PLOS ONE

RESEARCH ARTICLE

Test-retest reliability of myelin imaging in the human spinal cord: Measurement errors versus region- and aging-induced variations

Simon Lévy[1,2¤], Marie-Claude Guertin[3], Ali Khatibi[2,4,5,6], Aviv Mezer[7], Kristina Martinu[2], Jen-I Chen[2,8], Nikola Stikov[1,9], Pierre Rainville[2,8], Julien Cohen-Adad[1,10]*

1 NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC, Canada, **2** Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montréal, QC, Canada, **3** Montreal Health Innovations Coordinating Center (MHICC), Montreal Heart Institute, Montreal, QC, Canada, **4** Psychology Department, Bilkent University, Ankara, Turkey, **5** Interdisciplinary program in Neuroscience, Bilkent University, Ankara, Turkey, **6** National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey, **7** The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem, Israel, **8** Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada, **9** Montreal Heart Institute, Montreal, QC, Canada, **10** Functional Neuroimaging Unit, CRIUGM, Université de Montréal, Montreal, QC, Canada

¤ Current address: Centre d'Exploration Métabolique par Résonance Magnétique (CEMEREM), AP-HM, Hôpital de la Timone, Pôle d'imagerie médicale, Marseille, France
* jcohen@polymtl.ca

OPEN ACCESS

Citation: Lévy S, Guertin M-C, Khatibi A, Mezer A, Martinu K, Chen J-I, et al. (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

Editor: Fernando de Castro, Instituto Cajal-CSIC, SPAIN

Received: June 18, 2017

Accepted: December 5, 2017

Published: January 2, 2018

Data Availability Statement: The current Quebec law on the protection of private data prevents us from sharing all the MRI datasets this work is based on, even though they are de-identified, because they contain potentially sensitive information. This restriction has been imposed by the ethical review board of the Research Center of Montreal University Geriatric Institute (Comité mixte d'éthique de la recherche du RNQ, approval number CMER-RNQ_14-15-010). To contact the research ethics board of the RNQ, please refer to

Abstract

Purpose

To implement a statistical framework for assessing the precision of several quantitative MRI metrics sensitive to myelin in the human spinal cord: T_1, Magnetization Transfer Ratio (MTR), saturation imposed by an off-resonance pulse (MT_{sat}) and Macromolecular Tissue Volume (MTV).

Methods

Thirty-three healthy subjects within two age groups (young, elderly) were scanned at 3T. Among them, 16 underwent the protocol twice to assess repeatability. Statistical reliability indexes such as the Minimal Detectable Change (MDC) were compared across metrics quantified within different cervical levels and white matter (WM) sub-regions. The differences between pathways and age groups were quantified and interpreted in context of the test-retest repeatability of the measurements.

Results

The MDC was respectively 105.7ms, 2.77%, 0.37% and 4.08% for T_1, MTR, MT_{sat} and MTV when quantified over all WM, while the standard-deviation across subjects was 70.5ms, 1.34%, 0.20% and 2.44%. Even though particular WM regions did exhibit significant differences, these differences were on the same order as test-retest errors. No significant difference was found between age groups for all metrics.

Mrs Karima Bekhiti (phone: +1 514 527-9565 #3223; email: karima.bekhiti.iugm@ssss.gouv.qc.ca). Notwithstanding this limitation, we are willing to share the MRI data from four members of the lab, from which we have obtained explicit approval. Furthermore, we are also sharing the processing scripts used in this work, which will enable other researchers to re-run the analyses on the shared datasets. The data and scripts have now been uploaded to https://osf.io/ezmrj/.

Funding: This study was funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging (JCA), the Canadian Institutes of Health Research [CIHR FDN-143263] (JCA), the Canadian Institutes of Health Research [CIHR MOP-130341] (JCA and PR), the Fonds de Recherche du Québec—Santé [FRQS-28826] (JCA), the Fonds de Recherche du Québec—Nature et Technologies [2015-PR-182754] (JCA), the Natural Sciences and Engineering Research Council of Canada [NSERC-435897-2013] (JCA), the Natural Sciences and Engineering Research Council of Canada [NSERC 2016-06774] (NS), the Quebec BioImaging Network (JCA) and the Montreal Heart Institute Foundation (NS). Those funds were used for MRI data acquisition, computer and software resources and authors' funding.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

While T_1-based metrics (T_1 and MTV) exhibited better reliability than MT-based measurements (MTR and MT_{sat}), the observed differences between subjects or WM regions were comparable to (and often smaller than) the MDC. This makes it difficult to determine if observed changes are due to variations in myelin content, or simply due to measurement error. Measurement error remains a challenge in spinal cord myelin imaging, but this study provides statistical guidelines to standardize the field and make it possible to conduct large-scale multi-center studies.

1. Introduction

1.1. Quantitative MRI

Precise techniques are needed to monitor microstructural degeneration of the nervous tissue in clinics, especially for longitudinal follow up of white matter (WM) lesions in neurodegenerative pathologies, such as demyelination in multiple sclerosis. Rather than using MRI as a technique for simply viewing the anatomy, quantitative MRI (qMRI) aims to provide quantitative *metrics* related to some tissue properties. To date, several qMRI metrics have been proposed to characterize myelin content in the WM.

The longitudinal relaxation time T_1 has shown high correlation with the myelin volume quantified by histology [1–3]. However, T_1 is also affected by iron concentration [4], and it is difficult to disentangle the specific contribution of myelin and iron because of their co-localization [5]. The Magnetization Transfer Ratio (MTR) has also shown high correlation with histopathology of myelin in multiple sclerosis patients [2,3]. However, MTR consists of various contributions (T_1 and fraction F of exchanging protons bound to macromolecules) [6,7], which in some cases work against each other, reducing its sensitivity to myelin [2,8]. In this perspective, the quantification of the saturation imposed by an off-resonance pulse (MT_{sat}) has been proposed to minimize T_1 effects and increase the specificity to myelin [6].

Proton density (PD) is also a promising metric, as it measures the density of MRI-visible protons–i.e. protons with sufficiently long transversal relaxation time (T_2)–which are water (or liquid) protons. In the Central Nervous System (CNS), the complement of PD yields an estimate of the density of non-free protons, which are mostly bound to lipids and other macromolecules. Since myelin consists of 70 to 80% lipids and some macromolecules [9,10], this index can be expected to be a good marker of myelin content. Several PD estimation techniques and studies in the CNS have been published [11–23]. The complement of PD has been recently named *Macromolecular Tissue Volume* (MTV) [24,25] and its sensitivity and specificity to myelination was tested. MTV showed high accuracy and precision when quantifying the lipid content in phantoms. In addition, the MTV significantly decreased in the WM of multiple sclerosis patients compared to controls, showing independence from fiber geometry, unlike the Fractional Anisotropy (FA) from Diffusion Tensor Imaging (DTI). However, since MTV is defined as the fraction of non-liquid protons, it includes more than the volume occupied by myelin, raising the question of its specificity to myelin.

Myelin Water Imaging (MWI) using multi-echo T_2 [26] is another myelin mapping technique that has shown good sensitivity to myelin content in MS patients *post-mortem* [27] and *in vivo* [28]. While the earliest implementations of MWI were not clinically feasible, techniques such as Gradient- And Spin-Echo (GRASE [29,30]) were shown to speed up the acquisition [31]. Further investigations are ongoing.

The time constant of the transverse relaxation due to spin-spin interactions and local field inhomogeneities (T_2^*) has also exhibited sensitivity to myelin [32–34]. However, T_2^* includes important contributions from other factors, such as iron content [4,35], fiber orientation [36], blood vessels [37] and blood oxygen level [38].

Inhomogeneous Magnetization Transfer (ihMT) ratio is another recent metric [39] that is thought to be particularly sensitive and specific to myelin [40,41]. However, the measurement of this metric requires non-product sequence which are currently not available on clinical scanners.

1.2. Terminology

The above-mentioned metrics have their own advantages and limitations in quantifying myelin content in the CNS. To compare them, the relevant criteria for a myelin biomarker needs to be defined properly. *Sensitivity* and *specificity* are often the outstanding criteria. Here, sensitivity refers to the ability of the metric to monitor the variations in myelin content, while the specificity describes its exclusivity to myelin variations, i.e. to what extent the variations in the metric values are due to variations in the myelin content *only*. However, before tackling the sensitivity and specificity of a metric, it is essential to assess its *repeatability*. Indeed, sensitivity and specificity cannot be determined precisely if the metric values dramatically change between different scan sessions. The repeatability refers to the agreement (measurement precision) between two or more measurements made at different time points under the same conditions (e.g., same protocol, same scanner, same subjects, etc.) [42]. The repeatability must not be mistaken with *reproducibility*, which refers to the agreement between two or more measurements made at different time points under changing conditions. In both repeatability and reproducibility studies, the *reliability* is a relevant aspect to assess. The reliability compares the variability of scores due to measurement errors to the variability in the "true", error-free scores, i.e. to the variability induced by true variations of the measured feature (e.g., true variations in myelin content).

1.3. Review of past studies on qMRI metrics repeatability

The question of repeatability is even more relevant for spinal cord studies, where noise, motion and susceptibility artifacts make it difficult to acquire high quality images [43]. Previous studies investigated the repeatability of quantitative MRI metrics. Taso et al. [44] reported the repeatability of MTR, ihMTR and DTI (Diffusion Tensor Imaging) indexes within 3 healthy subjects at 3 time points by means of coefficients of variations (CV), defined as the ratio of the between-scans standard-deviation over the mean across scans. However, this index does not allow to properly compare between different metrics, as the means can differ drastically across metrics or even for a single metric across different studies (e.g., MTR [45]), yielding lower CVs for metrics with higher mean values. Smith et al. [46] also reported the test-retest repeatability of DTI and MT metrics within 9 healthy subjects at 2 time points using the normalized Bland-Altman difference (i.e. mean difference between scans divided by the mean across scans), which makes it harder to compare the repeatability between metrics with different means. Grussu et al. [47] reported the test-retest repeatability of NODDI (Neurite Orientation Dispersion and Density Imaging) indexes within 5 heathy subjects. The test-retest reliability was quantified by means of Intra-Class Correlation (ICC) coefficients defined as the ratio of the inter-subject variance over the total variance (i.e. the sum of the within- and between-subjects variances). Smith et al. [48] assessed the repeatability of MTR and F (fraction of exchanging protons bound to macromolecules) from quantitative magnetization transfer (qMT) imaging by means of the 95% confidence interval for the test-retest difference. However, this estimate

of the measurement error was not properly compared neither between metrics nor in the context of the differences observed between (expected) different myelin contents.

The test-retest repeatability has been studied extensively in research fields other than qMRI, notably in rehabilitation research [49–53]. Useful statistical indexes to quantify repeatability are provided. First, the existence of a systematic bias between test and retest measurements can be examined by the confidence interval for the test-retest difference (CI_d), as used in Smith et al. [48]. Then, the reliability can be assessed by the intra-class coefficient based on a two-way mixed effects model of analysis of variance. Finally, groups can be compared taking measurement errors into account (which is not done with usual statistical tests) using CI_d, showing whether the difference between groups is distinguishable from measurement errors or not. In the same vein, one can compute the Minimum Detectable Change (MDC) to quantify the minimum difference between two single metric values that is necessary to report a "true" error-free change, again taking the measurement errors into account. The MDC is particularly appropriate and intuitive for clinicians who would like to assess whether a treatment affects their patient or not.

In this work, we propose a statistical framework to quantify the test-retest reliability of qMRI metrics. We *(i)* quantify the repeatability of T_1, MTR, MT_{sat} and MTV in the spinal cord using a clinically-compatible protocol and *(ii)* evaluate the sensitivity of these metrics to myelin content across spinal pathways and age groups, in the context of the test-retest measurement errors.

2. Material and methods

2.1. Data acquisition

Thirty-three right-handed healthy subjects including 19 young (aged 24.9 ± 3.9, from 21 to 33 y.o.; 9 women, 10 men) and 14 elderly (aged 67.4 ± 4.0, from 61 to 73 y.o.; 6 women, 8 men) were recruited. A written consent form was obtained from each participant as supervised by the ethical review board of the Research Center of Montreal University Geriatric Institute (Comité mixte d'éthique de la recherche du RNQ, approval number CMER-RNQ_14-15-010).

