NeuroLibre
Reproducible Preprints

# Longitudinal stability of brain and spinal cord quantitative MRI measures

**Mathieu Boudreau** [1¶], **Agah Karakuzu** [1], **Arnaud Boré** [2,3], **Basile Pinsard** [2,3], **Kiril Zelenkovski** [4], **Eva Alonso-Ortiz** [1,5], **Julie Boyle** [2,3], **Pierre Bellec** [2,3,6], and **Julien Cohen-Adad** [1,3,5,7]

**1** NeuroPoly, Polytechnique Montreal, Montreal, QC, Canada **2** Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montreal, QC, Canada **3** Unité de Neuroimagerie Fonctionnelle (UNF), Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montreal, QC, Canada **4** Faculty of Computer Science and Engineering (FINKI), Skopje, Macedonia **5** Centre de recherche du CHU Sainte-Justine, Université de Montréal, Montreal, QC, Canada **6** Psychology Department, Université de Montréal, Montreal, QC, Canada **7** Mila - Quebec AI Institute, Montreal, QC, Canada ¶ Corresponding author

**DOI:** 10.xxxxxx/draft

**Reproducible Preprint**

- Jupyter Book

**Code**

- Technical Screening
- Submitted Repository

**Reproducibility Assets**

- Repository
- Dataset
- Jupyter Book
- Container

**Moderator:** NeuroLibre
**Screener(s):**

- @roboneuro

**Submitted:** 01 January 1970
**Published:** unpublished







## Summary

We present the initial data release of the Courtois project on neural modeling (CNeuroMod), with a specific focus on the quantitative MRI (qMRI) component. The primary objective of this study was to evaluate the longitudinal stability of qMRI measurements in both the brain and cervical spinal cord.

To achieve this, we conducted regular scanning sessions over a three-year period involving six participants Figure 1. Each participant underwent up to ten sessions, providing us with a robust dataset. Our brain qMRI imaging protocols consisted of T1, magnetization transfer (MTR, MTsat), and diffusion techniques. In addition to these, the spinal cord imaging protocol included T1w, T2w, and T2*w cross-sectional area (CSA) measurements.

The results of our study demonstrate the stability of the qMRI protocols used for both the brain and spinal cord. These findings offer valuable insights for the design of future longitudinal clinical studies in this domain. Furthermore, we have developed reproducible and reusable analysis pipelines for structural qMRI of the brain and spinal cord. These pipelines incorporate cutting-edge tools such as FSL, ANTs, qMRLab, and SCT, ensuring robust and accurate analysis.

To enhance the accessibility and dissemination of our work, we have presented our findings as an interactive article using Jupyter Book and Plotly. This format allows for seamless exploration and sharing of the curated findings within an integrated research object. We believe that this approach will facilitate collaboration and encourage further research in the field of qMRI analysis.

Overall, the initial data release of the Courtois project on neural modeling (CNeuroMod), specifically focusing on the quantitative MRI (qMRI) component, provides a significant contribution to the understanding of the longitudinal stability of qMRI measurements in the brain and spinal cord. The study offers valuable insights for future longitudinal clinical studies

and establishes reproducible analysis pipelines for structural qMRI. The interactive article format ensures easy accessibility and encourages collaboration among researchers.

# Figures

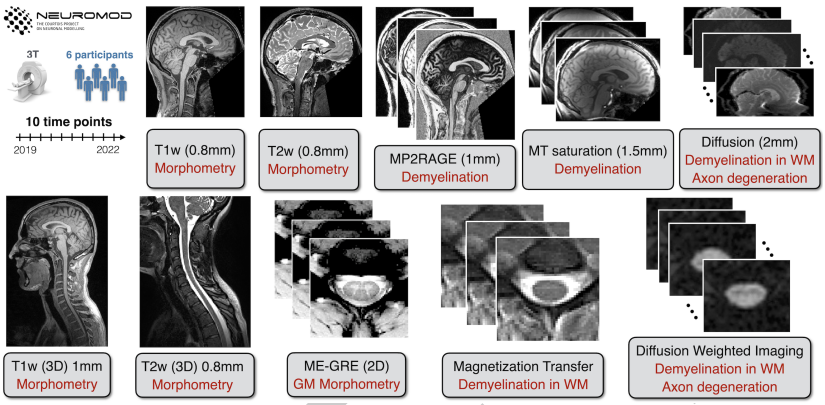


**Figure 1:** Overview of the structural dataset for the Courtois project on neural modelling (CNeuroMod). 6 participants were scanned up to ten times over three years; note that this is an initial data release for 2022, and more scans are regularly being acquired. The structural protocol consists of T1w, T2w and T2*w scans to quantify brain and SC (including grey matter, GM) morphometry, and MP2RAGE, magnetization transfer (MTR and MTsat), and diffusion-weighted sequences to compute metrics sensitive to demyelination in the white matter (WM).

# Acknowledgements

The Courtois project on neural modelling was made possible by a generous donation from the Courtois foundation. The Courtois NeuroMod team is based at "Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal", with several other institutions involved. See the CNeuromod documentation for an up-to-date list of contributors (https://docs.cneuromod.ca). This study was also funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging [950-230815], the Canadian Institute of Health Research [CIHR FDN-143263], the Canada Foundation for Innovation [32454, 34824], the Fonds de Recherche du Québec - Santé [322736], the Natural Sciences and Engineering Research Council of Canada [RGPIN-2019-07244], the Canada First Research Excellence Fund (IVADO and TransMedTech), and the Mila - Tech Transfer Funding Program.

**NOTE**

***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.

NeuroLibre Reproducible Preprints

(2022). NeuroLibre is sponsored by the Canadian Open Neuroscience Platform (CONP) (Harding et al., 2023).

Longitudinal stability of brain and spinal cord quantitative MRI measures

Mathieu Boudreau1, Agah Karakuzu1, Arnaud Boré2,3, Basile Pinsard2,3, Kiril Zelenkovski4, Eva Alonso-Ortiz1,5, Julie Boyle2,3, Pierre Bellec2,3,6, Julien Cohen-Adad1,3,5,7

1NeuroPoly, Polytechnique Montreal, Montreal, QC, Canada, 2Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montreal, QC, Canada 3Unité de Neuroimagerie Fonctionnelle (UNF), Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montreal, QC, Canada 4Faculty of Computer Science and Engineering (FINKI), Skopje, Macedonia 5Centre de recherche du CHU Sainte-Justine, Université de Montréal, Montreal, QC, Canada 6Psychology Department, Université de Montréal, Montreal, QC, Canada 7Mila - Quebec AI Institute, Montreal, QC, Canada

## Abstract

Quantitative MRI (qMRI) promises better specificity, accuracy, and stability relative to its clinically-used qualitative MRI counterpart. Longitudinal stability is particularly important in qMRI. The goal is to reliably quantify tissue properties that may be assessed in longitudinal clinical studies throughout disease progression or during treatment. In this work, we present the initial data release of the quantitative MRI portion of the Courtois project on neural modelling (CNeuroMod), where the brain and cervical spinal cord of six participants were scanned at regular intervals over the course of several years. This first release includes three years of data collection and up to ten sessions per participant using quantitative MRI imaging protocols (T1, magnetization transfer (MTR, MTsat), and diffusion). Coefficient of variations (COV) over this timeframe ranged between 0.6% to 2.3% (intrasubject) and 0.4% to 3.5% (intersubject) for T1/MTR/MTsat in whole-brain white matter (WM), and between 0.6% to 1.3% (intrasubject) and 3.0% to 10.3% (intersubject) for diffusion FA/MD/RD in the three corpus callosum regions. In the spine, COVs ranged between 2.3% and 4.5% (intrasubject) and 5.1% to 9.7% (intersubject) for measured spine WM cross-sectional area (CSA) across the C2 and C3 vertebral levels, and between 3.9% to 9.5% (intrasubject) and 4.0% to 8.4% (intersubject) in WM across the C2 and C5 vertebral levels for all qMRI metrics (T1, MTR, MTsat, FA, MD, RD). Results from this work show the level of stability that can be expected from qMRI protocols in the brain and spinal cord, and could help in the design of future longitudinal clinical studies.

