NeuroLibre
Reproducible Preprints

Results of the ISMRM 2020 joint Reproducible Research & Quantitative MR study groups reproducibility challenge on phantom and human brain T1 mapping

*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 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**[36], **Ben Statton**[37], **Christian Stehning**[30], **Stefano Tambalo**[25], **Ye Tian**[6], **Chenyang Wang**[38], **Kilian Weis**[30], **Niloufar Zakariaei**[39], **Shuo Zhang**[30], **Ziwei Zhao**[6], **Nikola Stikov**[1,2,40]

* These authors contributed equally.

> 🚀 **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 reliability of an independently-implemented image acquisition protocol using inversion recovery in a standardized phantom and healthy human brains. Participants acquired T1 mapping data from MRIs of three manufacturers at 3T, resulting in 38 phantom datasets and 56 datasets from healthy human subjects. The robust imaging protocol and fitting algorithm demonstrated good reproducibility in both phantom and human brain T1 measurements. However, variations in implementation led to higher variance in reported values compared to intra-submission variability. 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.

Statement of need

T1 mapping is a widely used quantitative MRI technique that provides valuable information about tissue properties. However, the field faces a significant challenge due to the inconsistency of tissue-specific T1 values across different imaging protocols, sites, and vendors. This inconsistency hampers the comparability and reliability of T1 measurements, limiting their utility in both research and clinical applications. To address this critical issue, the ISMRM Reproducible Research study group (RRSG) and Quantitative MR study group (qMRSG) collaborated to launch the T1 mapping reproducibility challenge.

The primary objective of the challenge was to investigate whether independently-implemented image acquisition protocols at multiple centers could reliably measure T1 using inversion

recovery in a standardized phantom and in the brains of healthy volunteers. By evaluating the reproducibility of a well-established T1 mapping protocol and fitting algorithm from a reputable publication (Barral et al., 2010), the challenge aimed to identify sources of variability and establish best practices for achieving consistent and accurate T1 measurements.

A diverse group of participants was invited to acquire T1 mapping data on a standard ISMRM/NIST phantom and/or in healthy human brains using MRI scanners from three different manufacturers (Siemens, GE, Philips) operating at 3T, with one submission acquired at 0.35T. To enhance reproducibility and transparency, data submission, pipeline development, and analysis were conducted using open-source platforms. For participants collecting data at multiple sites, both inter-submission and intra-submission comparisons were performed by selecting one dataset per submission.

The results of the challenge were promising, with a total of 18 submissions accepted, consisting of 38 phantom datasets and 56 datasets from healthy human subjects. The mean coefficient of variation (CoV) for inter-submission phantom measurements was 6.1%, nearly twice as high as the evaluated intra-submission CoV of 2.9%. A similar trend was observed in the human data, where the inter-submission CoV for the genu was 6.0% compared to the intra-submission CoV of 2.9%, and for the cortical gray matter, the inter-submission CoV was 16% while the intra-submission CoV was 6.9%.

The evaluation of the imaging protocol and fitting algorithm based on Barral et al. (2010) demonstrated good reproducibility of both phantom and human brain T1 measurements. However, variations in the implementation of the protocol among the submissions led to higher variance in reported values relative to the intra-submission variability. This finding underscores the importance of standardized protocols and consistent implementation to ensure reliable and comparable T1 measurements across different imaging centers.

One of the major outcomes of the challenge was the creation of a large open database of inversion recovery T1 mapping acquisitions, which encompasses data acquired from multiple sites and MRI vendors. This database, accessible at osf.io/ywc9g/, holds significant value for the wider research community, enabling researchers to explore and engage with a comprehensive collection of T1 mapping data. To further facilitate access and utilization of the dataset, an interactive dashboard (Figure 1) was developed, accessible at https://rrsg2020.db.neurolibre.org.

