# Towards quantitative intensity analysis of conventional T1-weighted images in multiple sclerosis

Tun Wiltgen [a,b,*] , Julian McGinnis [a,b,c] , Ronja Berg [a,b,d] , Cui Ci Voon [a,b] , Oula Puonti [e,f] , Katrin Giglhuber [a] , Carl Ganter [g] , Claus Zimmer [d] , Bernhard Hemmer [a,h] , Benedikt Wiestler [i,j] , Jan Kirschke [d] , Christine Preibisch [a,b,d] , Mark Mühlau [a,b]

[a] School of Medicine and Health, Department of Neurology, Technical University of Munich, Munich, Germany
[b] School of Medicine and Health, TUM-Neuroimaging Center, Technical University of Munich, Munich, Germany
[c] Chair for AI in Healthcare and Medicine, TUM University Hospital, Technical University of Munich, Munich, Germany
[d] School of Medicine and Health, Institute for Diagnostic and Interventional Neuroradiology, Technical University of Munich, Munich, Germany
[e] Danish Research Centre for Magnetic Resonance, Center for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Denmark
[f] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, USA
[g] School of Medicine and Health, Department of Diagnostic and Interventional Radiology, Technical University of Munich, Munich, Germany
[h] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
[i] School of Medicine and Health, AI for Image-Guided Diagnosis and Therapy, Technical University of Munich, Munich, Germany
[j] Munich Center for Machine Learning (MCML), Munich, Germany

## ARTICLE INFO

## ABSTRACT

Conventional T1-weighted (T1w) magnetic resonance imaging (MRI) is commonly used in multiple sclerosis (MS) morphometry and volumetry research. However, arbitrary intensity scales preclude interpretation of signal values across patients, sites, and time. This requires quantitative MRI techniques, which are not always available.

This study assessed T1w image intensity scaling methods, relying on extracerebral reference regions, for quantitative analysis of brain MRI in MS.

In total, 701 people with a diagnosis of radiologically isolated syndrome, clinically isolated syndrome, or MS were included. Four intensity scaling strategies were applied: 1) MRI signal modeling, 2) linear scaling with reference regions, 3) z-score standardization, and 4) none (only bias field correction). Methods were evaluated using variance analysis, R1 map comparison, and normal-appearing white matter (NAWM) intensity group comparison, using mean and coefficient of variation (CoV), between low ($\leq$3) and high ($>$3) expanded disability status scale (EDSS) scores. Statistical analysis was conducted using Pearson's r, two-sided Welch two-sample t-test, ANCOVA, and Cohen's d.

Linear scaling with temporal fatty tissue achieved the most consistent variance reduction and strong correlation with R1 maps ($r = 0.84$). R1 values in NAWM were significantly lower in people with high compared to low EDSS scores ($d = -0.351$). Similarly, group differences in mean NAWM intensity of fat-scaled images were significant ($d = -0.252$). The largest group differences were found in NAWM CoV in bias field-corrected T1w images ($d = 0.818$).

Linear scaling with fatty tissue most accurately reproduced the results obtained with R1 maps. Changes in MS NAWM appear to increase intensity variability detectable in conventional T1w images.

## 1. Introduction

In clinical practice, magnetic resonance imaging (MRI) is one of the main tools to diagnose and monitor neurological diseases such as multiple sclerosis (MS) (Thompson et al., 2018). Conventional MRI is mainly used to detect abnormalities such as demyelinating white matter (WM) lesions that appear hyperintense in T2-weighted (T2w) images and vary between hypo- and isointense in T1-weighted (T1w) images.

However, disease-related tissue damage (not only in MS) often extends beyond what can be detected by the human eye on conventional MRI. In MS, histological studies substantiated pathological processes in WM regions that appear normal in conventional MRI (Kutzelnigg et al., 2005; Laule et al., 2013; Moll et al., 2011). Analyzing normal-appearing white matter (NAWM) with respect to tissue damage requires image intensities on an absolute quantitative scale, making them interpretable and comparable between images and subjects. However, conventional T1w MRI is optimized for good visual contrast, with arbitrary image intensity scales influenced by multiple factors, such as hardware-specific settings (e.g., coils, manufacturer, signal amplifier gain), scanner software, and patient anatomy and positioning. Therefore, quantitative MRI sequences have been developed using mathematical models and calibration techniques. Numerous studies have shown that advanced quantitative MRI, such as R1 relaxometry, is sensitive to tissue damage in MS NAWM (Ladopoulos et al., 2022; Lommers et al., 2019; MacKay et al., 2009; Manfredonia et al., 2007; O'Muircheartaigh et al., 2019; Pontillo et al., 2023; Vrenken et al., 2006). Yet, acquiring quantitative MRI is not widely integrated into clinical routines because acquisition is time-consuming and off-line evaluation is frequently required. Given these limitations and that T1w data is widely available, recent studies have attempted to derive quantitative images from conventional T1w MRI using various intensity scaling approaches (Brown et al., 2020; Fortin et al., 2016; Ganzetti et al., 2014; Ghassemi et al., 2015; Hasse et al., 2022; Lavielle et al., 2024, 2023; Lee et al., 2015; Mejia et al., 2016; Moskovich et al., 2024; Shinohara et al., 2014; Wahid et al., 2021). Ideally, intensity scaling removes technically induced inter-image variance while preserving the biologically induced variance that contains information on the disease under investigation. Some approaches are based on reference regions that ideally should only reflect technical variance, which could then be modeled and compensated for independently of the biological variance. Most frequently, intensities or scaling factors are derived from reference regions within the brain, which can be critical in studies focusing on brain diseases. In the case of MS, regions outside of the brain are less likely to be directly affected by the disease and are, hence, attractive candidates for reference regions. In our recently published systematic review (Wiltgen et al., 2024b), we discussed studies relying on intensity scaling with extracerebral reference regions such as various fatty tissues, temporal muscle, and eyes (Brown et al., 2020; Ganzetti et al., 2014; Ghassemi et al., 2015; Gilmore et al., 2007; Lee et al., 2015) and found that these have shown promising results.

However, validation studies comparing different approaches or relating images with scaled intensities to quantitative MRI are sparse. Some studies aimed to validate intensity scaling methods through comparison with (semi-) quantitative MRI, such as the magnetization transfer ratio or diffusion-weighted imaging (Brown et al., 2020; Ganzetti et al., 2014; Gilmore et al., 2007). Other studies modeled the T1-dependency of T1w MRI signals to estimate T1 properties and validated them against actual T1 or R1 maps (Hasse et al., 2022; Lavielle et al., 2024, 2023; Mejia et al., 2016).

In this study, we aimed to compare and validate several approaches for intensity scaling of conventional T1w images: 1) MRI signal modeling, 2) linear scaling with extracerebral reference regions, 3) z-score standardization, and 4) none (only bias field correction). Technical validation assessed the reduction of inter-image variance and the correlation between scaled T1w images and R1 maps. Biological validation assessed each method's ability to detect differences in mean NAWM intensity between people with MS with low (≤3) and high scores (>3) on the expanded disability status scale (EDSS). Finally, we contrasted the group comparisons of mean intensities, a robust and global measure, with the intensity variability across NAWM using the coefficient of variation (CoV). This approach was inspired by the observation of Johnson et al. (Johnson et al., 2023) that the CoV of myelin water fraction (MWF) in NAWM better distinguishes people with MS from healthy controls than mean values. Building on this, and since the CoV is independent of the absolute intensity scale, we hypothesized that greater intensity variability in NAWM of conventional T1w images may already reflect a decreased tissue integrity.

## 2. Methods

This study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans (World Medical Association, 2001). Patients gave written informed consent to use their data for research purposes. An overview of the study workflow is presented in Fig. 1.

### 2.1. Data sets

In the total cohort, 1078 image data sets of 701 subjects with confirmed radiologically isolated syndrome (RIS), clinically isolated syndrome (CIS), or MS diagnosis according to the McDonald criteria (Thompson et al., 2018) from our in-house database were included. Further inclusion criteria were MRI acquisition on the same scanner and 18 years or older. The total cohort was used to identify valid reference regions.

A subset of 597 image data sets of 597 subjects (validation cohort) was used for the technical and biological validation of intensity scaling methods. It only included people with CIS or MS whose EDSS scores were acquired within 31 days before or after the MRI scan. When multiple image sessions were available for one subject, we selected the most recent data to increase sensitivity to disease-related effects. In the context of biological validation, the validation cohort was divided into two groups: low (≤3) and high (>3) EDSS (Leray et al., 2010). Details of the cohorts are presented in Table 1.

