, Los Angeles, CA, USA **33** Douglas Brain Imaging Centre, Montreal, Canada **34** Sunnybrook Research Institute, Toronto, Canada **35** Computer Science Department, Centro de Investigación en Matemáticas, A.C., Guanajuato, México **36** Medical Research Council, London Institute of Medical Sciences, Imperial College London, London, United Kingdom **37** Department of Radiation Oncology - CNS Service, The University of Texas MD Anderson Cancer Center, Texas, USA **38** Department of Biomedical Engineering, University of British Columbia, British Columbia, Canada **39** Center for Advanced Interdisciplinary Research, Ss. Cyril and Methodius University, Skopje, North Macedonia