Overall, this T1 mapping reproducibility challenge fills a critical need in the field by addressing the inconsistency of T1 values across different protocols, sites, and vendors. The findings and resources generated through this challenge will contribute to the standardization and improvement of T1 mapping techniques, promoting greater reliability and comparability of T1 measurements. Ultimately, these advancements will enhance the accuracy and clinical relevance of T1 mapping in various research and clinical applications, fostering advancements in precision medicine and improving patient care.
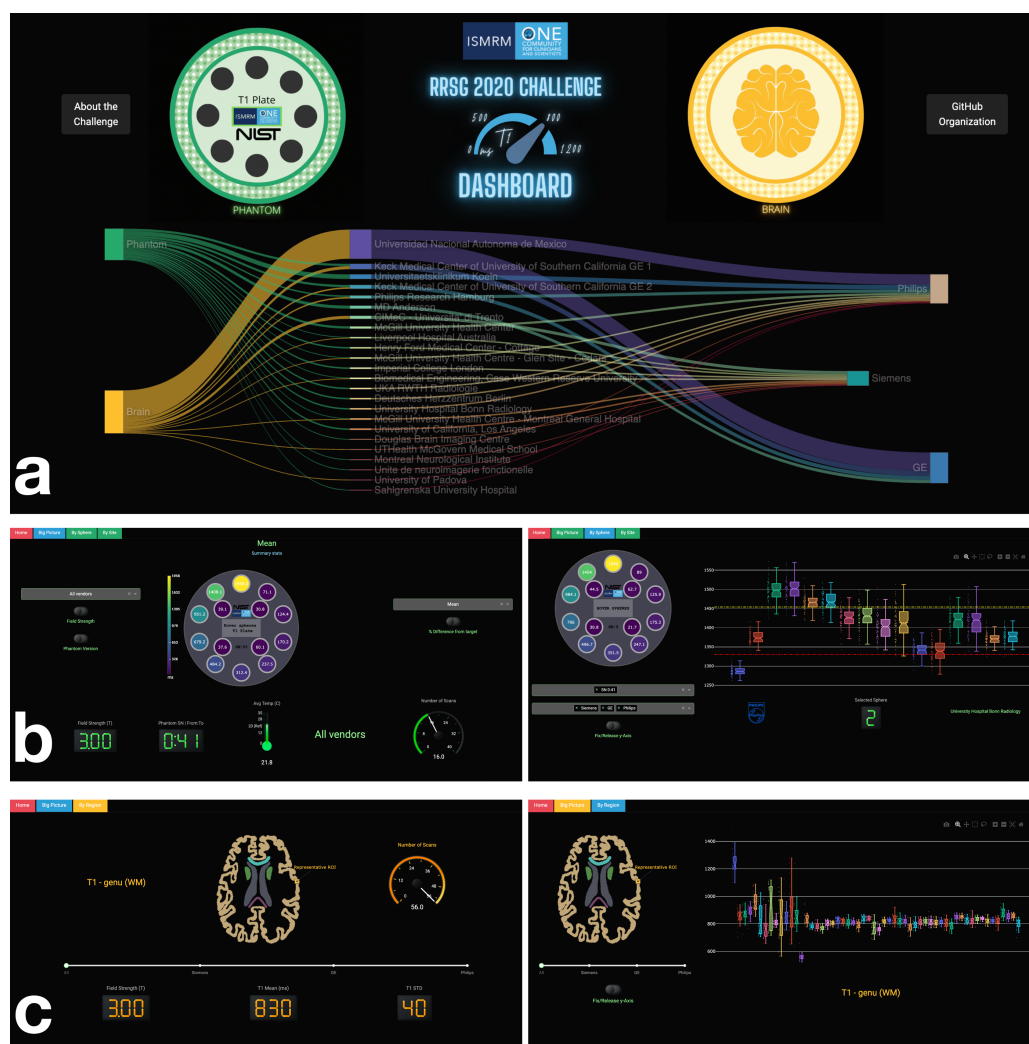
Figures

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

individuals who helped with the scanning at each imaging site.

> **! 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 quantitative T1 maps (Pykett & Mansfield, 1978) of tumors, but there is still disagreement in reported values for this fundamental parameter across different sites, vendors, and implementations (Stikov et al., 2015).

Among fundamental MRI parameters, T1 holds significant importance (Boudreau et al., 2020). It represents the time it takes for the longitudinal magnetization to recover after being disturbed by an RF pulse. The T1 value varies based on molecular mobility and magnetic field strength (Bottomley et al., 1984; Dieringer et al., 2014; Wansapura et al., 1999), making it a valuable parameter for distinguishing different tissue types. Accurate knowledge of T1 values is essential for optimizing clinical MRI pulse sequences for contrast and time efficiency (Ernst & Anderson, 1966; Redpath & Smith, 1994; Tofts, 1997) and as a calibration parameter for other quantitative MRI techniques (Sled & Pike, 2001; Yuan et al., 2012). Among the number of techniques to measure T1, inversion recovery (IR) (Drain, 1949; Hahn, 1949) is widely held as the gold standard technique, as it is robust against other effects (e.g. B1 inhomogeneity) and potential errors in measurements (e.g. insufficient spoiling) (Stikov et al., 2015). However, because the technique requires a long repetition time (TR > T1), it is very slow and impractical for whole-organ measurements, limiting its clinical use. In practice, it is mostly used as a reference to validate 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).