To assess the metrics repeatability, 8 young (aged 24.0 ± 3.9, from 21 to 31 y.o., 2 women, 6 men) and 8 elderly (aged 67 ± 4.5, from 61 to 72 y.o., 2 women, 6 men) subjects from the previously described cohort underwent two scanning sessions: 12 subjects were scanned twice within a 10-month interval, and 4 within the same session (with a 5-minute break out of the scanner between scan and rescan). All data were acquired on a 3T Siemens TIM TRIO scanner and with a standard 12-channels head coil and a standard 4-channels neck coil.

The protocol consisted of:

- One sagittal turbo-spin-echo 3D SPACE T_2-weigthed anatomic image (TR = 1500 ms; TE = 119 ms; flip angle = 120˚; BW = 723 Hz/voxel; matrix = 384x384x52; resolution = 1x1x1 mm; FOV = 384x384x52 mm) with a high contrast between cord and cerebrospinal fluid (CSF) to further take the curvature of the cord into account in the data processing;
- Four 3D FLASH acquisitions (TR = 35 ms; TE = 5.92 ms; BW = 260 Hz/voxel; matrix = 192x192x22; resolution = 0.9x0.9x5 mm; gap = 1 mm; FOV = 174x174x110 mm; R = 2 acceleration; phase encoding direction = right-left). The four FLASH scans consisted of:
    - One with a prior RF saturation pulse (Gaussian-shaped, duration = 9984 μs, offset frequency = 1.2 kHz) and an excitation flip angle of 10˚;
    - Three without a saturation pulse and flip angles of 4˚, 10˚, and 20˚;

- Two axial 2D segmented spin-echo EPI acquisitions (TR = 3000 ms; TE = 19 ms; BW = 1905 Hz/voxel; matrix = 64x64, 17 slices; resolution = 3.0x3.0x5.5 mm; FOV = 192x192 mm) with a flip angle of 60 and 120° respectively (for B_1^+ estimation purposes);

All images spanned at least C2 to C5 vertebral bodies. The duration of the protocol was 18 minutes.

2.2. Data processing

Analysis was performed using the *Spinal Cord Toolbox* (SCT) version 2.2.3 [54]. The four datasets were first co-registered, then metrics were calculated. For extracting metrics within specific pathways in the white matter (dorsal column, DC, lateral funiculi, LF, ventral funiculi, VF), data were registered to the MNI-Poly-AMU template [55], which includes an atlas of WM tracts [56]. For sake of clarity, details about the processing pipeline are included in the supplementary material (see S1 File in section 8. Supporting information).

2.3. Statistical analysis

Statistical analyses were performed using MATLAB R2014a (The MathWorks, Inc., Natick, Massachusetts, USA) and SPSS (IBM SPSS Statistics–Release 24.0.0.0) at the 0.05 significance level unless otherwise stated.

2.3.1. Repeatability. *Systematic change between test and retest*

The mean of the difference between test and retest ($\overline{d}$) across subjects was computed along with a 95% confidence interval for the true test-retest difference (CI_d) derived according to:

$$CI_d = \overline{d} \pm t_{n-1} \cdot SE$$

where $SE = SD_d/\sqrt{n}$ is the Standard Error, SD_d is the standard-deviation (SD) of the difference between test and retest across the subjects, n is the number of subjects and t_{n-1} is the t statistics with $n - 1$ degrees of freedom and type I error of 5% [57]. In our case, $t_{n-1} = 2.131$.

If zero is not included in CI_d, we can consider that a systematic change between test and retest has occurred [50]. In addition to assess the systematic bias between test and retest, the CI_d gives the minimum difference between two subjects groups that is distinguishable from measurement errors.

Absolute test-retest difference

The absolute difference between test and retest, termed $|d|$, and its mean across subjects ($\overline{|d|}$) were computed to give to the reader a basic and direct measure of the measurement errors magnitude.

Reliability

The Intra-Class Correlation (ICC) coefficient is an appropriate coefficient to assess the test-retest reliability [58]. It measures the proportion of variance that is attributable to the "true" error-free scores of subjects (inter-subject variance) compared to the total variance ("true" variance + variance due to measurement errors). The ICC is calculated from a 2-way mixed effects model of repeated-measures analysis of variance which particularly fits any kind of test-retest experiment designs: the total variance is partitioned between within- and between-objects (subjects) variances. A commonly used index to report repeatability is the Pearson's correlation coefficient. The ICC coefficient value is often close to the Pearson's correlation value. However, the ICC includes a penalization for a systematic error between measurements (in this case, the ICC would be lower than the Pearson's) and it can also assess the reliability of a measure based on more than two measurements by subjects (thanks to the model of analysis of variance used for computation). Moreover, the Pearson's coefficient normalizes each

measurement by its own mean and SD, whereas the ICC normalizes the variables by the pooled mean and SD of both measurements. So if the variables do not have a common unit and variance, the Pearson's is more appropriate. But, for test-retest measurements having the same units, the ICC is a better index [59].

The higher the ICC, the higher the reliability; the upper threshold above which the ICC would reflect a good reliability remains subjective and depends on the application but we can still refer to the scale proposed by Shrout and Fleiss [58], Fleiss [60] and Cicchetti [61]: poor $< 0.4 <$ fair $< 0.6 <$ good $< 0.75 <$ excellent ≤ 1. Chinn [62] suggests that measure needs to have at least an ICC coefficient of 0.6 to be useful. Contrary to the other repeatability indexes of this section, the ICC coefficient is a dimensionless index.

In this study, the ICC coefficient was computed according to the Matlab implementation of McGraw and Wong [59] (case 3A).

Minimal Detectable Change

Another useful index is the Minimal Detectable Change (MDC). It estimates the minimal difference between two scores that would reflect a "true" difference (i.e., not completely due to measurement error). It can be derived according to:

$$MDC = 1.96\sqrt{2} \cdot SEM$$

where $SEM = SD_{pooled}\sqrt{1 - ICC}$ is the Standard Error of Measurement and SD_{pooled} is the standard-deviation across all measurements [49,63]. The MDC can also be interpreted as an interval for repeated measures. If x is the score of a subject for a single measurement, there is a 95% chance that the score of a repeated measurement lies within $x \pm MDC$, assuming that the measurement errors are normally distributed. Any difference of ± MDC between two metric values can be considered as usual variation (due to measurement error); such a difference is not exceptional enough to be considered as a real change in the microstructure.

The MDC and the CI_d are based on the same idea of estimating the magnitude of the difference in metric values that can be only due to measurement errors. However, the MDC applies for two single metric values whereas CI_d, which takes into account the sign of the difference between test and retest, applies for group comparison where negative measurement errors compensate for positive ones.

Comparison of indexes with different units across studies

To allow the comparison between techniques having different measuring units, one can express the repeatability indexes as a percentage of the mean across all measures, similar to calculation of the coefficient of variation ($CV = 100 \cdot SD/mean$). This method works fine when the mean is similar between techniques, otherwise the comparison is biased by the mean. For example, it has been shown that MTR could lead to drastically different mean values when acquired with different offset saturation pulse parameters, e.g. from 9 to 51% in the healthy WM [45]. Hence, normalizing by the mean would yield lower indexes for techniques with higher mean value, whereas these techniques could have the same test-retest repeatability as other techniques with lower mean values. To avoid this while still being able to compare between techniques side by side, we expressed these reliability indexes as a percentage of the SD across subjects of the first MRI session values only ($SD_{subjects}$), i.e.:

$$Index_{\%\ of\ SD_{subjects}} = 100 \cdot \frac{Index}{SD_{subjects}}$$

where *Index* represents any reliability index expressed in the metric unit such as the MDC. Indeed, this manipulation enables us to compare metrics side by side while accounting for the property we are looking for. Here, we are looking for a metric that has low test-retest

variability relative to the inter-subject variability, i.e. relative to the dispersion of the sample this metric can offer. The SD across subjects is the most basic measure of the sample dispersion. In this way, we would like the $Index_{\%\ of\ SD_{subjects}}$ to be as low as possible (i.e., a low measurement error and a high SD across subjects) in order to observe differences between subjects that are higher than measurement errors.

2.3.2. Sensitivity to myelin content variations. To assess the metrics sensitivity to the variations in myelin content across vertebral levels/WM regions relative to the repeatability, differences in group mean (n = 33) between levels/regions were compared along with their measurement error (assessed by the CI_d).

Moreover, a one-way repeated measures ANOVA between levels/regions was performed independently for each metric (n = 33). The assumptions of normal distribution within each group (i.e., level or WM region) and of sphericity were checked using Lilliefors's test and Mauchly's test respectively. When the assumption of sphericity was not met, a Greenhouse-Geisser correction was used to compute the ANOVA. When the ANOVA detected a significant difference, a post hoc multiple comparison test using the Tukey's honestly significant difference criterion was performed in order to find which groups were significantly different from each other.

To test the metrics sensitivity to the demyelination with aging reported by histology in the literature [64–66], for each vertebral level/WM region, means across each age group were compared taking the measurement error (assessed by the CI_d from the previous analysis) into account in order to investigate whether the difference in means could reflect a "true" difference or whether it is indistinguishable from measurement errors.

In addition, to test for significant differences, we performed independently for each metric, on the larger sample (n = 33, n_{young} = 19, $n_{elderly}$ = 14), two-way repeated ANOVAs with the age group as between-subjects factor and, as within-subjects factor:

- vertebral levels to determine if this effect was consistent across levels (the metric being quantified in the whole WM);
- ROIs (WM, DC, LF, VF) to determine if this effect was consistent across ROIs (the metric being quantified from C2 to C4).

Finally, to complete this study, a power analysis was performed for two-tailed t-tests between young and elderly subjects based on whole WM values of each metric.

3. Results

3.1. Repeatability

Fig 1 shows test and retest multi-parametric maps by vertebral levels, for one single young and one single elderly subject, as well as for the group average (n = 33). The single subject data look noisy, however the average map shows clear distinction between WM and GM. Moreover, the symmetry that can be observed on the group average maps suggests no apparent differences in myelin content between left and right cord. In all metrics, the heterogeneity of values across WM regions suggests different microstructural compositions. For example, the fasciculus cuneatus shows higher MTV than the fasciculus gracilis, suggesting higher myelin content in agreement with previous histology studies [1,67]. Apart from MTR, all metrics show fairly stable values across vertebral levels.

A guide for reading (and understanding) figures and tables in the paper. Fig 2 shows intra- and inter-subject differences for metrics quantified in the WM. Fig 2 is a subset of Table 1, which quantifies the metrics repeatability over all WM at the different cervical levels

