White-matter microstructural properties of the corpus callosum: test–retest and repositioning effects in two parcellation schemes

Chaitali Anand[1,2], Andreas M. Brandmaier[3,4,5], Muzamil Arshad[1], Jonathan Lynn[1,2], Jeffrey A. Stanley[1], Naftali Raz[2,3,6]

[1] Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, USA

[2] Institute of Gerontology, Wayne State University, Detroit, MI, USA

[3] Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

[4] Max Planck, UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany

[5] Max Planck, UCL Centre for Computational Psychiatry and Ageing Research, London, UK

[6] Department of Psychology, Wayne State University, Detroit, MI, USA

E-mail addresses: jeffrey.stanley@wayne.edu *(Stanley, J. A.)* nraz@wayne.edu *(Raz, N.)*

Abstract

We investigated test–retest reliability of two MRI-derived indices of white-matter microstructural properties in the human corpus callosum (CC): myelin water fraction (MWF) and geometric mean T_2 relaxation time of intra/extracellular water (geomT$_{2IEW}$), using a 3D gradient and multi spin-echo sequence in 20 healthy adults (aged 24–69 years, 10 men). For each person, we acquired two back-to-back acquisitions in a single session, and the third after a break and repositioning the participant in the scanner. We assessed the contribution of session-related variance to reliability, using intra-class effect decomposition (ICED) while comparing two CC parcellation schemes that divided the CC into five and ten regions. We found high construct-level reliability of MWF and geomT$_{2IEW}$ in all regions of both schemes, except the posterior body—a slender region with a smaller number of large myelinated fibers. Only in that region, we observed significant session-specific variance in the MWF, interpreted as an effect of repositioning in the scanner. The geomT$_{2IEW}$ demonstrated higher reliability than MWF across both parcellation schemes and all CC regions. Thus, in both CC parcellation approaches, MWF and geomT$_{2IEW}$ have good test–retest reliability and are, therefore, suitable for longitudinal investigations in healthy adults. However, the five-region scheme appears more appropriate for MWF, whereas both schemes are suitable for geomT$_{2IEW}$ studies. Given the lower reliability in the posterior body, which may reflect sensitivity to the repositioning of the participant in the scanner, caution should be exercised in interpreting differential findings in that region.

Keywords Myelin, T_2 spin-spin relaxation time, MRI, Reliability, Commissural fibers, Multi-echo imaging

Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0005-1D01-A

Introduction

The corpus callosum (CC), arguably the largest white-matter tract in the human brain, comprises more than 300 million myelinated fibers, and primarily connects homologous regions across the cerebral hemispheres (Aboitiz and Montiel 2003). Although homogeneous with respect to fiber directionality, the CC is, nonetheless, regionally heterogeneous with regard to axon diameter, myelination status, and fiber density (Aboitiz et al. 1992a; Lamantia and Rakic 1990), along its anterior–posterior axis. The genu, posterior body, and anterior splenium, which connect the right and left heteromodal (prefrontal and posterior parietal) cortices, contain relatively few myelinated axons, but evidence a higher density of smaller, unmyelinated fibers. Relatively high myelin content of larger, sparse axons is observed in the anterior body and posterior splenium that connect the primary sensory and motor cortices across the hemispheres (Aboitiz et al. 1992b). For instance, multimodal association

regions of the cortex are primarily linked by fibers of smaller gauge (down to 0.4 μm) in comparison with those connecting primary sensory areas and exceeding diameters of 5 μm. The genu has the highest density of thin fibers (0.4–1 μm in diameter), whereas in the body, thicker fibers with diameters of up to 7 μm are more prominent. The fiber density increases again toward the splenium relative to the body, which has a wider variety of fiber gauges. Larger-diameter fibers show an inverted-U-shaped distribution across the CC, complementary to the distribution of the thin fibers. Unmyelinated fibers are scarce in the CC, except in the genu, where they comprise about 16–20% of total count (Aboitiz et al. 1992a; Lamantia and Rakic 1990). Thus, the in vivo white-matter microstructure is highly diverse with respect to myelin content, axon size, and packing density of large and small myelinated axons across the CC regions. At the same time, the CC lacks anatomical landmarks that could facilitate a clear demarcation of regions along its length.

Several geometric partitioning schemes have been designed to divide the CC into smaller regions according to functional specialization (Hofer and Frahm 2006) or histological characteristics of the fibers (Aboitiz et al. 1992b). The specialization of the CC regions is based on post-mortem studies of interhemispheric communication pathways that are also characterized by specific fiber composition (Witelson 1989) and diffusion properties (Hofer and Frahm 2006). In most studies, the MRI-derived parcellation of the CC is informed by regional histology and functional connectivity of the CC (Hofer and Frahm 2006; Witelson 1989). In actuality, however, the measurements rely on geometric heuristics, such as drawing a horizontal line between the most anterior and posterior points on the midsagittal outline of the CC, and dividing the area into segments by perpendiculars dropped onto that line (Hynd et al. 1991; O'Kusky et al. 1988; Rajapakse et al. 1996). A common parcellation scheme, for instance, generates five CC regions by dividing the horizontal line into five equal segments. In that five-region scheme, the anterior fifth corresponds to the rostrum and genu; the second, third, and fourth fifths to the isthmus and body; and the posterior fifth corresponds to the splenium (Fine et al. 2007; Weis et al. 1993). A seven-region software-driven parcellation of the CC was developed using factor analysis (Denenberg et al. 1991; Peters et al. 2002). However, in that method, a software designated the 'best possible' demarcation lines for the regions between the genu and the splenium, on a single slice and without references to anatomical landmarks.

Regional differences within the CC are not only spatial, but temporal as well. Heterochronicity of cortical myelination, with association regions lagging behind the primary sensory areas, has been established in earlier postmortem studies (Flechsig 1901; Kaes 1907). Alteration of myelin structure as well as reduction of myelin content have been tied to age-related cognitive decline (Bartzokis, 2004; Bowley et al. 2010; Peters 2009). Analyses of regional CC volumes suggest heterochronicity in the pattern of myelination that progresses from the splenium to the genu (Paus et al. 2001). Volumetric analyses suggest that although most of the CC matures by the end of the first year of life, its growth continues into adulthood (Paus et al. 2001). Notably, in CC development, relative stability of the absolute number of axons is accompanied by an increase in the proportion of large myelinated fibers (Aboitiz et al. 1996). Most studies elucidating the properties of the CC have been carried out on post-mortem material (de Lacoste et al. 1985; Flechsig 1901; Tomasch 1954; Witelson 1989). An obvious limitation of postmortem studies is their cross-sectional nature and resultant inability to capture dynamics of development and aging. In vivo neuroimaging studies have suggested that the CC demonstrates larger age-related differences in diffusion-based indices compared to most white-matter tracts (Lebel et al. 2012). Hence, a need for a reliable and valid in vivo measure that allows precise tracking of the developmental dynamics of regional myelin content as well as axon diameter and packing density in longitudinal studies.