# 1 | INTRODUCTION

## Quantitative MRI and the reproducibility crisis

Conventional MRI images used clinically stem from using the MRI machine as a non-invasive medical device and not as a scientific instrument (Cercignani et al., 2018; Tofts, 1998). Medical images produced from clinical MRI protocols must be interpreted by expert readers to extract useful diagnostic information, as the images alone lack biological specificity and reproducibility, due to underlying changes in biology and the electromagnetic fields the imaging hardware generates. Quantitative MRI (qMRI) techniques (Nicole Seiberlich et al., 2020) aim to produce measurements of biological or physical properties through a series of carefully planned conventional MRI images. Quantitative maps are calculated or fit from these measured datasets, which have voxelwise values that typically have physical units associated with them, for example, spin-lattice relaxation time (T1 [s]), spin-spin relaxation time (T2 [s]), myelin water fraction (MWF [%]), magnetization transfer ratio (MTR [%]), cerebral blood flow (CBF [ml/g/min]) and diffusion (restricted diffusion coefficients [mm2/s], eg. mean diffusivity (MD) and radial diffusivity (RD)). Some qMRI techniques are highly specific to certain biological

changes (eg, myelin loss (Mancini et al., 2020; Schmierer et al., 2007), cerebrovascular diseases and oxygen consumption disorders (Davis et al., 1998; Y. Ma et al., 2016; Mazerolle et al., 2018; Wang et al., 2017), iron deficiency (Lidén et al., 2021; Ropele et al., 2011), etc.). Because these measures either implicitly or explicitly account for effects that typically are unaccounted for in clinical MRI images, in principle they should have improved stability – this is one of the hallmark-promising features of qMRI. However, in practice, the field has fallen short of living up to this high bar. Even fundamental quantitative MRI techniques have been shown to vary widely amongst methods and sites; for example, despite the fact that T1 mapping is the first quantitative MRI technique to have been developed 45 years ago (Pykett & Mansfield, 1978), modern T1 mapping techniques have not consistently shown good accuracy in measuring T1 values in the brain across different sites or techniques (Stikov et al., 2015). A lot of work has been done recently to help quantify the accuracy and improve within-vendor stability of quantitative MR measurements, such as the development of quantitative MRI calibration phantoms (Golay & Oliver-Taylor, 2022; Keenan et al., 2018; Stupic et al., 2021) and increasing integration of quantitative MRI pulse sequences as stock sequences on commercial scanners (D. Ma et al., 2013; Marques et al., 2010; N. Seiberlich et al., 2012) or as vendor-neutral implementations (Herz et al., 2021; Karakuzu, Biswas, et al., 2022).

## Stability in qMRI: why is it needed?

The stability of a qMRI measurement is an important characteristic to consider when designing longitudinal studies, particularly when clinical features are expected to evolve over time (eg, worsening disease, or improvement through therapeutic intervention (Oh et al., 2021)). It is also important to know the anticipated variability of these metrics to find the minimum detectable effect size in a power analysis while designing your study. Same-day test-retest studies have shown that fundamental qMRI metrics (eg, T1, T2) exhibit low intra-scanner variability in vivo (on the order of 1-2%) (Gracien et al., 2020; Lee et al., 2019). However, test-retest studies are limited in their usefulness as a stability measure because they only consist of two measurements (leading to improper standard deviation calculations) and are done during the same day (same scanner operator, same scanner conditions), which are not realistic conditions experienced during longitudinal studies. Longitudinal stability is thus important to quantify, but can be challenging due to the potential confounds from actual changes of the subject's tissue properties over time, even from healthy volunteers. Quantitative MRI metrics in the brain have been shown to correlate with ageing through adulthood (Erramuzpe et al., 2021; Seiler et al., 2020), although changes appear to happen slowly (over decades) and thus short-term longitudinal studies (eg, 3-5 years) should in principle quantify longitudinal stability reliably.

## Stability in (q)MRI: what's been done

Many studies have investigated the stability of morphometrics and quantitative MRI measures. A recent landmark study investigated the longitudinal stability of clinical and functional MRI metrics of a single subject's brain measured on multiple vendors at multiple sites over the course of 15 years (73 sessions across 36 scanners) (Duchesne et al., 2019), finding poor reproducibility across MRI manufacturers for key clinical metrics (ie., white/grey matter contrast-to-noise ratio (CNR), FLAIR white matter hyperintensities volume). For qMRI metrics, there are a few longitudinal studies that have probed different aspects of their longitudinal stability. A 7-year scan-rescan brain ageing study explored the evolution of quantitative T1 values in different tissues using the variable flip angle (VFA) technique (which depends on an additional B1 map) (Gracien et al., 2017) and found T1 values were sensitive to ageing for this timespan. The stability of quantitative brain metrics when encountering MRI software and hardware upgrades was recently explored in a four time-point, seven-year repeatability and reproducibility study (Salluzzi et al., 2022), which reported the upgrades did not affect the effect size and stability of the tested MRI biomarkers. Stability has also been explored in non-brain anatomy. For spinal cord, inter-vendor variability was recently probed by a multi-center (19 sites) study

using a generic quantitative MRI spinal cord imaging protocol (Cohen-Adad et al., 2021a) on a single participant over the span of one year (Cohen-Adad, 2020). A test-retest quantitative MRI spine study has also been performed in two cohorts (young adult and elderly) over a ten month period (Simon Lévy et al., 2018), with minimal detectable changes reported for T1, MTR, MTsat, and macromolecular tissue volume (MTV) quantitative MRI measures.

### Study Objective and the CNeuroMod Project

The objective of this study was to measure and report the stability of quantitative microstructure MRI measurements across multiple time points in the brain and cervical spinal cord. To do this, two sets of qMRI protocols (brain and spinal cord) were integrated within the Courtois project on neural modelling (CNeuroMod)[1] for collecting longitudinal data on healthy subjects to train and improve artificial intelligence models on brain behaviour and activity. The qMRI measurements of the brain and spinal cord fell within the "anatomical" imaging branch of the CNeuroMod project, and additional branches of data acquired include deep scanning with functional MRI, biosignals (eg, cardiac, respiration, eye tracking), and magnetoencephalography (MEG). In addition, we developed reproducible and reusable analysis pipelines for structural qMRI of the brain and spinal cord. These pipelines are built using state-of-the-art tools in terms of pipeline management (NextFlow (Di Tommaso et al., 2017)), structural data analyses (FSL (Smith et al., 2004), ANTs (Avants et al., 2009), qMRLab (Cabana et al., 2015; Karakuzu et al., 2020), SCT (De Leener et al., 2017), etc.) and Jupyter notebooks (Beg et al., 2021) with Plotly (Plotly Technologies Inc., 2015) for presenting curated and interactive results.

## 2 | RESULTS

Six participants were repeatedly scanned on a 3T MRI scanner (Prisma Fit, Siemens, Erlangen, Germany) approximately four times a year (up to ten times for this initial 2022 data release, with more scans regularly being acquired). Custom headcases (Caseforge, Berkeley, USA) were used for each participant to minimise movements during the imaging sessions. Two sets of imaging protocols were acquired (Figure 1), one for the brain (T1w, T2w, MP2RAGE, MTsat, B1+, and diffusion) and one for the spinal cord (T1w, T2w, MTsat, and diffusion).