Efforts have been made to develop quantitative MRI phantoms to assist in standardizing T1 mapping methods (Keenan et al., 2018). A quantitative MRI standard system phantom was created in a joint project between the International Society for Magnetic Resonance in

Medicine (ISMRM) and the National Institute of Standards and Technology (NIST) (Stupic et al., 2021), and has since been commercialized (Premium System Phantom, CaliberMRI, Boulder, Colorado). The spherical phantom has a 57-element fiducial array containing spheres with doped liquids that model a wide range of T1, T2, and PD values. The reference values of each sphere were measured using NMR at 3T (Stupic et al., 2021). The standardized concentration for relaxometry values established as references by NIST are also used by another company for their relaxometry MRI phantoms (Gold Standard Phantoms Ltd., Rochester, England). The cardiac TIMES phantom (Captur et al., 2016) is another commercially available system phantom focusing on T1 and T2 values in blood and heart muscles, pre- and post-contrast. The ISMRM/NIST phantom has been used in several large multicenter studies already, for example in (Bane et al., 2018) where they compared measurements at eight sites on a single ISMRM/NIST phantom using the inversion recovery and VFA T1 mapping protocols recommended by NIST, as well as some site-specific imaging protocols used for dynamic contrast enhanced (DCE) imaging. Bane et al. (2018) concluded that the acquisition protocol, field strength, and T1 value of the sample impacted the level of accuracy, repeatability, and interplatform reproducibility that was observed. In another study led by NIST researchers (Keenan et al., 2021), T1 measurements were done at two clinical field strengths (1.5T and 3.0 T) and 27 MRI systems (three vendors) using the recommended NIST protocols. That study, which only investigated phantoms, found no significant relationship between T1 discrepancies of the measurements and the MRI vendors used.

The 2020 ISMRM reproducibility challenge [1] posed the following question:

> **Will an imaging protocol independently-implemented at multiple centers reliably measure what is considered one of the fundamental MR parameters (T1) using the most robust technique (inversion recovery) in a standardized phantom (ISMRM/NIST system phantom) and in the healthy human brain?**

More broadly, this challenge aimed at assessing the reproducibility of a qMRI method presented in a seminal paper, (Barral et al., 2010), by evaluating the variability in measurements observed by different research groups that implemented this imaging protocol. As the focus of this challenge was on reproducibility, the challenge design emphasized the use of reproducible research practices, such as sharing code, pipelines, data, and scripts to reproduce figures.

2 | METHODS

2.1 | Phantom and human data

The phantom portion of the challenge was launched for those with access to the ISMRM/NIST system phantom (Stupic et al., 2021) (Premium System Phantom, CaliberMRI, Boulder, Colorado). Two versions of the phantom have been produced with slightly different T1/T2/PD values in the liquid spheres, and both versions were used in this study. Phantoms with serial numbers 0042 or less are referred to as "Version 1", and those with 0043 or greater are "Version 2". The phantom has three plates containing sets of 14 spheres for ranges of proton density (PD), T1 (NiCl2), and T2 (MnCl2) values. Reference T1 values at 20 °C and 3T for the T1 plate are listed in {numref}`table1` for both versions of the phantom. Researchers that 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 after they had designed the phantom (prior to the challenge), and were provided to participants through the NIST website [2]. In brief, instructions included details about how to orient the phantom consistently at different sites, and how long the phantom should be in the scanner room prior to scanning so that a thermal equilibrium was achieved prior to scanning.

---

[1] ISMRM blog post announcingn the RRRSG challenge
[2] The website provided to the participants has since been removed from the NIST website.

Participants were also instructed to collect T1 maps in the brains of healthy human participants, if possible. To ensure consistency across datasets, single-slice positioning parallel to the anterior commissure - posterior commissure (AC-PC) line was recommended. Before the scanning process, the participants granted their consent [3] to share their de-identified data openly with the challenge organizers and on the website Open Science Framework (OSF.io). As the submitted single-slice inversion recovery images would be along the AC-PC line, they are unlikely to contain sufficient information for facial identification, and therefore participants were not instructed to de-face their data. The researchers who submitted human data for the challenge 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 sharing their data as described in the challenge.