The gold standard of assessing white-matter microstructure in the context of myelin content and axon diameter/packing density in vivo is multi-echo spin–spin relaxation (ME-T_2) imaging, which was introduced more than 2 decades ago (Alonso-Ortiz et al. 2015; MacKay et al. 1994). This method can estimate myelin content via a proxy-myelin water fraction (MWF) (Arshad et al. 2016; MacKay et al. 1994). MWF is computed as a fraction of the water signal attributed to the short-T_2 relaxation time constant (< 40 ms at 3T) relative to the total observed water signal. The short-T_2 relaxation component emanates from water molecules trapped between the phospholipid bilayers of the myelin sheath that wraps around the axon and is thus, considered a proxy for myelin content (MacKay et al. 1994; Menon and Allen 1991). The magnitude of the short-T_2 relaxation component is proportional to the number of wraps around axons (Curnes et al. 1988; Menon et al. 1992). In addition to the relatively small, short-T_2 relaxation component associated with myelin water, a larger water signal with an intermediate T_2 relaxation time constant (~ 50–80 ms at 3T), is associated with water molecules within the intracellular/extracellular space. This component is termed geometric mean T_2 or geomT$_{2IEW}$ (Arshad et al. 2017; MacKay et al. 1994). The remainder of the total T_2 decay signal, which consists of a long-T_2 relaxation time constant of about 1000 ms, reflects the contribution of the cerebrospinal fluid. A validation study in rodent spinal cord (Dula et al. 2010) found a positive association between the geomT$_{2IEW}$ and the diameter of axons in spinal white-matter tracts. In this study, vestibulospinal tracts containing large myelinated

axons were characterized by higher $geomT_{2IEW}$ values compared to the less myelinated dorsal corticospinal tracts composed of smaller axons, which exhibited lower $geomT_{2IEW}$ values (Dula et al. 2010). An increase in the axon diameter is accompanied by an expansion of the intracellular space and hence greater mobility of the intracellular water leading to higher $geomT_{2IEW}$ values (Arshad et al. 2017). Both MWF and $geomT_{2IEW}$ have been established as robust and reliable indices and are relatively insensitive to changes in participants' repositioning in the scanner during repeated testing (Arshad et al. 2017). In that study, however, test–retest reliability was evaluated for only two regions of the CC—the genu and the splenium. As theoretical advances in the field call for finer partitioning of the CC (Björnholm et al. 2017), investigating the reliability of MWF and $geomT_{2IEW}$ in more detailed partition schemes becomes a priority. Because of significant variation in the density and caliber of fibers along the length of the CC (Tomasch 1954), it is imperative that each region of the CC is assessed with high reliability at each measurement occasion of the longitudinal assessment. Moreover, it is important to demonstrate equal reliability across regions, as differences in reliability may jeopardize validity of inferences about differential associations of local CC properties and outcome variables. Thus, the main aim of this study was to evaluate, via analysis of previously collected MRI data (Arshad et al. 2017), the test–retest reliability of two common schemes of CC parcellation.

In Arshad et al.'s study, an ME-T_2 imaging sequence was used to collect data from two back-to-back runs and a third run, after repositioning the participant in the scanner. This study design allows separation of the sources of variance in MWF and $geomT_{2IEW}$ estimates into components that are independent from individual differences in the construct of interest; specifically, variance that can be attributed to scanner variables changing over identical runs and the variability introduced by repositioning the participant. Test–retest reliability has been examined in the context of track-based indices derived from diffusion tensor imaging (DTI) (Madhyastha et al. 2014), as well as cortical and subcortical volume and thickness measures obtained using FreeSurfer (Liem et al. 2015). However, to the best of our knowledge, no estimates of test–retest reliability are available for the indices of white-matter microstructure derived from the ME-T_2 method (MWF and $geomT_{2IEW}$) for CC regions. To estimate overall reliability and accomplish the partitioning of variance, with a specific attention to the effect of subject repositioning, we used a recently introduced framework based on structural equation modeling (SEM)—intra-class correlation effect decomposition [ICED; (Brandmaier et al. 2018)] that is described in detail below.

Methods

Participants

We used previously collected data that were applied to calculating reliability (intraclass correlations, ICCs) in six brain regions, including the entire genu and splenium of the CC (Arshad et al. 2017). These data were collected from 20 healthy adults (mean age $\pm$ SD $= 45.9 \pm 17.1$ years, range 24.4–69.5 years; 10 of each sex), who were recruited from the Metro Detroit area. All participants provided informed consent in accordance with Wayne State University Institutional Review Board guidelines, were screened via a telephone interview, and completed a detailed mail-in health questionnaire. The questionnaire was used to exclude individuals with a history of a neurological or major psychiatric disorder, cardiovascular disease (other than medically controlled hypertension), cerebrovascular disease, endocrine and metabolic disorders, cancer, or head trauma (with loss of consciousness for more than 5 min). Persons taking any anxiolytics, antidepressants, or anti-seizure medication were excluded from the study. Pregnant and lactating women were excluded from the study. All participants completed a Center for Epidemiologic Studies-Depression Scale (CES-D) to rule out current depression (CES-D; cut-off $= 15$) (Radloff 1977). Cognitive impairment was screened for, using the Mini Mental State Examination (MMSE; cutoff $= 26$) (Folstein et al. 1975). Men and women did not differ in the years of formal education, MMSE and CES-D scores. The participants underwent thorough MRI safety screening that excluded anybody with any metal (for example, pacemaker, bone/joint screws, metallic implants, braces, and permanent retainers) in their body. The participants were requested to avoid intake of caffeine and/or alcohol before the MRI scan. To keep the time of the scan consistent across participants, all data were collected in the morning around 9:00 am. This was done to avoid the confounding effects of time at which subjects were scanned, a significant contributor to within- and between-person differences (Karch et al. 2019). In addition, the intervals between the test, retest, and repositioning runs were kept constant among participants.

MRI acquisition

The data were collected on a 3T Siemens MAGNETOM Verio MRI system with a 12-channel receive-only volume head coil. The T_1-weighted images were acquired in the axial plane [repetition time (TR) $= 2400$ ms, echo time (TE) $= 2.63$ ms, flip angle (FA) $= 8°$, inversion

time (TI) = 1100 ms, matrix size = 256 × 256, number of slices = 160; GRAPPA factor = 2; voxel size = 1.0 × 1.0 × 1.0 mm^3]. The ME-T_2 images were acquired in the axial plane using the 3D gradient and spin-echo (GRASE) sequence: TR = 1100 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. Although acquiring the ME-T_2 images in the axial plane was not optimal for assessing the CC that is larger along the rostral–caudal rather than ventral–dorsal axis, our protocol was optimized for covering a broad range of white-matter tracts in the brain. These slices were further resampled to a 1 × 1 × 1 mm^3 resolution by interpolating each data set to 2.5 mm thickness, and co-registered to the structural T_1-weighted image. Acquisition time for a single GRASE sequence was 16:09 min. The study protocol included three acquisitions in a single visit. In the first scanning session, T_1-weighted MRI images were acquired, followed by two back-to-back scans of ME-T_2 images. At the end of this first session, participants were removed from the scanner, given a 5-min break, and placed back (repositioned) in the scanner for the second scanning session, which included a T_1-weighted and ME-T_2 scan.

For further details about the study design, image acquisition protocol, and preprocessing, see the previous publication (Arshad et al. 2017). Here, we focus on two outcome measures, MWF expressed as a percent of the total water signal, and geomT$_{2IEW}$ expressed in ms, of the CC regions parcellated according to two widely used schemes described below. The multi-session and multiple measurement aspects of the study allow for separation of the influences of repositioning (expressed as session-specific variance, SSV) from true-score variance (TSV), which represents between-person differences in the construct of interest, and is defined as the shared variance over all three repetitions, and residual error variance (REV).

Corpus callosum (CC) parcellation schemes

Although dividing the CC into smaller regions (e.g., genu, body, and splenium, as well as their smaller subdivisions) is common in multiple studies of CC (Denenberg et al. 1991; Fine et al. 2007; Peters et al. 2002); in the absence of clear anatomical borders, there is no consensus on CC parcellation. Most parcellation schemes infer the borders from known regional differences in function or histology (Cover et al. 2018), but are all based essentially on geometric heuristics and amount to a compromised solution that only partially matches histologically based CC parcellation, as noted in the introduction.