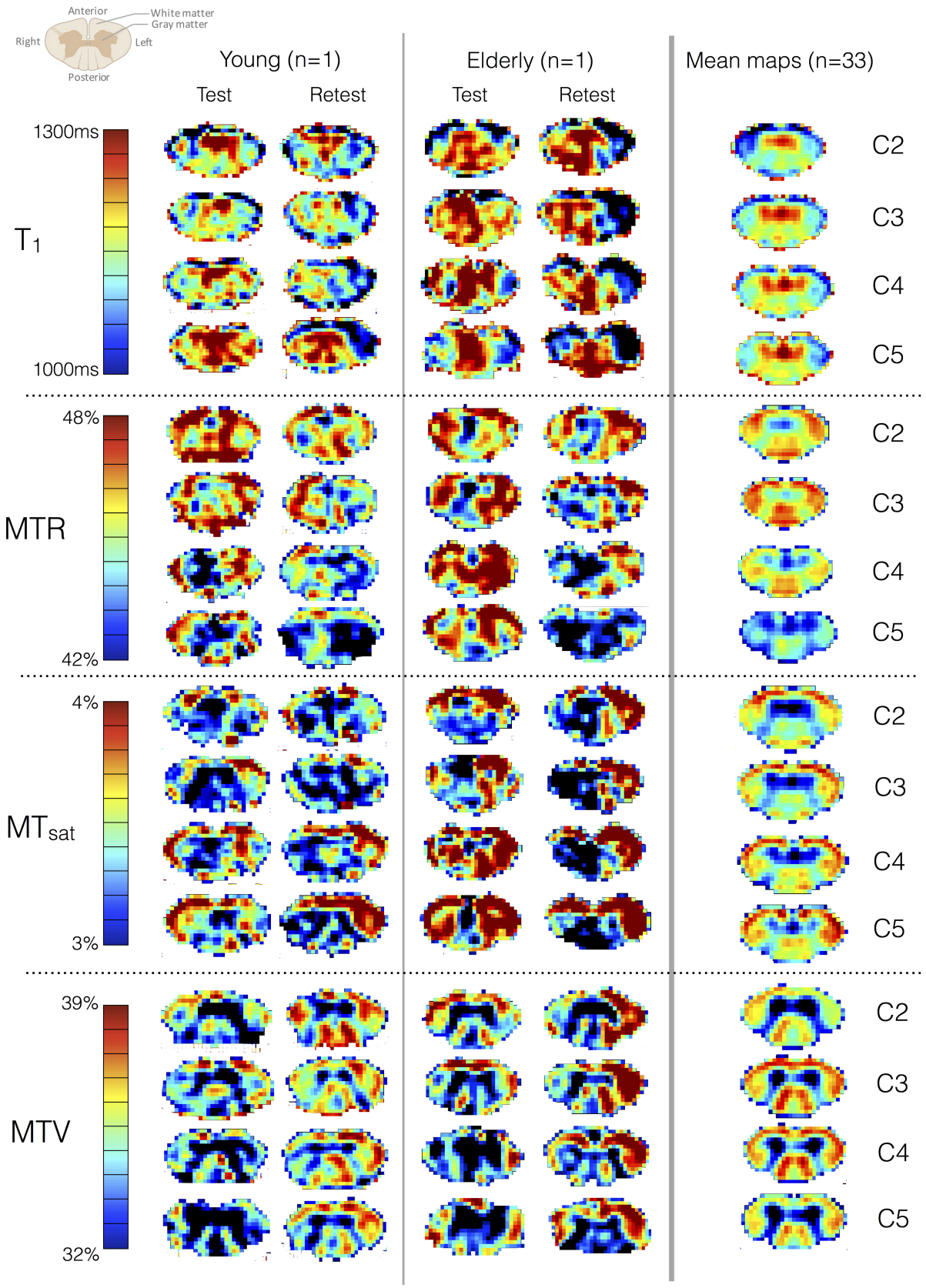
Anterior
White matter
Gray matter
Right
Left
Posterior
Young (n=1)
Elderly (n=1)
Mean maps (n=33)
Test
Retest
Test
Retest
1300ms
T1
1000ms
48%
MTR
42%
4%
MTsat
3%
39%
MTV
32%
C2
C3
C4
C5

Fig 1. Test and retest maps in a young and an elderly subject at each vertebral level (mean across levels) along with the mean maps across the 33 subjects. All these maps are in the template space. Note that the color bar scale has been adjusted to the mean maps contrast. On a single-basis subject, one can observe a somewhat poor test-retest repeatability, within and across slices. However, despite this poor repeatability, the average maps (here, n = 33) are more consistent in terms of symmetry and tract-specific variations. For example, we can clearly distinguish higher MTV in the fasciculus cuneatus versus in the gracilis (dorsal column), which is in agreement with previous histology work [1,67].

https://doi.org/10.1371/journal.pone.0189944.g001

(Fig 3 and Table 2 are their analogs quantifying the metrics repeatability over all reliable levels within the different WM sub-regions). Let's take an example to better explain how to use these repeatability indexes. Let's take the T_1 at C3. Regarding only one scan, the mean T_1 across the group is 1007.2ms and the SD is 74.3ms. A 95% confidence interval for the mean test-retest difference of [-38.5; 23.1]ms indicates that if we rescan the same group a second time, the mean is likely to lie between 968.7 and 1030.3ms (with 95% probability). Now, if we measure T_1 at C3 in a different group (e.g., a group of patients) and the resulting mean lies between 968.7 and 1030.3ms, we will not be able to report whether the difference in T_1 between the two groups is due to measurement errors or to a true difference in T_1. The MDC (113.2ms in our example case) will be useful for instance in a case where a clinician measures the T_1 in a new lesion of his patient at one time point t; say he gets a measure of $T_1(t) = x$ ms. If he re-measures it right after, there is 95% probability that $T_1(t + 30min)$ lies within $x \pm 113.2$ ms. Now, if he wants to control the evolution of the lesion one year later and he measures $T_1(t + 1year)$ still within $x \pm 113.2$ ms, he will not be able to say whether this change between $T_1(t)$ and $T_1(t + 1year)$ is due to an evolution of the tissue or to measurement errors.

The ICC and the MDC (expressed in percentage of the SD across subjects) are useful to compare repeatability across metrics (more extensively done in Fig 4). For example, if we compare T_1 to MTR at C3, the ICC is much higher for T_1 (0.72) than MTR (-0.3)–note here that the interpretation of a negative value for the ICC is the same as for a null value (very poor reliability). This is because T_1 has a lower test-retest variation ($\overline{|d|}_{C3}$ = 47.1ms in Fig 2) compared to the variation between subjects ($SD_{subjects}$ = 74.3ms in Table 1), whereas MTR has a high test-retest variation ($\overline{|d|}_{C3}$ = 1.43% in Fig 2) compared to the variation between subjects ($SD_{subjects}$ = 1.38% in Table 1). This also reflects in the MDC ($MDC = 1.96\sqrt{2} \cdot SD_{total}\sqrt{1 - ICC}$). For T_1 at C3, MDC = 113.2ms, which is 152.3% of $SD_{subjects}$ (Table 1), whereas for MTR at C2, MDC = 3.76%, which is 271.6% of $SD_{subjects}$. This result shows that measurement errors in MTR cover almost 3 times the standard variations between subjects, making it difficult to observe true differences in MTR.

The mean test-retest difference ($\overline{|d|}$, displayed in gray at the top left of each graph) is higher at C5 (Fig 2); however, one-way repeated ANOVAs testing the effect of vertebral levels on the absolute test-retest difference did not report significant results (p-values were 0.183, 0.195, 0.389 and 0.579 for T_1, MTR, MT_{sat} and MTV respectively). No clear test-retest difference between young and elderly subjects is observed on this graph.

For all metrics and all levels, no significant systematic bias between test and retest is detected (all CI_d include 0, see Table 1). When compared to other metrics, mean MT_{sat} shows minimal variations across vertebral levels (p-values of the repeated ANOVAs between levels were $<<0.0001$, $<<0.0001$, 0.02 and <0.0001 for T_1, MTR, MT_{sat} and MTV respectively). The ICC coefficient highlights a poor test-retest reliability, barely exceeding 0.5, especially for MTR and MT_{sat}. This point is supported by the MDC, which is generally around 2 times the SD across subjects.

Fig 3 shows repeatability results within sub-regions of the WM: dorsal column (DC), lateral funiculi (LF) and ventral funiculi (VF). Overall, the VF shows the largest test-retest differences.

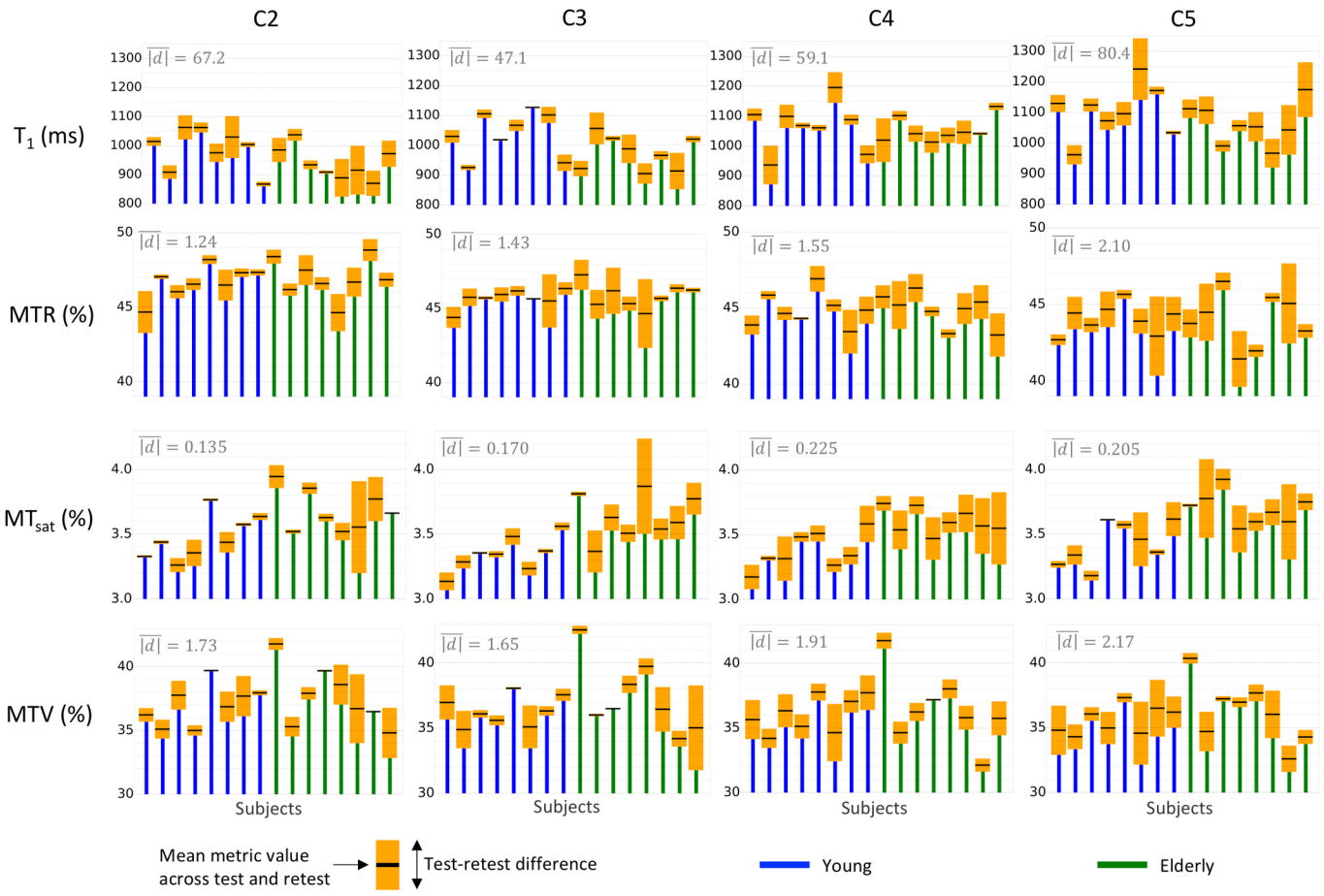

Fig 2. Subjects' distribution with test-retest differences quantified over all WM according to vertebral levels. The top and bottom of the orange boxes respectively represent the max and min among test and retest, while the black line in the middle of the box represents the mean. Note that the y-axis does not start from zero for the sake of clarity. The mean absolute difference between test and retest (mean height of orange boxes, $\overline{|d|}$) is displayed in the top left hand corner of each graph. This figure gives a comprehensive view of the repeatability compared to between-subject differences.

https://doi.org/10.1371/journal.pone.0189944.g002

These observations were confirmed (except for MT_{sat} which shows large test-retest differences in the DC) by one-way repeated ANOVAs performed between ROIs on the absolute test-retest difference (p-values <0.01, 0.01, 0.08, <0.01 for T_1, MTR, MT_{sat}, MTV respectively). In addition, similar repeatability is found when the metrics are estimated over all WM or within the DC or the LF.

Fig 3 is a subset of Table 2, which quantifies the metrics repeatability within sub-ROIs of the WM from C2 to C4. Interestingly, MT_{sat} performs really differently according to the ROI, yielding the worst repeatability result in the DC (ICC = 0.1, MDC $\approx$ 3 inter-subject SDs) and the best one in the LF (ICC = 0.82, MDC $\approx$ 1.2 inter-subject SDs). Note however that estimating the metric at several levels (here, C2 to C4) is not favorable to MT_{sat} given that its ICC in WM at C4 is half its ICC at C3 (Table 1). Overall, T_1 and MTV yield the best results. MTV regularly shows a fair repeatability whatever the ROI is, with a MDC about 1.5 to 2 times the inter-subject SD (which is equivalent to 87–95% of the sample distribution). In the level-wise analysis, MTV performs slightly better than T_1. We suspect that these results reflect the clearer delineation between the cord sub-regions and the more homogeneous values in those sub-regions that could be observed in MTV maps when compared to T_1 or even MT_{sat} maps (Fig

Table 1. Repeatability indexes used to assess the repeatability of metrics over all WM according to vertebral levels.

