◆ Human Brain Mapping 38:1780–1790 (2017) ◆

Test–Retest Reliability and Concurrent Validity of in Vivo Myelin Content Indices: Myelin Water Fraction and Calibrated T_1w/T_2w Image Ratio

Muzamil Arshad,[1,2] Jeffrey A. Stanley,[1] and Naftali Raz[2,3]*

[1]*Department of Psychiatry and Behavioral Neuroscience, School of Medicine, Wayne State University, Detroit, Michigan*
[2]*Institute of Gerontology, Wayne State University, Detroit, Michigan*
[3]*Department of Psychology, Wayne State University, Detroit, Michigan*

Abstract: In an age-heterogeneous sample of healthy adults, we examined test–retest reliability (with and without participant repositioning) of two popular MRI methods of estimating myelin content: modeling the short spin–spin (T_2) relaxation component of multi-echo imaging data and computing the ratio of T_1-weighted and T_2-weighted images (T_1w/T_2w). Taking the myelin water fraction (MWF) index of myelin content derived from the multi-component T_2 relaxation data as a standard, we evaluate the concurrent and differential validity of T_1w/T_2w ratio images. The results revealed high reliability of MWF and T_1w/T_2w ratio. However, we found significant correlations of low to moderate magnitude between MWF and the T_1w/T_2w ratio in only two of six examined regions of the cerebral white matter. Notably, significant correlations of the same or greater magnitude were observed for T_1w/T_2w ratio and the intermediate T_2 relaxation time constant, which is believed to reflect differences in the mobility of water between the intracellular and extracellular compartments. We conclude that although both methods are highly reliable and thus well-suited for longitudinal studies, T_1w/T_2w ratio has low criterion validity and may be not an optimal index of subcortical myelin content. *Hum Brain Mapp 38:1780–1790, 2017.*

Key words: white matter; MRI; T2 relaxation; Simpson paradox; brain water mobility

INTRODUCTION

Myelin, a core component of the central nervous system plays a critical role in ensuring speed and fidelity of neural transmission, efficiency of axonal energy metabolism, maintenance of membranes, formation of synapses and promotion of neuroplasticity [de Hoz and Simons, 2015; Jebelli et al., 2015; Saab et al., 2013]. Myelination of axons is one of the pivotal phenomena of brain development [Ansari and Loch, 1975; Barkovich et al., 1988; Davison and Dobbing, 1966], whereas reduction of myelin content and alteration of its structure have been proposed as neuroanatomic substrates of age-related cognitive declines [Bartzokis, 2004; Lu et al., 2013]. Therefore, valid in vivo evaluation of regional myelin content, reflecting the myelination of axons and reliable measurement of changes therein is a high priority. In recent years, several proxy indices of myelin content have been proposed of which two approaches have gained particular popularity: myelin water fraction (MWF)—derived from the modeling spin–spin (T_2) relaxation components of multi-echo T_2

Contract grant sponsor: National Institute on Aging; Contract grant number: R37-AG011230 (to N.R.); Contract grant sponsors: Wayne State University Office of the Vice-President for Research Grants Boost Program (to N.R.) and the Lycaki-Young Funds from the State of Michigan (to J.A.S.).

*Correspondence to: Naftali Raz, Institute of Gerontology, 87 East Ferry St. Detroit, MI 48202. E-mail: nraz@wayne.edu

Received for publication 30 August 2016; Revised 6 November 2016; Accepted 21 November 2016.

DOI: 10.1002/hbm.23481
Published online 23 December 2016 in Wiley Online Library (wileyonlinelibrary.com).

relaxation (ME-T_2) imaging data [Mackay et al., 1994; Whittal et al., 1997]—and the ratio of T_1-weighted and T_2-weighted MRI images (T_1w/T_2w) [Glasser and Van Essen, 2011]. Both methods are noninvasive and hence suitable for *in vivo* longitudinal investigations, which are necessary for studying time-dependent processes such as development and aging [Baltes and Nesselroade, 1979; Kramer et al., 2000; Lindenberger et al., 2011].

The first index, MWF, is based on modeling the multiple-echo water signal with multiple T_2 relaxation times and determining the contribution of the short T_2 relaxation time constant or component, which is proportional to the amount of the water trapped between the myelin sheaths [Curnes et al., 1988; Menon et al., 1992]. At present, it is considered a standard or reference method of myelin content estimation [Alonso-Ortiz et al., 2015; Billiet et al., 2015], but its main disadvantages are the need for special MRI sequence and relatively long acquisition times. In addition to the short T_2 component associated with myelin water, a larger intracellular/extracellular water compartment with an intermediate T_2 relaxation time constant can also be quantified in the ME-T_2 relaxation decay signal and its index, the geometric mean ($geomT_{2IEW}$), is typically reported [Mackay et al., 1994]. Although the precise neurobiological significance of the $geomT_{2IEW}$ with respect to the intracellular/extracellular compartment is not fully understood, according to a combined ME-T_2 imaging and quantitative histology study, the $geomT_{2IEW}$ correlates positively with the diameter of axons in white matter tracts of the rat spinal cord. That is, enlargement of axonal diameter is accompanied by expansion of the intracellular space and increased mobility of the intracellular water, with ensuing increase in $geomT_{2IEW.}$ Thus ME-T_2 derived indices have the potential to describe various microstructural properties of white matter [Dula et al., 2010].

The second putative myelin content index is based on standard T_1- and T_2-weighed sequences with relatively short acquisition times. Glasser and Van Essen [2011] argued that the pixel intensity on a T_1-weighted image is directly proportional to the myelin contrast, χ, and inversely proportional to it on a T_2-weighted image. They proposed that dividing the T_1-weighted image by the T_2-weighted image (and producing a T_1w/T_2w ratio) enhances the myelin contrast by χ^2 and can be used as a proxy for myelin content. However, the intensity nonuniformity (INU) on the images generated by the imperfection in the B_1 field differs between T_1- and T_2-weighted images, and thus does not adequately cancel when a ratio is taken [Ganzetti et al., 2014]. Recognition of that fact led to a revision of the method and adding independent INU correction for the T_1- and T_2-weighted images prior to computing the ratio [Ganzetti et al., 2014; Glasser and van Essen, 2014]. In addition, as a nonquantitative technique, T_1w/T_2w imaging is subjected to intensity scaling disparities across individuals and scanners. To overcome this, Ganzetti et al. [2014] have proposed a "calibration" method that involves linearly transforming the INU-corrected images prior to computing the ratio. Although it is acknowledged that T_1w/T_2w image ratio is not an index of myelin, it is still considered a useful proxy for myelin content [Shafee et al., 2015]. Nonetheless, it remains unclear how close the two methods are in their evaluation of myelin content as there are no extant studies comparing the two approaches. Given the references status of MWF, the calibrated T_1w/T_2w image ratio method needs to be evaluated against it, thus testing the concurrent validity of the latter [Cronbach and Meehl, 1955].