In this study, we compare reliability of two widely used CC parcellation schemes that vary in the degree of demarcation refinement. In the first scheme, adapted from Witelson (1989), the CC is parcellated into five regions, inspired by regional functional specialization of the CC, and is related to specific commissural tracts via diffusion tensor imaging (DTI)-based tractography (Hofer and Frahm 2006). Compared to the seven-region scheme (Witelson 1989), the five-region scheme combines the genu and the rostrum into a single region. The rostral body and the anterior mid-body are fused into one continuous anterior body as well, owing to an overlap in the cortical areas that the rostral body and the anterior mid-body fibers connect. The second, more refined approach, is based on regional histological characteristics of fibers (Aboitiz et al. 1992a; Björnholm et al. 2017). We decided to use the five- and ten-region schemes, as they are widely used in studies of the CC and are at the lowest and highest extremes of CC parcellation schemes.
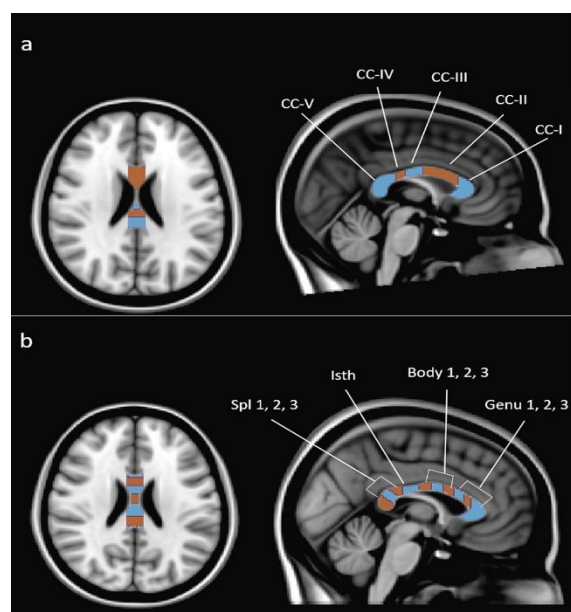
Fig. 1 Mapping of the CC regions on the MNI standard brain. **a** Axial and sagittal views of the CC parcellated into five regions along its anterior–posterior axis from the genu to the splenium, according to the scheme proposed by Hofer and Frahm (2006). The MNI template brain used for creating this scheme was rotated by 8º in the sagittal plane. **b** Axial and sagittal views of the CC parcellated into 10 regions along its anterior–posterior axis from the genu to the splenium, according to fiber composition (Aboitiz et al. 1992a; Björnholm et al. 2017). The region of interest (ROI) masks were generated in FSLview

Five-region scheme

As depicted in Fig. 1a, the CC is parcellated into five segments based on the cortical targets of commissural fibers: CC-I corresponds to fibers connecting prefrontal cortices, CC-II comprises the targets of premotor and supplementary motor areas, CC-III and CC-IV correspond to fibers linking primary motor and somatosensory cortices, respectively, and CC-V represents fibers connecting the parietal, temporal, and occipital cortical areas of the cerebral hemispheres (Hofer and Frahm 2006; Pandya and Seltzer 1986) and also described by Khan et al. (2012) (Khan 2012). The Hofer and Frahm scheme was adapted from the geometry-based parcellation (Witelson 1989), with differences confined to separation of the anterior tip (CC-I) and designation of a longer midbody area (CC-II). The genu and rostrum from Witelson's schemes have been combined into a single region, and so were the rostral and anterior mid-bodies. The five segments in the current study were demarcated with respect to the baseline, a straight line drawn between the most anterior point of the genu and the most posterior point of the splenium on the midsagittal slice. According to this scheme, the CC represents a topographical map of the specific cortical areas located along its sagittal axis. This scheme combines functional specialization as well as fiber characteristics to define the five regions.

Ten-region scheme

The scheme, depicted in Fig. 1b, is based on post-mortem histology with Holmes–Loyez staining on three adjacent midsagittal slices of the CC (Aboitiz et al. 1992a). The CC on these slices is divided into three main regions: genu, body, and isthmus-plus-splenium, which, based on fiber composition, were further segmented into three equal parts: anterior, middle, and posterior sections, except for the isthmus. The regions of the ten-region scheme are: genu (GENU-1, GENU-2, GENU-3), body (BODY-1, BODY-2, BODY-3), isth (ISTH), and splenium (SPL-1, SPL-2, SPL3). Note that in contrast to some studies (Clarke and Zaidel 1994; Fabri et al. 2014), the posterior body (BODY-3) did not include the isthmus.

In both the five- and ten-region schemes, the CC regions were created on the midsagittal slice of the Montreal Neurological Institute (MNI) standard brain as defined by Hofer and Frahm (2006), and Aboitiz et al. (1992a), respectively, using the FMRIB Software Library's FSLview (Jenkinson et al. 2012). The five-region scheme required an additional rotation in the MNI space by 8° in the sagittal plane to ensure that the axial plane coincided with the plane connecting the two endpoints of the CC (Hofer and Frahm 2006). The midline CC regions for the ten-region scheme were demarcated following Björnholm et al. (2017). In both parcellation schemes, CC regions defined in the MNI space were extended laterally by 7 mm from midline into each hemisphere to ensure a more comprehensive coverage of the CC in each region, as illustrated in Fig. 1. The 7-mm lateral extension was selected to correspond to a typical location of lateral longitudinal stria on the CC dissections (Di Ieva et al. 2015), thus adding anatomical rationale to the ROI demarcation and making inclusion of extra-callosal extensions—the forceps minor and major—less likely.

For each scan, the regions of both schemes in the MNI space were co-registered to the participant's GRASE images as follows. First, the T_1-weighted images were co-registered to the MNI standard brain using FSL's non-linear registration tool, FNIRT. Second, the volume of the first echo of the GRASE images was co-registered to the T_1-weighted images using FLIRT with 6 degrees of freedom. Third, the inverse of the warp-field transformation from step one was applied to each CC region (in MNI space) followed by the application of the inverse transformation from step two. The T_1-weighted images were segmented to generate tissue-probability maps of white matter, gray matter, and cerebrospinal fluid using the FSL tool FAST (Zhang et al. 2001), which assigns each voxel a probability of belonging to white matter, gray matter, or cerebrospinal fluid. The obtained white-matter probability maps were then co-registered to the same space as the GRASE images. To ensure that each CC region consisted of primarily white matter, the white-matter probability maps were thresholded and binarized to generate a mask reflecting probability values of 95% or greater, which was then multiplied to each CC region for each subject. The ROI masks in subject space were applied to the ME-T_2 data. For each CC region in each subject, a voxelwise ME-T_2 relaxation analysis was conducted using regularized non-negative least square algorithm (Whittall et al. 2002) as detailed by Arshad et al. (2017).

Statistical analyses

The SEM model underlying the ICED approach (Brandmaier et al. 2018) partitions observed between-person variance in each scan into several orthogonal error variance components that can be attributed to different measurement characteristics. We modeled the hierarchical structure of scans nested in sessions to identify an SSV component. Parameters in the SEM correspond to estimates of TSV, SSV, and REV components. Model specification and estimation were performed in Ωnyx (von Oertzen et al. 2015) and lavaan, an SEM package for the statistical programming language R (Rosseel 2012). The path diagram in Fig. 2 illustrates the ICED model for estimating the individual variance components of the total observed variance in MWF values in the GENU-1 region of the CC. The three ME-T_2 scans (scan 1, scan 2, and scan 3) of the MWF for GENU-1 are labeled in the
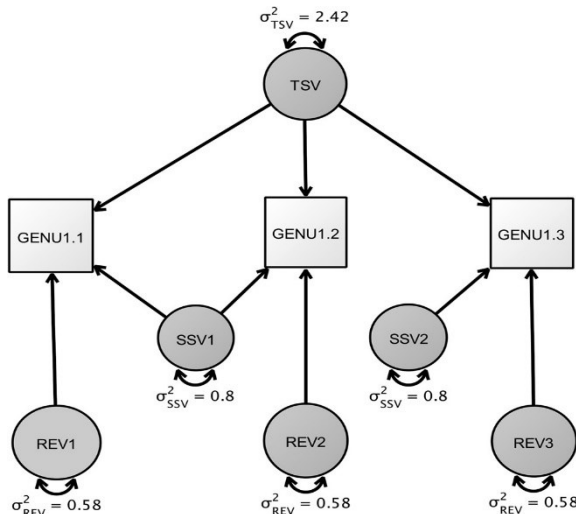
Fig. 2 Path diagram of a structural equation model derived from Ω_{nyx}. In this repeated-measures design, each participant is scanned three times, giving rise to three MWF values for GENU-1 at timepoints 1–3. TSV, SSV, and REV are represented by σ^2_{TSV}, σ^2_{SSV}, and σ^2_{REV}, respectively. The estimates in this illustration correspond to GENU-1 from the ten-region scheme. The diagram was designed in Ωnyx

path diagram as "GENU1.1", "GENU1.2", and "GENU1.3". The same labeling convention was applied to all regions in both parcellation schemes. An unconstrained SEM model was first generated, which freely estimated all three variance parameters related to the three sources of variance (TSV, SSV, and REV). Three additional null models were generated in which each source's (true, session-related, and error) variance was set to zero, one at a time. To assess the significance of the magnitudes of these separate sources of error, likelihood ratio tests were used to compare the unconstrained models against the respective null models.