2.2 | Protocol

Participants were instructed to acquire the T1 mapping data using the spin-echo inversion recovery protocol for T1 mapping as reported in (Barral et al., 2010), and detailed in {num-ref}table2. This protocol uses four inversion times optimized for human brain T1 values and uses a relatively short TR (2550 ms). It is important to note that this acquisition protocol is not suitable for T1 fitting models that assume TR > 5T1. Instead, more general models of inversion recovery, such as the Barral et al. (2010) fitting model described in Section 2.4.1, can be used to fit this data.

Researchers who participated in the challenge were advised to adhere to this protocol as closely as possible, and to report any differences in protocol parameters due to technical limitations of their scanners and/or software. It was recommended that participants submit complex data (magnitude and phase, or real and imaginary), but magnitude-only data was also accepted if complex data could not be conveniently exported.

2.3 | Data Submissions

Data submissions for the challenge were managed through a dedicated repository on GitHub, accessible at https://github.com/rrsg2020/data_submission. This allowed transparent and open review of the submissions, as well as standardization of the process. All datasets were converted to the NIfTI file format, and images from different TIs needed to be concatenated into the fourth (or "time") dimension. Magnitude-only datasets required one NIfTI file, while complex datasets required two files (magnitude and phase, or real and imaginary). Additionally, a YAML (*.yaml) configuration file containing submission, dataset, and acquisition details (such as data type, submitter name and email, site details, phantom or volunteer details, and imaging protocol details) was required for each submitted dataset to ensure that the information was standardized and easily found. Each submission was reviewed to confirm that guidelines were followed, and then datasets and configuration files were uploaded to OSF.io (osf.io/ywc9g). A Jupyter Notebook (Beg et al., 2021; Kluyver et al., 2016) pipeline was used to generate T1 maps using qMRLab (Cabana et al., 2015; Karakuzu et al., 2020) and quality-check the datasets prior to accepting the submissions; in particular, we assured that the NiCL2 array was imaged, that the DICOM images were correctly converted to NIfTI, the each images for each acquired TI were not renormalized (in particular, Philips platforms have different image export options that changes how the images are scaled, and a reconversion was necessary to ensure proper scaling for quantitative imaging in some cases – see the submissions GitHub issue #5 for one example [4]) for the purposes of quality assurance. Links to the Jupyter Notebook for reproducing the T1 map were shared using the MyBinder platform in each respective submission GitHub issue, ensuring that computational environments (eg, software dependencies and packages) could be reproduced to re-run the pipeline in a web browser.

---

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

[4]rrsg2020/data_submission issue #5

2.4 | Data Processing

2.4.1 | Fitting Model and Pipeline

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, introduced in , models the general T1 signal equation without the long-TR approximation. The a and b constants inherently factor TR in them, as well as other imaging parameters (eg, excitation and refocusing flip angles, TE, etc). Barral et al. (2010) shared the implementation of the fitting algorithm used in their paper [^their-paper](http://www-mrsrl.stanford.edu/~jbarral/t1map.html). Magnitude-only data were fitted to a modified-version of [1] (Eq. 15 of Barral et al. (2010)) with signal-polarity restoration. To facilitate its use in our pipelines, a wrapper was implemented around this code available in the open-source software qMRLab (Cabana et al., 2015; Karakuzu et al., 2020), which provides a commandline interface (CLI) to call the fitting in MATLAB/Octave scripts.

A Jupyter Notebook data processing pipeline was written using MATLAB/Octave. This pipeline automatically downloads all the datasets from the data-hosting platform osf.io (osf.io/ywc9g), loads each dataset configuration file, fits the T1 data voxel-wise, and exports the resulting T1 map to the NIfTI and PNG formats. This pipeline is available in a GitHub repository (https://github.com/rrsg2020/t1_fitting_pipeline, filename: RRSG_T1_fitting.ipynb). Once all submissions were collected and the pipeline was executed, the T1 maps were uploaded to OSF (osf.io/ywc9g).

2.4.2 | Image Labeling & Registration

A schematic of the phantom is shown in Figure 1-a. 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-b). 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.