Because of the central role allotted to myelin in theories of aging and development [e.g., Bartzokis, 2004] and the importance of longitudinal studies in these areas of research, it is important to ensure suitability of putative indices of the brain myelin content and appraise their test–retest reliability. In particular, sensitivity to changes of participants' position in the scanner that are inevitable in a study that involves multiple measurement occasions needs to be examined. It is also important to determine whether the estimates of myelin content are equally reliable across multiple regions as differential reliability would present a threat to validity of inferences pertaining to heterochronicity of brain aging that has been postulated in the extant literature [see Fjell et al., 2014; Raz and Rodrigue, 2006 for reviews]. High test–retest reliability of MWF has been assessed with Pearson correlations between two scans [Meyers et al., 2009], but unlike intraclass correlation (ICC), Pearson r is insensitive to linear changes in the index over occasions. In a subsequent study, Meyers et al. [2013] reported moderate test–retest reliability of MWF and high reliability of $geomT_{2IEW}$, as assessed by ICC. In two small sample studies of ME-T_2 repeatability, MWF was deemed to be less reliable than the $geomT_{2IEW}$ but no ICC values were reported [Levesque et al., 2010; Vavasour et al., 2006]. A recent study of test–retest reliability of T_1w/T_2w measurements in a sample of 99 young adults observed a relatively modest Pearson $r = 0.83$ across all vertices of the map generated by FreeSurfer, with no ICC reported [Shafee et al., 2015].

The main objectives of this study were, therefore, to evaluate the test–retest reliability of ME-T_2 and calibrated T_1w/T_2w indices in subcortical white matter tracts chosen as representatives of commissural, associative and projection fibers, and to establish concurrent validity of the latter index vis a vis the former. Based on previous work conducted on smaller samples, we hypothesized that MWF and $geomT_{2IEW}$ would meet the ICC ≥ 0.80 criterion of reliability. By adding INU correction and "calibration," we hypothesized that reliability of T_1w/T_2w ratios would also meet our reliability criterion. Finally, we tested myelin specificity of T_1w/T_2w ratio by comparing its correlations with MWF and $geomT_{2IEW}$, myelin-specific and intracellular/extracellular compartment indices derived from ME-T_2 relaxation imaging.

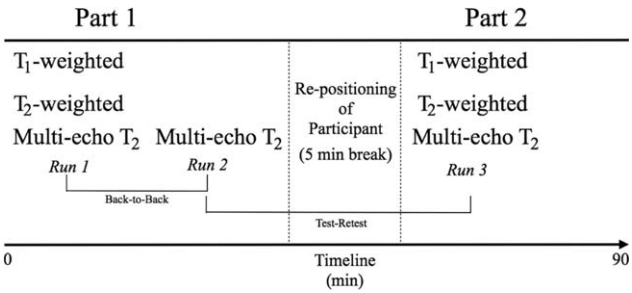

Figure 1.
Design diagram of the reliability trials.

METHODS

Participants

Twenty healthy adult participants were recruited from the Detroit metropolitan area. They were screened via a questionnaire for history of neurological and psychiatric disorders, cardiovascular disease other than medically treated hypertension, endocrine and metabolic disorders, head injury accompanied by loss of consciousness for more than 5 min, use of antiepileptic, anxiolytic and antidepressant medications. The participants were equally divided by sex and had the mean age ± SD = 45.9 ± 17.1 years, range of 24.4–69.5 years, no difference between men and women: t (18) = −0.81, P = 0.43.

Study Design

The MRI data were collected in a single session divided into two parts (see Fig. 1 for illustration). The first part was devoted to acquiring a 3D volume of structural T_1-weighted and T_2 -weighted MRI images and a ME-T_2 relaxation images followed by repeating only the acquisition of the ME-T_2 relaxation images, performed back-to-back, without repositioning the participant in the MRI scanner. The participants were then removed from the scanner and given a 5-min break, following which they were repositioned in the scanner for part 2 of the session that included a repeat of the T_1-weighted, T_2-weighted and ME-T_2 relaxation MRI sequences. All images were aligned at acquisition along the anterior–posterior commissure (AC-PC line). Study design is outlined in Figure 1.

MRI Acquisition Protocol

The data were collected on a 3T Siemens MAGNETOM Verio^{TM} MRI system equipped with a 12-channel volume head coil. The T_1-weighted images were acquired in the axial plane with isotropic voxels (1 mm^3) using the magnetization prepared gradient-echo (MPRAGE) sequence with a repetition time (TR) = 2,400 ms, echo time (TE) = 2.63 ms, flip angle (FA) = 8°, inversion time (TI) = 1,100 ms, matrix size = 256 × 256, number of slices = 160; GRAPPA factor = 2; acquisition time (TA) = 8:07 min. The T_2-weighted sequence was also collected in the axial plane with isotropic voxel (1 mm^3) using the sampling perfection with application-optimized contrast with different flip-angle evolutions (SPACE) sequence with a TR = 3,200 ms, TE = 449 ms, echo space = 3.52 ms, turbo factor = 141, matrix size = 250 × 250, number of slices = 176 and GRAPPA factor = 2; TA = 4:43 min. ME-T_2 relaxation images were acquired in the axial plane using the 3D gradient and spin echo (GRASE) sequence, developed by Dr. Jongho Lee, Seoul National University, Republic of Korea). Acquisition parameters for the GRASE sequence were as follows: TR = 1,100 ms, number of echoes = 32, first echo = 11 ms, inter-echo spacing = 11 ms, FOV = 190 × 220 mm^2, matrix size = 165 × 192, slice thickness = 5 mm, number of slices = 24, slice oversampling = 0, in-plane resolution = 1.1 × 1.1 mm^2; TA = 17 min. Parallel acquisition was not available for the GRASE sequence.

Processing of the ME-T_2 Data

To generate ME-T_2 indices, MWF and the geomT_{2IEW}, for each region of interest (ROI) in the subjects space, we mapped ROIs from standard space (MNI152) to subject space followed by the voxel-wise ME-T_2 relaxation analysis for each ROI. ME-T_2 relaxation imaging data were analyzed using a combination of FMRIB Software Library (FSL), in-house Linux shell scripts and MATLAB (MathWorks, Natick, MA) programs. Each dataset was interpolated to 2.5-mm thickness and coregistered to the structural T_1-weighted volume from part 1 using the FSL FLIRT with 6 degrees of freedom [Jenkinson and Smith, 2001; Jenkinson et al., 2002]. The procedure to generate the ROIs will be described in a separate section below and additional details are available in the previous publication [Arshad et al., 2016].