### 2.2. MRI acquisition

All subjects underwent conventional MRI and multiparameter mapping (MPM) (Berg et al., 2022; Leutritz et al., 2020; Tabelow et al., 2019), acquired on a Philips 3.0 T Ingenia scanner (Philips Medical Systems, Best, The Netherlands) within a time window of 21 months, during which no hard- or software upgrades were applied. Conventional imaging included a 3D T1w magnetization-prepared rapid gradient echo (MPRAGE) sequence and a 3D T2w fluid-attenuated inversion recovery (FLAIR) sequence. MPM data were acquired with a 3D radiofrequency-spoiled multi-echo gradient-echo sequence (fast field echo) with six echoes, using a Compressed SENSE acceleration factor of 6 (Berg et al., 2022; Geerts-Ossevoort et al., 2018). Within each MPM acquisition, four protocols were included: 1) B1 mapping for voxel-wise bias field correction, 2) T1w, 3) proton density-weighted (PDw), and 4) magnetization transfer-weighted (MTw). Further technical details have been described in a previous study (Berg et al., 2022). MRI parameters are presented in Table 2.

### 2.3. Image processing

First, N4 bias field correction (Tustison et al., 2010), implemented in the Advanced Normalization Tools (ANTs) software (https://github.com/ANTsX/ANTs), was applied to all T1w images. Sequence Adaptive Multimodal SEGmentation (SAMSEG) software integrated into Freesurfer version 7.3.2 (Cerri et al., 2021; Fischl, 2012; Puonti et al., 2016) was used to segment WM, cortical gray matter (GM), cerebellum GM and WM, accumbens, amygdala, caudate, hippocampus, putamen, and thalamus. The inputs were pairs of registered T1w and FLAIR images, co-registered using "mri_coreg" and "mri_vol2vol" integrated into Freesurfer. NAWM masks were created by first applying WM lesion segmentation with LST-AI (Wiltgen et al., 2024a) using non-registered T1w and FLAIR images as input, as registration is integrated into LST-AI. Next, the lesion mask was subtracted from the SAMSEG WM segmentation. The mask was further eroded using a $3 \times 3 \times 3$ cube as a
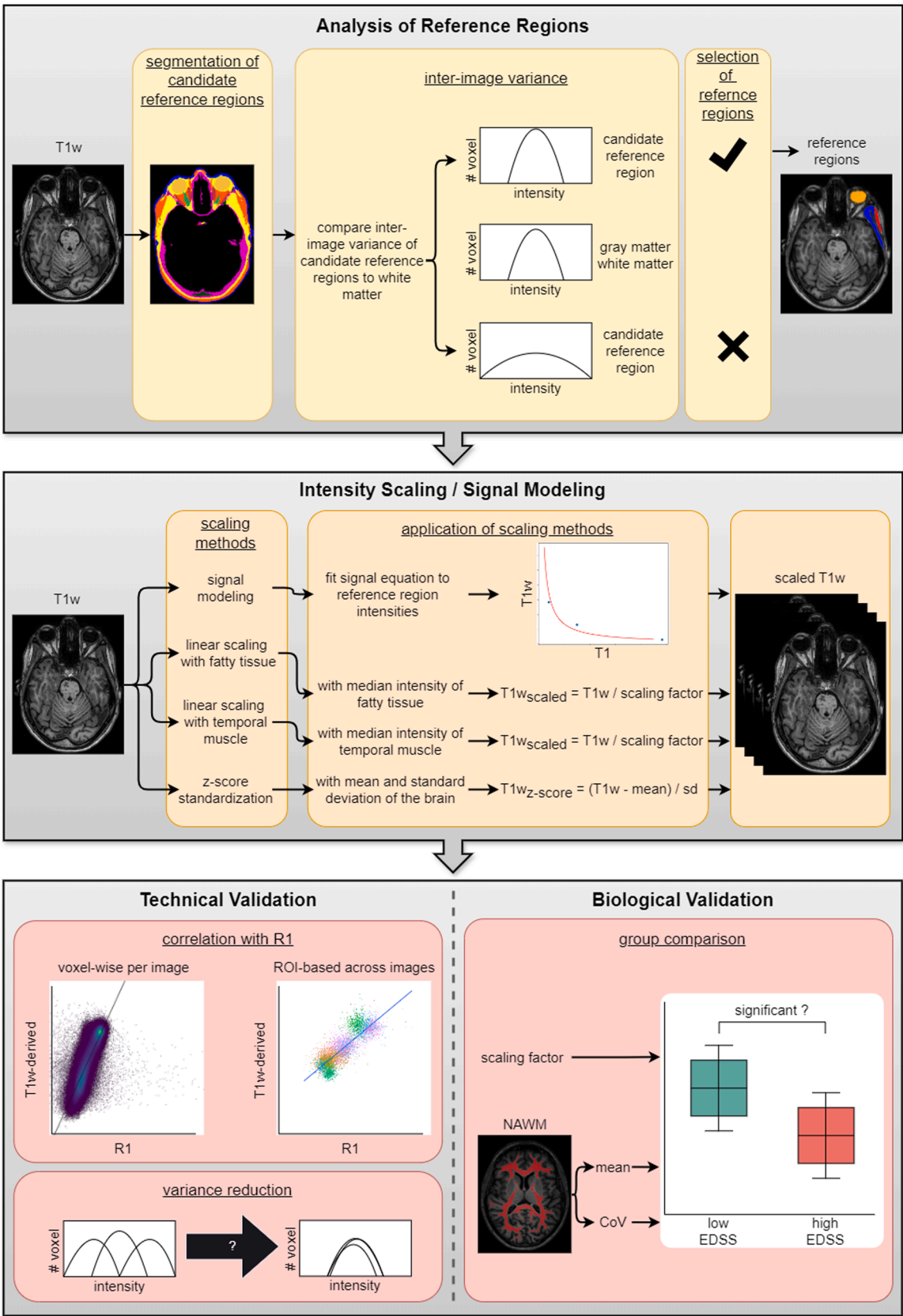
*(caption on next page)*

**Fig. 1.** Workflow overview.

Overview of the workflow. First (top panel), potential extracerebral reference regions were selected with a data-driven approach. Multiple candidate reference regions were segmented, and inter-image variance was quantified with the CoV and compared with GM and WM. Only regions with equal or lower CoVs were selected. Second (middle panel), different scaling methods were applied to bias-corrected T1w images. These methods included MRI signal modeling, linear scaling with a reference region, and z-score standardization. Third (bottom panel), scaling methods were evaluated through technical and biological validation. For technical validation (left), scaled T1w images were compared to quantitative R1 maps, and the inter-image variance was analyzed. Biological validation (right) comprised group comparisons of NAWM intensities (with mean and CoV) between subjects with low (≤3) and high (>3) EDSS scores. In addition, scaling factors were tested for group differences to confirm that scaling methods are not influenced by disease severity. Abbreviations: CoV: coefficient of variation, EDSS: expanded disability status scale, GM: gray matter, MRI: magnetic resonance imaging, NAWM: normal-appearing white matter, ROI: region of interest, sd: standard deviation, T1w: T1-weighted, WM: white matter.

structuring element, resulting in a conservative segmentation of NAWM. In addition, whole-brain binary masks were generated with the HD-BET brain extraction tool (Isensee et al., 2019) using T1w images as input.

Candidate reference regions were segmented using SAMSEG combined with an atlas comprising extracerebral tissues, similar to the approach by Puonti et al. (Puonti et al., 2020). Details on the atlas generation are provided in the supplementary material. The extracerebral tissues included cancellous bone, cortical bone, eyes, optic nerves, rectus muscles, and skin. In addition, temporal fatty tissue and temporal muscle were segmented as they are distinctly visible in conventional T1w MRI, and previous studies have also used fatty tissue and temporal muscle for scaling (Brown et al., 2020; Ganzetti et al., 2014; Ghassemi et al., 2015; Gilmore et al., 2007; Lee et al., 2015). For precise and efficient segmentation of fatty tissue and temporal muscle, we trained

two 3D nnUNets (Isensee et al., 2021) with a training data set of 166 manually segmented masks (see supplementary material), conducted with 3D Slicer version 5.6.2 (Fedorov et al., 2012, https://www.slicer.org/). Masks were generated for the total cohort through inference using the trained models and manually corrected if necessary. In general, all images and segmentation masks underwent a thorough visual quality check.