Figure 1 Schematic of the system phantom (a) used in this challenge. Reproduced and cropped from Stupic et al. 2021 (Stupic et al. 2021) (Creative Commons CC BY license). ROI selection for the ISMRM/NIST phantom (b) and the human brain (c). b) The 14 phantom ROIs (shades of blue/green) were automatically generated using a script provided by NIST. In yellow are the three reference pins in the phantom, i.e. these are not ROIs or spheres. c) Human brain ROIs were manually segmented in four regions: the genu (yellow, 5 5 voxels), splenium (green, 5 5 voxels), deep gray matter (blue, 5 5 voxels), and cortical gray matter (red, three sets of 3 3 voxels). Note: due to differences in slice positioning from the single-slice datasets provided by certain sites, for some datasets it was not possible to manually segment an ROI in the genu or deep gray matter. In the case of the missing genu, left or right frontal white matter (WM) was selected; for deep gray matter (GM), it was omitted entirely for those cases.

Manual ROIs were segmented by a single researcher (M.B., 11+ years of neuroimaging experience) using FSLeyes (McCarthy, 2022) in four regions for the human datasets Figure 1-c): 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.

2.4.3 | Analysis and Statistics

Analysis code and scripts were developed and shared in a version-tracked public GitHub repository [^public-repo](https://github.com/rrsg2020/analysis). T1 fitting and main data analysis was performed for all datasets by one of the challenge organizers (M.B.). Python-based Jupyter Notebooks were used for both the quality assurance and main analysis workflows. The computational environment requirements were containerized in Docker (Boettiger, 2015; Merkel, 2014), allowing for an executable environment that can reproduce the analysis in a web browser through MyBinder [5] (**Project?**_Jupyter2018-ll). Python scripts handled reference data, database handling, ROI masking, and general analysis tools, while configuration files managed the dataset information which were downloaded and pooled using a script (make_pooled_datasets.py). The databases were created using a reproducible Jupyter Notebook and subsequently saved in the repository.

For the ISMRM/NIST phantom data, mean T1 values 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). This comparison was repeated for individual measurements at each site and for all measurements grouped together. Temperature correction was carried out via nonlinear interpolation [6] of the set of reference NIST T1 values between 16 °C and 26 °C (2 °C intervals), listed in the phantom technical specifications. For the human datasets, a notebook was created to plot the mean and standard deviations for each tissue ROI from all submissions from all sites. All quality assurance and analysis plot images were saved to the repository for ease-of-access and a timestamped version-controlled record of the state of the analysis figures. The database files of ROI values and acquisition details for all submissions were also saved to the repository.

An interactive dashboard [7] was developed in Dash by Plotly (Plotly Technologies Inc. 2015) and hosted by NeuroLibre (Karakuzu, DuPre, et al., 2022) to enable real-time exploration of the data, analysis, and statistics of the challenge results. The dashboard reports descriptive statistics for a variety of alternative looks at phantom and brain data, as well as some statistical comparisons (e.g., the hierarchical shift function [8]). The data was collected from the pre-prepared databases of masked ROI values and incorporated other database information, such as phantom version, temperature, MRI system, and reference values. The interactive dashboard displays these results for all measurements at all sites.

4 | DISCUSSION

4.1 | Achievements of the challenge

The challenge focused on exploring the reproducibility of the gold standard inversion recovery T1 mapping method reported in a seminal paper (Barral et al., 2010). Eighteen submissions independently implemented the inversion recovery T1 mapping acquisition protocol as outlined in Barral et al. (2010)(which is optimized for the T1 values observed in brain tissue), 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. A standardized T1 processing pipeline was developed for different dataset types, including magnitude-only and complex data. Additionally, Jupyter notebooks that can be executed in containerized environments

---

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

[6]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.

[7]Interactive dashboard: https://rrsg2020.db.neurolibre.org, code repository: https://github.com/rrsg2020/rrsg2020-dashboard

[8]The hierarchical shift function compares distributions throughout their range across multiple dependent measurements. More information can be found in this article, (Wilcox & Rousselet, 2023), and in this blog post.

were developed for quality assurance, visualization, and analyses. An interactive web-based dashboard was also developed to allow for easy exploration of the challenge results in a web-browser.