The covariance matrix is set up, as presented in Table 1, and demonstrates the variances shared between two scans. For instance, since there was no repositioning between scan 1 and scan 2 (see Table 1), they comprise the same session, and thus, the correlation between scan 1 and scan 2 is given as $\frac{\sigma^2_{TSV}+\sigma^2_{SSV}}{\sigma^2_{TSV}+\sigma^2_{SSV}+\sigma_{REV}}$, where the covariance between scan 1 and scan 2 is divided by the square-root of the sum of the total variances at each scan.

The variance components were used to calculate ICC, an index of reliability between two MRI scans within the study. In addition, ICC2 was calculated to assess the reliability of the entire study including all three measurement occasions (test, retest, and repositioning). Bootstrapped 95% confidence intervals, using 1000 samples, were generated for ICC and ICC2 values of the MWF and geomT2IEW in both parcellation schemes, as well as for each variance component. Acceptable target of reliability was set at ICC and ICC2 ≥ 0.80 (Shrout and Fleiss 1979). These analyses were conducted on each region of the CC in each parcellation scheme. All statistical analyses were conducted in R *lavaan* (Rosseel 2012) for the SEM, and applied *boot.ci* for bootstrapping. The reliabilities of MWF and geomT2IEW across all regions of the CC, from both parcellation schemes, were compared with a continuity-corrected Wilcoxon signed-rank test.

Results

Mean MWF and geomT2IEW values across the CC regions

The means and the coefficient of variation (CV) for MWF (percent of the total) and geomT2IEW (ms) for both parcellation schemes are presented in Table 2.

Reliability and variance explained for parcellation schemes

In Table 3, we present ICC and ICC2 values for MWF and geomT2IEW for both parcellation schemes.
The ICC represents reliability achieved in a test–retest scenario with a single run. On the other hand, ICC2

Table 1 Covariance matrix

	Scan 1	Scan 2	Scan 3
Scan 1	$\sigma^2_{TSV} + \sigma^2_{SSV} + \sigma^2_{REV}$	$\sigma^2_{TSV} + \sigma^2_{SSV}$	σ^2_{TSV}
Scan 2	$\sigma^2_{TSV} + \sigma^2_{SSV}$	$\sigma^2_{TSV} + \sigma^2_{SSV} + \sigma^2_{REV}$	σ^2_{TSV}
Scan 3	σ^2_{TSV}	σ^2_{TSV}	$\sigma^2_{TSV} + \sigma^2_{SSV} + \sigma^2_{REV}$

Rows and columns correspond to the three measurement occasions (brain scans) distributed over two sessions on the same day
Parameters of the covariance matrix include the individual differences in the construct of interest (σ^2_{TSV}), the session-specific (σ^2_{SSV}), and the residual error variance (σ^2_{REV})

Table 2 Mean and coefficient of variation (CV) for MWF and geomT$_{2IEW}$ for the CC regions of interest, according to the two parcellation schemes

Region of interest	MWF (CV)	GeomT$_{2IEW}$ (CV)
Five-region scheme		
CC-I	11.3 (0.19)	61.8 (0.04)
CC-II	9.6 (0.20)	66.2 (0.04)
CC-III	12.3 (0.17)	69.8 (0.03)
CC-IV	12.4 (0.23)	69.6 (0.05)
CC-V	16.2 (0.23)	67.2 (0.04)
Ten-region scheme		
GENU-1	11.4 (0.18)	61.6 (0.04)
GENU-2	8.9 (0.23)	64.9 (0.04)
GENU-3	8.9 (0.22)	65.5 (0.04)
BODY-1	9.7 (0.22)	66.3 (0.03)
BODY-2	11.2 (0.22)	68.7 (0.04)
BODY-3	12.4 (0.17)	70.0 (0.03)
ISTH	12.5 (0.23)	69.2 (0.05)
SPL-1	13.8 (0.26)	67.3 (0.04)
SPL-2	15.5 (0.25)	66.2 (0.05)
SPL-3	19.0 (0.23)	67.6 (0.04)

captures the reliability of an entire experimental session such as a nested design with three runs reported here, with all runs considered simultaneously. It is important to note that ICC2, rather than ICC, is a true indicator of the construct-level reliability, that is, reliability of the entire study design. ICC is lower than the ICC2 (see Table 3), since adding a third source of variability in ICC2 increases the reliability of study design.

In Table 4a, b, we present percentage of variance explained by three sources (TSV, SSV, and REV) for ROIs in both parcellation schemes.

Variance in Myelin Water Fraction

Although ICC2 estimates ranged between 0.79 and 0.98 for the five-region scheme and between 0.80 and 0.97 for the ten-region scheme (see Table 3), all confidence intervals overlapped considerably, thus indicating equivalently high reliability among all regions of the CC. The 95% confidence intervals around the raw scores for the sources of variance (see Table 4a) indicate that the SSV was not significantly different from zero for all regions except CC-III of the five-region scheme and the corresponding region, BODY-3, of the ten-region scheme. In comparison with other callosal ROIs, this region yielded relatively low, though acceptable ICC2 values (0.79 in the five-region scheme and 0.81 in the ten-region scheme), both falling within 95% confidence limits of the ICC2 s for other ROIs. As depicted in Fig. 3a, b, across all CC regions, for both parcellation schemes, progressively greater TSV was noted from the anterior to the posterior regions along the sagittal axis: five-region scheme: $\chi^2(4) = 14.4$, $p = 0.006$; ten-region scheme $\chi^2(9) = 28.8$, $p = 0.0006$.

Table 3 Point estimates and 95% confidence intervals for ICC and ICC2 values for MWF and geomT_{2IEW} for the CC regions of interest according to the two parcellation schemes

| Region of interest	ICC For MWF	ICC For GeomT$_{2IEW}$	ICC2 For MWF	ICC2 For GeomT$_{2IEW}$
Five-region scheme				
CC-I	0.85 [0.75–0.95]	0.98 [0.97–1.00]	0.81 [0.61–0.99]	0.98 [0.96–1.00]
CC-II	0.80 [0.62–0.98]	0.92 [0.77–1.07]	0.91 [0.81–1.00]	0.99 [0.84–1.13]
CC-III	0.86 [0.75–0.97]	0.92 [0.82–1.00]	0.79 [0.59–0.99]	0.88 [0.79–0.97]
CC-IV	0.88 [0.79–0.96]	0.88 [0.84–0.91]	0.85 [0.72–0.97]	0.90 [0.84–0.97]
CC-V	0.88 [0.79–0.97]	0.94 [0.78–1.09]	0.98 [0.95–1.00]	0.99 [0.92–1.06]
Ten-region scheme				
GENU-1	0.85 [0.74–0.95]	0.98 [0.97–0.99]	0.80 [0.59–1.00]	0.98 [0.95–1.00]
GENU-2	0.71 [0.49–0.93]	0.93 [0.83–1.03]	0.91 [0.77–1.04]	0.99 [0.96–1.01]
GENU-3	0.72 [0.50–0.95]	0.89 [0.75–1.03]	0.86 [0.72–1.00]	0.97 [0.94–0.99]
BODY-1	0.77 [0.58–0.96]	0.83 [0.61–1.05]	0.87 [0.73–1.00]	0.94 [0.88–1.00]
BODY-2	0.73 [0.46–0.99]	0.83 [0.61–1.03]	0.86 [0.64–1.07]	0.90 [0.76–1.05]
BODY-3	0.87 [0.75–0.98]	0.89 [0.79–0.99]	0.81 [0.63–0.99]	0.88 [0.79–0.98]
ISTH	0.86 [0.75–0.96]	0.93 [0.85–1.02]	0.89 [0.78–0.98]	0.93 [0.86–0.99]
SPL-1	0.83 [0.66–0.99]	0.96 [0.82–1.12]	0.91 [0.78–1.02]	0.99 [0.91–1.05]
SPL-2	0.87 [0.78–0.96]	0.93 [0.79–1.07]	0.96 [0.91–1.00]	0.99 [0.94–1.04]
SPL-3	0.89 [0.82–0.95]	0.91 [0.82–1.00]	0.97 [0.94–0.99]	0.98 [0.94–1.00]