FIGURE 1 Overview of the structural dataset for the Courtois project on neural modelling (CNeuroMod). 6 participants were scanned up to ten times over three years; note that this is an initial data release for 2022, and more scans are regularly being acquired. The structural protocol consists of T1w, T2w and T2*w scans to quantify brain and SC (including grey matter, GM) morphometry, and MP2RAGE, magnetization transfer (MTR and MTsat), and diffusion-weighted sequences to compute metrics sensitive to demyelination in the white matter (WM).

### 2.1 | Brain

Average quantitative MRI (excluding diffusion) values for the segmented whole-brain white matter (WM) and grey matter (GM) for each subject and session are shown in Figure 2. Missing data points are either unacquired sessions or because they were excluded after doing quality control, more details are listed in the "Quality Control" section. Note that MTR is calculated from a subset of the MTsat measurements, and B1 is not shown because it is only used as a transmit radiofrequency (RF) field correction factor for the MTsat measurement, and does not have biological specificity.

FIGURE 2 Brain qMRI metrics (excluding diffusion). Each point represents the mean metric within the WM or GM for one subject and one session. Missing data points are due to unacquired sessions, the pipelines failing to produce an output, or were excluded due to quality control (see Quality Control section for more details). The intra- and inter- subject COVs for

[1] Please see https://www.cneuromod.ca.

these metrics in WM and GM are shown inside each respective plot. Note: subject 4 stopped participating after their fifth session for reasons out of our control.

From Figure 2, it is evident that mean T1 values measured with the MP2RAGE pulse sequence (calculated from 2 images) generally showed less intrasubject variation than T1 values measured with MTsat (calculated from five images: three for MTsat calculation and two for B1 calculation). Intrasubject COV means for WM T1 measured using MP2RAGE was 0.6 %, which is four times lower than for T1 measured using MTsat. Intrasubject COVs for WM MTR (calculated from two images) were similar to those from MP2RAGE, and three times lower than MTsat (MTR is a subset of MTsat measurements, with two out of the five MTsat measurements being shared). Intrasubject COV standard deviations (STD) (not displayed in figure [2]) were low for all metrics in WM ($< 1\%$). Intersubject mean COV was highest for WM T1 calculated from MTsat at 3.5%, and lowest for MTR at 0.4 %. GM intrasubject and intersubject COVs followed similar trends to those for WM, with the same order of magnitude COV mean and STD values. The very low intrasubject COVs and larger intersubject COV for T1 (MP2RAGE) is also expressed as each subject having specific mean whole-brain WM and GM T1 values distinct from each other, and that these values were stable longitudinally (Figure 2); this can also be seen to a lesser extent for T1 (MTsat) and MTsat, but not for MTR which had intrasubject COVs on the order or higher than the intersubject COVs.

FIGURE 3 The mean diffusion metrics (FA, MD, and RD) for each acquired session are shown for three atlas-based regions of the corpus callosum (genu in blue, body in yellow, splenium in green) of each subject.

Figure 3 displays the three calculated diffusion metrics (fractional anisotropy: FA, mean diffusivity: MD, and radial diffusivity: RD) within the three corpus callosum regions (genu, body, splenium). All three metrics exhibited high intersubject mean COVs ($> 3\%$) and low intrasubject COV means ($< 1.3\%$). The lowest intrasubject COV means are reported for FA in the body and splenium (0.6%), and the lowest intersubject mean COV was reported in the body and splenium for MD (3.0% and 3.1%, respectively). Intrasubject COV standard deviations (STD) (not displayed in figure) were low for all metrics and regions ($< 0.6\%$), and FA in the splenium had the lowest value (0.1%). The substantially higher intersubject mean COVs than intrasubject mean COVs also indicates, like for the T1 (MP2RAGE) earlier, that each subject and region had specific diffusion metric values which were distinct from each other and were relatively stable as can be seen in Figure 3.

## 2.2 | Spinal cord

Figure 4 displays the results for the spinal cord cross-sectional area calculated for WM (using T1w and T2w images) and GM (using T2*w images). WM cross-sectional area (CSA) across the C2 and C3 vertebral levels calculated with T2w images resulted in intrasubject COVs of 2.3%, half of that found using T1w images (4.5%). For intersubject COVs, the trend is inverted; T2w had nearly double the intersubject COVs value (9.1 %) than T1w (5.2 %). The intrasubject standard deviations were on the order of the means (3.3% for WM using T1w, 1.7% for WM using T2w, and 10.4% for GM using T2w). We notice a particularly high COV for CSA (WM, T1w) for subject 2, which is due to high subject motion, resulting in unreliable spinal cord segmentation. In order to avoid rater bias in the intra- and inter-subject statistics, the analysis pipeline was fully automated, and no mask was manually edited.

FIGURE 4 Spinal cord cross-sectional area (CSA) for each acquired subject and session in WM (using either the T1w or T2w images) and in GM (using the T2*w images).

Figure 5 shows the scatter plots of all qMRI metric means calculated in the WM across the C2 and C5 vertebral levels of the spinal cord. As also observed in the brain, MTR resulted in lower intrasubject COV means (5.1%) than MTsat (7.9%, which is a superset of the MTR measurements plus one additional no-MT SPGR measurement and a B1 map). T1 had the

[2] Standard deviation values of the intrasubject COVs are reported in the interactive figures.

better mean intersubject COV (7.9%) relative to its two concomitant metrics (MTR - 4.6%, MTsat - 4.0 %), demonstrating unique mean quantitative T1 values in WM for the set of subjects for this timeframe. For diffusion, FA resulted in the lowest intrasubject COV means (3.9%), and MD and RD were substantially higher (5-9%) in contrast to the observations in the brain (0.6-1.3%).

FIGURE 5 Spinal cord qMRI metrics (T1, MTR, MTsat, FA, MD, RD). Each point represents the mean metric within the WM across C2 and C5 levels, for one subject and one session.

# 3 | DISCUSSION

Longitudinal stability of quantitative MRI measures is an important feature for clinical and research studies that intend to use the MRI scanner as a scientific instrument. Here, we report on the stability of a fundamental MR parameter (T1) and of microstructural biomarkers (MTR, MTsat, diffusion) in the central nervous system (brain and spinal cord) over the course of three years at a single imaging site. The concept of the "stability" of quantitative MR measures must be considered carefully; long-term biological changes in brain tissue also occur naturally in healthy people due to macro- and microstructural effects associated with normal ageing (MacDonald & Pike, 2021). Because this study was limited to three years and only investigated adults in mid-adulthood (ages 31 to 47 at initial scan date), the naturally-occurring effects of ageing in the brain (eg, myelin generation/degradation, ventricular enlargement, etc) are expected to occur slowly during this timespan (Ge et al., 2002; Hagiwara et al., 2021; Steen et al., 1995). The results of this initial data release, which can be made available upon request, may be used as a benchmark for the development of other analytical methods, as has been done using other large MRI data studies (Cohen-Adad et al., 2021b; Seif et al., 2022). This work is also a small piece of a larger ongoing project, CNeuroMod, and this long-term database of quantitative MRI measurements may be valuable information to incorporate in deep learning training models of other longitudinal measurements (eg, fMRI, MEG) to account for confounding changes in the brains of these subjects.