Once the ROI masks were in the subject space, they were applied to the multi-echo data. Voxel-wise ME-T_2 relaxation analysis was conducted for each ROI using a regularized non-negative least squares (rNNLS) algorithm [Whittall et al., 2002]. The optimal value for the regularization parameter was determined via Generalized Cross Validation [Golub et al., 1979]. Nonideal refocusing flip angles were accounted for in the fitting procedure using the extended phase graph (EPG) algorithm as part of the rNNLS fitting [Hennig, 1988; Prasloski et al., 2012]. T_2 distributions were generated using 200 logarithmically spaced T_2 relaxation values ranging from 10 to 2,000 ms. The MWF was calculated as the discrete integral of T_2 relaxation times between 10 and 40 ms and between 40 and 200 ms for the intracellular/extracellular water signal. MWF values were normalized to the discrete integral of T_2 relaxation times between 10 and 2,000 ms, which reflected the total observed water signal. The geomT_{2IEW} was calculated on the logarithmic scale [Whittal et al., 2002].

Processing of the Calibrated T_1w/T_2w Ratio Imaging Data

Similar to the ME-T_2 analysis, the calibrated T_1w/T_2w ratio values were generated in the subjects' space for each ROI. Because the INU differs for T_1- and T_2-weighted images, we corrected both images independently and used the "calibration" method proposed by Ganzetti et al. to account for intensity scaling differences across both subjects and within subject (due to repositioning). The T_1- and T_2-weighted images were first skull-stripped using the FSL BET tool [Smith, 2002], with a value of 0.3 for the fractional intensity threshold, and binary brain masks were saved. Skull-stripped images were co-registered to the T_1-weighted image from part 1 using FLIRT 6 degrees of freedom and the transformation matrices were saved. Coregistration between T_1- and T_2 -weighted images were conducted using mutual information cost function in FLIRT. The saved transformation matrices, which were previously generated, were applied to the brain masks and to the non-deskulled images. After application of the transformation matrix to the brain masks, they were binarized. Each non-deskulled image was corrected for INU using the nonuniform intensity normalization (N3) algorithm [Sled and Zijdenbos, 1998]. Brain masks were provided to the N3 along with the non-deskulled images. Although the N3 method is widely used for INU correction [Zheng et al., 2009], its default parameters are optimized for the 1.5 T field strength. Previous studies have shown that setting of the parameter controlling the smoothness of the bias field at the default value of 200 mm is not optimal for 3T scanners [Zheng et al., 2009]. We used a value of 100 mm for the smoothness of the bias field because the wavelength would be expected to decrease by a factor of 2 when going from 1.5 to 3T field strength.

Following INU correction we "calibrated" the images using the linear transformation method proposed by Ganzetti et al. [2014]. The "calibration" requires intensity values from the eye and temporalis muscle from both the T_1- and T_2-weighted images. These regions were defined in the MNI152 space and mapped to the subjects' space. The procedure to map the eye and temporalis muscle from the template to subject space is described below. The masks were then applied to the T_1- and T_2-weighted images. Histograms for each mask were generated and intensity values below the 10th and greater than the 90th percentile were removed, followed by calculating the median intensity value. These values serve as the person-specific values for the eye and temporalis muscle for the T_1- and T_2-weighted images. Similar to Ganzetti et al. [2014], we used the following reference values for the eye mask: 28.2 arbitrary units for T_1-weighted image and 99.9for the T_2-weighted image. For the temporalis muscle mask the reference values were: 58.6 for the T_1-weighted image and 21.1 for the T_2-weighted image. Ratio images were calculated after calibration.

Regional Parcellation and Postprocessing

Regions of interest (ROIs), were defined in FSL [Jenkinson et al., 2002] using the Johns Hopkins University (JHU) and ICBM-DTI-81 white matter atlas [Hua et al., 2008; Wakana et al., 2007], in template space, and included two midline commissural tracts: genu (genu CC) and splenium (splenium CC) of the corpus callosum; two right and left association tracts: superior longitudinal fasciculus (SLF) and the inferior fronto-occipital fasciculus (IFOF); and two right and left projection tracts: anterior (ALIC) and the posterior (PLIC) limbs of the internal capsule. The overarching goal was to convert these ROIs from the template space into the subjects' space and then apply these masks to the ME-T_2 and T_1w/T_2w images. Because we compare ME-T_2 indices to the calibrated T_1w/T_2w ratio, we wanted to ensure that the ROIs used for both analyses were identical.

To generate the warp field for mapping from the template to the subjects' space, we used the FSL nonlinear registration tool FNIRT to register the T_1-weighted image from part 1 to the MNI152 template. The resultant warp field was saved, inverted and used to map all ROIs, including the eye and temporalis muscle masks, from template to subject space. To reduce partial voluming artifacts, we first segmented the T_1-weighted image into white matter, gray matter and cerebrospinal fluid using the FSL tool FAST [Zhang et al., 2001]. We then applied the masks to the segmented white matter and set a threshold of 0.95 followed by binarizing the masks. This ensured that our ROI masks consisted of white matter with a probability of 0.95 or greater. These masks were applied to INU corrected and "calibrated" T_1w/T_2w images to yield T_1w/T_2w ratios for each ROI.

For the ME-T_2 analysis, these masks were downsampled to 2.5 mm thickness and applied to ME-T_2 data. The MWF and $geomT_{2IEW}$ were averaged within an ROI. To minimize rounding errors, MWF values were multiplied by 100, resulting in units of percentage for MWF. The $geomT_{2IEW}$ values were multiplied by 1,000, giving units of milliseconds (ms). The T_1w/T_2w ratios were multiplied by 100 and are unit-less. Bilateral ROI values were averaged, yielding single measures, and thus reducing the number of comparisons. All ROIs in the ME-T_2 and T_1w/T_2w datasets were inspected for gross registration errors.