Table 4 Point estimates and 95% confidence intervals for true score, session-specific, and residual variances for (a) MWF and (b) geomT$_{2IEW}$ for the CC regions of interest according to the two parcellation schemes

| Region of interest	True-score variance (TSV)	Session-specific variance (SSV)	Residual error variance (REV)
(a) MWF			
Five-region scheme			
CC-I	2.39 [0.46–4.31]	0.76 [– 0.02–1.54]	0.54 [0.20–0.88]
CC-II	2.84 [0.83–4.84]	0.10 [– 0.44 to 0.64]	0.73 [0.26–1.19]
CC-III	2.56 [0.46–4.65]	0.93 [0.03–1.82]	0.57 [0.21–0.93]
CC-IV	5.54 [1.31–9.76]	1.29 [– 0.06 to 2.66]	0.98 [0.37–1.59]
CC-V	10.81 [3.71–17.91]	– 0.41 [– 1.11 to 0.31]	1.4 [0.57–2.22]
Ten-region scheme			
GENU-1	2.42 [0.46–4.40]	0.79 [– 0.03 to 1.61]	0.58 [0.21–0.93]
GENU-2	3.16 [0.92–5.40]	– 0.15 [– 0.87 to 0.57]	1.23 [0.47–1.97]
GENU-3	2.90 [0.75–5.05]	0.13 [– 0.73–0.99]	1.16 [0.39–1.92]
BODY-1	3.17 [0.83–5.49]	0.24 [– 0.55 to 1.02]	1.02 [0.37–1.67]
BODY-2	3.47 [0.86–6.08]	0.21 [– 0.80 to 1.21]	1.38 [0.51–2.25]
BODY-3	2.91 [0.58–5.25]	0.95 [0.02–1.87]	0.60 [0.22–0.97]
ISTH	5.78 [1.58–9.97]	0.72 [– 0.37 to 1.81]	1.11 [0.42–1.79]
SPL-1	8.02 [2.35–13.70]	0.44 [– 0.91 to 1.79]	1.75 [0.67–2.83]
SPL-2	10.36 [3.42–17.30]	– 0.08 [– 1.01 to 0.85]	1.53 [0.61–2.44]
SPL-3	14.04 [4.72–23.37]	– 0.22 [– 1.21 to 0.77]	1.75 [0.72–2.78]
(b) geomT$_{2IEW}$			
Five-region scheme			
CC-I	5.28 [1.86–8.70]	0.13 [0.00–0.25]	0.08 [0.03–0.13]
CC-II	5.07 [1.80–8.34]	– 0.13 [– 0.36 to 0.11]	0.43 [0.16–0.69]
CC-III	3.16 [0.87–5.44]	0.58 [0.02–1.14]	0.35 [0.12–0.57]
CC-IV	6.95 [2.01–11.91]	0.69 [– 0.37 to 1.76]	1.09 [0.43–1.76]
CC-V	7.84 [2.82–12.85]	– 0.19 [– 0.49 to 0.09]	0.53 [0.18–0.87]
Ten-region scheme			
GENU-1	5.28 [1.86–8.70]	0.14 [0.00–0.27]	0.09 [0.03–0.15]
GENU-2	7.18 [2.54–11.82]	– 0.16 [– 0.43–0.12]	0.52 [0.21–0.84]
GENU-3	5.08 [1.75–8.41]	– 0.08 [– 0.45–0.29]	0.61 [0.22–0.99]
BODY-1	4.25 [1.38–7.12]	– 0.06 [– 0.59–0.47]	0.85 [0.32–1.38]
BODY-2	5.31 [1.56–9.07]	0.29 [– 0.59 to 1.18]	1.15 [0.44–1.86]
BODY-3	3.16 [0.88–5.44]	0.49 [– 0.09 to 1.06]	0.46 [0.17–0.76]
ISTH	6.89 [2.16–11.64	0.62 [– 0.07 to 1.32]	0.55 [0.21–0.88]
SPL-1	8.89 [3.16–14.62]	0.02 [– 0.23 to 0.26]	0.35 [0.12–0.57]
SPL-2	11.31 [4.07–18.55]	– 0.30 [– 0.71 to 0.10]	0.78 [0.29–1.27]
SPL-3	7.20 [2.49–11.85]	– 0.10 [– 0.57 to 0.36]	0.70 [0.22–1.18]

Variance in geomT$_{2IEW}$

Reliability of geomT$_{2IEW}$ was high for both schemes (see Table 3) and ranged between ICC2 of 0.88 and 0.99. SSV and REV values (see Table 4b) were very small and confirmed the excellent test–retest reliability of the of the ME-T$_2$-imaging sequence to measure indices white-matter microstructure reflecting axon diameter and packing density. As depicted in Fig. 4a, b, the TSV did not differ significantly across the regions of the CC for the five-region scheme [$\chi^2(4) = 3.84, p = 0.428$] and ten-region scheme [$\chi^2(9) = 10.58, p = 0.306$].