## Stability of qMRI measures

The reported intrasubject COV means indicate good stability of all quantitative metrics measured in the brain ($<$ 2.3 % in WM, $<$ 3.1 % in GM) throughout the ten structural sessions acquired over three years. Several metrics (T1 (MP2RAGE) and MTsat in Figure 2 and FA/MD/RD in Figure 3), also had higher intersubject mean COVs than intrasubject COV means, which suggests that the quantitative metrics were specific to the individuals and are stable enough to monitor longitudinal differences. The qMRI metrics that exhibited the lowest intrasubject COVs (MTR and T1 (MP2RAGE)) were also the metrics that used the lowest number of raw MRI images to calculate the metrics (MTR and MP2RAGE only need two, versus whereasMTsat and T1 (MTsat) need three), suggesting that quantitative MRI metric stability may degrade if they need substantially more measurements than simpler alternatives (MTR and T1 (MP2RAGE), calculated from two images). Another potential reason for the improved stability is that MP2RAGE is inherently optimised to reduce sensitivity of B1 effects (Marques et al., 2010), and future work should explore if quantitative techniques with good robustness against field inhomogeneities provide better long term stability than techniques necessitating additional measurements to correct for these effects. The longitudinal stability of a different implementation of T1 mapping (variable flip angle: VFA, which uses two measurements plus a B1 map) was reported in a healthy cohort at two timepoints acquired seven years apart (Gracien et al., 2017). Good stability was reported in WM T1 values, as well as a decrease in T1 values in cortical GM, the magnitude of which was proportional to the subject's age. The age range of the study was 51-77 at the initial time point, thus a higher overall cohort age than the CNeuroMod cohort. Another recent longitudinal study (York et al., 2022) investigated the longitudinal trends of quantitative MRI myelin measures (MTR, MTsat, and diffusion) in a cohort of both healthy and MS patients, and found that MTsat was more

sensitive to subtle changes in normal appearing white matter (NAWM) than MTR. However, only the MS cohort was investigated longitudinally over one year; the healthy cohort was a scan-rescan over two weeks. The longitudinal stability measures we reported in a healthy cohort (and in particular our open-source datasets) could be used to further support studies such as this one. In recent months, another longitudinal study (Salluzzi et al., 2022) investigated the short-term repeatability and long-term reproducibility in a healthy cohort over a 5 year interval with a different set of quantitative MRI metrics (T2/T2*, quantitative susceptibility, cerebral blood flow, and diffusivity). Their work, though investigating mostly different metrics, is complementary to our study in that its main objective was to assess the potential impacts of both software and hardware MRI upgrades on the repeatability and reproducibility of this set of qMRI metrics. They reported intrasubject COVs on the order of 1% or less for diffusion metrics (FA/MD/RD) in the three corpus callosum regions, in agreement with the observations reported in our study.

Spinal cord CSA had an intrasubject COV mean of 4.5 % and 2.3 % for CSA calculated from T1w and T2w scans, respectively. The almost twice smaller intrasubject COV for CSA computed on the T2w scan is likely due to the higher robustness to subject motion and/or spinal cord pulsatile motion for the T2w fast spin echo sequence vs. the T1w MPRAGE. This is consistent with a recent study (Bautin & Cohen-Adad, 2021), where intrasubject CSA COVs where 0.8% for T1w images and 0.57% for T2w images. Note that the Bautin & Cohen-Adad (2021) study was based on in-silico generation of scan-rescan using random affine transformations, hence the variability was highly under-estimated compared to the present study. In the present study, the reported COVs are likely closer to a realistic longitudinal scenario and suggest good long term stability for this quantitative metric in the spinal cord, and that T2w is the better choice for CSA quantification stability. In another related multi-site and multi-manufacturer study (Cohen-Adad et al., 2021b), were one subject was scanned in 19 different imaging centers over a period of 77 days,, they reported intra-site COVs for MTR and MTsat were below 3.6% and 11% respectively, on the order of our reported longitudinally measured values (5.1% and 7.9%). Intrasite FA COVs were reported on the order of or below 5.9%, higher than our mean longitudinal intrasubject COV value of 3.9%. These overall agreements between a multi-center snapshot in time and a single-centre longitudinal study provide encouraging evidence for the longitudinal stability when imaging the spinal cord.

## Limitations

Some limitations related to this study are important to highlight. Foremost, all measurements in this work were done on a single MRI scanner, and thus a single MRI vendor. Recent work (Cohen-Adad et al., 2021a, 2021b) done in the spinal cord suggests that while quantitative MR values differ across vendors, the COVs compare well. Multi-vendor harmonisation can only go so far; key differences in proprietary vendor pulse sequence implementations will always introduce differences out of the control of the user-researchers. However there is a lot of recent work on open-source pulse sequence frameworks (Cordes et al., 2020; Karakuzu, Biswas, et al., 2022; Layton et al., 2017) aiming to minimise these differences and give more control to the user researchers that may provide a solution to this limitation. Alternatively, inter-vendor biases can be accounted for in the statistics analysis step (Hagiwara et al., 2019), or by using a standard system phantom (Keenan et al., 2021). Our work reported on the longitudinal stability of mostly coarse regions-of-interest in the brain and spinal cord (whole-brain WM and GM mean values, in-plane WM and GM spinal cord means), except for the brain diffusion metrics which were averaged for the three corpus callosum regions (as was similarly done in (Salluzzi et al., 2022)). More granular masking methods exist for both the brain and spinal cord (eg. white & grey matter (Desikan et al., 2006; S. Lévy et al., 2015; Oishi et al., 2009)), tractometry (Catani & Thiebaut de Schotten, 2008)), and may be explored in the future. Another important point is that the processing pipelines were all only automatic, and no manual interventions were done during the segmentation steps of the pipeline. Manual corrections or more robust tools would likely improve the reliability of the reported metrics in both brain and

spinal cord. Although outside of the scope of this current study, the stability of quantitative morphometry in the brain (eg. cortical thickness) could also be explored and compared against the quantitative MRI metrics using this open dataset.

# 4 | METHODS

## Data acquisition

Six healthy participants (three females) were recruited in 2018 (aged 31 to 47 at initial scan date) and consented to be scanned regularly as part of the on-going CNeuroMod project (Boyle et al., 2020). The anatomical imaging protocol is run on each participant at a rate of approximately four times / year, for three years for this initial 2022 data release; more scans are regularly being acquired as the CNeuroMod project is ongoing. The participation of the subject labelled number 4 was unable to continue participating after their fifth session, and other participants occasionally were unable to attend their scheduled scans thus the total number of scans per participant varied. Each subject had the following number of scans at the time of data processing: subject 1 – 8 scans, subject 2 – 10 scans, subject 3 – 10 scans, subject 4 – 5 scans, subject 5 – 8 scans, subject 6 – 9 scans. All imaging sessions were performed at the same site on a 3.0 T whole-body MRI scanner (Prisma Fit, Siemens, Erlangen, Germany) with a 64-channel head/neck receive coil and 2-channel body transmit coil. Custom headcases (Caseforge, Berkeley, USA) were used for each participant to minimise movements during the imaging sessions; inter-scan motion is particularly important to be minimised for quantitative MRI as the actual fields in the imaging volume change with different anatomical positioning and cannot be easily corrected for using image registration techniques (Balbastre et al., 2022; Papp et al., 2016). Up to ten imaging sessions were acquired in total, and the same imaging protocol was used for each subject and session. Two sets of imaging protocols were implemented, one for the brain and one for the spinal cord, the details of which are summarised next, but are also documented on the CNeuroMod project documentation [3], including the Siemens MRI exam card PDFs exported from the scanner [4].