Statistical Analyses

Reliability was quantified by the intraclass correlation coefficient, namely the formula assuming random effects, ICC(1,1) [Shrout and Fleiss, 1979]. According to Shrout and Fleiss [1979], there are three cases of intraclass correlations applicable to assessment of reliability. Case 3 that assumes that each method or rater or, in our study, scanning run are fixed and the results can be generalized only if the identical runs are repeated. This assumption is unrealistic and precludes generalizing to other runs on other MRI scanners and on other days. Case 1 assumes that each run is different, and is drawn from a set of all possible

TABLE I. Reliability of ME-T_2 and ratio indices on back-to-back and repositioning scans: Intraclass correlations (top line) with 95% confidence intervals (bottom line)

	MWF (%)		$geomT_{2\text{ - IEW}}$ (ms)		T_1w/T_2w Ratio
ROI	Back-to-Back	Repositioning	Back-to-Back	Repositioning	Repositioning
ALIC	0.94 0.85–0.97	0.83 0.64−0.91	0.98 0.92–0.99	0.93 0.75−0.97	0.80 0.53, 0.91
PLIC	0.90 0.74–0.96	0.86 0.75−0.92	0.95 0.84–0.98	0.94 0.81−0.98	0.76 0.43, 0.92
Genu CC	0.94 0.87–0.97	0.83 0.64−0.93	0.99 0.98–0.99	0.99 0.96–0.99	0.89 0.69, 0.97
Splenium CC	0.95 0.89–0.98	0.88 **0.77**, 0.94	0.98 0.94–0.99	0.97 0.92−0.98	0.90 0.66, 0.97
SLF	0.95 0.90–0.97	0.79[a] **0.55**, 0.93	0.98 0.90–0.99	0.96 0.75−0.99	0.89 0.69, 0.98
IFOF	0.95 0.87–0.98	0.81 **0.60**, 0.91	0.96 0.93–0.98	0.90 0.80−0.97	0.90 0.75, 0.96
Aggregated ROIs	0.97 0.96–0.98	0.93 0.90, 0.95	0.99 0.98–1.0	0.98 0.98–0.99	0.91 0.86, 0.94

Note. ALIC – anterior limb of the internal capsule; PLIC – posterior limb of the internal capsule; Genu CC – genu of the corpus callosum; Splenium CC – splenium of the corpus callosum; SLF – superior longitudinal fasciculus; IFOF – inferior fronto-occipital fasciculus; ROI – region of interest.
[a]Dropping one subject increases the ICC value to 0.84.

runs at random. A run being application, in a randomly determined scheduling time, of a certain scanner configuration and scanner hardware state to produce imaging data is presumed randomly selected from a set of MRI scanners and their hardware/software configurations available for imaging. Hence, our selection of ICC(1,1) formula for comparing back-to-back runs for the ME-T_2 data. For assessing the effect of participant repositioning in the scanner, we compared Runs 2 and 3 for the ME-T_2 data and Runs 1 and 2 for the T_1w/T_2w data (across part 1 and 2 for both datasets). In this comparison, participant's deviation on Run 3 from their position in the scanner on Run 2 was assumed random and ICC(1,1) was selected again.

Bootstrapped 95% confidence intervals, using 5,000 samples, were generated for the analysis on the ME-T_2 indices and T_1w/T_2w data. ICC values ≥ 0.80 were considered reliable. These analyses were conducted by combining all ROIs and for each ROI independently. ICC values along with 95% confidence intervals are reported for all ROIs combined as well as individual ROIs. Finally, the association between the T_1w/T_2w ratio and the ME-T_2 indices (MWF and $geomT_{2IEW}$) were evaluated using linear regression. ROIs from the T_1w/T_2w data were averaged across part 1 and 2 while the ME-T_2 ROIs were averaged across Runs 2 and 3. These runs were chosen, for the ME-T_2 data, because they were collected with repositioning of the participant in between, similar to the T_1w/T_2w data. The averaged values were used in the linear regression. All statistical analyses were conducted using MATLAB version 2012a (MathWorks, Natick, MA).

RESULTS

Reliability of Indices Derived from the ME-T_2 Relaxation Imaging

For both indices, the individual regions' ICC values exceeded the set threshold of reliability, except for the SLF whose MWF ICC value was 0.79 after repositioning. However, dropping one observation with an extreme MWF value increased the ICC to 0.84. Moreover, the overlapping 95% confidence intervals, suggest uniform reliability across the sampled regions, see Table I. With all ROIs aggregated into a single global region, both MWF and $geomT_{2IEW}$ yielded ICC values > 0.80 for the back-to-back run and after repositioning.

Reliability of Calibrated T_1w/T_2w Ratio Imaging

Two participants had wrap-around artifacts on their T_2-weighted images, and one participant had motion artifacts on the T_1-weighted images. These participants were excluded from T_1w/T_2w analyses, which were thus conducted on $N = 17$. When all ROIs were combined, the T_1w/T_2w ratio met our reliability criterion. However, the regional analysis revealed that for PLIC, reliability fell somewhat short of the set level of 0.80 (see Table I).

Associations between ME-T_2 Indices and T_1w/T_2w Ratio

The bivariate distributions of ME-T_2 indices and T_1w/T_2w ratio, and regression lines for aggregated and separate ROIs as well as bivariate correlations between each pair of indices are presented in Figure 2.

MWF versus T_1w/T_2w ratio

The correlations between regional MWF and T_1w/T_2w ratios were *positive* and ranged between 0.21 and 0.65 (median $r = 0.44$), with only the ALIC reaching the adjusted significance level, and the Genu reaching the unadjusted significance level. However, when all ROIs are combined, the correlation between MWF and T_1w/T_2w ratio was *negative:* $r = -0.26$, $P < 0.05$. Significance of the