Longitudinal relaxation rate (R1 = 1/T1) maps were computed with the hMRI (Tabelow et al., 2019) toolbox (version: v0.1.3-dev; RRID: SCR_017682; https://github.com/tleutritz-cbs/hMRI-toolbox) implemented in the Statistical Parametric Mapping (SPM12) framework (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), using the same workflow as in (Berg et al., 2022). The quantitative R1 maps were rigidly registered to the Montreal Neurological Institute (MNI) space using the

**Table 1**
Cohort characteristics.

| | | Number of subjects | Number of imaging data sets | Female/ Male | Diagnosis | Age in years mean ± sd | Lesion volume in mm3 median (IQR) | EDSS median (IQR) | Disease duration |
|---|---|---|---|---|---|---|---|---|---|
| **Total cohort** | | 701 | 1078 | 682 / 396 | CIS: 39 RRMS: 960 SPMS: 22 PPMS: 40 RIS: 17 | 39.90 ± 11.03 | 2543.7 (859.8 - 7098.78) | 1.5 (0.0 - 2.0) | 6.7 (2.5 - 11.5) |
| **Validation cohort** | **Total** | 597 | 597 | 376 / 221 | CIS: 22 RRMS: 541 SPMS: 14 PPMS: 20 | 40.48 ± 11.12 | 2667.1 (866.1 - 7121.7) | 1.5 (0.0 - 2.0) | 7.1 (3.1 - 11.5) |
| | **Low EDSS** | 526 | 526 | 330 / 196 | CIS: 22 RRMS: 499 SPMS: 0 PPMS: 5 | 39.47 ± 10.61 | 2309.3 (781.2 - 5995.1) | 1.0 (0.0 - 2.0) | 6.7 (2.9 - 11.0) |
| | **High EDSS** | 71 | 71 | 46 / 25 | CIS: 0 RRMS: 42 SPMS: 14 PPMS: 15 | 47.94 ± 12.04 | 7752.8 (3499.9 - 18,147.6) | 4.5 (3.5 - 6.0) | 11.8 (5.3 - 21.7) |

Abbreviations: CIS: clinically isolated syndrome, EDSS: expanded disability status scale, IQR: interquartile range, MS: multiple sclerosis, PPMS: primary progressive multiple sclerosis, RRMS: relapsing-remitting multiple sclerosis, RIS: radiologically isolated syndrome, sd: standard deviation, SPMS: secondary progressive multiple sclerosis.

**Table 2**
MRI acquisition parameters.

| | 3D T1w (MPRAGE) | 3D FLAIR | Multiparameter mapping | | | |
|---|---|---|---|---|---|---|
| | | | T1w | PDw | MTw | B1 map |
| Reconstruction voxel size | 0.75×0.75×0.75 mm$^3$ | 0.75×0.75×0.75 mm$^3$ | 1 × 1 × 1 mm$^3$ | 1 × 1 × 1 mm$^3$ | 1 × 1 × 1 mm$^3$ | 1 × 1 × 2.5 mm$^3$ |
| Acquisition voxel size | 1 × 1 × 1 mm$^3$ | 1 × 1 × 1 mm$^3$ | 1 × 1 × 1 mm$^3$ | 1 × 1 × 1 mm$^3$ | 1 × 1 × 1 mm$^3$ | 3.5 × 3.5 × 5 mm$^3$ |
| TR | 9 ms | 4800 ms | 18 ms | 18 ms | 48 ms | 30/150 ms |
| TE | 4 ms | shortest | 6 echoes first: 2.4 ms echo spacing: 2.4 ms | 6 echoes first: 2.4 ms echo spacing: 2.4 ms | 6 echoes first: 2.4 ms echo spacing: 2.4 ms | shortest |
| TI | 1000ms | 1650 ms | / | / | / | / |
| Flip angle | 8° | 90° | 25° | 4° | 6° | 60° |
| Acquisition mode | turbo field echo | turbo spin echo | fast field echo | fast field echo | fast field echo | fast field echo |

Abbreviations: FLAIR: fluid-attenuated inversion recovery, MPRAGE: magnetization-prepared rapid gradient echo, MTw: magnetization transfer-weighted, PDw: proton density-weighted, T1w: T1-weighted, TE: echo time, TI: inversion time, TR: repetition time.

"Auto-Reorient" application included in the hMRI toolbox.

T1w images were rigidly registered to R1 maps in MNI space, using the Greedy command line tool (Yushkevich et al., 2016, 2006; https://github.com/pyushkevich/greedy), allowing for comparison of intensity-scaled T1w images with R1 maps. The resulting transformation matrices were used to register the intensity-scaled T1w images and segmentations to the respective R1 maps.

### 2.4. Scaling of conventional MRI

### 2.4.1. MRI signal modeling

This method builds on a physical model of MRI signal generation, similar to approaches in previous studies (Hasse et al., 2022; Lavielle et al., 2023; Moskovich et al., 2024). The signal equation relates the acquired MRI signal to physical properties, such as T1 relaxation times (see detailed derivation and sequence diagram in supplementary material). For an MPRAGE sequence with echo time TE, flip angle $\alpha$, repetition time TR, recovery time $T_{rec}$, and inversion time TI, the signal after the i$^{th}$ inversion pulse – obtained from a sample with longitudinal and effective transverse relaxation time T1 and T2*, respectively – is given by (Brant-Zawadzki et al., 1992; Deichmann et al., 2000; Deichmann and Haase, 1992; Haase, 1990; Vernier et al., 2021):

$$S_i = \sin(\alpha)*M_i^N*\exp\left(-TE / T2^*\right) \tag{1}$$

with $M_i^N$ being the longitudinal magnetization after the N$^{th}$ radio frequency pulse in the rapid gradient echo block:

$$M_i^N = M_0\left(1 - e^{-\frac{TR}{T1}}\right)*\frac{1 - \left(\cos(\alpha)*e^{-\frac{TR}{T1}}\right)^N}{1 - \cos(\alpha)*e^{-\frac{TR}{T1}}}$$
$$+ \left(M_0\left(1 - e^{-\frac{TI}{T1}}\right) - M^{eq}*e^{-\frac{TI}{T1}}\right)*\left(\cos(\alpha)*e^{-\frac{TR}{T1}}\right)^N \tag{2}$$

with the steady state equilibrium magnetization $M^{eq}$:

$$M^{eq} = M_0\left(*\frac{1 - e^{-\frac{Trec}{T1}} + \left(1 - e^{-\frac{TR}{T1}}\right)*\frac{1 - \left(\cos(\alpha)*e^{-\frac{TR}{T1}}\right)^{N_{total}}}{1 - \cos(\alpha)*e^{-\frac{TR}{T1}}}*e^{-\frac{Trec}{T1}}}{1 + \left(\cos(\alpha)*e^{-\frac{TR}{T1}}\right)^{N_{total}}*e^{-\frac{TI}{T1}}*e^{-\frac{Trec}{T1}}} + \frac{\left(1 - e^{-\frac{TI}{T1}}\right)*\left(\cos(\alpha)*e^{-\frac{TR}{T1}}\right)^{N_{total}}*e^{-\frac{Trec}{T1}}}{1 + \left(\cos(\alpha)*e^{-\frac{TR}{T1}}\right)^{N_{total}}*e^{-\frac{TI}{T1}}*e^{-\frac{Trec}{T1}}}\right) \tag{3}$$

$M_0$ represents the acquisition-specific equilibrium magnetization, which depends on the spin density ($\rho$) and on technical properties ($k$) such as receiver amplifier gain and coil design. This can be expressed as: $M_0 = k*\rho$. For simplification, we relied on the MPRAGE sequence's characteristics, which leverage strong T1 contrast and minimize the impact of $\rho$ differences across tissues, allowing for the approximation of $\rho \approx constant$ without compromising contrast accuracy (Haase, 1990). Consequently, we consider $M_0$ as uniform across the entire image.

Based on equations [1–3], the theoretically expected T1w signal could be calculated if the equilibrium magnetization $M_0$, relaxation times (T1, T2*), and imaging parameters (TE, $\alpha$, TR, $T_{rec}$, and TI) were known. We neglected the influence of T2* relaxation and flip angle inhomogeneity because TE is sufficiently short (TE = 4 ms) and we applied

bias field correction of T1w images. Since equation [1] is a formally non-invertible function, $M_0$ was numerically estimated using pairs of T1 and T1w values. First, for each reference region, a fixed T1 value was determined by extracting the median T1 value from that region in each individual R1 map (T1 = 1/R1) and then computing the median of these values across all data sets ($n = 1078$). Second, median T1w intensities were extracted from reference regions for each bias-corrected T1w image. The signal equation was fitted to these data pairs using orthogonal distance regression, estimating each image's $M_0$ constant. The image-specific $M_0$ value was then used to calculate the 1-to-1 mapping between T1w intensities and T1 values. This was realized by sampling T1 values across a range covering all relevant tissues, from 200 ms (lower than fatty tissue) to 4500 ms (approximating cerebrospinal fluid) (Bojorquez et al., 2017), and calculating the corresponding T1w intensities. The 1-to-1 mapping was then used to generate estimated R1 maps, where each voxel was assigned the inverse T1 value (R1 = 1/T1) based on its actual T1w intensity.

### 2.4.2. Linear scaling with reference regions

Linear scaling divides the entire image by a scaling factor, defined as the median intensity of a reference region, ideally removing technically induced inter-image variance while preserving biological variance:

$$T1w_{scaled} = T1w/scaling\ factor$$

### 2.4.3. z-score standardization

Z-score standardization is popular in medical image processing and eliminates intensity scale discrepancies by centering the data around zero and transforming it to units of standard deviation:

$$T1w_{zscore} = (T1w - mean)/standard\ deviation$$

We calculated the mean intensity and standard deviation within the binary brain masks generated with HD-BET. The brain masks cover cerebral and cerebellar WM and GM, and CSF.