	Level	Mean ± $SD_{subjects}$	CI_d	ICC	MDC [% of $SD_{subjects}$]
T_1 (ms)	C2	964.9 ± 70.7	-17.2 to 66.0	0.46	± 158.4 [224.1]
	C3	1007.2 ± 74.3	-38.5 to 23.1	0.72	± 113.2 [152.3]
	C4	1060.0 ± 69.5	-63.6 to 4.8	0.53	± 135.5 [195.0]
	C5	1083.6 ± 95.2	-68.0 to 33.4	0.43	± 189.1 [198.7]
MTR (%)	C2	46.83 ± 1.52	-0.99 to 0.54	0.43	± 2.85 [186.7]
	C3	45.78 ± 1.38	-1.54 to 0.42	-0.3	± 3.76 [271.6]
	C4	44.87 ± 1.55	-1.14 to 0.75	0.16	± 3.53 [228.1]
	C5	44.02 ± 1.9	-2.0 to 0.66	0.05	± 5.06 [265.9]
MT_{sat} (%)	C2	3.579 ± 0.194	-0.12 to 0.113	0.5	± 0.429 [220.6]
	C3	3.492 ± 0.184	-0.058 to 0.189	0.51	± 0.466 [253.6]
	C4	3.49 ± 0.21	-0.003 to 0.247	0.27	± 0.501 [238.1]
	C5	3.562 ± 0.266	-0.132 to 0.162	0.33	± 0.544 [204.9]
MTV (%)	C2	37.36 ± 2.38	-1.58 to 0.81	0.48	± 4.46 [187.4]
	C3	36.84 ± 2.55	-0.94 to 1.53	0.52	± 4.57 [178.8]
	C4	36.25 ± 2.44	-0.64 to 1.57	0.6	± 4.15 [169.8]
	C5	35.92 ± 2.5	-0.98 to 1.73	0.33	± 5.05 [202.2]

From left to right, the columns correspond to the mean ± SD across subjects (n = 16) based on values from the first scan session only, the 95% confidence interval for the true test-retest difference, the ICC coefficient, the MDC. All numbers are in the metric unit except those in square brackets, which are expressed as a percentage of the SD across subjects to quantify the repeatability relative to the inter-subject difference, i.e. the reliability. Fig 2 is a subset of this table.

https://doi.org/10.1371/journal.pone.0189944.t001

1). Furthermore, as expected, MTR regularly performs worst, in part because of the low contrast between subjects it exhibits, whatever the ROI is.

Fig 4 compares three main repeatability indexes (absolute test-retest difference, ICC and MDC) between the different metrics. While no particular metric stands out from this comparison, MTR seems to be the least reliable at every level. For most of the vertebral levels, $\overline{|d|}$ of MTR is on the same order as the inter-subject SD (which is equivalent to 68% of the population if we assume a Normal distribution for the sample), the ICC is below 0.4 at every level and the MDC exceeds 2.5 inter-subjects SDs (equivalent to 98.8% of the population) at 2 levels over 4. When considering the effect of vertebral level, C5 seems to be the least reliable (ICC < 0.5 for all metrics). Regarding the effect of WM regions (Fig 4B), some differences are observed. For instance, MT_{sat} yields the best ICC score in the LF (0.82) and the worst in the DC (0.1).

3.2. Sensitivity to myelin content

This section deals with the larger sample (n = 33 subjects).

3.2.1. Effects of vertebral levels and WM regions. Fig 5 plots the group mean along with the measurement error magnitude (CI_d) in order to allow the reader to assess whether differences between vertebral levels or WM regions can be distinguished from measurement errors or not. Individual subjects data are also plotted to see if differences between subjects can be carried out despite the measurement error. However, for individual comparison, measurement errors are assessed by the MDC, which is much larger than the CI_d (as negative and positive errors do not compensate for each other). Only T_1 and MTV seem to allow the comparison between some healthy subjects.

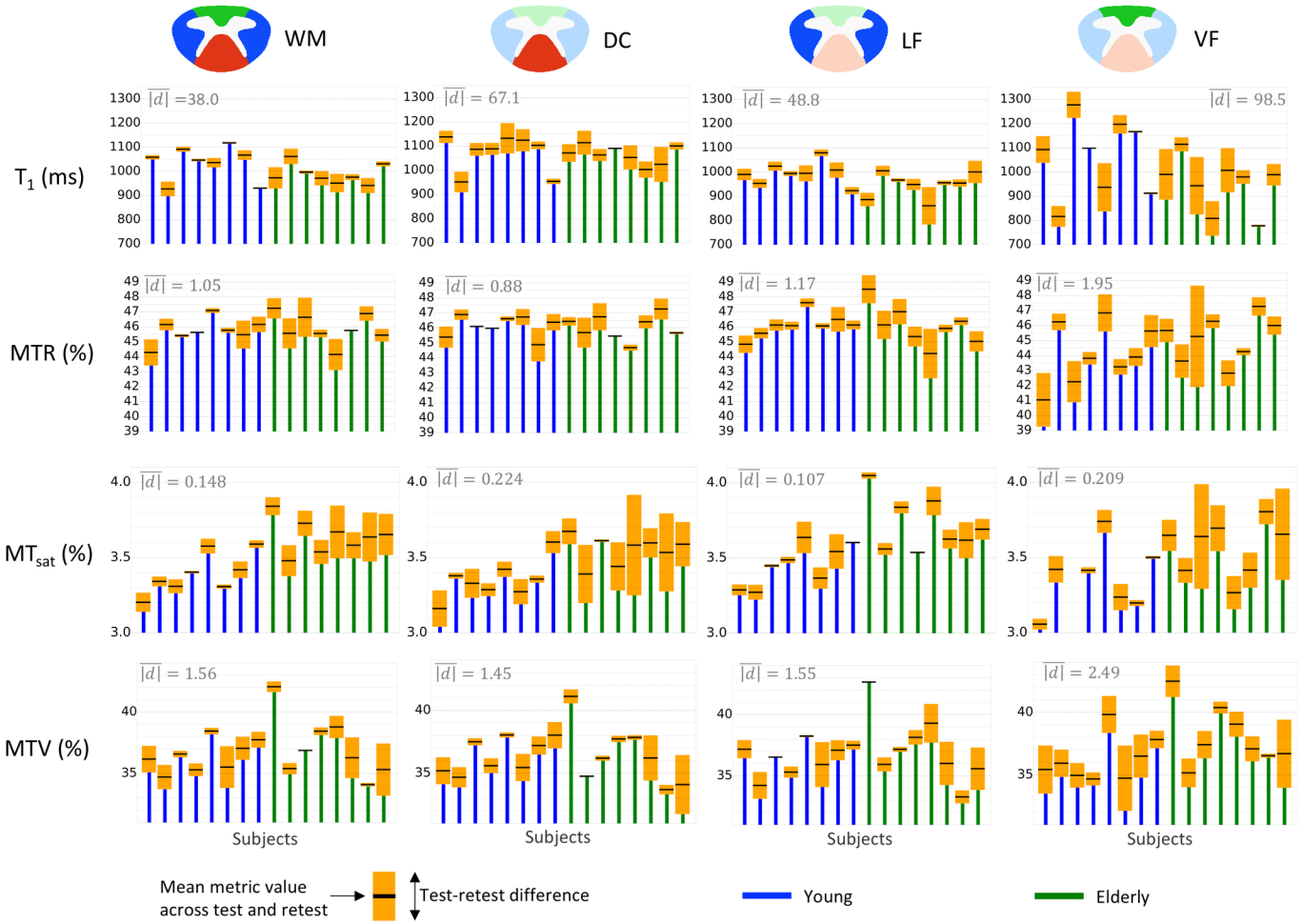

Fig 3. Subjects' distribution along with the test-retest difference for each metric in the four ROIs. The top and bottom of the orange boxes are respectively the max and min among test and retest, while the black line in the middle of the box is the mean. The mean absolute test-retest difference (mean height of orange boxes, $\overline{|d|}$) across subjects is displayed in the top left hand corner of each graph. Due to its tiny size and its border location between GM and CSF, the VF yields the largest test-retest variations.

https://doi.org/10.1371/journal.pone.0189944.g003

The differences that are distinguishable from measurement errors were sum up in Table 3, along with the results of the one-way repeated ANOVAs. One can observe that some cases show significant differences but those differences are too small to be distinguished from measurement errors. This is the case for the MTR which is significantly different between every vertebral level but only C2 and C5 show a difference large enough to be due to something else than measurement errors. Also, significant differences between WM regions are found with MTR and T_1 but none of them are larger than measurement errors.

3.2.2. Effect of age. Fig 6 compares the differences between young and elderly to the measurement errors assessed by the CI_d. With all metrics within every spinal cord region (vertebral level or WM region), the difference between young and elderly can always be explained by measurement errors only. Moreover, the repeated ANOVAs did not report any significant effect of age for all metrics, neither level-wise nor ROI-wise. However, we can still notice some general trends: T_1, MTR and MTV generally support the demyelination with aging histologically reported in the literature, whereas MT_{sat} constantly shows the reverse trend.

Table 2. Repeatability indexes used to assess the repeatability of metrics in different sub-regions of the WM.

	ROI	Mean ± $SD_{subjects}$	CI_d	ICC	MDC [% of $SD_{subjects}$]
T_1 (ms)	WM	1011.2 ± 60.8	-29.4 to 19.3	0.74	± 89.3 [146.8]
	DC	1068.3 ± 63.7	-69.0 to 5.4	0.41	± 146.8 [230.5]
	LF	971.6 ± 64.5	-22.9 to 39.6	0.52	± 115.6 [179.2]
	VF	1006.8 ± 168.0	-60.0 to 71.5	0.68	± 240.8 [143.4]
MTR (%)	WM	45.82 ± 1.3	-0.99 to 0.37	0.28	± 2.55 [196.2]
	DC	46.08 ± 0.87	-1.02 to 0.07	0.29	± 2.15 [247.3]
	LF	46.09 ± 1.54	-0.94 to 0.53	0.38	± 2.73 [178.1]
	VF	44.64 ± 2.48	-1.65 to 0.93	0.35	± 4.8 [193.1]
MT_{sat} (%)	WM	3.517 ± 0.177	-0.022 to 0.152	0.6	± 0.34 [192.4]
	DC	3.452 ± 0.181	-0.029 to 0.25	0.1	± 0.543 [299.5]
	LF	3.59 ± 0.202	-0.009 to 0.118	0.82	± 0.252 [124.9]
	VF	3.438 ± 0.283	-0.124 to 0.172	0.54	± 0.546 [192.6]
MTV (%)	WM	36.79 ± 2.3	-0.96 to 1.13	0.6	± 3.83 [166.4]
	DC	36.46 ± 2.24	-0.65 to 1.34	0.6	± 3.7 [164.7]
	LF	36.88 ± 2.41	-1.21 to 0.89	0.65	± 3.87 [160.6]
	VF	37.17 ± 2.95	-1.2 to 1.81	0.42	± 5.58 [189.1]

From left to right, the columns correspond to the mean ± SD across subjects (n = 16) based on values from the first scan session only, the 95% confidence interval of the true test-retest difference, the ICC coefficient, the MDC. All numbers are in the metric unit except those in square brackets, which are expressed as a percentage of the SD across subjects to quantify the repeatability with respect to the inter-subject difference, i.e. the reliability. Fig 3 is a subset of this table.

https://doi.org/10.1371/journal.pone.0189944.t002

To complete this study, Table 4 reports the statistical power analysis. From this analysis, one can compare the difference that can be detected given the metrics test-retest errors (length of the CI_d, 2nd column) to the minimum difference in the true metric values required to detect a significant difference (1st column) between young and elderly (with a fair test power). We can notice for example that, given the measurement errors of MTR (1.36%), even if the difference in means were large enough (≥1.27%) to yield a significant result, the imprecision of measurement is too large to detect such a difference. It is not the case with the other metrics. Moreover, we can notice that the observed differences in means (3rd column) are very low compared to the difference needed to obtain significant results (1st column), yielding very low statistical power for those tests (4th column). Finally, given the large sample size required to obtain a significant difference (5th column), T_1 and MTV do not seem sensitive to age groups (based on their mean WM values in this study).

4. Discussion

This study proposes a statistical framework for comparing clinically feasible myelin imaging techniques (T_1, MTR, MT_{sat} and MTV) in the cervical spinal cord.