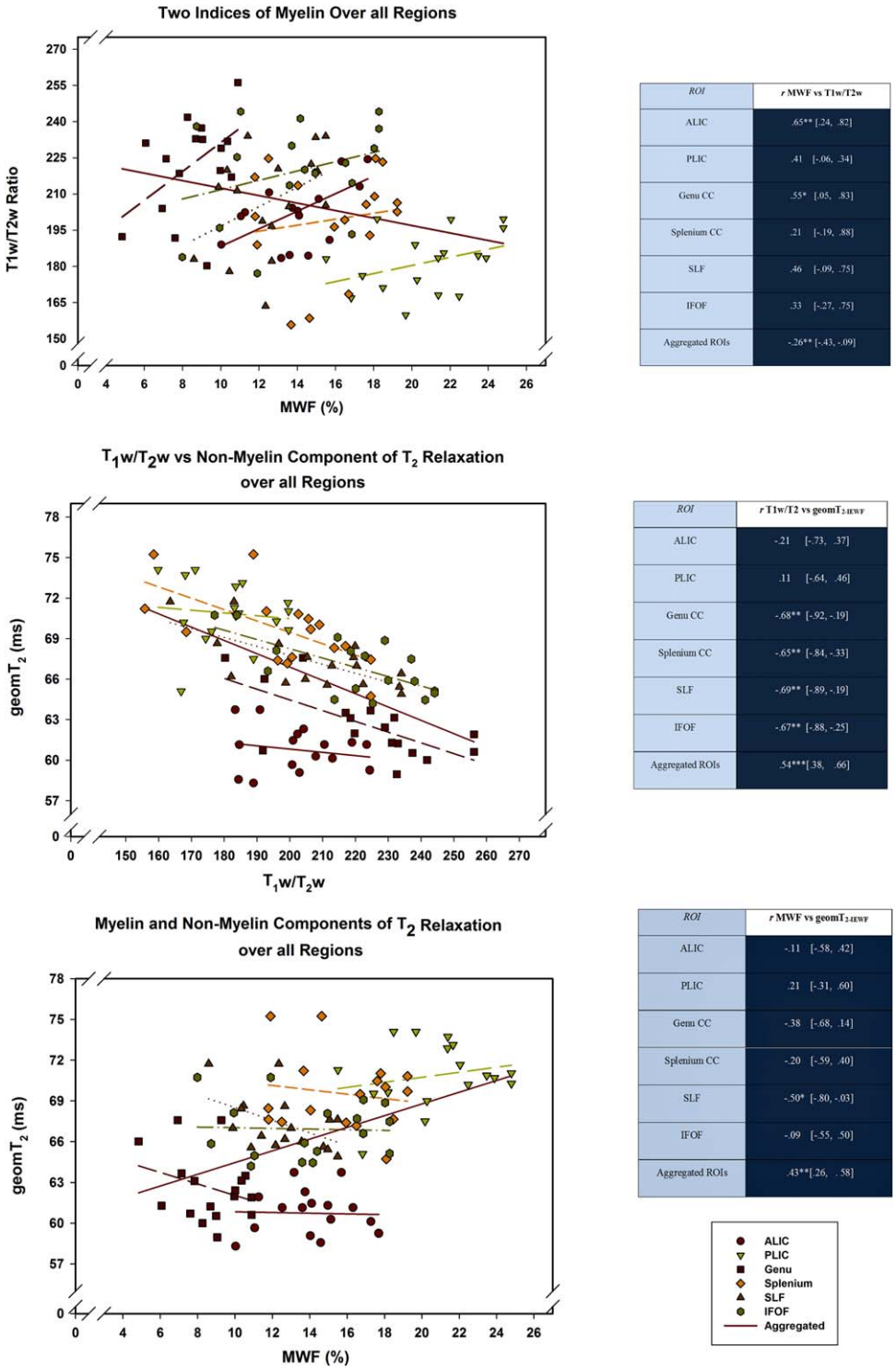

Figure 2.

Plots of associations between Myelin Water Fraction, $geomT_{2iew}$ and T_1w/T_2w ratio (multiplied by 100) for all regions of interest, and their aggregation. Values are based on the mean of the right and left side except for the genu and splenium. $^*P < 0.05$; $^{**}P < 0.008$. [Color figure can be viewed at wileyonlinelibrary.com]

sign reversal in the correlation between two putative indices of myelin is discussed below.

T_1w/T_2w ratio versus geomT_{2IEW}

The correlations between geomT_{2IEW} and calibrated T_1w/T_2w ranged between −0.11 and −0.69 (median $r=-0.66$, see Fig. 2). Regional correlations were significant at the adjusted level of significance for the Genu and Splenium, SLF and IFOF. Similar to the regional correlations, when all ROIs were combined we found a statistically significant correlation, with the negative sign retained ($r=-0.54$, $P<0.05$).

MWF and geomT_{2IEW}

The correlations between MWF and geomT_{2IEW} ranged between −0.50 and 0.21 (median $r=-0.16$), with all but one being negative (see Fig. 2). Regional correlations were significant at the unadjusted level of significance only for the SLF. As in the case of associations between MWF and T_1w/T_2w, for aggregated ROIs we found a statically significant correlation, with the sign reversed to positive: $r=0.43$, $P<0.05$).

DISCUSSION

ME-T_2 Derived Indices of Myelin Content and Intracellular/Extracellular Water Are Reliable

The indices of myelin content (MWF) and intracellular/extracellular water compartment (geomT_{2IEW}) derived from ME-T_2 imaging showed adequate reliability and stability making them suitable for longitudinal studies. Of note, MWF and geomT_{2IEW} were equally reliable across the ROIs investigated. This uniformity mitigates concerns about differential unreliability that could threaten the validity of conclusions based on differences among associations between regional myelin content and variables of interest such as age. Given the wide confidence intervals this conclusion must be qualified by a caveat that in a larger sample size small differences in reliability across the tracts could be detected. Nonetheless, our study confirms high reliability of ME-T_2 indices (and greater reliability of geomT_{2IEW} compared to MWF) in a much larger sample and wider age range than previously reported [Levesque et al., 2010; Meyers et al., 2009, 2013; Vavasour et al., 2006].

Reliability of Calibrated T_1w/T_2w Imaging

We found that across the white matter tracts investigated here, calibrated T_1w/T_2w ratio had high test–retest reliability. Although for one region (PLIC) the ICC did not meet the set reliability criterion of 0.80, the overlapping 95% confidence intervals suggest that we do not have differential reliability across ROIs. Therefore, calibrated T_1w/T_2w ratio is of sufficient reliability in white matter tracts to make it useful in longitudinal studies.

Validity of Calibrated T_1w/T_2w Ratio as a Proxy for Myelin Content

Reliability is the upper limit of validity and having reliable measures constitutes a necessary but not a sufficient condition for producing valid estimates of myelin content. With reliability of the examined indices of myelin established, we now address the issue of validity. Although T_1w/T_2w imaging has multiple advantages—high spatial resolution (1 mm^3), reasonably straightforward processing of the images for calibration and no need for custom MRI acquisition sequences, it is unclear whether it yields a valid proxy for myelination, that is, has good concurrent validity. Although ME-T_2-based calculation of MWF is hampered by a smaller signal (and hence lower spatial resolution) relative to T_1w/T_2w, and requires special acquisition sequences followed by more complex modeling, it has been histologically validated as an index of myelin content in cortical and subcortical white matter tracts [Laule et al., 2006; McCreary et al., 2009; Webb et al., 2003]. Finding low to moderate correlations between MWF and T_1w/T_2w with none of them accounting for even 50% of the intermethod variance suggests that at least within the examined regions, the validity of T_1w/T_2w is unsatisfactory. In other words, at least in subcortical white matter tracts, the T_1w/T_2w ratio is a poor proxy for myelination in spite of its high reliability.