The reliabilities of measures for each region in the five-region scheme were numerically greater for geomT$_{2IEW}$ than for MWF, thus achieving the maximal possible rank-sum Wilcoxon signed test statistic ($W = 15, p = 0.063$). For five observations, the rank sum cannot exceed the observed value. Therefore, this hypothesis was essentially un-testable within the conventional hypothesis-testing framework. For the ten-region scheme, statistical inference was possible, and indeed, reliabilities differed significantly ($W = 55, p = 0.002$).
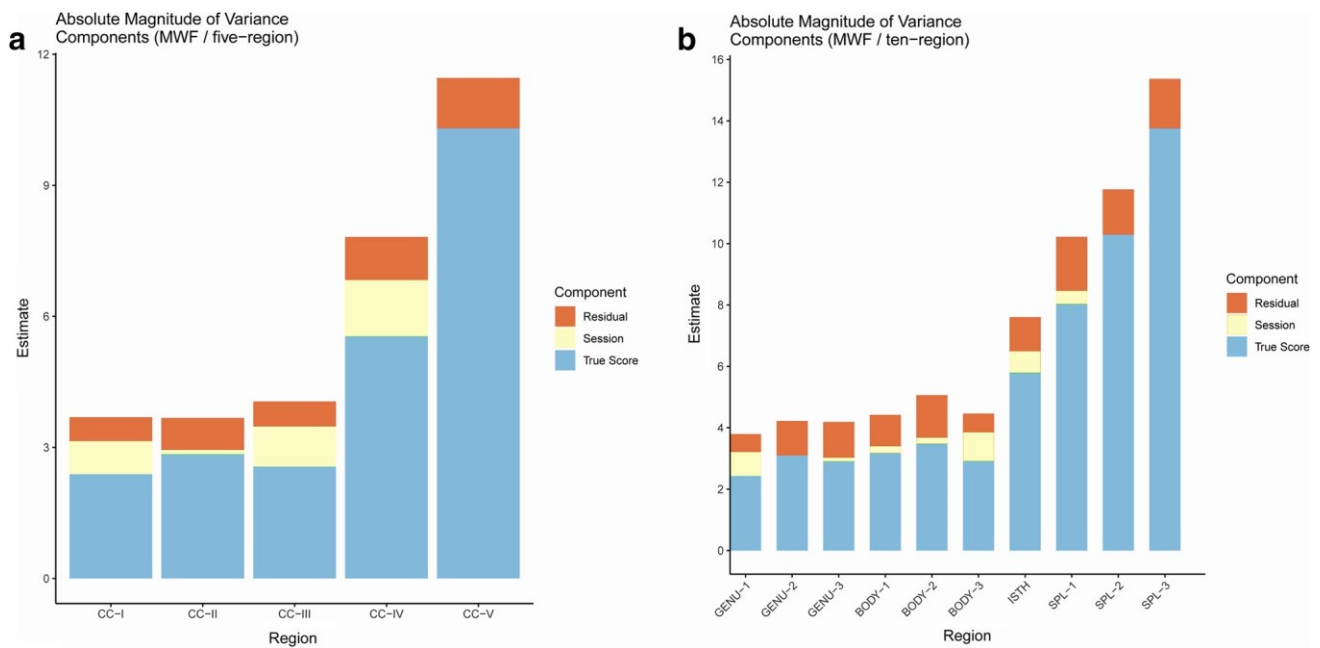
Fig. 3 a Distribution of absolute magnitudes of sources of variance for MWF across the five regions of the CC. **b** Distribution of absolute magnitudes of sources of variance for MWF across the ten regions of the CC. Note: instances of non-significant negative SSVs were constrained to zero in this illustration (CC-V, GENU-2, SPL-2, and SPL 3). The plots were generated using the *ggplot2* package in R
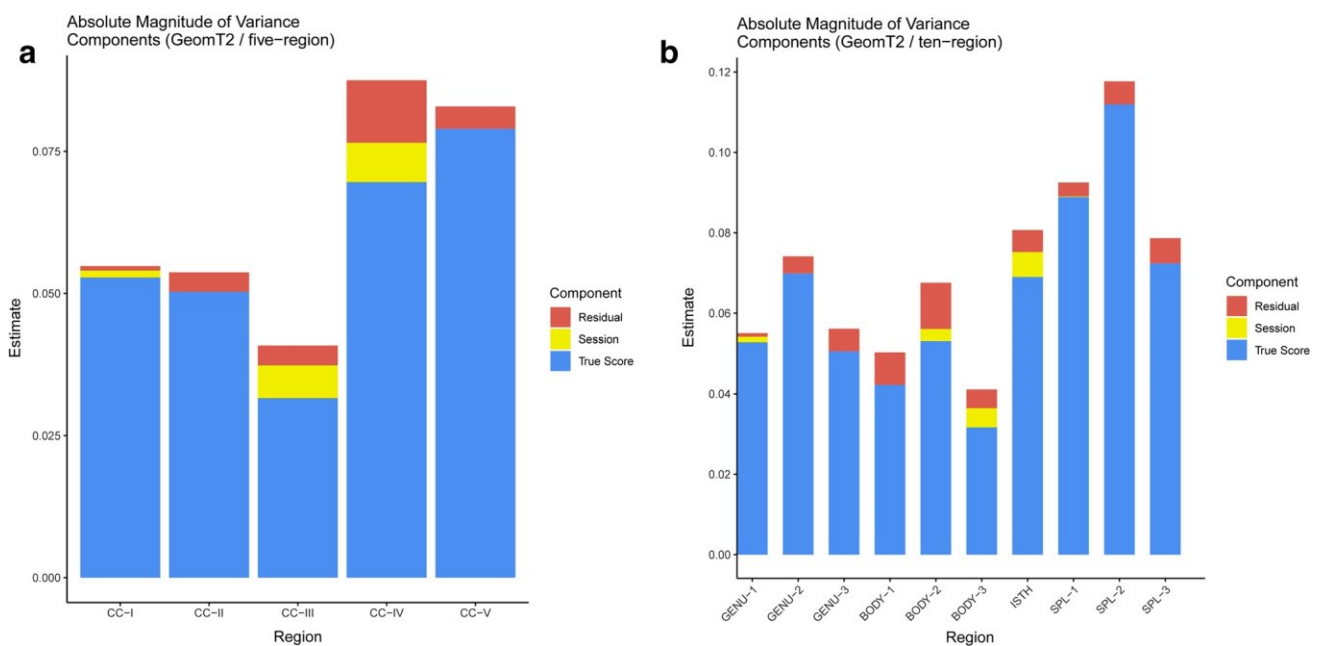
Fig. 4 a Distribution of absolute magnitudes of sources of variance for geomT$_{2IEW}$ across the five regions of the CC. **b** Distribution of absolute magnitudes of sources of variance for geomT_{2IEW} across the ten regions of the CC. As shown in Fig. 3, the SSV values were constrained to be positive for a more effective illustration the of the variance components partitioning for each ROI (CC-II, CC-V, GENU-2, SPL-2, and SPL-3 in this case). The plots were generated using the *ggplot2* package in R

Discussion

In this study, we evaluated test–retest reliability of MRI-derived indices of the regional microstructural characteristics in the callosal white matter. We updated the extant parcellation schemes (Aboitiz et al. 1992a; Hofer and Frahm 2006) using neuroanatomical criteria for lateral expansion of the CC into each hemisphere. To examine the effects of participant's repositioning in the scanner on MWF and geomT$_{2IEW}$, we used a method (ICED) (Brandmaier et al. 2018) that enables simultaneous estimation and partitioning of multiple sources of unreliability.

Both CC parcellation schemes evaluated in this study evidenced high test–retest reliability and were robust vis-à-vis participant repositioning in all CC regions, with a possible exception in the posterior body-CC-III and BODY-3, for the five-region scheme and ten-region scheme, respectively. Overall, the reliability was higher in the geomT$_{2IEW}$ than in the MWF measurements, with the difference being statistically significant in the ten-region scheme and not statistically testable in the five-region scheme.

Across both schemes, we observed the following. First, the percent of total variance in the MWF attributed to the TSV increased from the anterior to the posterior regions along the sagittal axis of the CC. Second, the percent of total variance in the geomT$_{2IEW}$ attributed to the TSV, was the lowest in the posterior body (CC-III and BODY-3), which was the region that also showed the lowest reliability. Third, the posterior body demonstrated a significant effect of scanning session, that is, the lower reliability was attributed to the effect of repositioning the participant in the scanner. Last, the construct-level reliability, ICC2, for the geomT$_{2IEW}$ was higher than that of MWF for both schemes and all CC regions. Below, we interpret this differential reliability in the context of the current knowledge about CC neuroanatomy.

Corpus callosum topography of myelin content and fiber structure

The magnitude of the water signal from the short-T$_2$ relaxation component on which the MWF is based reflects the contribution of highly restricted water molecules trapped between the myelin sheaths. With every additional myelin wrap, the axonal diameter increases proportionately, which also leads to increased MWF. The genu has fewer myelinated axons (about 70–80%) than the splenium, which is 93% myelinated (Lamantia and Rakic 1990; Pandya and Seltzer 1986; Tomasch 1954). The general trend of increasing MWF values from the genu to the splenium as demonstrated in Table 2 is consistent with these neuroanatomical characteristics of the CC.

In contrast, geomT$_{2IEW}$ values along the CC, from the genu to the splenium, demonstrate an inverted-U pattern (see Table 2). As noted above, higher geomT$_{2IEW}$ values are associated with intracellular water in more sparsely packed axons of larger diameter, and intra- and extra-cellular water has lesser mobility in regions with densely packed smaller caliber axons (Arshad et al. 2017). The inverted-U pattern is consistent with a previous MRI study of the CC (Björnholm et al. 2017), and most likely reflects the increasing proportion of larger myelinated fibers (> 5 μm diameter) along the anterior-to-posterior dimension of the CC, peaking in its body and declining thereafter (Aboitiz et al. 1992a).