### Brain imaging protocol

The brain imaging protocol (Figure 1, top) consisted of the following set of MRI measurements: T1-weighted, T2-weighted, diffusion, MP2RAGE, B1 mapping, and magnetization transfer (MT) saturation. The T1-weighted image consisted of a 3D MPRAGE acquisition using a repetition time (TR) = 2.4 s, echo time (TE) = 2.2 ms, excitation flip angle (FA) = 8 deg, 0.8 mm isotropic resolution, and parallel imaging acceleration factor (R) = 2. The T2-weighted pulse sequence was a 3D fast spin-echo (FSE) acquisition with TR = 3.2 s, TE = 563 ms, 0.8 mm isotropic resolution, and R = 2. The diffusion-weighted protocol used a 2D axial EPI sequence (TR = 2.3 s, TE = 82 ms, FA = 78 deg, 2 mm3 isotropic resolution, simultaneous multi-slice (SMS) factor of 3, two-shells, minimum b-value = 1500 s/mm2, maximum b-value = 3000 s/mm2), and was acquired twice using either P-A or A-P phase-encoding directions, to correct for susceptibility-induced distortion. The MP2RAGE 3D protocol produced two images with different inversion times (TI) = 700 ms and 1500 ms, TR = 4s, TE = 1.51 ms FA = 7 deg and 5 deg for each TI respectively, 1.2 mm isotropic resolution, and R = 2. B1 maps were acquired using the default Siemens B1 mapping sequence based on a gradient echo sequence with ultrafast turbo-FLASH readout (6mm isotropic resolution) (Chung et al., 2010). Lastly, the MT saturation protocol consists of a set of three 3D spoiled gradient echo images: an MT-weighted (MTw) image (TR = 28 ms, TE = 3.3 ms, FA = 6 deg, 1.5 mm isotropic resolution, R = 2, and a Gaussian-shaped MT preparation pulse with an off-resonance frequency = 1.2 kHz), a proton-density-weighted (PDw) image (same protocol as the MTw, with the omission of the MT preparation pulse), and a T1-weighted (T1w) image (same

[3] Brain anatomical sequences
[4] Anatomical protocol PDF

protocol as the PDw, except TR = 18 ms and FA = 20 deg).

### Spinal cord imaging protocol

The spinal cord imaging protocol (Figure 1, bottom) consisted of the following set of MRI measurements: T1-weighted, T2-weighted, diffusion, and magnetization transfer (MT) saturation. The T1-weighted image consisted of a 3D MPRAGE acquisition with TR = 2 s, TE = 3.72 ms, FA = 9 deg, 1 mm isotropic resolution, and R = 2. The T2-weighted pulse sequence was a 3D fast spin-echo (FSE) acquisition with TR = 1.5 s, TE = 120 ms, FA = 120 deg, 0.8 mm isotropic resolution, and R = 3. The diffusion-weighted protocol used a 2D axial EPI sequence that was cardiac-gated with a pulse oximeter and TR ~ 620 ms, TE = 60 ms, 0.9 mm in-plane resolution, 5 mm slice resolution, phase encoding in the A-P direction, and a maximum b-value of 800 s/mm2). Lastly, the MT saturation protocol consisted of an MTw acquisition (TR = 35 ms, TE = 3.13 ms, FA = 9 deg, 0.9 mm2 in-plane resolution, 0.5 mm slice resolution, R = 2, and a Gaussian-shaped MT preparation pulse with an off-resonance frequency = 1.2 kHz), a proton-density-weighted (PDw) image (same protocol as the MTw, with the omission of the MT preparation pulse), and a T1-weighted (T1w) image (same protocol as the PDw, except TR = 15 ms and FA = 15 deg).

## Data preparation

All datasets acquired within the CNeuroMod project were prepared with the intention to be shared. Data were anonymized and defaced by masking out face, teeth, and ears. Datasets were prepared and organised in the BIDS (Brain Imaging Data Structure) format (Gorgolewski et al., 2016). Quantitative image acquisitions were prepared according to the BEP001 specification (Karakuzu, Appelhoff, et al., 2022), and spinal cord data used the "bp-cspine" tag as proposed in BEP025 to distinguish against the brain datasets for the same subject. Datasets were managed using Datalad (Halchenko et al., 2021) and git-annex in a databank; access to this databank is made available through the CNeuroMod website [5]. Session numbers in the database that are missing for some subjects are omitted datasets from scanning sessions that were aborted due to various scanning issues. sMRIprep (Esteban et al., 2022) was executed on the T1w brain scans from the first two sessions of each subject, which were later published on GitHub using git-annex as part of the CNeuroMod project. These outputs were used solely for the brain diffusion pipeline.

## Analysis pipeline

Two separate post-processing and analysis pipelines were developed for the brain and spinal cord data. Figure 6 shows an overview of both pipelines with the outcome metrics.

The brain pipelines were managed using Nextflow (Di Tommaso et al., 2017), a container management tool for data processing pipelines. Two Docker container images were prebuilt and used for this pipeline: `dockerhub.io/qmrlab/antsfl:latest` (digest: `597de3e6e1aa`) and `dockerhub.io/qmrlab/minimal:v2.5.0b` (digest: `40270330e7b5`). Image registration was performed using the Advanced Normalization Tools (ANTS; `version 2.1.0`) (Avants et al., 2009). Brain extraction was done using the brain extraction tool (BET) tool in the FMRIB Software Library (FSL; `version 5.0`) (Smith, 2002; Smith et al., 2004), and whole-brain WM and GM segmentation were done using the FMRIB's Automated Segmentation Tool (FAST) in FSL (Zhang et al., 2001). With the exception of diffusion, for all quantitative MRI methods the core data fitting algorithms used in this pipeline are from the open-source qMRLab software (version tag `2.5.0b`) (Cabana et al., 2015; Karakuzu et al., 2020). For diffusion, the TractoFlow pipeline (`version 2.4.1`) was used (Theaud et al., 2020), which uses DIPY (Garyfallidis et al., 2014) and MRtrix3 (Tournier et al., 2019) for the core diffusion processing functionalities, and FSL and ANTs for the image processing tools. The diffusion pipeline

[5] Neuromod data access

consists of a denoising step (MRtrix3), TOPUP (using the two phase encoding directions diffusion images) and eddy current corrections (FSL), DTIs (DIPY), brain tissue segmentation (ANTs), and lastly tractography maps (Cousineau et al., 2017); the full processing diagram is shown in Figure 6. DTI metrics were calculated using the 1500 s/mm2 b-value shell. In addition to the diffusion images as inputs, TractoFlow also used the average of the T1w structural images of the first two sessions (for each subject) that was registered to the MNI152 atlas, which is the output of another standard pipeline, sMRIprep (Esteban et al., 2022), that consists [6] of intensity non-uniformity corrections, alignment and fusion of the images, skull stripping, and non-linear registration to the template. The three regions-of-interests (ROIs) of the corpus callosum (genu/body/splenium) were extracted using the John Hopkins University ICBM-DTI-81 WM labels provided by FSL. The labels were first transformed from MNI152 space to the average T1w space (with transformations files available from the sMRIprep outputs [7]), and then from the average T1w space to the diffusion space using the affine matrix files provided as outputs of TractoFlow.

For the spinal cord data, the pipeline was developed in a shell script [8] using all tools available through the Spinal Cord Toolbox (SCT) `v5.6` (De Leener et al., 2017). The script was run through all the available subjects and sessions using the pipeline management tool `sct_run_batch`. The SC was segmented on T2w images using `sct_deepseg_sc` (Gros et al., 2019), then vertebral levels were identified (Ullmann et al., 2014). The SC was then registered to the adult PAM50 template (De Leener et al., 2018). T1w images were analysed similarly: the SC was segmented and then registered to the PAM50 template using the transformation T2w-PAM50 calculated earlier. The ME-GRE images were analysed using `sct_deepseg_gm` (Perone et al., 2018) to segment the grey matter. MT images were processed as follows. The SC was segmented on the GRE-MT1 scan, followed by registration to the PAM50 template via the T2w-PAM50 transformation. GRE-MT0 and GRE-T1w scans were then registered to the GRE-MT1 scans. Magnetization transfer ratio (MTR) and MTsat were computed. DWI images were motion-corrected using a mask centred around the SC for more robustness, then registered to the PAM50 template using the initial transformation. DTI metrics were computed using `sct_compute_dti` (powered by DIPY (Garyfallidis et al., 2014)).