Moreover, T_1w/T_2w showed closer association with the other index derived from ME-T_2 imaging, geomT_{2IEW}, which is not expected to be directly related to myelin content. In four of the examined six ROIs, T_1w/T_2w was associated with the geomT_{2IEW} rather than with MWF, a measure of myelin content. Given the established high reliability of each index, low correlations of T_1w/T_2w with MWF combined with its higher correlations with an index that does not represent myelin content suggest that T_1w/T_2w measures something else rather than myelin content. What are the properties of the white matter that are tapped by T_1w/T_2w remains to be discovered. We can only offer a speculative account of its possible substrates.

The short T_2 relaxation component (~10–40 ms at 3T) that is used to compute MWF reflects highly restricted mobility of water molecules trapped between the myelin sheaths. The number of myelin lamina (and thus myelin thickness) is related to axonal diameter in an almost linear fashion [Walhovd et al., 2014], and with increasing number of myelin wraps around axons, MWF also rises. The water signal associated with the intermediate T_2 relaxation component, geomT_{2IEW}, corresponds to less restricted water molecules in the intracellular and extracellular space and thus is not a faithful representative of myelin content. However, two factors should be considered in discussing

the neurobiological meaning of the geomT_{2IEW} value: axon size and axon packing density.

As the axon caliber increases proportionately to thickness of its myelin sheath, the intracellular space expands and subsequent increase in the geomT_{2IEW} ensues. Therefore, white matter tracts that contain axons of larger caliber are expected to yield longer geomT_{2IEW} values. Indeed, PLIC, a region that contains some of the largest axons originating from and projecting to somatosensory and motor regions [Tomasi et al., 2012] evidenced (along with splenium) the longest geomT_{2IEW} in our sample. In rat spinal cord, larger axonal diameter was associated with longer geomT_{2IEW} values [Dula et al., 2010]. Thus, in a given region, geomT_{2IEW} is positively correlated with axonal caliber.

The signal from subcortical white matter tracts is composed of contributions from water trapped between the myelin sheaths and water located in the intracellular and extracellular space. Although the multi-T_2 relaxation components differentiate between these distinct water environments, the measured signal from T_1- and T_2-weighted images does not allow such intercompartmental resolution. Therefore, while myelin may influence the T_1w/T_2w ratio to some extent, the magnitude of this association is rather modest and inconsistent across the white matter tracts. Given that the typical TE time for a T_2w sequence used in T_1w/T_2w ratio imaging is relatively long, the signal contrast from the T_2-weighted images cannot be influenced by the short T_2 relaxation component attributable to the water between the myelin sheaths. Therefore, this suggests that the IE compartment is the primary source of the observed signal on the T_2-weighted image. This is consistent with the observed relatively higher correlation of T_1w/T_2w ratios with the geomT_{2IEW} rather than with MWF.

Correlations between T_1w/T_2w ratio with geomT_{2IEW} in four out of six examined regions, including two of the commissural tracts, genu and splenium may reflect neurobiological structural differences among these regions. Along corpus callosum, from genu to splenium, axon packing density gradually declines and axon caliber increases [Aboitiz et al., 1992; De Santis et al., 2016; LaMantia and Rakic, 1990; Riise and Pakkenberg, 2011]. As splenium has more myelin compared to genu, if T_1w/T_2w ratio were indeed a proxy for myelin, it would be expected to be greater in the former compared to the latter. Contrary to that expectation, we find a smaller T_1w/T_2w ratio in the splenium—a result that is expected if the axon density/diameter rather than myelin is the dominant influence on that index. Similar results were found in a post-mortem study that measured myelin content and the T_1w/T_2w ratio along the corpus callosum [Sandrone et al., 2015]. Thus, T_1w/T_2w ratio may reflect variation in caliber and packing density of the axons more than myelin content. This conclusion is in line with a recent observation of larger-caliber axons having longer T1 thus suggesting that the T_1w/T_2w ratio may be sensitive to variations in axonal diameter [Harkins et al., 2015].

In examining the bivariate associations between the white matter measures, we noted a curious reversal of sign in associations of MWF with T_1w/T_2w ratio and geomT_{2IEW} that occurred when the data were analyzed by individual ROIs versus their sum. Such discrepancy between the associations observed in aggregated data versus findings for specific subgroups or regions is known as Simpson paradox, which strictly speaking is not a paradox at all but a reflection of regional data structure that is obscured by aggregation [Kievit et al., 2013; Simpson, 1951]. Nevertheless, in the case of associations between the two indices of myelin content, the observed pattern may reflect neurobiological reality and reinforce the conjecture that T_1w/T_2w ratio reflects mainly axonal caliber and packing density rather than myelin content.

When the analysis is restricted to a specific homogeneous region, the variation in the axonal diameters between subjects maybe smaller than the differences in myelin content and therefore the T_1w/T_2w ratio is influenced by myelin to a greater extent, hence a positive (although still small) correlation. However, when all ROIs are combined, the variation in the axonal diameters across the tracts increases. Regions containing axons of larger caliber have a longer geomT_{2IEW}, (and a smaller T_1w/T_2w ratio), and this effect seems to overpower the effect of myelin content on the T_1w/T_2w ratio, hence resulting in a negative correlation. In sum, strong circumstantial evidence indicates that the T_1w/T_2w ratio is influenced by axonal diameter and packing density. Because the goal of this paper was not to investigate the relationship between the axonal diameter and the T_1w/T_2w ratio, we do not have independent measures of that property. Nonetheless, these promising results suggest that future investigations should combine more advanced diffusion imaging techniques, capable of estimating axonal diameter, with the T_1w/T_2w imaging to directly test this hypothesis.

The findings described here may have a significant impact on understanding of white matter properties under pinning cognitive performance and individual differences therein. Myelin loss has been hypothesized as a potential substrate of reduced processing speed in both normal aging and in neurodegenerative diseases such as multiple sclerosis. Reliable assessment of individual differences in myelin content and changes over time via MWF will allow a more focused evaluation of the myelination-speed of processing relationship than previously reported. If both the geomT_{2IEW} and the T_1w/T_2w ratio are indeed more sensitive to axonal diameter and axonal density than they are to myelin content, their role in processing speed and other cognitive functions can be evaluated and dissociated from that of myelin. There potential applications should provide impetus to further exploration of diverse aspects of white matter structure via assessment of distinct components of the T_2 relaxation.

Limitations

First, because of time constraints, we could not add another run before repositioning and compare Runs 3 and

4 (repositioned participant) instead of comparing Runs 2 and 3 (repositioned participant). Run 2 being already a part of the back-to-back runs comparison introduces dependence in estimation of repositioning effects.