Increases in proportion of TSV in the MWF from genu to splenium

As indicated in Table 4a, the proportion of TSV in the MWF generally increased along the CC, from the genu to the splenium. This gradual increase in the share of myelinated axons along the CC is likely to produce a less noisy estimate of the individual differences in regional MWF and give rise to the observed genu-to-splenium gradient of reliability. Wider 95% confidence intervals of the ICC2 (see Table 3) in the genu compared to splenium are also in accord with this explanation. However, a significant reduction in the TSV and increase in SSV were noticed in the posterior body (BODY-3, CC-III) region. We speculate upon the reasons for this finding below.

Significant contribution of SSV in MWF and geomT$_{2IEW}$ in the posterior body

The posterior body (CC-III and BODY-3 from the five- and ten-region schemes, respectively), in addition to lower ICC2, also evidenced the greatest share of SSV, i.e., repositioning effect in MWF as well as in geomT$_{2IEW}$. The posterior body shows the greatest individual differences in size, which may be in part due to the frequent occurrence of a depression on its superior surface (Witelson 1989), possibly making it even more sensitive to spatial changes introduced by repositioning. Between-subject variations in its shape may increase the sensitivity of this region to the effects of repositioning, thus reducing the reliability of the MWF estimates across the sample. Along with being the second smallest region of the CC (as indicated by the number of pixels; see Online Resources Figures S1a, S1b, and Table S1), the posterior body is also characterized by the lowest fiber density and, therefore, greater inter-fiber distance (Tomasch 1954), with the fibers mostly being large myelinated ones (> 3 μm) (Aboitiz et al. 1992a). The overall fiber density drops by 25% from the genu to the posterior body (from approximately

400,000 to about 300,000 fibers/mm^2), possibly leaving 25% of BODY-3 (CC-III) without fibers or with only a few large-diameter fibers (Aboitiz et al. 1992a). The lower fiber density in this region leaves more room for extracellular water, which is characterized by a higher geomT$_{2IEW}$ (see Table 2). We speculate that the relative fiber sparsity may result in a greater heterogeneity of the voxels' composition every time a scan is collected. We surmise that even a slight change in head angle upon repositioning of the participant in the scanner, a more heterogeneous voxel composition, vis à vis myelin content, may translate into lesser reliability of MWF estimates. Thus, a relative sparseness of fibers and higher inter-fiber distance observed in the posterior body of the CC (Aboitiz et al. 1992a) along with the fact that it is the second smallest region of the CC, may result in greater sensitivity to the partial volume effect introduced by repositioning in the scanner. Finally, because MWF and geomT$_{2IEW}$ were computed from the data collected at 5-mm-thick axial slices, a region that is thinner in the dorsal–ventral dimension is expected to have greater partial voluming and, therefore, be more susceptible to changes in angulation than thicker ROIs.

Higher ICC2 values for geomT$_{2IEW}$ compared to MWF

The ICC2 values for geomT$_{2IEW}$ across all regions and schemes were higher than those for MWF (Table 3). This difference reflects the higher precision of geomT$_{2IEW}$, as indicated by a lower REVs for that index (see Table 4b) and is consistent with moderate test–retest reliability of MWF compared to the geomT$_{2IEW}$ in the genu, splenium, major and minor forceps reported in a previous scan–rescan cross-site study (Meyers et al. 2013).

We stress that ICC2 is the primary outcome of this study, because it captures the reliability in the context of our test–retest and repositioning study design. In contrast, ICC estimates the reliability of a single measurement. In other words, ICC2 estimates the reliability researchers should expect if they adopt this exact scheme of measuring two back-to-back runs in a single session and a third run in a second session, whereas ICC indicates the reliability researchers should expect if they adopt a single measurement. The ICC values have been derived by computing the relative proportion of TSV and SSV to total variance, that is, the sum of TSV, SSV, and REV (see (Brandmaier et al. 2018).

Limitations

First, in this study, we employed a relatively small and heterogeneous sample of healthy adults. It is unclear, therefore, how these highly reliable measures would behave in more homogeneous samples and in patients with white-matter lesions. Second, the statistical power required for demonstrating the same effect size varies with respect to the brain region of interest and post-processing choices (Liem et al. 2015), and therefore, it is possible that differential unreliability might have been obscured in this relatively small sample. That said, one must keep in mind, however, that the very goal of ICED and ICC analyses is the evaluation of the measurement's stability in capturing individual and group differences, as well as changes therein over time. Third, the small sample size precludes meaningful analyses of age and sex differences in the reliability of MWF and geomT$_{2IEW}$ measurements. Sex and age are associated with the fiber composition in the regions of the CC (Aboitiz et al. 1992a, 1996; Khan 2012; Kim and Juraska, 1997; Witelson 1989). Both effects need to be examined in studies with greater statistical power. Fourth, the interval between scans was very short, and it is unclear how well the indices in question will hold for measurement occasions separated by months and years. For example, significant changes in body composition and size, especially in children, may result in greater discrepancies in scanner placement, thus eroding the true-score variance and boosting measurement error.

Summary and conclusions

We report high reliability of MWF and geomT$_{2IEW}$, in a test–retest design that involved participant's repositioning in the scanner, in CC regions generated by two parcellation approaches. For geomT$_{2IEW}$, reliability was uniformly high across all examined regions in both parcellation schemes. By comparison, a measure of myelin content, MWF, which also showed very good overall reliability, was, however, differentially less reliable and differentially more sensitive to repositioning of the participant in a scanner in some of the regions. The ten-region scheme seems suitable for longitudinal studies of geomT$_{2IEW}$, whereas for MWF, a five-region parcellation approach should be preferred. Other CC parcellation schemes (Fine et al. 2007; Peters et al. 2002) that are based on purely anatomical, geometrical, or a mixture thereof can be examined for reliability using the approach implemented here. Moreover, such comparisons are important in the context of diverse processing pipelines and ROI definitions that can affect the reliability of the metrics, at least in other imaging modalities, such as DTI (Madhyastha et al. 2014). Finally, generalizability of the conclusions regarding reliability of the methods examined here remains to be tested in neuropathological populations and vis à vis age and sex differences.

Author contributions NR, AB, JS, and CA conceptualized the study. CA, NR, and JS drafted the manuscript. CA analyzed the data and designed the figures. AB developed the statistical model and supervised the statistical analyses. MA collected and processed the MRI data. JS, JL, and CA processed the MRI data. NR secured the funding. All authors contributed to editing multiple drafts of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study contains an extended analysis of the MRI data that were previously collected by our group. Non-overlapping analyses have been published in Arshad et al. (2017) and Brandmaier et al. (2018). The data collection was conducted following the protocol approved by the Wayne State University Institutional Review Board, in accordance with the Declaration of Helsinki (1964).

Informed consent All participants provided informed consent.

References

Aboitiz F, Montiel J (2003) One hundred million years of interhemispheric communication: the history of the corpus callosum. Braz J Med Biol Res 36(4):409–420

Aboitiz F, Scheibel AB, Fisher RS, Zaidel E (1992a) Fiber composition of the human corpus callosum. Brain Res 598(1–2):143–153

Aboitiz F, Scheibel AB, Fisher RS, Zaidel E (1992b) Individual differences in brain asymmetries and fiber composition in the human corpus callosum. Brain Res 598(1–2):154–161

Aboitiz F, Rodriguez E, Olivares R, Zaidel E (1996) Age-related changes in fibre composition of the human corpus callosum: sex differences. NeuroReport 7(11):1761–1764

Alonso-Ortiz E, Levesque IR, Pike GB (2015) MRI-based myelin water imaging: a technical review. Magn Reson Med 73(1):70–81. https://doi.org/10.1002/mrm.25198

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. Neuroimage 143:26–39. https://doi.org/10.1016/j.neuroimage.2016.08.047

Arshad M, Stanley JA, Raz N (2017) Test-retest reliability and concurrent validity of in vivo myelin content indices: myelin water fraction and calibrated T1 w/T2 w image ratio. Hum Brain Mapp 38(4):1780–1790. https://doi.org/10.1002/hbm.23481