4.1. Myelin-sensitive metrics values in the spinal cord

The resulting mean values across subjects are in agreement with previous studies. Stikov et al. [68] observed a T_1 around 1000ms in the brain, which is comparable to the T_1 in the spinal cord WM *in-vivo* at 3T [69,70]. The same holds for our MTV measurements which are in agreement with reported PD values [12,18–23,69,71]. There is no gold-standard for clinically feasible MT-based protocols due to their dependence on pulse sequence parameters. However,

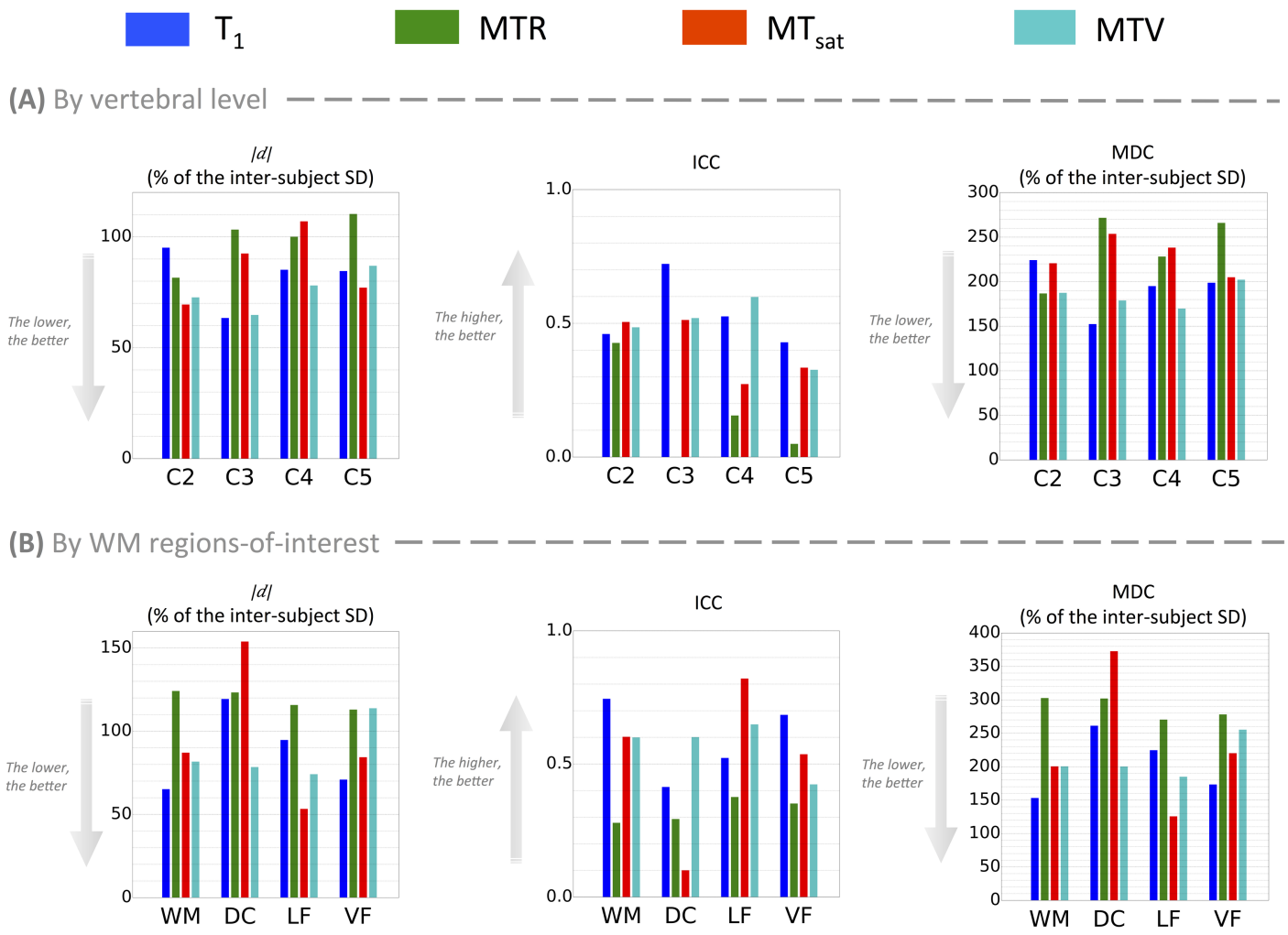

Fig 4. Comparison between the repeatability of the four myelin-sensitive metrics when the metric is estimated (A) in the whole WM by vertebral level and (B) from C2 to C4 within WM sub-ROIs. Repeatability indexes from left to right: mean absolute test-retest difference ($\overline{|d|}$), Intra-Class Correlation (ICC) coefficient, Minimal Detectable Change (MDC). $\overline{|d|}$ and MDC are expressed in percentage of inter-subject SD in order to assess the repeatability relative to the differentiation between subjects (i.e., the reliability), despite the different units of the metrics.

https://doi.org/10.1371/journal.pone.0189944.g004

the values for MTR and MT_{sat} we observed are also in agreement with literature [6,45,48,72–75].

4.2. Repeatability

Even for the most reliable metrics (T_1 and MTV, see Fig 4), the ICC is moderate (around 0.5) and the MDC is on the order of two inter-subject SDs. Given the test-retest variations, the minimal difference between individual healthy subjects that can be detected with these metrics (MDC) is much larger than the usual variations we observed (see Fig 5). Looking at groups of subjects, significant differences between spinal cord regions stand out but still, they are not large enough to be distinguished from measurement errors (quantified by the CI_d in this case, as shown in Fig 5).

In comparison with the brain, repeatability in the spinal cord is hampered by multiple sources of artifacts (motion, susceptibility) and low SNR [43]. Better repeatability might be

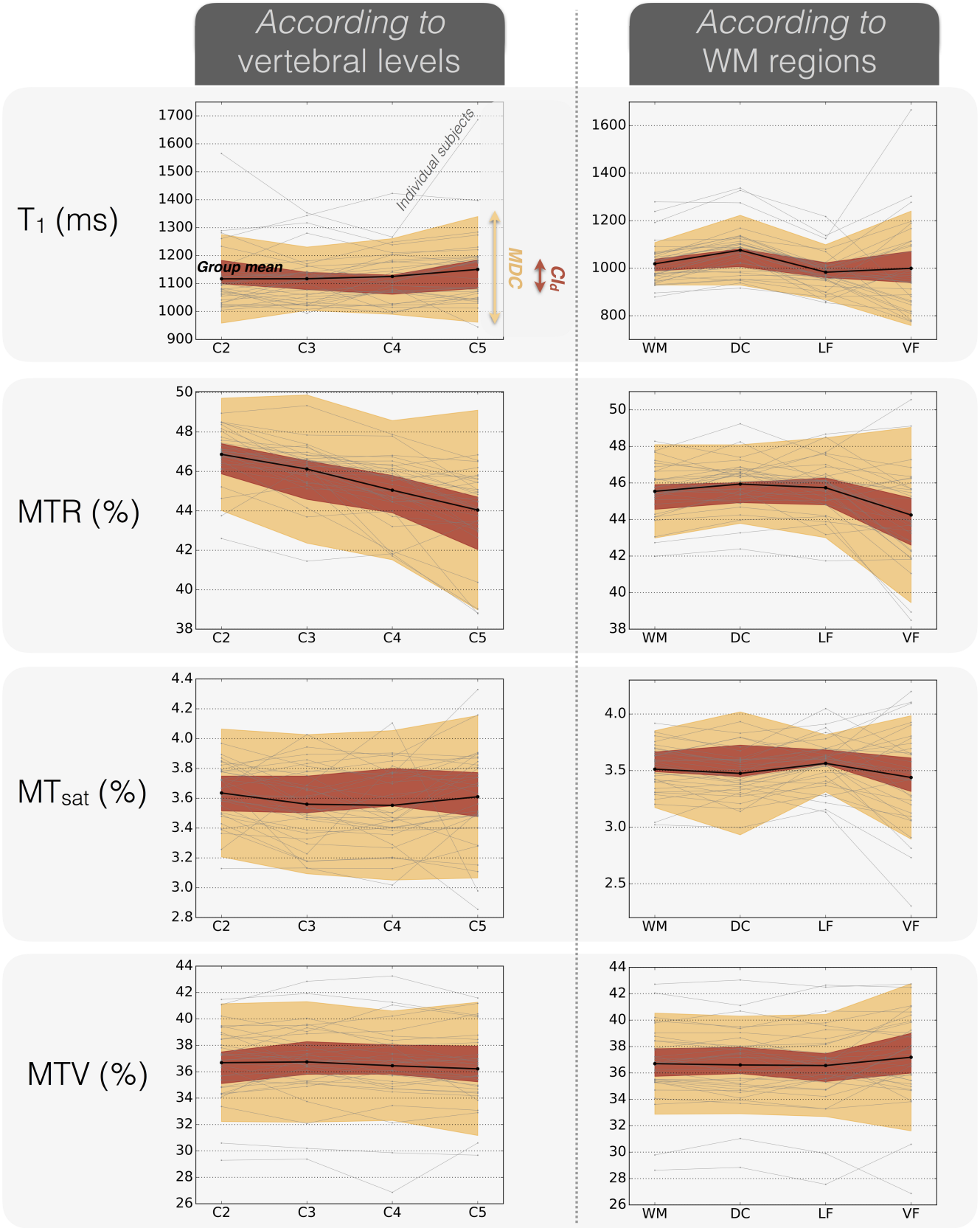
According to vertebral levels
According to WM regions
T1 (ms)
MTR (%)
MTsat (%)
MTV (%)
Individual subjects
Group mean
MDC
CId
C2
C3
C4
C5
WM
DC
LF
VF

Fig 5. Comparison across vertebral levels and WM regions along with the measurement errors for the group mean (n = 33) and individual subjects. The red envelope represents the 95% confidence interval for the test-retest difference (CI_d), which assesses the measurement error magnitude of the group mean (in black). The orange envelope represents the MDC (Minimum Detectable Change), difference required to compare individual subjects (faded gray lines). Note that the group mean approaching the edges of the CI_d (red envelope) reflects an asymmetric confidence interval due to a non-null offset between test and retest (non-null mean test-retest difference, $\overline{d}$). However, no offset was large enough to report a significant systematic bias between test and retest (see section 3.1. Repeatability, Table 1 and Table 2).

https://doi.org/10.1371/journal.pone.0189944.g005

achieved with coarser resolution and/or more averaging, though at the cost of longer acquisition times, which could be associated with more subject motion.

Taso et al. [44] reported results for myelin-related metrics in the spinal cord WM: a CV of 5.3% for MTR and 2.9% for ihMT ratio. However, this study reported the repeatability in terms of CVs, which are misleading when comparing metrics with different units and/or dynamic ranges (as mentioned in section 2.3.1. Repeatability). Smith et al. [48] reported a CI_d of [− 3%, +5%] for MTR over all WM from C2 to C5 within 10 young healthy subjects. Even if the repeatability of the metrics reported in our study is not good enough to differentiate between WM regions or age groups, it is still much better (CI_d of [− 0.99%, +0.54%] for MTR). This may suggest that significant differences not accounting for precision of measurements might have been reported in the literature, whereas they could be only explained by measurement errors.

Looking at the metrics individually, T_1-based metrics (MTV and T_1) generally show the best reliability (Fig 4). Regarding sensitivity to myelin, MTV shows clearer delineation of the GM and smooth variations in the WM (Fig 1), but no difference between WM regions stood out when compared to the measurement error. When looking at individual maps, T_1 seems particularly affected by cord movements and compressions occurring during respiratory and cardiac cycles (Fig 1), which produces statistically significant differences (see Table 3), but those differences are not larger than measurement errors. The same applies for MTR, which emerges as the less reliable metric due to its very small variation between subjects (Fig 4). However, MTR is the only metric exhibiting a significant effect that accounts for measurement error (difference between vertebral levels C2 and C5 in Table 3). This decrease in MTR towards lower levels could reflect a true decrease in myelin content, but could also be due to B_1^+ inhomogeneity. MTR variations due to B_1 errors have already been reported in the brain [76] and correcting for them should be further investigated in the spinal cord. MT_{sat} minimizes the T_1 contribution included in MTR, and is thereby less variable across vertebral levels.

Table 3. Comparison of significantly different vertebral levels (A) or WM regions (B) with differences larger than measurement errors.

	(A) Analysis by vertebral levels			(B) Analysis by WM regions		
	p	***Significantly different levels (p<0.05)***	***Differences larger than measurement errors***	***p***	***Significantly different ROIs (p<0.05)***	***Differences larger than measurement errors***
T_1 (ms)	0.041	• C2 vs. C5	None.	<0.01	• DC vs. LF • DC vs. VF	None.
MTR (%)	$<<10^{-4}$	• All levels are significantly different from each other.	• C2 vs. C5	$<<10^{-4}$	• DC vs. VF • LF vs. VF	None.
MT_{sat} (%)	0.189		None.	0.076		None.
MTV (%)	0.081		None.	0.085		None.