### 2.5. Analysis and validation

### 2.5.1. Analysis of reference regions

A data-driven approach was used to identify suitable reference regions. As intensity scaling methods aim to convert image intensities into comparable, quantitative units, especially within the brain, the reference regions should not exhibit a higher inter-image variance than brain regions such as the cerebral cortex or WM. Regions with higher variance are expected to contain confounding biological variance, hampering the estimation of technical variance. The inter-image variance of candidate reference regions was measured with the CoV of median intensities across all bias-corrected T1w images ($n = 1078$) and compared to WM and cortical GM.

### 2.5.2. Technical validation

The first step assessed the potential of intensity scaling methods to

reduce inter-image variance. Therefore, for each region of interest (ROI) within the brain, the variance of median intensities was calculated for each scaling method using the CoV across images ($n = 597$).

The second step assessed the ability to convert conventional T1w image intensities into quantitative units with correlation analysis between intensity-scaled T1w images and quantitative R1 maps ($n = 597$). Two approaches were used: 1) calculating the median intensity of each brain ROI per image and computing Pearson's r across all ROIs and images, and 2) computing voxel-wise Pearson's r between R1 maps and intensity-scaled T1w images per data set, then averaging across Pearson's r values.

### 2.5.3. Biological validation
The difference of each scaling factor between low ($n = 526$) and high ($n = 71$) EDSS groups was assessed to ensure their independence from disease severity. Then, for each intensity scaling method, the mean intensity and the CoV were extracted from NAWM and compared between low and high EDSS groups. The two-sided Welch's *t*-test was used for group comparisons. Cohen's d was calculated to measure the effect size of the group differences (low vs. high EDSS). For the NAWM comparison, we also applied ANCOVA, including age, sex, disease duration, and total lesion volume as covariates.

## 3. Results

### 3.1. Analysis of reference regions

Only median intensities, extracted from unscaled bias-corrected T1w images, in the temporal fatty tissue (CoV = 0.139) and temporal muscle (CoV = 0.139) were in the range of cortical GM (CoV = 0.139) and WM (CoV = 0.130). All other candidate reference regions yielded CoVs ranging between 0.19 and 0.45. Consequently, we only implemented linear scaling with the median intensities of 1) fatty tissue and 2) temporal muscle as scaling factors. Regarding MPRAGE signal modeling, fixed T1 values for fatty tissue (T1 = 378.62 ms) and temporal muscle (T1 = 1491.44 ms) were calculated using all R1 maps ($n = 1078$). Additionally, we included the eye's vitreous body (T1 = 4838.19 ms), which was the only accurately segmented region in the low T1w intensity range with reasonably low absolute T1w intensity variability (standard deviation: ±425.89 [a.u.]) compared to GM and WM

(standard deviation: ± 1893.65 [a.u.] and ± 3191.54 [a.u.]), providing a stable reference point for orthogonal distance regression. Fig. 2 shows a representative illustration of the fitted MPRAGE signal modeling curve.

### 3.2. Variance analysis

Signal modeling and linear scaling with fatty tissue consistently resulted in CoVs closest to those of R1 maps. Z-score standardization resulted in the lowest CoVs in NAWM and cerebellum WM but produced outliers (highest CoVs) in all other regions. Apart from these outliers, unprocessed and bias-corrected T1w images showed the largest CoVs in all regions. The CoVs per ROI of all image types are presented in Fig. 3.

### 3.3. Correlations with R1

Correlations between scaled T1w-based images and R1 maps are provided based on median intensities from multiple ROIs and for average voxel-wise correlations (see Fig. 4).

Across multiple brain regions, the highest correlation between scaled T1w images and R1 maps was observed for z-score standardization ($r = 0.86$), followed by linear scaling with fatty tissue ($r = 0.84$) and MPRAGE signal modeling ($r = 0.83$). The correlation plots of the different methods are shown in Fig. 4A, and Pearson's r correlation values of all methods are shown in Fig. 4B

The averages of voxel-wise correlation values were similar for all image types ($r = 0.84 \pm 0.04$) except for unprocessed T1w images ($r = 0.76 \pm 0.05$). As expected, linear scaling and z-score standardization preserved the linear relationship of unscaled bias-corrected T1w image intensities with R1 values. This is also approximately true for signal modeling, as the conversion of T1w image intensities into estimated R1 ($= 1/T1$) values is a near-to-linear transformation being the inversion of the hyperbolic relation between T1 values and T1w signal intensities illustrated in Fig. 2. Exemplary voxel-wise correlation plots of T1w image intensities with R1 maps are shown in Fig. 4C.

### 3.4. Group comparisons based on EDSS

#### 3.4.1. Scaling factors
Only the scaling factor derived from z-score standardization (this method relies on the standard deviation of whole-brain intensities) differed significantly between high EDSS and low EDSS. Consequently, a group comparison of NAWM intensities was not conducted with z-score standardized T1w images.

#### 3.4.2. NAWM intensity
As expected, mean R1 values in NAWM demonstrated a significant difference between low and high EDSS. Besides R1 maps, only linear scaling with fatty tissue revealed a significant group difference in mean NAWM intensities (see Table 3); of note, the effect size was similarly high (see Fig. 5).

Considering NAWM CoVs, significant group differences were observed across all image types. The most significant group differences were demonstrated by merely bias-corrected T1w images. Since CoVs are invariant under any linear scaling transformation, both linear scaling methods yielded the same results, while signal modeling yielded a very similar result. When age, sex, disease duration, and lesion volume were added as covariates, the differences only remained significant for bias-corrected T1w images. Detailed results are presented in Table 3. Additionally, Fig. 5 shows that effect sizes of group differences are much larger for CoVs than for mean intensities. Of note, the effect size of CoVs derived from bias-corrected T1w images (Cohen's $d = 0.818$) is greater than any effect size from R1 maps (Cohen's d of mean: $-0.351$, Cohen's d of CoV: 0.414).
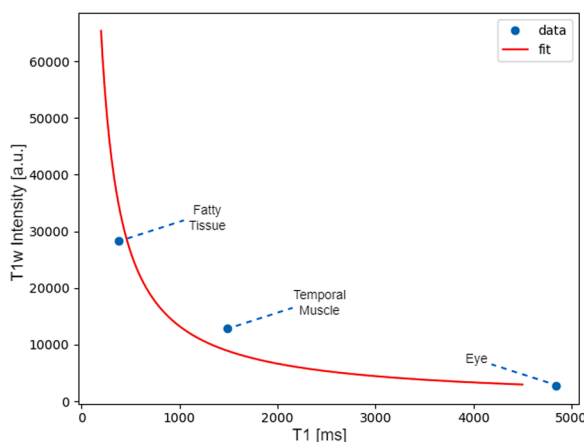


**Fig. 2.** MPRAGE signal modeling.
The red curve shows how the MPRAGE signal was fitted to data from reference regions (fatty tissue, temporal muscle, and eye). The blue points represent, for each reference region, data points defined by pairs of fixed median intensity extracted from T1 maps and median intensity extracted from a specific bias-corrected T1w image. Notice how each data point provides a reference in a different intensity range, sampling the MRI signal spectrum in the low-, mid-, and high-intensity range. Abbreviations: MPRAGE: magnetization prepared rapid gradient echo, MRI: magnetic resonance imaging, T1w: T1-weighted.
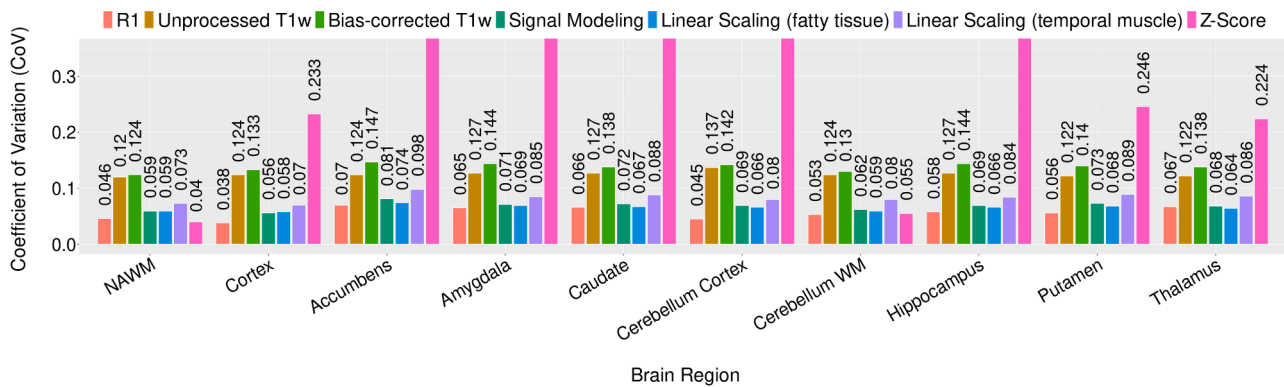
**Fig. 3.** Inter-image variance in brain regions.
The inter-image intensity variance is presented for each brain region and image type in terms of CoV. Linear scaling with fatty tissue and signal modeling consistently yields CoVs similar to R1 maps. Linear scaling with temporal muscle results in slightly larger CoVs; Z-score standardization provides the lowest CoVs in WM but the largest CoVs in all other regions (CoVs in the Accumbens, Amygdala, Caudate, Cerebellum Cortex, and Hippocampus are not completely displayed, and range between CoV = 0.6 and CoV = 3.2). Abbreviations: CoV: coefficient of variation, NAWM: normal-appearing white matter, T1w: T1-weighted, WM: white matter.