Second, although we show the discrepancies between MWF and the T_1w/T_2w ratio in evaluating subcortical white matter, we did not investigate this relationship in the cortex. The myelin content in cortical gray matter is too low for reliable detection by current in vivo ME-T_2 imaging. A recent publication observed the "inverted U" relationship between age and the T_1w/T_2w ratio in the cortex [Grydeland et al., 2013], in accord with post mortem literature [Kaes, 1907] and may be in line with the hypothesized role of the T_1w/T_2w ratio as myelin content index. Therefore, it is imperative to evaluate the relationship between the T_1w/T_2w ratio and cortical myelin content with histological methods and improved ME-T_2 imaging.

Third, potential effects of iron on the estimates of myelin content may be considered. Such influence, however, is highly unlikely, for even though oligodendrocytes are indeed significant repositories of iron [Connor and Benkovic,1992], the refocusing pulses of the ME-T_2 acquisition are insensitive to any parametric effects of ferric iron [Chavhan et al., 2009]. Therefore, unlike gradient-echo acquisition schemes [Oh et al., 2013], the modeling of the multiple T_2 relaxation components of ME-T_2 data is not affected by iron composition, as confirmed by ex vivo studies [Li et al., 2009]. Nonetheless, new T2*-based methods may offer improvements in MWF mapping while overcoming the abovementioned limitations [Lee et al., 2016].

CONCLUSIONS

Both MWF and geomT_{2IEW} indices produce very reliable measurements over multiple white matter areas. Thus, ME-T_2 imaging is suitable for in vivo longitudinal investigations of myelin content. In contrast, while also showing high and regionally uniform reliability, the T_1w/T_2w ratio revealed limited validity as an index of myelin content. It does not appear a sufficiently specific index of subcortical myelin content as it appears to correlate more with the intermediate T_2 relaxation component that reflects axonal caliber and density than with the short T_2 relaxation component that represents water trapped in myelin.

DISCLOSE STATEMENT

The authors of this publication have no conflicts to disclose.

ACKNOWLEDGMENTS

We gratefully acknowledge generous help of Dr. Jongho Lee (Laboratory for Imaging Science and Technology, Department of Electrical and Computer Engineering, Seoul National University, Republic of Korea) who developed the 3D GRASE sequence used in this study.

REFERENCES

Aboitiz F, Scheibel AB, Fisher RS, Zaidel E (1992): Fiber composition of the human corpus callosum. Brain Res 598:143–153.

Alonso-Ortiz E, Levesque IR, Pike GB (2015): MRI-based myelin water imaging: A technical review. Magn Reson Med 73:70–81.

Ansari KA, Loch J (1975): Decreased myelin basic protein content of the aged human brain. Neurology 25:1045–1050.

Arshad M, Stanley JA, Raz N (2016): Adult age differences in subcortical myelin content are consistent with protracted myelination and unrelated to Diffusion Tensor Imaging indices 143:26–39.

Baltes PB, Nesselroade, JR (1979): History and rational of longitudinal research. In: Nesselroade JR, Baltes PB, editors. Longitudinal research in the study of behavior and development. New York: Academic Press. pp 1–39.

Barkovich AJ, Kjos BO, Jackson DE Jr, Norman D (1988): Normal maturation of the neonatal and infant brain: MR imaging at 1.5 T. Radiology 166:173–180.

Bartzokis G (2004): Age-related myelin breakdown: A developmental model of cognitive decline and Alzheimer's disease. Neurobiol Aging 25:5–18.

Billiet T, Vandenbulcke M, Mädler B, Peeters R, Dhollander T, Zhang H, Deprez S, Van den Bergh BR, Sunaert S, Emsell L (2015): Age-related microstructural differences quantified using myelin water imaging and advanced diffusion MRI. Neurobiol Aging 36:2107–2121.

Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM (2009): Principles, techniques, and applications of T2*-based MR imaging and its special applications. Radiographics 29:1433–1449.

Connor JR, Benkovic SA (1992): Iron regulation in the brain: Histochemical, biochemical, and molecular considerations. Ann Neurol 32 Suppl:S51–S61.

Cronbach LJ, Meehl PE (1955): Construct validity in psychological tests. Psychol Bull 52:281–302.

Curnes JT, Burger PC, Djang WT, Boyko OB (1988): MR imaging of compact white matter pathways. AJNR Am J Neuroradiol 9: 1061–1068.

Davison AN, Dobbing J (1966): Myelination as a vulnerable period in brain development. Br Med Bull 22:40–44.

De Santis S, Jones DK, Roebroeck A (2016): Including diffusion time dependence in the extra-axonal space improves in vivo estimates of axonal diameter and density in human white matter. Neuroimage 130:91–103.

Dula AN, Gochber DF, Valentine HL, Valentine WM, Does MD (2010): Multi-exponential T2, magnetization transfer and quantitative histology in white matter tracts of rat spinal cord. Magn Reon Med 63:902–909.

Fjell AM, McEvoy L, Holland D, Dale AM, Walhovd KB; Alzheimer's Disease Neuroimaging Initiative (2014): What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. Prog Neurobiol 117:20–40.

Ganzetti M, Wenderoth N, Mantini D (2014): Whole brain myelin mapping using T1- and- T2 weighted MR imaging data. Front Hum Neurosci 8:671.

Ganzetti M, Wenderoth N, Mantini D (2016): Quantitative evaluation of intensity inhomogeneity correction methods for structural MR brain images. Neuroinformatics 14:5–21.

Glasser MF, Van Essen DC (2011): Mapping human cortical areas in vivo based on myelin content as revealed by T1-and-T2 weighted MRI. J Neurosci 31:11597–11616.

Glasser MF, Goyal MS, Preuss TM, Raichle ME, Van Essen DC (2014): Trends and properties of human cerebral cortex: Correlations with cortical myelin content. Neuroimage 93:165–175.

Golub GH, Heath M, Wahba G (1979): Generalized cross validation as a method for choosing a good ridge parameter. Technometrics 21:215–223.

Grydeland H, Walhovd KB, Tamnes CK, Westlye LT, Fjell AM (2013): Intracortical myelin links with performance variability across the human lifespan: Results from T1- and T2- weighted MRI myelin mapping and diffusion tensor imaging. J Neurosci 33:18618–18630.

Harkins KD, Xu J, Dula AN, Li K, Valentine WM, Gochberg DF, Gore JC, Does MD (2015): The microstructural correlates of T1 in white matter. Magn Reson Med 75:1341–1345.

Hennig J (1988): Multiecho imaging sequences with low refocusing flip angles. J Magn Reson 78:397–407.

de Hoz L, Simons M (2015): The emerging functions of oligodendrocytes in regulating neuronal network behaviour. Bioessays 37:60–69.

Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PC, Mori S (2008): Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. NeuroImage 39:336–347.