The computed metrics are as follows: SC CSA averaged between C2-C3 levels from the T1w and T2w scans (using sct_process_segmentation), GM CSA averaged between C3-C4 from the ME-GRE scan, MTR, MTsat, T1 and DTI metrics extracted in the WM between levels C2-C5.

FIGURE 6 Overview of the three analysis pipelines used in this project: qMRLab (top row), Tractoflow (middle row), Spinal Cord Toolbox (bottom row). The human datasets were processed using NextFlow-based pipelines (qMRLab for qMRI processing, and Tractoflow for diffusion processing), whereas spine datasets used a bash script-based pipeline using the Spinal Cord Toolbox software.

## Quality control

For brain qMRI data processing (excluding diffusion), quality assurance was done manually with the assistance of the Nextflow log, which provides a report on success/failure of each processing step for all subjects and sessions. The resulting maps and masks were also visually verified manually, which resulted in some subsequent corrections to how the tissue masks were calculated [9] and the removal of parts of the MTsat acquisition volume due to slab profile effects [10]. Five data points were omitted due to missing B1 maps in the CNeuroMod database at the time of processing for these subject's sessions: `sub-03_ses-003`, `sub-06_ses-001`, `sub-06_ses-002`, `sub-06_ses-003`, `sub-06_ses-005`.

[6] The pipeline diagram for the external tool sMRIprep is available in their documentation
[7] Neuromod sMRIprep
[8] Neuromod process spinal cord data
[9] Release r20220916
[10] Release r20220921

For brain diffusion data processing, a report was generated from the TractoFlow tool dmriqc_flow (`v0.2.0` - (Theaud & Descoteaux, 2022)). Each step of the pipeline has been manually validated without any reported issues. Two sessions were excluded due to corrupted initial acquisitions (`sub-03_ses-002`, `sub-03_ses-003`). For the spinal cord data processing pipeline, a QC report showing various steps of the analysis (segmentation, vertebral labelling, registration) was generated and made publicly available on the GitHub project repository, release version `r20220804`). Following expert readings, some data points were excluded due to factors such as excessive motion (`sub-05_ses-007` [T2w]), poor shimming (`sub-03_ses-010` [T1w] and `sub-05_ses-007` [T1w]), and incorrect volume placement or incorrect b-values (`sub-02_ses-001` [DWI], `sub-03_ses-003` [DWI], `sub-06_ses-008`): details are listed in GitHub issues. In addition, the pipeline failed to produce an output for two data points (`sub-04_ses-001`, `sub-06_ses-005`).

# ACKNOWLEDGEMENT

The Courtois project on neural modelling was made possible by a generous donation from the Courtois foundation. The Courtois NeuroMod team is based at "Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal", with several other institutions involved. See the CNeuromod documentation for an up-to-date list of contributors (https://docs.cneuromod.ca). This study was also funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging [950-230815], the Canadian Institute of Health Research [CIHR FDN-143263], the Canada Foundation for Innovation [32454, 34824], the Fonds de Recherche du Québec - Santé [322736], the Natural Sciences and Engineering Research Council of Canada [RGPIN-2019-07244], the Canada First Research Excellence Fund (IVADO and TransMedTech), and the Mila - Tech Transfer Funding Program.

# DATA AVAILABILITY STATEMENT

In the aim of better reproducibility and transparency in research, all the data, processing pipelines, containers, and analysis code have been made available online. The anonymized and defaced datasets are in BIDS format and managed using Datalad and git-annex in a GitHub repository, https://github.com/courtois-neuromod/anat (commit: 5a5f687), and the data itself is hosted on an self-hosted S3 server. The sMRIPrep pipeline outputs for each subjects are also managed using git-annex and GitHub, https://github.com/courtois-neuromod/anat.smriprep (commit: b055f52). To request access to this data, we invite researchers to fill out an application form on our website https://www.cneuromod.ca/access/access/. The brain quantitative MRI processing pipeline was written in Nextflow (brain) and shell (spine) and are available in this repository: https://github.com/courtois-neuromod/anat-processing. The TractoFlow pipeline is built using open-source tools and is available on GitHub: https://github.com/scilus/tractoflow combined with the container image on Dockerhub: dockerhub.io/scilus/scilus:1.4.2 (digest: 25415e45ea7f, https://hub.docker.com/repository/docker/scilus/scilus) . The qMRI brain pipeline used two Docker containers which have been made available as saved container images on Dockerhub: dockerhub.io/qmrlab/antsfl:latest (digest: 597de3e6e1aa, https://hub.docker.com/repository/docker/qmrlab/antsfsl) and dockerhub.io/qmrlab/minimal:v2.5.0b (digest: 40270330e7b5, https://hub.docker.com/repository/docker/qmrlab/minimal)). The condensed outputs of these pipelines (eg, masked and averaged values for each tissue) are shared in GitHub releases of this repository, which can be found here: https://github.com/courtois-neuromod/anat-processing/releases/. The data figures and tables in this article were produced using analysis code integrated in an interactive Jupyter Book and powered by Plotly, which is available here, https://courtois-neuromod.github.io/anat-processing-paper/, and the code repository for this book is https://github.com/courtois-neuromod/anat-processing-paper.

# References

Avants, Tustison, & Song. (2009). Advanced normalization tools (ANTS). *Insight J.* https://doi.org/10.54294/uvnhin

Balbastre, Y., Aghaeifar, A., Corbin, N., Brudfors, M., Ashburner, J., & Callaghan, M. F. (2022). Correcting inter-scan motion artifacts in quantitative $R_1$ mapping at 7T. In *Magn. Reson. Med.* (No. 1; Vol. 88, pp. 280–291). https://doi.org/10.1002/mrm.29216

Bautin, P., & Cohen-Adad, J. (2021). Minimum detectable spinal cord atrophy with automatic segmentation: Investigations using an open-access dataset of healthy participants. *NeuroImage. Clinical*, *32*, 102849. https://doi.org/10.1016/j.nicl.2021.102849

Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiéry, N. M., & Fangohr, H. (2021). Using jupyter for reproducible scientific workflows. *Computing in Science & Engineering*, *23*(2), 36–46. https://doi.org/10.1109/MCSE.2021.3052101

Boyle, J. A., Pinsard, B., Boukhdhir, A., Belleville, S., Brambatti, S., Chen, J., Cohen-Adad, J., Cyr, A., Fuente, A., Rainville, P., & Bellec, P. (2020). The Courtois project on neuronal modelling - 2020 data release. *Annual Meeting of the Organization for Human Brain Mapping*, 1939.

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

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *44*(8), 1105–1132. https://doi.org/10.1016/j.cortex.2008.05.004

Cercignani, M., Dowell, N. G., & Tofts, P. S. (2018). *Quantitative MRI of the brain: Principles of physical measurement, second edition*. CRC Press.

Chung, S., Kim, D., Breton, E., & Axel, L. (2010). Rapid B1+ mapping using a preconditioning RF pulse with TurboFLASH readout. *Magnetic Resonance in Medicine*, *64*(2), 439–446. https://doi.org/10.1002/mrm.22423

Cohen-Adad, J. (2020). *Spine generic public database (single subject)*.