Bartzokis G (2004) Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer's disease. Neurobiol Aging 25(1):5–18 **(author reply 49-62)**

Björnholm L, Nikkinen J, Kiviniemi V, Nordstrom T, Niemela S, Drakesmith M, Paus T (2017) Structural properties of the human corpus callosum: multimodal assessment and sex differences. Neuroimage 152:108–118. https://doi.org/10.1016/j.neuroimage.2017.02.056

Bowley MP, Cabral H, Rosene DL, Peters A (2010) Age changes in myelinated nerve fibers of the cingulate bundle and corpus callosum in the rhesus monkey. J Comp Neurol 518(15):3046–3064. https://doi.org/10.1002/cne.22379

Brandmaier AM, Wenger E, Bodammer NC, Kuhn S, Raz N, Lindenberger U (2018) Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). Elife. https://doi.org/10.7554/eLife.35718

Clarke JM, Zaidel E (1994) Anatomical-behavioral relationships: corpus callosum morphometry and hemispheric specialization. Behav Brain Res 64(1–2):185–202

Cover GS, Herrera WG, Bento MP, Appenzeller S, Rittner L (2018) Computational methods for corpus callosum segmentation on MRI: a systematic literature review. Comput Methods Progr Biomed 154:25–35. https://doi.org/10.1016/j.cmpb.2017.10.025

Curnes JT, Burger PC, Djang WT, Boyko OB (1988) MR imaging of compact white matter pathways. AJNR Am J Neuroradiol 9(6):1061–1068

de Lacoste MC, Kirkpatrick JB, Ross ED (1985) Topography of the human corpus callosum. J Neuropathol Exp Neurol 44(6):578–591

Denenberg VH, Kertesz A, Cowell PE (1991) A factor analysis of the human's corpus callosum. Brain Res 548(1–2):126–132. https://doi.org/10.1016/0006-8993(91)91113-f

Di Ieva A, Fathalla H, Cusimano MD, Tschabitscher M (2015) The indusium griseum and the longitudinal striae of the corpus callosum. Cortex 62:34–40. https://doi.org/10.1016/j.cortex.2014.06.016

Dula AN, Gochberg DF, Valentine HL, Valentine WM, Does MD (2010) Multiexponential T2, magnetization transfer, and quantitative histology in white matter tracts of rat spinal cord. Magn Reson Med 63(4):902–909. https://doi.org/10.1002/mrm.22267

Fabri M, Pierpaoli C, Barbaresi P, Polonara G (2014) Functional topography of the corpus callosum investigated by DTI and fMRI. World J Radiol 6(12):895–906. https://doi.org/10.4329/wjr.v6.i12.895

Fine JG, Semrud-Clikeman M, Keith TZ, Stapleton LM, Hynd GW (2007) Reading and the corpus callosum: an MRI family study of volume and area. Neuropsychology 21(2):235–241. https://doi.org/10.1037/0894-4105.21.2.235

Flechsig, (1901) Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. Lancet 2(1207):1021–1096

Folstein MF, Folstein SE, McHugh PR (1975) Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12(3):10

Hofer S, Frahm J (2006) Topography of the human corpus callosum revisited–comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. Neuroimage 32(3):989–994. https://doi.org/10.1016/j.neuroimage.2006.05.044

Hynd GW, Semrud-Clikeman M, Lorys AR, Novey ES, Eliopulos D, Lyytinen H (1991) Corpus callosum morphology in attention deficit-hyperactivity disorder: morphometric analysis of MRI. J Learn Disab 24:6

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) Fsl Neuroimage 62(2):782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

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

Karch JD, Filevich E, Wenger E, Lisofsky N, Becker M, Butler O, Kühn S (2019) Identifying predictors of within-person variance in MRI-based brain volume estimates. Neuroimage **(in press)**

Khan IN (2012) Sex differences and the effects of sex hormones on the structure of the corpus callosum, Ph.D. Thesis. University of Nottingham

Kim JH, Juraska JM (1997) Sex differences in the development of axon number in the splenium of the rat corpus callosum from postnatal day 15 through 60. Brain Res Dev Brain Res 102(1):77–85. https://doi.org/10.1016/s0165-3806(97)00080-1

Lamantia AS, Rakic P (1990) Cytological and quantitative characteristics of four cerebral commissures in the rhesus monkey. J Comp Neurol 291(4):520–537. https://doi.org/10.1002/cne.902910404

Lebel C, Gee M, Camicioli R, Wieler M, Martin W, Beaulieu C (2012) Diffusion tensor imaging of white matter tract evolution over the lifespan. Neuroimage 60(1):340–352. https://doi.org/10.1016/j.neuroimage.2011.11.094

Liem F, Merillat S, Bezzola L, Hirsiger S, Philipp M, Madhyastha T, Jäncke L (2015) Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. Neuroimage 108:95–109. https://doi.org/10.1016/j.neuroimage.2014.12.035

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(6):673–677

Madhyastha T, Merillat S, Hirsiger S, Bezzola L, Liem F, Grabowski T, Jäncke L (2014) Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. Hum Brain Mapp 35(9):4544–4555. https://doi.org/10.1002/hbm.22493

Menon RS, Allen PS (1991) Application of continuous relaxation time distributions to the fitting of data from model systems and excised tissue. Magn Reson Med 20(2):214–227

Menon RS, Rusinko MS, Allen PS (1992) Proton relaxation studies of water compartmentalization in a model neurological system. Magn Reson Med 28(2):264–274

Meyers SM, Vavasour IM, Madler B, Harris T, Fu E, Li DK, Laule C (2013) Multicenter measurements of myelin water fraction and geometric mean T2: intra- and intersite reproducibility. J Magn Reson Imaging 38(6):1445–1453. https://doi.org/10.1002/jmri.24106

O'Kusky J, Strauss E, Kosaka B (1988) The corpus callosum is larger with right-hemisphere cerebral speech dominance. Ann Neurol 24:4

Pandya DN, Seltzer B (1986) The topography of commissural fibers. In: Leporé F, Ptito M, Jasper HH (eds) Two hemispheres-one brain: functions of the corpus callosum. Alan Liss, New York, pp 47–73

Paus T, Collins DL, Evans AC, Leonard G, Pike B, Zijdenbos A (2001) Maturation of white matter in the human brain: a review of magnetic resonance studies. Brain Res Bull 54(3):255–266

Peters A (2009) The effects of normal aging on myelinated nerve fibers in monkey central nervous system. Front Neuroanat 3:11. https://doi.org/10.3389/neuro.05.011.2009

Peters M, Oeltze S, Seminowicz D, Steinmetz H, Koeneke S, Jäncke L (2002) Division of the corpus callosum into subregions. Brain Cognit 50:11

Radloff LS (1977) The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1(3):7

Rajapakse JC, Giedd JN, Rumsey JM, Vaituzis AC, Hamburger SD, Rapoport JL (1996) Regional MRI measurements of the corpus callosum: a methodological and developmental study. Brain Dev 18(5):379–388

Rosseel (2012) lavaan: an R package for structural equation modeling. J Stat Softw 48:36

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86(2):420–428

Tomasch J (1954) Size, distribution, and number of fibres in the human corpus callosum. Anat Rec 119(1):119–135

von Oertzen T, Brandmaier AM, Tsang S (2015) Structural equation modeling with Onyx. Struct Eq Model Multidiscip J 22:12

Weis S, Kimbacher M, Wenger E, Neuhold A (1993) Morphometric analysis of the corpus callosum using MR: correlation of measurements with aging in healthy individuals. AJNR Am J Neuroradiol 14(3):637–645

Whittall KP, MacKay AL, Li DK, Vavasour IM, Jones CK, Paty DW (2002) Normal-appearing white matter in multiple sclerosis has heterogeneous, diffusely prolonged T(2). Magn Reson Med 47(2):403–408

Witelson SF (1989) Hand and sex differences in the isthmus and genu of the human corpus callosum. A postmortem morphological study. Brain 112(Pt 3):799–835

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