For each analysis (A, B), the left column is the results of the one-way repeated ANOVAs whereas the right column reports the vertebral levels/WM regions showing differences larger than measurement errors (see also Fig 5).

https://doi.org/10.1371/journal.pone.0189944.t003

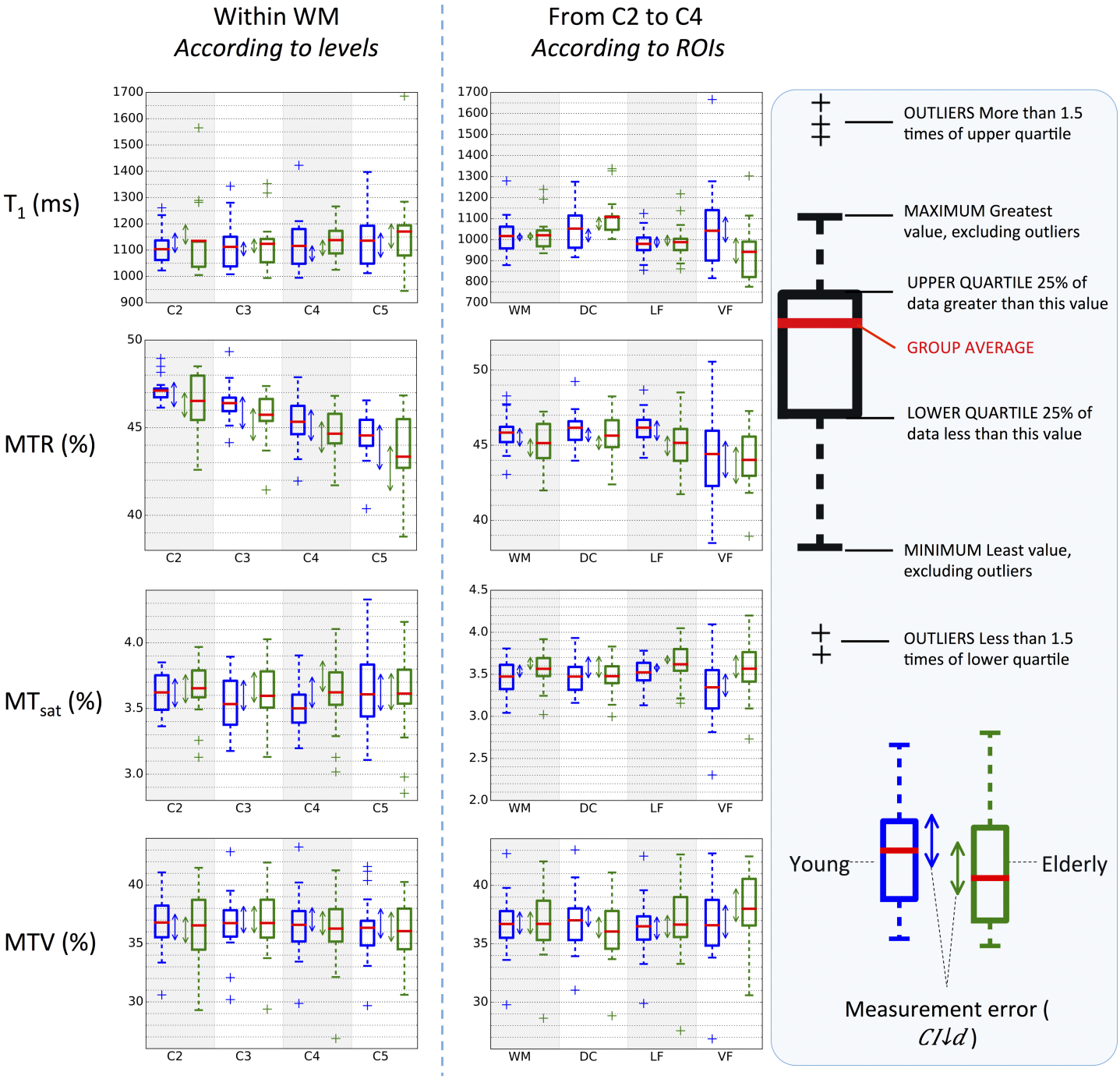

Fig 6. Comparison between young (n_{young} = 19) and elderly ($n_{elderly}$ = 14) subjects along with measurement errors. For each case, the corresponding 95% confidence interval for the mean test-retest difference (CI_d), estimated from the test-retest analysis (see section 3.1. Repeatability) was centered at the mean of each group, in order to assess whether the difference between young and elderly is larger than the test-retest errors or not. With all metrics within every spinal cord region (vertebral level or WM region), the difference in means between young and elderly was undistinguishable from measurement errors.

https://doi.org/10.1371/journal.pone.0189944.g006

Table 4. Power analysis based on each metric WM values for a two-sample t-test between young and elderly subjects with a significance level of 5%.

	Minimum difference required to detect a significant difference with such a sample and 80% probability (effect size)	Length of CI_d	Observed difference in means (*young – elderly*)	Power (probability to detect a significant difference with such a sample)	Sample size required to detect a significant difference with such means and 80% probability
T_1 *(ms)*	93.8	48.7	-3.6	5.1%	10394
MTR (%)	1.27	1.36	0.70	33.5%	52
MT_{sat} *(%)*	0.203	0.174	-0.092	24.5%	75
MTV (%)	2.67	2.09	-0.01	5.0%	1690133

https://doi.org/10.1371/journal.pone.0189944.t004

4.3. Sensitivity and specificity to myelin with MRI

The assessment of the sensitivity of metrics to myelin content remains difficult, due to the lack of a ground truth. A loss of myelinated fibers with aging (mainly the small caliber ones) was observed histologically in the brain [77] and cervical spinal cord [64–66] but it remains unclear if these variations can be detected by clinical MRI nowadays. Age effects have been reported in the brain with MTR [78] and DTI [79–82]. In the spinal cord, most age effects are reported with DTI [83–85]. One study investigated MTR evolution in the spinal cord during aging, but no significant effect was reported [44]. The same study reported a decrease in ihMT ratio between subjects aged 35 to 50 and subjects aged over 50, not accounting for measurement errors however. Our study did not observe any difference between age groups, with or without accounting for measurement error (Fig 6). This lack of sensitivity to aging could be due to the choice of acquisition parameters, the small effect/sample size, or simply due to a lack of true differences in myelination.

As noted in the introduction, some of the myelin-sensitive techniques are also hampered by confounding factors. For example, T_2^* is affected by iron content, fiber orientation, blood vessels and blood oxygen level. MTR is affected by T_1 and B_1 field, and more generally, magnetization transfer and MTV are sensitive to macromolecules (i.e., not only myelin). For each of these techniques, there are ways to mitigate those confounds. For example, quantitative susceptibility maps could inform T_2^* maps, or T_1 and B_1^+ fields could be acquired to correct MTR maps [76]. All these strategies come at the cost of additional scan time, and possibly larger output variance (due to the introduction of yet other noisy measures).

While DTI has some intrinsic limitations, other techniques also based on diffusion-weighted imaging might offer more sensitivity to myelin. It is important to note, however, that because water protons trapped between myelin sheaths have a short T_2 (around 10 ms at 3T, which could be quantified using myelin water fraction techniques) and that protons from bound molecules have an even shorter T_2 (order of μs, which could be quantified with ultra-short TE imaging or magnetization transfer techniques), diffusion-weighted protocols typically use a TE ($>$ 60ms) too long to be sensitive to signal coming from the myelin (and from water trapped in it). Some advanced diffusion-weighted techniques include NODDI [47,86], which can notably estimate the intra-cellular volume fraction and CHARMED/AxCaliber [87–89], which can notably estimate the hindered (extra-cellular) and restricted (intra-cellular) water fraction. All these metrics are thus indirectly related to the myelin volume fraction, although additional information would be required to be able to quantify absolute myelin content.

To improve specificity to myelin, combining several metrics, using for example independent component analysis, or acquiring maps of confounding factors for a posteriori corrections, might be advisable [90]. Future work will be undertaken in this direction [91].

4.4. Perspective of repeatability assessment

Repeatability assessment is crucial for the development of qMRI biomarkers. Our results show that significant differences between groups can be reported with standard statistical tests, yet these differences are comparable to (or even smaller than) test-retest measurement errors. Controlling for both aspects (statistical significance and measurement errors) is necessary for qMRI studies.

The indexes reported in this work (95% confidence interval for the test-retest difference (CI_d), ICC and MDC) are useful for quantifying repeatability and allowing comparisons across studies. As mentioned before, the coefficient of variation depends on the magnitude of the metric, and should not be the primary index for assessing repeatability, especially if metrics have different means or units. The CI_d first allows to control for the existence of a potential systematic bias between measurements (i.e. scan sessions). In addition, it gives an estimation of the measurement error for group averages. In the same vein, the MDC provides a measure of the minimum difference between two individual measurements to report a true difference, taking into account the measurement errors. For example, the CI_d would be useful for researchers comparing different populations, whereas the MDC would be useful for a clinician needing to assess the evolution of a WM lesion within a single patient. Furthermore, the ICC coefficient has the advantage to be dimensionless, and can thus be easily compared to assess reliability across metrics, studies, vendors or sites. Aside from providing a robust quantification of the repeatability with two measurements (test-retest studies), the ICC coefficient (and consequently, the MDC) can also be consistently used with more than two measurements. Those reliability indexes have already been extensively used in test-retest studies from other research fields, such as rehabilitation, where the precision of tests is crucial [49–53]. In this work, the absolute test-retest difference ($|d|$) was reported to provide the reader with a direct and basic measure of measurement errors; however, this index is not sufficient to estimate the repeatability and compare it across studies.

Finally, the assessment of the repeatability needs to be adapted to the study goals. Indeed, the ICC depends on the sample homogeneity. Therefore, if the goal is to differentiate between the microstructure of healthy subjects, including patients in the sample will artificially increase the between-subjects variability and overestimate the ICC. In this study, we can confidently assert that the ICC is lower (and the MDC is higher) than it would have been for a sample that includes patients and controls. Therefore, if the goal is to distinguish between pathological cases, we recommend including the different types of tissue (healthy and pathological tissues, with different stages of the disease) in the cohort. This way, the MDC and ICC would integrate the associated between-subjects variability.

4.5. Data sharing

Due to IRB restrictions, all data used here could not be publicly shared. However, we obtained specific consent for sharing MRI data from four young volunteers. Three of them were part of the tested and retested group. Along with those datasets, we provide the batch scripts used to produce the myelin-sensitive metric maps and to register them to spinal cord template and white matter atlas. Also available is a Microsoft Excel spreadsheet gathering all results of the metric estimations within each region of interest for every scan session and every volunteer of the cohort. The 1st tab of the sheet corresponds to the tested and retested cohort only (n = 16),

and the 2nd tab corresponds to the whole cohort (n = 33). Finally, also shared are the scripts to extract these metrics values, to compute the statistical indices for reliability assessment and to produce the figures presented in this work. All these data and code are available at: https://osf.io/ezmrj/.

5. Conclusion

In this study, we assessed the repeatability and distribution of myelin-sensitive metrics (T_1, MTR, MT_{sat} and MTV) in the spinal cord. T_1 and MTV (1 – *proton density*) showed the best reliability regarding the inter-subject variations, but the measurement error remains too large to detect differences between healthy individuals. T_1, MTR and MTV showed trends consistent with the hypothesis of demyelination with aging, but again the differences were not large enough to be distinguishable from measurement errors, or to be significant.

This study used a range of statistical tools to explore the differences between myelin-sensitive metrics. We show that even though statistically significant differences can be reported using standard statistical tests, an important proportion of these differences can be attributed to measurement error. In particular, the coefficient of variation is a misleading index when comparing metrics with different units, and we recommend using the MDC when comparing individual measurements, and the 95% confidence interval of the test-retest difference when comparing groups. The indexes explored in this study allow for a fair comparison of qMRI metrics across studies, MRI vendors and sites, leading toward standardizing the field of myelin imaging and increasing its clinical relevance.

Supporting information

S1 File. Data processing pipeline. This section describes the data processing steps performed to estimate MTR, MT_{sat}, T_1 and MTV maps and to register those maps to the MNI-Poly-AMU template [55] and WM atlas [56].
(PDF)

Acknowledgments

The authors would like to sincerely thank Robert Brown for the helpful discussions.