## 4. Discussion

This study compared and validated methods for converting conventional T1w MRI image intensities into (semi-) quantitative values. Besides quantitative R1 maps, the scaled T1w images were also compared to unprocessed T1w and bias-corrected T1w images to demonstrate the advantage over unscaled image intensities. To prevent the interference of brain pathology, such as in MS, we focussed on scaling methods avoiding brain regions as reference regions, which has only been attempted in a few studies (Brown et al., 2020; Ganzetti et al., 2014; Ghassemi et al., 2015; Lee et al., 2015; Wahid et al., 2021; Wiltgen et al., 2024b). We found that linear scaling with temporal fatty tissue yielded the best results concerning variance reduction, correlation with R1 maps, and group comparison (mild vs. severe MS-related disability) of NAWM intensity. Similarly, Ghassemi et al. (Ghassemi et al., 2015) found that a scaling method including fatty tissue yielded the best variance reduction and image harmonization results compared to unprocessed T1w images and Nyul's method (Nyúl and Udupa, 1999).

We tested eight candidate regions with a data-driven approach to identify suitable reference regions outside the brain. The assumption was that the intensities of MRI images are affected by biological and technical factors and that a single linear factor can account for the latter (Brown et al., 2020; Haase, 1990). The biological inter-image variance in healthy brain WM and GM is generally low; however, both can be influenced by tissue damage in diseases affecting the brain. Nonetheless, this biological inter-image variance represents only a small fraction of the total inter-image variance compared to the technical inter-image variance. Therefore, intensities of regions outside the brain, suitable for estimating the scaling factor, should not exhibit a variance considerably higher than that of brain WM and GM. Only temporal fatty tissue and temporal muscle fulfilled this criterion. For the MRI signal modeling method, eyes were also selected due to their low standard deviation, making them suitable for orthogonal distance regression. To our knowledge, previous studies have not applied this approach; however, we consider the systematic identification of reference regions a crucial step to ensure reliable implementation of intensity scaling.

Intensity scaling methods can only be applied reliably if validated technically, serving as proof-of-concept, and biologically, confirming the appropriateness of scaling methods in the specific use case (in our case, MS). R1 maps served as a reference for validation since they are known to be sensitive to pathology in MS (Ladopoulos et al., 2022; Lommers et al., 2019; MacKay et al., 2009; Manfredonia et al., 2007; O'Muircheartaigh et al., 2019; Pontillo et al., 2023; Schmierer et al., 2004; Vrenken et al., 2006).

As part of the technical validation, reducing inter-image variance is a

prerequisite for intensity scaling. However, it should not be the sole quality criterion, as minimizing variance can also lead to removing biological variance. For instance, Brown et al. showed that scaling methods with the highest inter-image variance reduction can remove biologically meaningful variance (Brown et al., 2020). The current study shows that linear scaling with fatty tissue and signal modeling yielded CoVs similar to those of R1 maps, achieving the most effective variance reduction across all brain ROIs (Fig. 3). Their slightly increased CoVs in comparison to R1 maps in bigger regions (e.g., NAWM and cortex) may be caused by less effective bias field correction in comparison to the measurement-based correction of the low-frequency bias field in MPM. Regarding the correlation with R1 maps, we found that bias correction expectedly increases the voxel-wise correlation compared to unprocessed T1w images (Fig. 4C), well compatible with the notion that bias correction decreases technical variance in the individual image while ignoring the absolute intensity scale. Interestingly, ROI-based correlations of scaled T1w images with R1 maps (Fig. 4A and Fig. 4B) were virtually the same for all scaling methods ($R^2 = 0.66$–$0.73$), not allowing for a differentiated quality assessment. Brown et al. found similar correlation values ($R^2 = 0.63$) between fat-scaled T1w images and MTR, a measurement that is sensitive to similar properties as R1 maps (e.g., density of macromolecules) (Brown et al., 2020). Our results indicate that scaling with regions outside the brain can yield correlations with R1 maps comparable to scaling methods based on WM, GM, and CSF, as (Hasse et al., 2022) reported in healthy subjects ($R^2 = 0.72$).

Z-score standardization yielded seemingly promising results regarding the correlation with R1 maps and variance reduction in WM, but it also resulted in the largest CoVs in all other regions. In addition, the standard deviation used as a scaling factor in z-score standardization showed a significant difference between the two MS groups (mild and severe disability), indicating that the efficient reduction of technical variance goes along with removing biological variance. This is conceivable since the standard deviation is calculated across the entire brain and is, therefore, influenced by WM lesion load, GM atrophy, and NAWM damage. Hence, z-score standardization is not a reliable approach for intensity conversion of T1w images in the specific case of MS.

Regarding biological validation, results in Table 3 show that mean R1 values in NAWM significantly differed between groups (mild vs. severe disability), well in accordance with the literature, where a decrease in R1 values in NAWM has been associated with MS pathology (Ladopoulos et al., 2022; Lommers et al., 2019; MacKay et al., 2009; Manfredonia et al., 2007; O'Muircheartaigh et al., 2019; Pontillo et al., 2023; Schmierer et al., 2004; Vrenken et al., 2006). Notably, this group difference could also be demonstrated with scaled T1w images using
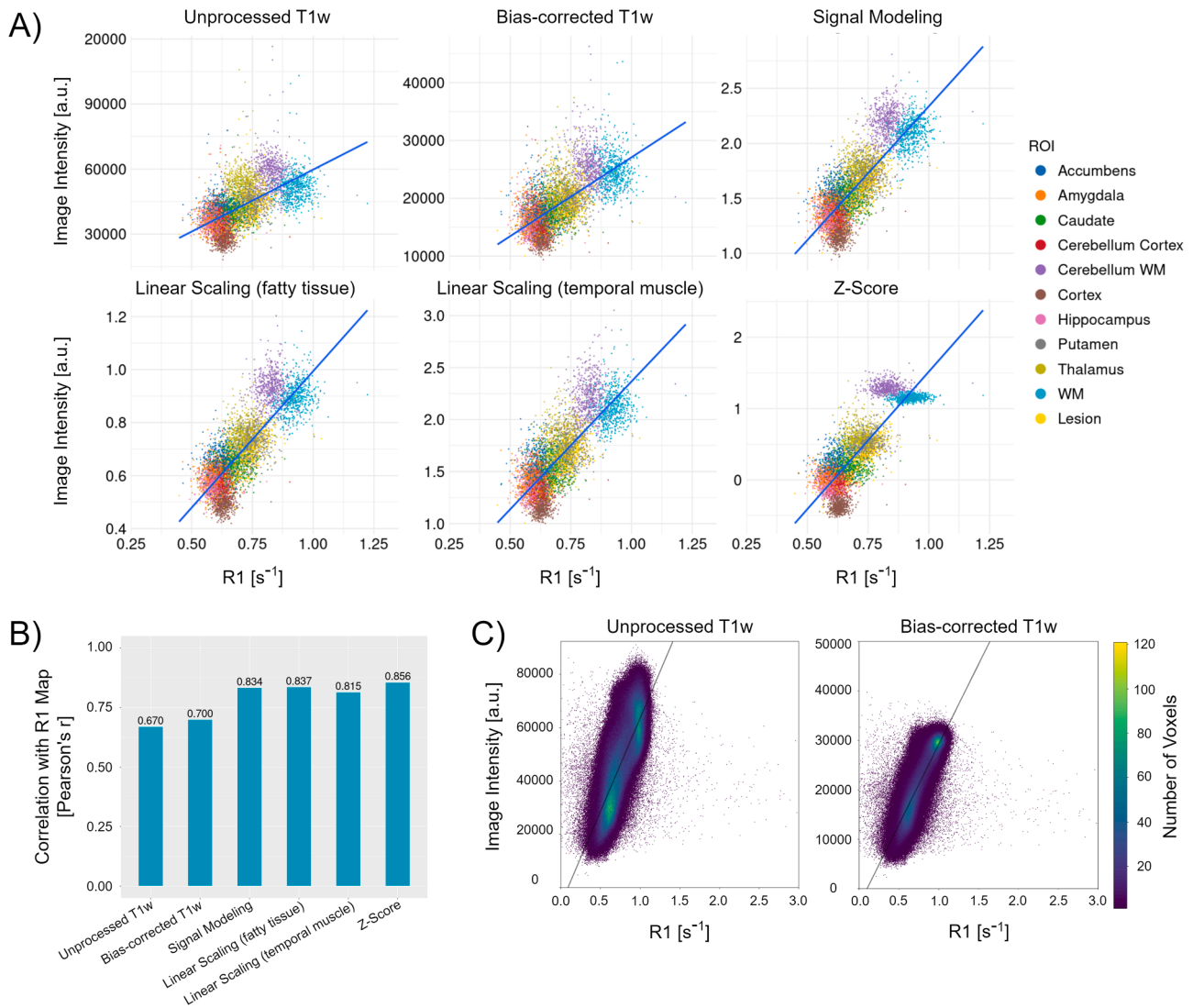
**Fig. 4.** Correlation of scaled T1w image intensities with R1 values.
The graphs showcase the correlation of the differently scaled T1w images with R1 maps. A) Correlation plots across brain regions. Each dot represents the median R1 value (x-axis) and the median intensity according to the selected image type (y-axis) within a brain region; brain regions are highlighted in different colors. The blue linear regression line shows the relation between the corresponding scaled T1w image intensities and R1 values. B) Pearson's r values associated with the linear regressions from A). The strongest correlation with R1 was observed for z-score standardization ($r = 0.86$), followed by linear scaling with fatty tissue ($r = 0.84$) and MRI signal modeling ($r = 0.83$). C) Example of voxel-wise correlation. The voxel-wise correlation (Pearson's r) with R1 is calculated per image and averaged across all images. The mean correlation with R1 was the same for each image type derived from bias-corrected T1w images ($r = 0.84 \pm 0.04$) and was higher than that of unprocessed T1w images ($r = 0.76 \pm 0.05$). The correlation plot of the unprocessed T1w image shows a much larger spread (y-axis) than that of the bias-corrected T1w image. The correlation plot was virtually the same for all other intensity scaling methods, apart from different scales of the y-axis, since the voxel-wise correlation within an image is invariant under linear transformation (i.e., intensity scaling). Abbreviations: ROI: region of interest, T1w: T1-weighted, WM: white matter.