To evaluate the accuracy of the resulting T1 values, the challenge used the standard ISMRM/NIST phantom with fiducial spheres having T1 values in the range of human brain tissue, from 500 to 2000 ms (see Figure 5). As anticipated for this protocol, there was a decrease in the accuracy in measurements for spheres with T1 below 300 ms. Overall, the majority of the independently implemented imaging protocols from various sites are consistent with the temperature-corrected reference values, with only a few exceptions. Using the NIST phantom, we report that sites that independently implemented the imaging protocol resulted in an inter-submission mean CoV (6.1 %) that was twice as high as the intra-submission mean CoV measured at seven sites (2.9 %). A similar trend was observed in vivo. Inter-submission CoV for WM (genu) was 6.0 % and for GM (cortex) was 16.5 % vs the intra-submission CoV that was 2.9 % and 6.9%, with generally higher CoVs relative to the phantom measurements likely due to biological variability (Piechnik et al., 2013; Stanisz et al., 2005).

4.2 | Comparison with other studies

The work done during this challenge involved a multi-center quantitative T1 mapping study using the NIST phantom across various sites. This work overlaps with two recent studies (Bane et al., 2018; Keenan et al., 2021). Bane et al. (2018) focused on the reproducibility of two standard quantitative T1 techniques (inversion recovery and variable flip angle) and a wide variety of site-specific T1 mapping protocols for DCE, mostly VFA protocols with fewer flip angles, which were implemented at eight imaging centers covering the same 3 MRI vendors featured in this challenge (GE/Philips/Siemens). The inter-platform coefficient of variation for the standard inversion recovery T1 protocol was 5.46% at 3 T in (Bane et al., 2018), which was substantially lower than what they observed for their standard VFA protocol (22.87%). However, Bane et al.'s work differed from the challenge in several ways. First, the standard imaging protocol for inversion recovery used by Bane et al. (2018) had more inversion times (14 compared to the challenge's 4) to cover the entire range of T1 values of the phantom. Secondly, Bane et al. (2018) used a single traveling phantom for all sites, whereas the challenge used a total of 8 different phantoms (some were shared amongst people who participated independently). Thirdly, Bane et al. (2018) averaged the signals within each ROI of each sphere prior to fitting for the T1 values, whereas the challenge pipeline fits the T1 values on a per-voxel basis and only subsequently calculates the mean/median/std. They also only acquired magnitude data, in contrast to the challenge where participants were encouraged to submit both complex and magnitude-only data. Lastly, in Bane et al. (2018), the implementations of the common inversion recovery protocols were fully standardized (full protocol) across all the platforms (except for two cases where one manufacturer couldn't achieve the lowest TI) and imposed and coordinated by the principal researchers. In contrast, the challenge sought to explore the variations that would occur for a less-restricted protocol (Table 2) that is independently-implemented at multiple centers, which more closely emulates the quantitative MR research flow (publication of a technique and protocol → independently implement the pulse sequence and/or protocol → use the new implementation independently in a study → publish). Of note, in the challenge, one participating group coordinated a large multicenter dataset that mirrors the study by Bane et al. (2018) by imaging a single phantom across 7 different imaging sites, albeit doing so on a single manufacturer. Using this subset, the mean cross-site CoV was 2.9 % (range: 1.6 - 4.9 %) for the first five spheres, which is in agreement with the range of observations for all spheres by Bane et al. (Bane et al. 2018) at 3T using their full inversion recovery protocol (CoV = 5.46 %; range: 0.99 - 14.6 %).

Another study by Bane et al. (2018); Keenan et al. (2021) also investigated the accuracy of T1 mapping techniques using a single ISMRM/NIST system phantom at multiple sites and on multiple platforms. Like Bane et al. (2018) they used an inversion recovery imaging protocol optimized for the full range of T1 values represented in the ISMRM/NIST phantom, which

consisted of 9 to 10 inversion times and a TR of 4500 ms (TR ~ 5T1 of WM at 3T). They reported no consistent pattern of differences in measured inversion recovery T1 values across MRI vendors for the two T1 mapping techniques they used (inversion recovery and VFA). They observed relative errors between their T1 measurements and the reference values of the phantom to be below 10% for all T1 values and the larger errors were observed at the lowest and highest T1 values of the phantom.

4.3 | Lessons Learned and Future Directions

There are some important things to note about this challenge. Firstly, the submissions for this challenge were due in March 2020, which was impacted by the COVID-19 pandemic lockdowns, thereby reducing repeated experiments due to access limitations. Nevertheless, a substantial number of participants submitted their datasets. Some groups intended on acquiring more data, and others intended on re-scanning volunteers, but could no longer do so due to local pandemic restrictions.