Jebelli J, Su W, Hopkins S, Pocock J, Garden GA (2015): Glia: Guardians, gluttons, or guides for the maintenance of neuronal connectivity?. Ann N Y Acad Sci 1351:1–10.

Jenkinson M, Smith S (2001): A global optimization method for robust affine registration of brain images. Med Image Anal 5: 143–156.

Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Kaes T. (1907). Die Grosshirnrinde des Menschen in ihren Massen und in ihrem Fasergehalt. Ein Gehirnanatomischer Atlas. Jena: Gustav Fischer.

Kievit RA, Frankenhuis WE, Waldorp LJ, Borsboom D (2013): Simpson's paradox in psychological science: A practical guide. Front Psychol 4:1–14.

Kramer HC, Yesavage JA, Taylor JL, Kupfer D (2000): How can we learn about developmental processes from cross-sectional studies, or can we? Am J Psychiatry 157:163–171.

Lamantia AS, Rakic P (1990): Cytological and quantitative characteristics of four cerebral commisures in the rhesus monkey. J Comp Neurol 291:520–537.

Laule C, Leung E, Li DKB, Traboulsee AL, Paty DW, MacKay AL, Moore GRW (2006): Myelin water imaging in multiple sclerosis: Quantitative correlations with histopathology. Mult Scler 12:747–753.

Lee J, Nam Y, Choi JY, Kim EY, Oh SH, Kim DH (2016): Mechanisms of T2* anisotropy and gradient echo myelin water imaging. NMR Biomed. doi: 10.1002/nbm.3513. [Epub ahead of print]

Levesque IR, Chia CL, Pike GB (2010): Reproducibility of in vivo magnetic resonance imaging-based measurement of myelin water. J Magn Reson Imaging 32:60–68.

Li TQ, Yao B, van Gelderen P, Merkle H, Dodd S, Talagala L, Koretsky AP, Duyn J (2009): Characterization of T_2* heterogeneity in human brain white matter. Magn Reson Med 62: 1652–1657.

Lindenberger U, von Oertzen T, Ghisletta P, Hertzog C (2011): Cross-sectional age variance extraction: What's change got to do with it?. Psychol Aging 26:34–47.

Lu PH, Lee GJ, Tishler TA, Meghpara M, Thompson PM, Bartzokis G (2013): Myelin breakdown mediates age-related slowing in cognitive processing speed in healthy elderly men. Brain Cogn 81:131–138.

MacKay A, Whittall K, Adler J, Li D, Paty D, Graeb D (1994): *In vivo* visualization of myelin water in brain by magnetic resonance. Magn Reson Med 31:673–677.

McCreary CR, Bjarnason TA, Skihar V, Mitchell JR, Yong VW, Dunn JF (2009): Multiexponential T2 and magnetization transfer MRI of demyelination and remyelination in murine spinal cord. Neuroimage 45:1173–1182.

Menon RS, Rusinko MS, Allen PS (1992): Proton relaxation studies of water compartmentalization in a model neurological system. Magn Reson Med 28:264–274.

Meyers SM, Laule C, Vavasour IM, Kolind SH, Mädler B, Tam R, Traboulsee AL, Lee J, Li DK, MacKay AL (2009): Reproducibility of myelin water fraction analysis: A comparison of region of interest and voxel-based analysis methods. Magn Reson Imaging 27:1096–1103.

Meyers SM, Vavasour IM, Mädler B, Harris T, Fu E, Li DK, Traboulsee AL, MacKay AL, Laule C (2013): Multicenter measurements of myelin water fraction and geometric mean T2: Intra- and intersite reproducibility. J Magn Reson Imaging 38: 1445–1453.

Oh SH, Bilello M, Schindler M, Markowitz CE, Detre JA, Lee J (2013): Direct visualization of short transverse relaxation time component (ViSTa). Neuroimage 83:485–492.

Prasloski T, Mädler B, Xiang QS, MacKay AL, Jones C (2012): Applications of stimulated echo correction to multicomponent T2 analysis. Magn Reson Med 67:1803–1814.

Raz N, Rodrigue KM (2006): Differential aging of the brain: Patterns, cognitive correlates and modifiers. Neurosci Biobehav Rev 30:730–748.

Riise J, Pakkenberg B (2011): Stereological estimation of the total number of myelinated callosal fibers in human subjects. J Anat 218:277–284.

Saab AS, Tzvetanova ID, Nave KA (2013): The role of myelin and oligodendrocytes in axonal energy metabolism. Curr Opin Neurobiol 23:1065–1072.

Sandrone S, Thiebaut de Schotten M, Reimann K, Murphy D, Geyer S, Catani M, Dell'Acqua. (2015): T1-weighted/T2-weighted MRI myelin mapping does not map myelin in the human corpus callosum. In: Poster presented at the Organization of Human Brain Mapping, Honolulu, Hawaii.

Shafee R, Buckner RL, Fischl B (2015): Gray matter myelination of 1555 human brains using partial volume corrected MRI images. Neuroimage 105:473–485.

Shrout PE, Fleiss JL (1979): Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86:420–428.

Simpson EH (1951): The interpretation of interaction in contingency tables. J Royal Statist Soc Ser B 13:238–241.

Sled JG, Zijdenbos AP (1998): A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17:87–97.

Smith SM (2002): Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Tomasi S, Caminiti R, Innocenti GM (2012): Areal differences in diameter and length of corticofugal projections. Cereb Cortex 22:1463–1472.

Vavasour IM, Clark CM, Li DK, Mackay AL (2006): Reproducibility and reliability of MR measurements in white matter: Clinical implications. NeuroImage 32:637–642.

Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, Hua K, Zhang J, Jiang H, Dubey P, Blitz A, van Zijl P, Mori S (2007): Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage 36: 630–644.

Walhovd KB, Johansen-Berg H, Káradóttir RT (2014): Unraveling the secrets of white matter–bridging the gap between cellular, animal and human imaging studies. Neuroscience 276:2–13.

Webb S, Munro CA, Midha R, Stanisz GJ (2003): Is multicomponent T2 a good measure of myelin content in peripheral nerve? Magn Reson Med 49:638–645.

Whittal KP, MacKay AL, Graeb DA, Nugent RA, Li DKB, Paty DW (1997): In vivo measurement of T2 distributions and water content in normal human brain. Magn Reson Med 37:34–43.

Whittall KP, MacKay AL, Li DK, Vavasour IM, Jones CK, Paty DW (2002): Normal- appearing white matter in multiple sclerosis has heterogeneous, diffusely prolonged T2. Magn Reson Med 47:403–408.

Zhang Y, Brady M, Smith S (2001): Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20:45–57.

Zheng W, Chee MW, Zagorodnov V (2009): Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. Neuroimage 48:73–83.