**Table 3**
NAWM group comparison.

| | NAWM Mean intensity | | | NAWM CoV | | |
|---|---|---|---|---|---|---|
| | Low EDSS | High EDSS | p-value | Low EDSS | High EDSS | p-value |
| **R1** | $0.972 \pm 0.044$ | $0.956 \pm 0.047$ | **0.0081** | $0.070 \pm 0.013$ | $0.076 \pm 0.015$ | **0.0027** |
| **Unprocessed T1w** | $53,873.9 \pm 6154.9$ | $52,025.8 \pm 8156.6$ | 0.0695 | $0.109 \pm 0.008$ | $0.114 \pm 0.012$ | **0.0031** |
| **Bias-corrected T1w** | $25,187.8 \pm 3084.7$ | $25,610.6 \pm 3391.1$ | 0.3218 | $0.062 \pm 0.010$ | $0.072 \pm 0.015$ | **1.636E-07**\* |
| **Signal modeling** | $2.155 \pm 0.128$ | $2.136 \pm 0.107$ | 0.1833 | $0.062 \pm 0.010$ | $0.073 \pm 0.015$ | **1.648E-07**\* |
| **Linear scaling (fatty tissue)** | $0.918 \pm 0.055$ | $0.905 \pm 0.046$ | **0.0344** | $0.062 \pm 0.010$ | $0.072 \pm 0.015$ | **1.636E-07**\* |
| **Linear scaling (temporal muscle)** | $2.174 \pm 0.160$ | $2.197 \pm 0.137$ | 0.1989 | $0.062 \pm 0.010$ | $0.072 \pm 0.015$ | **1.636E-07**\* |

Bold numbers indicate p-values < 0.05, i.e., statistically significant results.

Abbreviations: CoV: coefficient of variation, EDSS: expanded disability status scale, NAWM: normal-appearing white matter, T1w: T1-weighted.

\* remains significant in ANCOVA with age, sex, disease duration, and lesion volume as covariates.
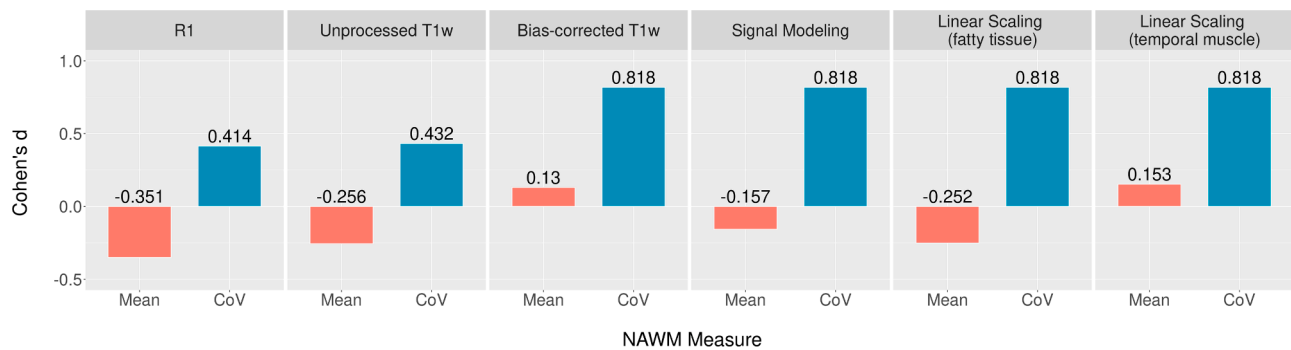
**Fig. 5.** Group comparison effect size.

The effect size (Cohen's d) is presented for group comparison of mean and CoV of NAWM R1 and image intensities between low and high EDSS groups. Differences between low EDSS and high EDSS are more pronounced for CoVs than mean values and in opposite directions. Noticeably, CoVs extracted from bias-corrected T1w images yield a larger effect ($d = 0.818$) than measures extracted from R1 maps (Mean: $d = -0.351$, CoV: $d = 0.414$). Besides R1 maps, only unprocessed T1w images and linear scaling with fatty tissue show a small effect size regarding mean NAWM image intensity. Note that the difference in mean NAWM intensities was not statistically significant for unprocessed T1w images (see Table 3). All other image types exhibited negligible effect sizes with regard to mean NAWM intensity. Abbreviations: CoV: coefficient of variation, EDSS: expanded disability status scale, NAWM: normal-appearing white matter, T1w: T1-weighted.

fatty tissue, showing sensitivity with regard to MS-related disability.

In contrast to the mean NAWM intensity, a rather global tissue integrity measure, the NAWM CoV could reflect local alterations within one subject. The CoV is a scale-independent measure that can be directly derived from bias-corrected T1w images. To our surprise, the group difference between people with low (≤3) and high (>3) EDSS in NAWM was most significant when using the CoV instead of the mean values for all image types. Even after including age, sex, disease duration, and lesion volume as covariates, the CoV from bias-corrected T1w images yielded a significant difference. This was not the case in R1 maps since, in contrast to conventional T1w images, R1 maps are optimized for generating reliable quantitative intensities, which comes at the expense of a lower image resolution and tissue contrast. Similar results were found by Johnson et al. (Johnson et al., 2023), with CoVs of the MWF in NAWM differentiating more accurately between healthy controls and MS than the mean. Taken together, CoV and mean values within NAWM could provide complementary information on tissue integrity. On the one hand, CoV could be used to indicate disease-related widespread inhomogeneous tissue alterations. On the other hand, intensity scaling allows for calculating mean intensity values that could help quantify the global level of tissue damage (e.g., demyelination in MS). Overall, group differences in mean and CoVs of NAWM from scaled T1w images show the potential for retrospective intensity-based analyses of conventional T1w images to identify new imaging biomarkers associated with the progression of diseases such as MS.

We acknowledge the limitations of our study. The MRI images were acquired on the same scanner and with the same protocol. T1w sequences are optimized for brain tissues; coverage of regions outside the brain may vary between scanners and protocols, and so may their suitability as reference regions. Finally, our study was limited to brain MRI of people with MS. Consequently, we regard a validation of intensity scaling based on temporal fatty tissue in different T1w sequences in a multi-center study, ideally across multiple neuropsychiatric diseases, as a reasonable next step. In the same context, the advantage of using a single reference region (temporal fatty tissue) over multiple reference regions (signal modeling) needs to be confirmed.

## 5. Conclusion

Temporal fatty tissue is a promising candidate for a valuable reference region to convert conventional T1w image intensities into (semi-)quantitative values with high inter-image comparability when quantitative T1 maps are not available. In addition to absolute intensity measures, we found that changes in MS NAWM appear to increase intensity variability detectable in conventional T1w images. This

observation is noteworthy and warrants further investigation.