This reproducibility challenge aimed to compare differences between independently-implemented protocols. Crowning a winner was not an aim of this challenge, due to concerns that participants would have changed their protocols to get closer to the reference T1 values, leading to a broader difference in protocol implementations across MRI sites. Instead, we focused on building consensus by creating an open data repository, sharing reproducible workflows, and presenting the results through interactive visualizations. Future work warrants the study of inter-site differences in a vendor-neutral workflow (Karakuzu, Biswas, et al., 2022) by adhering to the latest Brain Imaging Data Structure (BIDS) community data standard on qMRI (Karakuzu, Appelhoff, et al., 2022).

References

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

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

Captur, G., Gatehouse, P., Keenan, K. E., Heslinga, F. G., Bruehl, R., Prothmann, M., Graves, M. J., Eames, R. J., Torlasco, C., Benedetti, G., Donovan, J., Ittermann, B., Boubertakh, R., Bathgate, A., Royet, C., Pang, W., Nezafat, R., Salerno, M., Kellman, P., & Moon, J. C. (2016). A medical device-grade T1 and ECV phantom for global T1 mapping quality assurance-the T1 mapping and ECV standardization in cardiovascular magnetic resonance (T1MES) program. *J. Cardiovasc. Magn. Reson.*, *18*(1), 58. https://doi.org/10.1186/s12968-016-0280-z

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

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

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., Appelhoff, S., Auer, T., Boudreau, M., Feingold, F., Khan, A. R., Lazari, A., Markiewicz, C., Mulder, M., Phillips, C., & others. (2022). qMRI-BIDS: An extension

to the brain imaging data structure for quantitative magnetic resonance imaging data. *Scientific Data*, *9*(1), 517. https://doi.org/10.1038/s41597-022-01571-4

Karakuzu, A., Biswas, L., Cohen-Adad, J., & Stikov, N. (2022). Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. *Magnetic Resonance in Medicine*, *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

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

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

McCarthy, P. (2022). *FSLeyes* (Version 1.4.0). Zenodo. https://doi.org/10.5281/zenodo.6511596

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

Piechnik, S. K., Ferreira, V. M., Lewandowski, A. J., Ntusi, N. A. B., Banerjee, R., Holloway, C., Hofman, M. B. M., Sado, D. M., Maestrini, V., White, S. K., Lazdam, M., Karamitsos, T., Moon, J. C., Neubauer, S., Leeson, P., & Robson, M. D. (2013). Normal variation of magnetic resonance T1 relaxation times in the human population at 1.5 T using ShMOLLI. *J. Cardiovasc. Magn. Reson.*, *15*(1), 13. https://doi.org/10.1186/1532-429X-15-13

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

Stanisz, G. J., Odrobina, E. E., Pun, J., Escaravage, M., Graham, S. J., Bronskill, M. J., & Henkelman, R. M. (2005). T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magn. Reson. Med.*, *54*(3), 507–512. https://doi.org/10.1002/mrm.20605

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. *Magnetic Resonance in Medicine*, *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

Wilcox, R. R., & Rousselet, G. A. (2023). An updated guide to robust statistical methods in neuroscience. *Curr Protoc*, *3*(3), e719. https://doi.org/10.1002/cpz1.719

Yuan, J., Chow, S. K. K., Yeung, D. K. W., Ahuja, A. T., & King, A. D. (2012). Quantitative evaluation of dual-flip-angle T1 mapping on DCE-MRI kinetic parameter estimation in head and neck. *Quant. Imaging Med. Surg.*, *2*(4), 245–253. https://doi.org/10.3978/j.issn.2223-4292.2012.11.04

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** 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** National Laboratory for Magnetic Resonance Imaging, Institute of Neurobiology, Universidad Nacional Autónoma de México (UNAM), Juriquilla, Mexico **36** Computer Science Department, Centro de Investigación en Matemáticas, A.C., Guanajuato, México **37** Medical Research Council, London Institute of Medical Sciences, Imperial College London, London, United Kingdom **38** Department of Radiation Oncology - CNS Service, The University of Texas MD Anderson Cancer Center, Texas, USA **39** Department of Biomedical Engineering, University of British Columbia, British Columbia, Canada **40** Center for Advanced Interdisciplinary Research, Ss. Cyril and Methodius University, Skopje, North Macedonia