Cohen-Adad, J., Alonso-Ortiz, E., Abramovic, M., Arneitz, C., Atcheson, N., Barlow, L., Barry, R. L., Barth, M., Battiston, M., Büchel, C., & others. (2021a). Generic acquisition protocol for quantitative MRI of the spinal cord. *Nature Protocols*, *16*(10), 4611–4632. https://doi.org/10.1038/s41596-021-00588-0

Cohen-Adad, J., Alonso-Ortiz, E., Abramovic, M., Arneitz, C., Atcheson, N., Barlow, L., Barry, R. L., Barth, M., Battiston, M., Büchel, C., & others. (2021b). Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. *Scientific Data*, *8*(1), 219. https://doi.org/10.1038/s41597-021-01044-0

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

Cousineau, M., Jodoin, P.-M., Garyfallidis, E., Côté, M.-A., Morency, F. C., Rozanski, V., Grand'Maison, M., Bedell, B. J., & Descoteaux, M. (2017). A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical*, *16*, 222–233. https://doi.org/10.1016/j.nicl.2017.07.020

Davis, T. L., Kwong, K. K., Weisskoff, R. M., & Rosen, B. R. (1998). Calibrated functional MRI: Mapping the dynamics of oxidative metabolism. *Proceedings of the National Academy of Sciences*, *95*(4), 1834–1839. https://doi.org/10.1073/pnas.95.4.1834

De Leener, B., Fonov, V. S., Collins, D. L., Callot, V., Stikov, N., & Cohen-Adad, J. (2018). PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. *Neuroimage*, *165*, 170–179. https://doi.org/10.1016/j.neuroimage.2017.10.041

De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Louis Collins, D., Callot, V., & Cohen-Adad, J. (2017). SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage*, *145*, 24–43. https://doi.org/10.1016/j.neuroimage.2016.10.009

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Duchesne, S., Dieumegarde, L., Chouinard, I., Farokhian, F., Badhwar, A., Bellec, P., Tétreault, P., Descoteaux, M., Boré, A., Houde, J.-C., Beaulieu, C., & Potvin, O. (2019). Structural and functional multi-platform MRI series of a single human volunteer over more than fifteen years. *Scientific Data*, *6*(1), 1–9. https://doi.org/10.1038/s41597-019-0262-8

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

Erramuzpe, A., Schurr, R., Yeatman, J. D., Gotlib, I. H., Sacchet, M. D., Travis, K. E., Feldman, H. M., & Mezer, A. A. (2021). A comparison of quantitative R1 and cortical thickness in identifying age, lifespan dynamics, and disease states of the human cortex. *Cerebral Cortex*, *31*(2), 1211–1226. https://doi.org/10.1093/cercor/bhaa288

Esteban, O., Markiewicz, C. J., Blair, R., Poldrack, R. A., & Gorgolewski, K. J. (2022). *sMRIPrep: Structural MRI PREProcessing workflows*.

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., Nimmo-Smith, I., & Dipy Contributors. (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, *8*, 8. https://doi.org/10.3389/fninf.2014.00008

Ge, Y., Grossman, R. I., Babb, J. S., Rabin, M. L., Mannon, L. J., & Kolson, D. L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part II: Quantitative magnetization transfer ratio histogram analysis. *AJNR: American Journal of Neuroradiology*, *23*(8), 1334–1341.

Golay, X., & Oliver-Taylor, A. (2022). *Phantom for Multi-Parametric calibration in magnetic resonance imaging* (11,391,804 B2).

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.44

Gracien, R.-M., Maiworm, M., Brüche, N., Shrestha, M., Nöth, U., Hattingen, E., Wagner, M., & Deichmann, R. (2020). How stable is quantitative MRI? - Assessment of intra- and inter-scanner-model reproducibility using identical acquisition sequences and data analysis programs. *Neuroimage*, *207*, 116364. https://doi.org/10.1016/j.neuroimage.2019.116364

Gracien, R.-M., Nürnberger, L., Hok, P., Hof, S.-M., Reitz, S. C., Rüb, U., Steinmetz, H., Hilker-Roggendorf, R., Klein, J. C., Deichmann, R., & Baudrexel, S. (2017). Evaluation of brain ageing: A quantitative longitudinal MRI study over 7 years. *European Radiology*, *27*(4), 1568–1576. https://doi.org/10.1007/s00330-016-4485-1

Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., Talbott, J., Zhuoquiong, R., Liu, Y., Granberg, T., Ouellette, R., Tachibana, Y., Hori, M., Kamiya, K., Chougar, L., Stawiarz, L., Hillert, J., Bannier, E., Kerbrat, A., … Cohen-Adad, J. (2019). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage*, *184*, 901–915. https://doi.org/10.1016/j.neuroimage.2018.09.081

Hagiwara, A., Fujimoto, K., Kamagata, K., Murata, S., Irie, R., Kaga, H., Someya, Y., Andica, C., Fujita, S., Kato, S., Fukunaga, I., Wada, A., Hori, M., Tamura, Y., Kawamori, R., Watada, H., & Aoki, S. (2021). Age-Related changes in relaxation times, proton density, myelin, and tissue volumes in adult brain analyzed by 2-Dimensional quantitative synthetic magnetic resonance imaging. *Investigative Radiology*, *56*(3), 163. https://doi.org/10.1097/RLI.0000000000000720

Hagiwara, A., Hori, M., Cohen-Adad, J., Nakazawa, M., Suzuki, Y., Kasahara, A., Horita, M., Haruyama, T., Andica, C., Maekawa, T., Kamagata, K., Kumamaru, K. K., Abe, O., & Aoki, S. (2019). Linearity, bias, intrascanner repeatability, and interscanner reproducibility of quantitative multidynamic multiecho sequence for rapid simultaneous relaxometry at 3 T: A validation study with a standardized phantom and healthy controls. *Investigative Radiology*, *54*(1), 39–47. https://doi.org/10.1097/RLI.0000000000000510

Halchenko, Y., Meyer, K., Poldrack, B., Solanky, D., Wagner, A., Gors, J., MacFarlane, D., Pustina, D., Sochat, V., Ghosh, S., Mönch, C., Markiewicz, C., Waite, L., Shlyakhter, I., de la Vega, A., Hayashi, S., Häusler, C., Poline, J.-B., Kadelka, T., … Hanke, M. (2021). DataLad: Distributed system for joint management of code, data, and their relationship. *J. Open Source Softw.*, *6*(63), 3262. https://doi.org/10.21105/joss.03262

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

Herz, K., Mueller, S., Perlman, O., Zaitsev, M., Knutsson, L., Sun, P. Z., Zhou, J., van Zijl, P., Heinecke, K., Schuenke, P., & others. (2021). Pulseq-CEST: Towards multi-site multi-vendor compatibility and reproducibility of CEST experiments using an open-source sequence standard. *Magnetic Resonance in Medicine*, *86*(4), 1845–1858. https://doi.org/10.1002/mrm.28825

Karakuzu, A., Appelhoff, S., Auer, T., Boudreau, M., Feingold, F., Khan, A. R., Lazari, A., Markiewicz, C., Mulder, M., Phillips, C., Salo, T., Stikov, N., Whitaker, K., & de Hollander, G. (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. *Magnetic Resonance in Medicine*, *79*(1), 48–61. https://doi.org/10.1002/mrm.26982

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

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. *Magnetic Resonance in Medicine*, *77*(4), 1544–1552. https://doi.org/10.1002/mrm.26235

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. *Magnetic Resonance in Medicine*, *81*(1), 454–465. https://doi.org/10.1002/mrm.27421

Lévy, S., Benhamou, M., Naaman, C., Rainville, P., Callot, V., & Cohen-Adad, J. (2015). White matter atlas of the human spinal cord with estimation of partial volume effect. *Neuroimage*, *119*, 262–271. https://doi.org/10.1016/j.neuroimage.2015.06.040

Lévy, Simon, Guertin, M.-C., Khatibi, A., Mezer, A., Martinu, K., Chen, J.-I., Stikov, N., Rainville, P., & Cohen-Adad, J. (2018). Test-retest reliability of myelin imaging in the human spinal cord: Measurement errors versus region- and aging-induced variations. *PLoS One*, *13*(1), e0189944. https://doi.org/10.1371/journal.pone.0189944