**Data availability**

Data used and generated during this study are private and not publicly available, could be made available upon reasonable request.

**CRediT authorship contribution statement**

**Tun Wiltgen:** Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Julian McGinnis:** Writing – original draft, Software, Resources, Methodology, Data curation. **Ronja Berg:** Writing – original draft, Software, Methodology. **Cui Ci Voon:** Writing – original draft, Resources, Data curation. **Oula Puonti:** Writing – original draft, Software, Methodology. **Katrin Giglhuber:** Writing – original draft, Resources. **Carl Ganter:** Writing – original draft, Methodology, Formal analysis. **Claus Zimmer:** Writing – original draft, Resources. **Bernhard Hemmer:** Writing – original draft, Resources. **Benedikt Wiestler:** Writing – original draft, Supervision, Resources. **Jan Kirschke:** Writing – original draft, Resources, Data curation. **Christine Preibisch:** Writing – original draft, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Mark Mühlau:** Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2025.121395.

## References

Berg, R.C., Leutritz, T., Weiskopf, N., Preibisch, C., 2022. Multi-parameter quantitative mapping of R1, R2*, PD, and MTsat is reproducible when accelerated with Compressed SENSE. Neuroimage 253, 119092. https://doi.org/10.1016/j.neuroimage.2022.119092.

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

Brant-Zawadzki, M., Gillan, G.D., Nitz, W.R., 1992. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. Radiology. 182, 769–775. https://doi.org/10.1148/radiology.182.3.1535892.

Brown, R.A., Fetco, D., Fratila, R., Fadda, G., Jiang, S., Alkhawajah, N.M., Yeh, E.A., Banwell, B., Bar-Or, A., Arnold, D.L., 2020. Deep learning segmentation of orbital fat to calibrate conventional MRI for longitudinal studies. Neuroimage 208, 116442. https://doi.org/10.1016/j.neuroimage.2019.116442.

Cerri, S., Puonti, O., Meier, D.S., Wuerfel, J., Mühlau, M., Siebner, H.R., Van Leemput, K., 2021. A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. Neuroimage 225, 117471. https://doi.org/10.1016/j.neuroimage.2020.117471.

Deichmann, R., Good, C.D., Josephs, O., Ashburner, J., Turner, R., 2000. Optimization of 3-D MP-RAGE sequences for structural brain imaging. Neuroimage 12, 112–127. https://doi.org/10.1006/nimg.2000.0601.

Deichmann, R., Haase, A., 1992. Quantification of *T1* values by SNAPSHOT-FLASH NMR imaging. J. Magnetic Resonance 96, 608–612. https://doi.org/10.1016/0022-2364(92)90347-A, 1969.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J.V., Pieper, S., Kikinis, R., 2012. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imag., Quantit. Imag. Cancer 30, 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001.

Fischl, B., 2012. FreeSurfer. NeuroImage, 20 YEARS OF fMRI 62, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.

Fortin, J.-P., Sweeney, E.M., Muschelli, J., Crainiceanu, C.M., Shinohara, R.T., Alzheimer's Disease Neuroimaging Initiative, 2016. Removing inter-subject technical variability in magnetic resonance imaging studies. Neuroimage 132, 198–212. https://doi.org/10.1016/j.neuroimage.2016.02.036.

Ganzetti, M., Wenderoth, N., Mantini, D., 2014. Whole brain myelin mapping using T1- and T2-weighted MR imaging data. Front. Hum. Neurosci. 8. https://doi.org/10.3389/fnhum.2014.00671.

Geerts-Ossevoort, L., de Weerdt, E., Duijndam, A., van IJperen, G., Peeters, H., Doneva, M., Nijenhuis, M., Huang, A., 2018. Speed done right. Every time.

Ghassemi, R., Brown, R., Narayanan, S., Banwell, B., Nakamura, K., Arnold, D.L., 2015. Normalization of white matter intensity on T1-weighted images of patients with acquired Central nervous system demyelination. J. Neuroimag. 25, 184–190. https://doi.org/10.1111/jon.12129.

Gilmore, J.H., Lin, W., Corouge, I., Vetsa, Y.S.K., Smith, J.K., Kang, C., Gu, H., Hamer, R. M., Lieberman, J.A., Gerig, G., 2007. Early postnatal development of Corpus Callosum and corticospinal white matter assessed with quantitative tractography. AJNR Am. J. Neuroradiol. 28, 1789–1795. https://doi.org/10.3174/ajnr.A0751.

Haase, A., 1990. Snapshot flash mri. Applications to t1, t2, and chemical-shift imaging. Magn. Reson. Med. 13, 77–89. https://doi.org/10.1002/mrm.1910130109.

Hasse, A., Bertini, J., Foxley, S., Jeong, Y., Javed, A., Carroll, T.J., 2022. Application of a novel T1 retrospective quantification using internal references (T1-REQUIRE) algorithm to derive quantitative T1 relaxation maps of the brain. Int. J. ImAging Syst. Technol. 32, 1903–1915. https://doi.org/10.1002/ima.22768.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18, 203–211. https://doi.org/10.1038/s41592-020-01008-z.

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.-P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K.H., Kickingereder, P., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. Hum. Brain Mapp. 40, 4952–4964. https://doi.org/10.1002/hbm.24750.

Johnson, P., Vavasour, I.M., Stojkova, B.J., Abel, S., Lee, L.E., Laule, C., Tam, R., Li, D.K.B., Ackermans, N., Schabas, A.J., Chan, J., Cross, H., Sayao, A.-L., Devonshire, V., Carruthers, R., Traboulsee, A., Kolind, S.H., 2023. Myelin heterogeneity for assessing normal appearing white matter myelin damage in multiple sclerosis. J. Neuroimaging 33, 227–234. https://doi.org/10.1111/jon.13069.

Kutzelnigg, A., Lucchinetti, C.F., Stadelmann, C., Brück, W., Rauschka, H., Bergmann, M., Schmidbauer, M., Parisi, J.E., Lassmann, H., 2005. Cortical demyelination and diffuse white matter injury in multiple sclerosis. Brain 128, 2705–2712. https://doi.org/10.1093/brain/awh641.

Ladopoulos, T., Matusche, B., Bellenberg, B., Heuser, F., Gold, R., Lukas, C., Schneider, R., 2022. Relaxometry and brain myelin quantification with synthetic MRI in MS subtypes and their associations with spinal cord atrophy. NeuroImage: Clin. 36, 103166. https://doi.org/10.1016/j.nicl.2022.103166.

Laule, C., Pavlova, V., Leung, E., Zhao, G., MacKay, A.L., Kozlowski, P., Traboulsee, A.L., Li, D.K.B., Moore, G.R.W., 2013. Diffusely abnormal white matter in multiple sclerosis: further histologic studies provide evidence for a primary lipid abnormality with neurodegeneration. J. Neuropathol. Experim. Neurol. 72, 42–52. https://doi.org/10.1097/NEN.0b013e31827bced3.

Lavielle, A., Boux, F., Deborne, J., Pinaud, N., Dufort, S., Verry, C., Grand, S., Troprès, I., Vecco-Garda, C., Le Duc, G., Mornet, S., Crémillieux, Y., 2023. T1 Mapping from MPRAGE acquisitions: application to the measurement of the concentration of nanoparticles in tumors for theranostic use. J. Magnetic Reson. Imaging 58, 313–323. https://doi.org/10.1002/jmri.28509.

Lavielle, A., Pinaud, N., Zhang, B., Crémillieux, Y., 2024. Quantitative brain T1 maps derived from T1-weighted MRI acquisitions: a proof-of-concept study. Eur. Radiol. Exp. 8, 109. https://doi.org/10.1186/s41747-024-00517-2.

Lee, K., Cherel, M., Budin, F., Gilmore, J., Consing, K.Z., Rasmussen, J., Wadhwa, P.D., Entringer, S., Glasser, M.F., Van Essen, D.C., Buss, C., Styner, M., 2015. Early postnatal myelin content estimate of white matter via T1w/T2w ratio. Proc. SPIE Int. Soc. Opt. Eng. 9417. https://doi.org/10.1117/12.2082198, 94171R.

Leray, E., Yaouanq, J., Le Page, E., Coustans, M., Laplaud, D., Oger, J., Edan, G., 2010. Evidence for a two-stage disability progression in multiple sclerosis. Brain 133, 1900–1913. https://doi.org/10.1093/brain/awq076.

Leutritz, T., Seif, M., Helms, G., Samson, R.S., Curt, A., Freund, P., Weiskopf, N., 2020. Multiparameter mapping of relaxation (R1, R2*), proton density and magnetization transfer saturation at 3 T: a multicenter dual-vendor reproducibility and repeatability study. Hum. Brain Mapp. 41, 4232–4247. https://doi.org/10.1002/hbm.25122.