Author Contributions

Conceptualization: Simon Lévy, Nikola Stikov.

Data curation: Simon Lévy.

Formal analysis: Simon Lévy.

Funding acquisition: Pierre Rainville, Julien Cohen-Adad.

Investigation: Simon Lévy, Ali Khatibi, Kristina Martinu, Jen-I Chen.

Methodology: Simon Lévy, Aviv Mezer.

Project administration: Simon Lévy, Ali Khatibi, Kristina Martinu, Jen-I Chen, Julien Cohen-Adad.

Resources: Julien Cohen-Adad.

Software: Simon Lévy, Aviv Mezer.

Supervision: Nikola Stikov, Pierre Rainville, Julien Cohen-Adad.

Validation: Simon Lévy, Marie-Claude Guertin, Nikola Stikov.

Visualization: Simon Lévy, Nikola Stikov.

Writing – original draft: Simon Lévy.

Writing – review & editing: Marie-Claude Guertin, Aviv Mezer, Nikola Stikov, Pierre Rainville, Julien Cohen-Adad.

References

1. Bot JCJ, Blezer ELA, Kamphorst W, Nijeholt GJLà, Ader HJ, Castelijns JA, et al. (2004) The Spinal Cord in Multiple Sclerosis: Relationship of High-Spatial-Resolution Quantitative MR Imaging Findings to Histopathologic Results. Radiology 233: 531–540. https://doi.org/10.1148/radiol.2332031572 PMID: 15385682
2. Mottershead JP, Schmierer K, Clemence M, Thornton JS, Scaravilli F, Barker GJ, et al. (2003) High field MRI correlates of myelin content and axonal density in multiple sclerosis. Journal of Neurology 250: 1293–1301. https://doi.org/10.1007/s00415-003-0192-3 PMID: 14648144
3. Schmierer K, Scaravilli F, Altmann DR, Barker GJ, Miller DH (2004) Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. Annals of Neurology 56: 407–415. https://doi.org/10.1002/ana.20202 PMID: 15349868
4. Stüber C, Morawski M, Schäfer A, Labadie C, Wähnert M, Leuze C, et al. (2014) Myelin and iron concentration in the human brain: A quantitative study of MRI contrast. NeuroImage 93, Part 1: 95–106.
5. Fukunaga M, Li T-Q, van Gelderen P, de Zwart JA, Shmueli K, Yao B, et al. (2010) Layer-specific variation of iron content in cerebral cortex as a source of MRI contrast. Proceedings of the National Academy of Sciences 107: 3834–3839.
6. Helms G, Dathe H, Kallenberg K, Dechent P (2008) High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. Magnetic Resonance in Medicine 60: 1396–1407. https://doi.org/10.1002/mrm.21732 PMID: 19025906
7. Stikov N, Keenan KE, Pauly JM, Smith RL, Dougherty RF, Gold GE (2011) Cross-relaxation imaging of human articular cartilage. Magnetic Resonance in Medicine 66: 725–734. https://doi.org/10.1002/mrm.22865 PMID: 21416504
8. Schmierer K, Tozer DJ, Scaravilli F, Altmann DR, Barker GJ, Tofts PS, et al. (2007) Quantitative magnetization transfer imaging in postmortem multiple sclerosis brain. Journal of Magnetic Resonance Imaging 26: 41–51. https://doi.org/10.1002/jmri.20984 PMID: 17659567
9. Norton WT, Autilio LA (1966) The lipid composition of purified bovine brain myelin. Journal of Neurochemistry 13: 213–222. PMID: 5937889
10. Laule C, Vavasour IM, Kolind SH, Li DKB, Traboulsee TL, Moore GRW, et al. (2007) Magnetic Resonance Imaging of Myelin. Neurotherapeutics 4: 460–484. https://doi.org/10.1016/j.nurt.2007.05.004 PMID: 17599712
11. Neeb H, Zilles K, Shah NJ (2006) A new method for fast quantitative mapping of absolute water content in vivo. NeuroImage 31: 1156–1168. https://doi.org/10.1016/j.neuroimage.2005.12.063 PMID: 16650780
12. Whittall KP, Mackay AL, Graeb DA, Nugent RA, Li DKB, Paty DW (1997) In vivo measurement of T2 distributions and water contents in normal human brain. Magnetic Resonance in Medicine 37: 34–43. PMID: 8978630
13. Volz S, Nöth U, Deichmann R (2012) Correction of systematic errors in quantitative proton density mapping. Magnetic Resonance in Medicine 68: 74–85. https://doi.org/10.1002/mrm.23206 PMID: 22144171
14. Volz S, Nöth U, Jurcoane A, Ziemann U, Hattingen E, Deichmann R (2012) Quantitative proton density mapping: correcting the receiver sensitivity bias via pseudo proton densities. NeuroImage 63: 540–552. https://doi.org/10.1016/j.neuroimage.2012.06.076 PMID: 22796988
15. Abbas Z, Gras V, Möllenhoff K, Keil F, Oros-Peusquens A-M, Shah NJ (2014) Analysis of proton-density bias corrections based on T1 measurement for robust quantification of water content in the brain at 3 Tesla. Magnetic Resonance in Medicine 72: 1735–1745. https://doi.org/10.1002/mrm.25086 PMID: 24436248
16. Abbas Z, Gras V, Möllenhoff K, Oros-Peusquens A-M, Shah NJ (2015) Quantitative water content mapping at clinically relevant field strengths: A comparative study at 1.5 T and 3 T. NeuroImage 106: 404–413. https://doi.org/10.1016/j.neuroimage.2014.11.017 PMID: 25463455

17. Olivier N, Mark A, Fergus G, Michael B Sir (2009) Intensity correction with a pair of spoiled gradient recalled echo images. Physics in Medicine and Biology 54: 3473. https://doi.org/10.1088/0031-9155/54/11/013 PMID: 19436101

18. Wehrli FW, BREGER RK, MacFALL JR, DANIELS DL, HAUGHTON VM, CHARLES HC, et al. (1985) Quantification of Contrast in Clinical MR Brain Imaging at High Magnetic Field. Investigative Radiology 20: 360–369. PMID: 4044176

19. Farace P, Pontalti R, Cristoforetti L, Antolini R, Scarpa M (1997) An automated method for mapping human tissue permittivities by MRI in hyperthermia treatment planning. Physics in Medicine and Biology 42: 2159. PMID: 9394404

20. Gutteridge S, Ramanathan C, Bowtell R (2002) Mapping the absolute value of M0 using dipolar field effects. Magnetic Resonance in Medicine 47: 871–879. https://doi.org/10.1002/mrm.10142 PMID: 11979565

21. Ernst T, Kreis R, Ross BD (1993) Absolute Quantitation of Water and Metabolites in the Human Brain. I. Compartments and Water. Journal of Magnetic Resonance, Series B 102: 1–8.

22. Danielsen ER, Henriksen O (1994) Absolute quantitative proton NMR spectroscopy based on the amplitude of the local water suppression pulse. Quantification of brain water and metabolites. NMR in Biomedicine 7: 311–318. PMID: 7718431

23. Helms G (2000) A precise and user-independent quantification technique for regional comparison of single volume proton MR spectroscopy of the human brain. NMR in Biomedicine 13: 398–406. PMID: 11114063

24. Mezer A, Rokem A, Berman S, Hastie T, Wandell BA (2016) Evaluating quantitative proton-density-mapping methods. Human Brain Mapping 37: 3623–3635. https://doi.org/10.1002/hbm.23264 PMID: 27273015

25. Mezer A, Yeatman JD, Stikov N, Kay KN, Cho N-J, Dougherty RF, et al. (2013) Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. Nat Med 19: 1667–1672. https://doi.org/10.1038/nm.3390 PMID: 24185694

26. Mackay A, Whittall K, Adler J, Li D, Paty D, Graeb D (1994) In vivo visualization of myelin water in brain by magnetic resonance. Magnetic Resonance in Medicine 31: 673–677. PMID: 8057820

27. Laule C, Kozlowski P, Leung E, Li DKB, MacKay AL, Moore GRW (2008) Myelin water imaging of multiple sclerosis at 7 T: Correlations with histopathology. NeuroImage 40: 1575–1580. https://doi.org/10.1016/j.neuroimage.2007.12.008 PMID: 18321730

28. Vavasour IM, Laule C, Li DKB, Oger J, Moore GRW, Traboulsee A, et al. (2009) Longitudinal changes in myelin water fraction in two MS patients with active disease. Journal of the Neurological Sciences 276: 49–53. https://doi.org/10.1016/j.jns.2008.08.022 PMID: 18822435

29. Feinberg DA, Oshio K (1991) GRASE (gradient-and spin-echo) MR imaging: a new fast clinical imaging technique. Radiology 181: 597–602. https://doi.org/10.1148/radiology.181.2.1924811 PMID: 1924811

30. Prasloski T, Rauscher A, MacKay AL, Hodgson M, Vavasour IM, Laule C, et al. (2012) Rapid whole cerebrum myelin water imaging using a 3D GRASE sequence. NeuroImage 63: 533–539. https://doi.org/10.1016/j.neuroimage.2012.06.064 PMID: 22776448

31. Emil Ljungberg IV, Roger Tam, Youngjin Yoo, Alexander Rauscher, David Li, Anthony Traboulsee, Alex MacKay, Shannon Kolind. Rapid Myelin Water Imaging in Human Cervical Spinal Cord; 2016 Tuesday, May 10, 2016 Singapore, Singapore.

32. Pitt D, Boster A, Pei W, et al. (2010) IMaging cortical lesions in multiple sclerosis with ultra–high-field magnetic resonance imaging. Archives of Neurology 67: 812–818. https://doi.org/10.1001/archneurol.2010.148 PMID: 20625086

33. Mainero C, Louapre C, Govindarajan ST, Giannì C, Nielsen AS, Cohen-Adad J, et al. (2015) A gradient in cortical pathology in multiple sclerosis by in vivo quantitative 7 T imaging. Brain 138: 932–945. https://doi.org/10.1093/brain/awv011 PMID: 25681411

34. Cohen-Adad J, Benner T, Greve D, Kinkel RP, Radding A, Fischl B, et al. (2011) In vivo evidence of disseminated subpial T2* signal changes in multiple sclerosis at 7 T: A surface-based analysis. NeuroImage 57: 55–62. https://doi.org/10.1016/j.neuroimage.2011.04.009 PMID: 21511042

35. Lee J, Shmueli K, Kang B-T, Yao B, Fukunaga M, van Gelderen P, et al. (2012) The contribution of myelin to magnetic susceptibility-weighted contrasts in high-field MRI of the brain. NeuroImage 59: 3967–3975. https://doi.org/10.1016/j.neuroimage.2011.10.076 PMID: 22056461

36. Cohen-Adad J, Polimeni JR, Helmer KG, Benner T, McNab JA, Wald LL, et al. (2012) T2* mapping and B0 orientation-dependence at 7 T reveal cyto- and myeloarchitecture organization of the human cortex. NeuroImage 60: 1006–1014. https://doi.org/10.1016/j.neuroimage.2012.01.053 PMID: 22270354

37. Spees WM, Yablonskiy DA, Oswood MC, Ackerman JJH (2001) Water proton MR properties of human blood at 1.5 Tesla: Magnetic susceptibility, T1, T2, T *2, and non-Lorentzian signal behavior. Magnetic Resonance in Medicine 45: 533–542. PMID: 11283978

38. Li D, Wang Y, Waight DJ (1998) Blood oxygen saturation assessment in vivo using T2 * estimation. Magnetic Resonance in Medicine 39: 685–690. PMID: 9581597

39. Alsop D, de Bazelaire C, Garcia D, Duhamel G. Inhomogenous magnetization transfer imaging: a potentially specific marker for myelin; 2005; Miami, Florida, USA. pp. 2224.

40. Alsop D, Dandamudi R, Bakshi R. Inhomogeneous magnetization transfer imaging of myelin concentration in multiple sclerosis; 2007. pp. 2188.

41. Duhamel GLT, A; Prevost, V; Varma, G; Guye, M; Ranjeva, JP; Pelletier, J; Alsop, DC; Girard, OM. Magnetization transfer from inhomogeneously broadened lines (ihMT): application on multiple sclerosis; 2015 2015, June 3rd; Toronto, ON, Canada. pp. 4346.

42. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, et al. (2015) The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Statistical Methods in Medical Research 24: 9–26. https://doi.org/10.1177/0962280214537333 PMID: 24919826

43. Stroman PW, Wheeler-Kingshott C, Bacon M, Schwab JM, Bosma R, Brooks J, et al. (2014) The current state-of-the-art of spinal cord imaging: Methods. NeuroImage 84: 1070–1081. https://doi.org/10.1016/j.neuroimage.2013.04.124 PMID: 23685159

44. Taso M, Girard OM, Duhamel G, Le Troter A, Feiweier T, Guye M, et al. (2016) Tract-specific and age-related variations of the spinal cord microstructure: a multi-parametric MRI study using diffusion tensor imaging (DTI) and inhomogeneous magnetization transfer (ihMT). NMR in Biomedicine 29: 817–832. https://doi.org/10.1002/nbm.3530 PMID: 27100385

45. Berry I, Barker GJ, Barkhof F, Campi A, Dousset V, Franconi J-M, et al. (1999) A multicenter measurement of magnetization transfer ratio in normal white matter. Journal of Magnetic Resonance Imaging 9: 441–446. PMID: 10194715

46. Smith SA, Jones CK, Gifford A, Belegu V, Chodkowski B, Farrell JAD, et al. (2010) Reproducibility of tract-specific magnetization transfer and diffusion tensor imaging in the cervical spinal cord at 3 tesla. NMR in Biomedicine 23: 207–217. https://doi.org/10.1002/nbm.1447 PMID: 19924726

47. Grussu F, Schneider T, Zhang H, Alexander DC, Wheeler–Kingshott CAM (2015) Neurite orientation dispersion and density imaging of the healthy cervical spinal cord in vivo. NeuroImage 111: 590–601. https://doi.org/10.1016/j.neuroimage.2015.01.045 PMID: 25652391

48. Smith AK, Dortch RD, Dethrage LM, Smith SA (2014) Rapid, high-resolution quantitative magnetization transfer MRI of the human spinal cord. NeuroImage 95: 106–116. https://doi.org/10.1016/j.neuroimage.2014.03.005 PMID: 24632465

49. Carter R, Lubinsky J (2015) Rehabilitation research: principles and applications: Elsevier Health Sciences.

50. Lexell JE, Downham DY (2005) How to Assess the Reliability of Measurements in Rehabilitation. American Journal of Physical Medicine & Rehabilitation 84: 719–723.

51. Bashardoust Tajali S, MacDermid JC, Grewal R, Young C (2016) Reliability and Validity of Electro-Goniometric Range of Motion Measurements in Patients with Hand and Wrist Limitations. The Open Orthopaedics Journal 10: 190–205. https://doi.org/10.2174/1874325001610010190 PMID: 27398107

52. James S, Ziviani J, Ware RS, Boyd RN (2016) Test–retest Reproducibility of the Assessment of Motor and Process Skills in Children with Unilateral Cerebral Palsy. Physical & Occupational Therapy In Pediatrics 36: 144–154.

53. Sakzewski L, Lewis M, Ziviani J (2016) Test–retest reproducibility of the Assessment of Motor and Process Skills for school-aged children with acquired brain injuries. Scandinavian Journal of Occupational Therapy: 1–6.

54. De Leener B, Lévy S, Dupont SM, Fonov VS, Stikov N, Louis Collins D, et al. (2016) SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. NeuroImage.

55. Fonov VS, Le Troter A, Taso M, De Leener B, Lévêque G, Benhamou M, et al. (2014) Framework for integrated MRI average of the spinal cord white and gray matter: The MNI–Poly–AMU template. NeuroImage 102, Part 2: 817–827.

56. 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 PMID: 26099457

57. Bland JM, Altman DG (1986) Originally published as Volume 1, Issue 8476STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet 327: 307–310.

58. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychological bulletin 86: 420. PMID: 18839484

59. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. Psychological methods 1: 30.

60. Fleiss J (1986) Book Reviews. Journal of Applied Statistics 13: 231–231.

61. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 6: 284–290.

62. Chinn S (1991) Statistics in respiratory medicine. 2. Repeatability and method comparison. Thorax 46: 454–456. PMID: 1858087

63. Stratford PW (2004) Getting more from the literature: estimating the standard error of measurement from reliability studies. Physiotherapy Canada 56: 27–30.

64. Nakanishi R, Goto J, Ezure H, Motoura H, Ayabe S-i, Atsumi T (2004) Morphometric Analyses of Axons in the Human Lateral Corticospinal Tract: Cervical/Lumbar Level Comparison and Relation to the Ageing Process. Okajimas Folia Anatomica Japonica 81: 1–4. PMID: 15248559

65. Ohnishi A, O'Brien PC, Okazaki H, Dyck PJ (1976) Morphometry of myelinated fibers of fasciculus gracilis of man. Journal of the Neurological Sciences 27: 163–172. PMID: 1249584

66. Terao S-i, Sobue G, Hashizume Y, Shimada N, Mitsuma T (1994) Age-related changes of the myelinated fibers in the human corticospinal tract: a quantitative analysis. Acta Neuropathologica 88: 137–142. PMID: 7985494

67. Nijeholt GJLà, Bergers E, Kamphorst W, Bot J, Nicolay K, Castelijns JA, et al. (2001) Post-mortem high-resolution MRI of the spinal cord in multiple sclerosisA correlative study with conventional MRI, histopathology and clinical phenotype. Brain 124: 154–166. PMID: 11133795

68. Stikov N, Boudreau M, Levesque IR, Tardif CL, Barral JK, Pike GB (2015) On the accuracy of T1 mapping: Searching for common ground. Magnetic Resonance in Medicine 73: 514–522. https://doi.org/10.1002/mrm.25135 PMID: 24578189

69. Duval T, Lévy S, Stikov N, Campbell J, Mezer A, Witzel T, et al. (2017) g-Ratio weighted imaging of the human spinal cord in vivo. NeuroImage 145, Part A: 11–23.

70. Smith SA, Edden RAE, Farrell JAD, Barker PB, Van Zijl PCM (2008) Measurement of T1 and T2 in the cervical spinal cord at 3 tesla. Magnetic Resonance in Medicine 60: 213–219. https://doi.org/10.1002/mrm.21596 PMID: 18581383

71. Duval T, Lévy S, Stikov N, Campbell J, Mezer A, Witzel T, et al. g-Ratio weighted imaging of the human spinal cord in vivo. NeuroImage.

72. Samson RS, Ciccarelli O, Kachramanoglou C, Brightman L, Lutti A, Thomas DL, et al. (2013) Tissue- and column-specific measurements from multi-parameter mapping of the human cervical spinal cord at 3 T. NMR in Biomedicine 26: 1823–1830. https://doi.org/10.1002/nbm.3022 PMID: 24105923

73. Yiannakas MC, Kearney H, Samson RS, Chard DT, Ciccarelli O, Miller DH, et al. (2012) Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: A pilot study with application to magnetisation transfer measurements. NeuroImage 63: 1054–1059. https://doi.org/10.1016/j.neuroimage.2012.07.048 PMID: 22850571

74. Hickman SJ, Hadjiprocopis A, Coulon O, Miller DH, Barker GJ (2004) Cervical spinal cord MTR histogram analysis in multiple sclerosis using a 3D acquisition and a B-spline active surface segmentation technique. Magnetic Resonance Imaging 22: 891–895. https://doi.org/10.1016/j.mri.2004.01.056 PMID: 15234459

75. Rovaris M, Judica E, Ceccarelli A, Ghezzi A, Martinelli V, Comi G, et al. (2008) Absence of diffuse cervical cord tissue damage in early, non-disabling relapsing-remitting MS: a preliminary study. Multiple Sclerosis Journal 14: 853–856. https://doi.org/10.1177/1352458507088103 PMID: 18611991

76. Ropele S, Filippi M, Valsasina P, Korteweg T, Barkhof F, Tofts PS, et al. (2005) Assessment and correction of B1-induced errors in magnetization transfer ratio measurements. Magnetic Resonance in Medicine 53: 134–140. https://doi.org/10.1002/mrm.20310 PMID: 15690512

77. Tang Y, Nyengaard JR, Pakkenberg B, Gundersen HJG (1997) Age-Induced White Matter Changes in the Human Brain: A Stereological Investigation. Neurobiology of Aging 18: 609–615. PMID: 9461058

78. Ge Y, Grossman RI, Babb JS, Rabin ML, Mannon LJ, Kolson DL (2002) Age-Related Total Gray Matter and White Matter Changes in Normal Adult Brain. Part II: Quantitative Magnetization Transfer Ratio Histogram Analysis. American Journal of Neuroradiology 23: 1334–1341. PMID: 12223374

79. Barrick TR, Charlton RA, Clark CA, Markus HS (2010) White matter structural decline in normal ageing: A prospective longitudinal study using tract-based spatial statistics. NeuroImage 51: 565–577. https://doi.org/10.1016/j.neuroimage.2010.02.033 PMID: 20178850

80. Likitjaroen Y, Meindl T, Friese U, Wagner M, Buerger K, Hampel H, et al. (2012) Longitudinal changes of fractional anisotropy in Alzheimer's disease patients treated with galantamine: a 12-month randomized, placebo-controlled, double-blinded study. European Archives of Psychiatry and Clinical Neuroscience 262: 341–350. https://doi.org/10.1007/s00406-011-0234-2 PMID: 21818628

81. Teipel SJ, Meindl T, Wagner M, Stieltjes B, Reuter S, Hauenstein K-H, et al. (2009) Longitudinal changes in fiber tract integrity in healthy aging and mild cognitive impairment: a DTI follow-up study. Journal of Alzheimer's disease: JAD 22: 507–522.

82. Kochunov P, Thompson PM, Lancaster JL, Bartzokis G, Smith S, Coyle T, et al. (2007) Relationship between white matter fractional anisotropy and other indices of cerebral health in normal aging: Tract-based spatial statistics study of aging. NeuroImage 35: 478–487. https://doi.org/10.1016/j.neuroimage.2006.12.021 PMID: 17292629

83. Wang K, Song Q, Zhang F, Chen Z, Hou C, Tang Y, et al. (2014) Age-related changes of the diffusion tensor imaging parameters of the normal cervical spinal cord. European Journal of Radiology 83: 2196–2202. https://doi.org/10.1016/j.ejrad.2014.09.010 PMID: 25287960

84. Chan T-Y, Li X, Mak K-C, Cheung J-y, Luk K-K, Hu Y (2015) Normal values of cervical spinal cord diffusion tensor in young and middle-aged healthy Chinese. European Spine Journal 24: 2991–2998. https://doi.org/10.1007/s00586-015-4144-2 PMID: 26208941

85. Agosta F, Laganà M, Valsasina P, Sala S, Dall'Occhio L, Sormani MP, et al. (2007) Evidence for cervical cord tissue disorganisation with aging by diffusion tensor MRI. NeuroImage 36: 728–735. https://doi.org/10.1016/j.neuroimage.2007.03.048 PMID: 17490894

86. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC (2012) NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage 61: 1000–1016. https://doi.org/10.1016/j.neuroimage.2012.03.072 PMID: 22484410

87. Assaf Y, Blumenfeld-Katzir T, Yovel Y, Basser PJ (2008) Axcaliber: A method for measuring axon diameter distribution from diffusion MRI. Magnetic Resonance in Medicine 59: 1347–1354. https://doi.org/10.1002/mrm.21577 PMID: 18506799

88. Assaf Y, Basser PJ (2005) Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. NeuroImage 27: 48–58. https://doi.org/10.1016/j.neuroimage.2005.03.042 PMID: 15979342

89. Duval T, McNab JA, Setsompop K, Witzel T, Schneider T, Huang SY, et al. (2015) In vivo mapping of human spinal cord microstructure at 300mT/m. NeuroImage 118: 494–507. https://doi.org/10.1016/j.neuroimage.2015.06.038 PMID: 26095093

90. Mangeat G, Govindarajan ST, Mainero C, Cohen-Adad J (2015) Multivariate combination of magnetization transfer, T2* and B0 orientation to study the myelo-architecture of the in vivo human cortex. NeuroImage 119: 89–102. https://doi.org/10.1016/j.neuroimage.2015.06.033 PMID: 26095090

91. Lévy S, Khatibi A, Mangeat G, Chen J-I, Martinu K, Rainville P, et al. Statistical combinations of T1, MTR, MTsat and Macromolecular Tissue Volume to improve myelin content estimation in the human spinal cord at 3T; 2017 April 26, 2017; Honolulu, USA.