Lidén, M., Adrian, D., Widell, J., Uggla, B., & Thunberg, P. (2021). Quantitative T2* imaging of iron overload in a non-dedicated center - Normal variation, repeatability and reader variation. *European Journal of Radiology Open*, *8*, 100357. https://doi.org/10.1016/j.ejro.2021.100357

Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., & Griswold, M. A. (2013). Magnetic resonance fingerprinting. *Nature*, *495*(7440), 187–192. https://doi.org/10.1038/nature11971

Ma, Y., Berman, A. J. L., & Pike, G. B. (2016). The effect of dissolved oxygen on the relaxation rates of blood plasma: Implications for hyperoxia calibrated BOLD. *Magnetic Resonance in Medicine*, *76*(6), 1905–1911. https://doi.org/10.1002/mrm.26069

MacDonald, M. E., & Pike, G. B. (2021). MRI of healthy brain aging: A review. *NMR in Biomedicine*, *34*(9), e4564. https://doi.org/10.1002/nbm.4564

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/10.7554/eLife.61523

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., 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

Mazerolle, E. L., Ma, Y., Sinclair, D., & Pike, G. B. (2018). Impact of abnormal cerebrovascular reactivity on BOLD fMRI: A preliminary investigation of moyamoya disease. *Clinical Physiology and Functional Imaging*, *38*(1), 87–92. https://doi.org/10.1111/cpf.12387

Oh, J., Chen, M., Cybulsky, K., Suthiphosuwan, S., Seyman, E., Dewey, B., Diener-West, M., van Zijl, P., Prince, J., Reich, D. S., & Calabresi, P. A. (2021). Five-year longitudinal changes in quantitative spinal cord MRI in multiple sclerosis. *Multiple Sclerosis*, *27*(4), 549–558. https://doi.org/10.1177/1352458520923970

Oishi, K., Faria, A., Jiang, H., Li, X., Akhter, K., Zhang, J., Hsu, J. T., Miller, M. I., van Zijl, P. C. M., Albert, M., Lyketsos, C. G., Woods, R., Toga, A. W., Pike, G. B., Rosa-Neto, P., Evans, A., Mazziotta, J., & Mori, S. (2009). Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: Application to normal elderly and Alzheimer's disease participants. *Neuroimage*, *46*(2), 486–499. https://doi.org/10.1016/j.neuroimage.2009.01.002

Papp, D., Callaghan, M. F., Meyer, H., Buckley, C., & Weiskopf, N. (2016). Correction of inter-scan motion artifacts in quantitative R1 mapping by accounting for receive coil sensitivity effects. In *Magnetic Resonance in Medicine* (No. 5; Vol. 76, pp. 1478–1485). https://doi.org/10.1002/mrm.26058

Perone, C. S., Calabrese, E., & Cohen-Adad, J. (2018). Spinal cord gray matter segmentation using deep dilated convolutions. *Scientific Reports*, *8*(1), 5966. https://doi.org/10.1038/s41598-018-24304-3

Pykett, I. L., & Mansfield, P. (1978). A line scan image study of a tumorous rat leg by NMR. *Physics in Medicine and Biology*, *23*(5), 961–967. https://doi.org/10.1088/0031-9155/23/5/012

Ropele, S., de Graaf, W., Khalil, M., Wattjes, M. P., Langkammer, C., Rocca, M. A., Rovira, A., Palace, J., Barkhof, F., Filippi, M., & Fazekas, F. (2011). MRI assessment of iron deposition in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, *34*(1), 13–21. https://doi.org/10.1002/jmri.22590

Salluzzi, M., McCreary, C. R., Gobbi, D. G., Lauzon, M. L., & Frayne, R. (2022). Short-term repeatability and long-term reproducibility of quantitative MR imaging biomarkers in a single centre longitudinal study. *NeuroImage*, *260*, 119488. https://doi.org/10.1016/j.neuroimage.2022.119488

Schmierer, K., Tozer, D. J., Scaravilli, F., Altmann, D. R., Barker, G. J., Tofts, P. S., & Miller, D. H. (2007). Quantitative magnetization transfer imaging in postmortem multiple sclerosis brain. *Journal of Magnetic Resonance Imaging*, *26*(1), 41–51. https://doi.org/10.1002/jmri.20984

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

Seiberlich, N., Ma, D., Gulani, V., & Griswold, M. (2012). *Nuclear magnetic resonance (NMR) fingerprinting* (20120235678 A1).

Seif, M., Leutritz, T., Schading, S., Emmengger, T., Curt, A., Weiskopf, N., & Freund, P. (2022). Reliability of multi-parameter mapping (MPM) in the cervical cord: A multi-center multi-vendor quantitative MRI study. *NeuroImage*, *264*, 119751. https://doi.org/10.1016/j.neuroimage.2022.119751

Seiler, A., Schöngrundner, S., Stock, B., Nöth, U., Hattingen, E., Steinmetz, H., Klein, J. C., Baudrexel, S., Wagner, M., Deichmann, R., & Gracien, R.-M. (2020). Cortical aging - new insights with multiparametric quantitative MRI. *Stress and The Aging Brain*, *12*(16), 16195–16210. https://doi.org/10.18632/aging.103629

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23 Suppl 1*, S208–19. https://doi.org/10.1016/j.neuroimage.2004.07.051

Steen, R. G., Gronemeyer, S. A., & Taylor, J. S. (1995). Age-related changes in proton T1 values of normal human brain. *Journal of Magnetic Resonance Imaging*, *5*(1), 43–48. https://doi.org/10.1002/jmri.1880050111

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. *Magnetic Resonance in Medicine*, *86*(3), 1194–1211. https://doi.org/10.1002/mrm.28779

Theaud, G., & Descoteaux, M. (2022). dMRIQCpy: A python-based toolbox for diffusion MRI quality control and beyond. *International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting*, 3906. https://doi.org/10.58530/2022/3906

Theaud, G., Houde, J.-C., Boré, A., Rheault, F., Morency, F., & Descoteaux, M. (2020). TractoFlow: A robust, efficient and reproducible diffusion MRI pipeline leveraging Nextflow & Singularity. *Neuroimage*, *218*, 116889. https://doi.org/10.1016/j.neuroimage.2020.116889

Tofts, P. S. (1998). Standardisation and optimisation of magnetic resonance techniques for multicentre studies. *Journal of Neurology, Neurosurgery and Psychiatry*, *64 Suppl 1*, S37–43.

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*, *202*, 116137. https://doi.org/10.1016/j.neuroimage.2019.116137

Ullmann, E., Pelletier Paquette, J. F., Thong, W. E., & Cohen-Adad, J. (2014). Automatic labeling of vertebral levels using a robust template-based approach. *International Journal of Biomedical Imaging*, *2014*, 719520. https://doi.org/10.1155/2014/719520

Wang, R., Xie, G., Zhai, M., Zhang, Z., Wu, B., Zheng, D., Hong, N., Jiang, T., Wen, B., & Cheng, J. (2017). Stability of R2* and quantitative susceptibility mapping of the brain tissue in a large scale multi-center study. *Scientific Reports*, *7*, 45261. https://doi.org/10.1038/srep45261

York, E. N., Meijboom, R., Thrippleton, M. J., Bastin, M. E., Kampaite, A., White, N., Chandran, S., & Waldman, A. D. (2022). Longitudinal microstructural MRI markers of demyelination and neurodegeneration in early relapsing-remitting multiple sclerosis: Magnetisation transfer, water diffusion and g-ratio. *NeuroImage: Clinical*, *36*, 103228. https://doi.org/10.1016/j.nicl.2022.103228

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424