Lommers, E., Simon, J., Reuter, G., Delrue, G., Dive, D., Degueldre, C., Balteau, E., Phillips, C., Maquet, P., 2019. Multiparameter MRI quantification of microstructural tissue alterations in multiple sclerosis. NeuroImage: Clinical 23, 101879. https://doi.org/10.1016/j.nicl.2019.101879.

MacKay, A.L., Vavasour, I.M., Rauscher, A., Kolind, S.H., Mädler, B., Moore, G.R.W., Traboulsee, A.L., Li, D.K.B., Laule, C., 2009. MR Relaxation in Multiple sclerosis. Neuroimag. Clinics North Am., Multiple Sclerosis, Part II: Nonconvent. MRI Techniq. 19, 1–26. https://doi.org/10.1016/j.nic.2008.09.007.

Manfredonia, F., Ciccarelli, O., Khaleeli, Z., Tozer, D.J., Sastre-Garriga, J., Miller, D.H., Thompson, A.J., 2007. Normal-appearing brain T1 relaxation time predicts disability in early primary progressive Multiple sclerosis. Arch. Neurol. 64, 411–415. https://doi.org/10.1001/archneur.64.3.411.

Mejia, A.F., Sweeney, E.M., Dewey, B., Nair, G., Sati, P., Shea, C., Reich, D.S., Shinohara, R.T., 2016. Statistical estimation of T1 relaxation times using conventional magnetic resonance imaging. Neuroimage 133, 176–188. https://doi.org/10.1016/j.neuroimage.2015.12.037.

Moll, N.M., Rietsch, A.M., Thomas, S., Ransohoff, A.J., Lee, J.-C., Fox, R., Chang, A., Ransohoff, R.M., Fisher, E., 2011. Multiple sclerosis normal-appearing white matter: pathology–imaging correlations. Ann. Neurol. 70, 764–773. https://doi.org/10.1002/ana.22521.

Moskovich, S., Shtangel, O., Mezer, A.A., 2024. Approximating R1 and R2: a quantitative approach to clinical weighted MRI. Hum. Brain Mapp. 45, e70102. https://doi.org/10.1002/hbm.70102.

Nyúl, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn. Reson. Med. 42, 1072–1081. https://doi.org/10.1002/(SICI)1522-2594(199912)42,6<1072::AID−MRM11>3.0.CO;2-M.

O'Muircheartaigh, J., Vavasour, I., Ljungberg, E., Li, D.K.B., Rauscher, A., Levesque, V., Garren, H., Clayton, D., Tam, R., Traboulsee, A., Kolind, S., 2019. Quantitative neuroimaging measures of myelin in the healthy brain and in multiple sclerosis. Hum. Brain Mapp. 40, 2104–2116. https://doi.org/10.1002/hbm.24510.

Pontillo, G., Petracca, M., Monti, S., Quarantelli, M., Lanzillo, R., Costabile, T., Carotenuto, A., Tortora, F., Elefante, A., Morra, V.B., Brunetti, A., Palma, G., Cocozza, S., 2023. Clinical correlates of R1 relaxometry and magnetic susceptibility changes in multiple sclerosis: a multi-parameter quantitative MRI study of brain iron and myelin. Eur. Radiol. 33, 2185–2194. https://doi.org/10.1007/s00330-022-09154-y.

Puonti, O., Iglesias, J.E., Van Leemput, K., 2016. Fast and sequence-adaptive whole-brain segmentation using parametric bayesian modeling. Neuroimage 143, 235–249. https://doi.org/10.1016/j.neuroimage.2016.09.011.

Puonti, O., Van Leemput, K., Saturnino, G.B., Siebner, H.R., Madsen, K.H., Thielscher, A., 2020. Accurate and robust whole-head segmentation from magnetic resonance images for individualized head modeling. Neuroimage 219, 117044. https://doi.org/10.1016/j.neuroimage.2020.117044.

Schmierer, K., Scaravilli, F., Altmann, D.R., Barker, G.J., Miller, D.H., 2004. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. Ann. Neurol. 56, 407–415. https://doi.org/10.1002/ana.20202.

Shinohara, R.T., Sweeney, E.M., Goldsmith, J., Shiee, N., Mateen, F.J., Calabresi, P.A., Jarso, S., Pham, D.L., Reich, D.S., Crainiceanu, C.M., Ageing, for the A.I.B.L.F.S. of, Initiative, the A.D.N., 2014. Statistical normalization techniques for magnetic resonance imaging. NeuroImage : Clin. 6, 9. https://doi.org/10.1016/j.nicl.2014.08.008.

Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M.F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., Mohammadi, S., 2019. hMRI – A toolbox for quantitative

MRI in neuroscience and clinical research. Neuroimage 194, 191–210. https://doi.org/10.1016/j.neuroimage.2019.01.029.

Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., Fujihara, K., Galetta, S.L., Hartung, H.P., Kappos, L., Lublin, F.D., Marrie, R.A., Miller, A.E., Miller, D.H., Montalban, X., Mowry, E.M., Sorensen, P.S., Tintoré, M., Traboulsee, A.L., Trojano, M., Uitdehaag, B.M.J., Vukusic, S., Waubant, E., Weinshenker, B.G., Reingold, S.C., Cohen, J.A., 2018. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 17, 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEe Trans. Med. ImAging 29, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908.

Vernier, B., Van Reeth, E., Pilleul, F., Beuf, O., Ratiney, H., 2021. MRI contrast enhancement of magnetization prepared steady State sequence: an optimal control framework. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Presented at the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 1709–1713. https://doi.org/10.1109/ISBI48211.2021.9434151.

Vrenken, H., Geurts, J.J.G., Knol, D.L., van Dijk, L.N., Dattola, V., Jasperse, B., van Schijndel, R.A., Polman, C.H., Castelijns, J.A., Barkhof, F., Pouwels, P.J.W., 2006. Whole-brain T1 mapping in multiple sclerosis: global changes of normal-appearing gray and white matter. Radiology. 240, 811–820. https://doi.org/10.1148/radiol.2403050569.

Wahid, K.A., He, R., McDonald, B.A., Anderson, B.M., Salzillo, T., Mulder, S., Wang, Jarey, Sharafi, C.S., McCoy, L.A., Naser, M.A., Ahmed, S., Sanders, K.L., Mohamed, A.S.R., Ding, Y., Wang, Jihong, Hutcheson, K., Lai, S.Y., Fuller, C.D., van Dijk, L.V., 2021. Intensity standardization methods in magnetic resonance imaging of head and neck cancer. Phys. ImAging Radiat. Oncol. 20, 88–93. https://doi.org/10.1016/j.phro.2021.11.001.

Wiltgen, T., McGinnis, J., Schlaeger, S., Kofler, F., Voon, C., Berthele, A., Bischl, D., Grundl, L., Will, N., Metz, M., Schinz, D., Sepp, D., Prucker, P., Schmitz-Koep, B., Zimmer, C., Menze, B., Rueckert, D., Hemmer, B., Kirschke, J., Mühlau, M., Wiestler, B., 2024a. LST-AI: a deep learning ensemble for accurate MS lesion segmentation. NeuroImage: Clinical 42, 103611. https://doi.org/10.1016/j.nicl.2024.103611.

Wiltgen, T., Voon, C., Van Leemput, K., Wiestler, B., Mühlau, M., 2024b. Intensity scaling of conventional brain magnetic resonance images avoiding cerebral reference regions: a systematic review. PLoS. One 19, e0298642. https://doi.org/10.1371/journal.pone.0298642.

World Medical Association, 2001. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Bull. World Health Organ. 79, 373–374.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015.

Yushkevich, P.A., Pluta, J., Wang, H., Wisse, L.E.M., Das, S., Wolk, D., 2016. Fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions In 3 tesla and 7 tesla T2-weighted MRI. Alzheimer's Dementia 12, P126–P127. https://doi.org/10.1016/j.jalz.2016.06.205.

## Glossary

*CIS:* clinically isolated syndrome
*CoV:* coefficient of variation
*EDSS:* expanded disability status scale
*FLAIR:* fluid-attenuated inversion recovery
*GM:* gray matter
*LST-AI:* lesion segmentation tool – artificial intelligence
*MNI:* Montreal Neurological Institute
*MPM:* multiparameter mapping
*MPRAGE:* magnetization-prepared rapid gradient echo
*MRI:* magnetic resonance imaging
*MS:* multiple sclerosis
*MTw:* magnetization transfer-weighted
*MWF:* myelin water fraction
*NAWM:* normal-appearing white matter
*PDw:* proton density-weighted
*ROI:* region of interest
*RIS:* radiologically isolated syndrome
*SAMSEG:* sequence adaptive multimodal segmentation
*T1w:* T1-weighted
*T2w:* T2-weighted
*TE:* echo time
*TI:* inversion time
*TR:* repetition time
*WM:* white matter