Received: 15 March 2022 | Revised: 8 August 2022 | Accepted: 21 August 2022

DOI: 10.1002/hbm.26064

RESEARCH ARTICLE

WILEY

Test–retest reliability of diffusion tensor imaging scalars in 5-year-olds

Aylin Rosberg[1,2,3] | Jetro J. Tuulari[1,2,4] | Venla Kumpulainen[1] | Minna Lukkarinen[1,5] | Elmo P. Pulli[1] | Eero Silver[1] | Anni Copeland[1] | Ekaterina Saukko[3] | Jani Saunavaara[6] | John D. Lewis[7] | Linnea Karlsson[1,2,5,8] | Hasse Karlsson[1,2,8] | Harri Merisaari[1,3]

[1]FinnBrain Birth Cohort Study, Turku Brain and Mind Centre, Department of Clinical Medicine, University of Turku, Turku, Finland

[2]Department of Psychiatry, Turku University Hospital and University of Turku, Turku, Finland

[3]Department of Radiology, Turku University Hospital, Turku, Finland

[4]Turku Collegium for Science, Medicine and Technology, University of Turku, Turku, Finland

[5]Department of Pediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland

[6]Department of Medical Physics, Turku University Hospital and University of Turku, Turku, Finland

[7]Montreal Neurological Institute, McGill University, Montreal, Canada

[8]Centre for Population Health Research, Turku University Hospital and University of Turku, Turku, Finland

Correspondence
Aylin Rosberg, FinnBrain, Teutori building, 2nd floor, Lemminkäisenkatu 3, 21520, Turku, Finland.
Email: ayrosb@utu.fi

Funding information
Academy of Finland, Grant/Award Numbers: 325292, 308176, 134950, 253270, 2608098; Emil Aaltosen Säätiö; Jane ja Aatos Erkon Säätiö; Juho Vainion Säätiö; Signe ja Ane Gyllenbergin Säätiö; Sigrid Juséliuksen Säätiö; Suomen Aivosäätiö; Suomen Lääketieteen Säätiö; Turunmaa Duodecim Society; Varsinais-Suomen Rahasto; Varsinais-Suomen Sairaanhoitopiiri

Abstract

Diffusion tensor imaging (DTI) has provided great insights into the microstructural features of the developing brain. However, DTI images are prone to several artifacts and the reliability of DTI scalars is of paramount importance for interpreting and generalizing the findings of DTI studies, especially in the younger population. In this study, we investigated the intrascan test–retest repeatability of four DTI scalars: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) in 5-year-old children ($N = 67$) with two different data preprocessing approaches: a volume censoring pipeline and an outlier replacement pipeline. We applied a region of interest (ROI) and a voxelwise analysis after careful quality control, tensor fitting and tract-based spatial statistics. The data had three subsets and each subset included 31, 32, or 33 directions thus a total of 96 unique uniformly distributed diffusion encoding directions per subject. The repeatability of DTI scalars was evaluated with intraclass correlation coefficient (ICC(3,1)) and the variability between test and retest subsets. The results of both pipelines yielded good to excellent (ICC(3,1) > 0.75) reliability for most of the ROIs and an overall low variability (<10%). In the voxelwise analysis, FA and RD had higher ICC(3,1) values compared to AD and MD and the variability remained low (<12%) across all scalars. Our results suggest high intrascan repeatability in pediatric DTI and lend confidence to the use of the data in future cross-sectional and longitudinal studies.

KEYWORDS
analysis workflows, ICC, pediatric DTI, TBSS, test–retest repeatability

1 | INTRODUCTION

Five years of age is a timepoint of great interest in terms of brain development, because the brain starts to resemble an adult brain. By the age of five; the brain already weighs almost as much as an adult brain (Dekaban & Sadowsky, 1978; Lenroot & Giedd, 2006) and no age-dependent total brain volume changes are observed after the age of five (Faria et al., 2010). In addition, the brain undergoes complex microstructural changes including myelination, axonal packing, and other alterations in the first few years of life (Croteau-Chonka et al., 2016; Muircheartaigh et al., 2014). A rapid white matter (WM) maturation period takes place until about age five. Even though, the brain continues to rewire and grow, the changes are never as extensive as they are in the first five years of life (Lebel et al., 2017; Lebel & Deoni, 2018). Because of this, high quality brain imaging at this timepoint is of great importance.

Diffusion tensor imaging (DTI) is a special magnetic resonance imaging (MRI) technique that utilizes the molecular motion of water to noninvasively estimate the microstructure of the brain (Le Bihan & Johansen-Berg, 2012) by measuring diffusion in multiple predefined directions. The images acquired with this technique can be used to study the microstructural changes in neurodevelopment, to assess the differences in group comparison studies, to detect pathological and incidental findings in the data or to reconstruct the WM pathways. Furthermore, DTI is more sensitive to subject motion compared to conventional MRI (Kreilkamp et al., 2016; Taylor et al., 2016), which creates ghosting and coregistration errors (Mori & Zhang, 2006).

Previous studies suggest that children who are five years of age or younger have difficulties in cooperating during the imaging process and staying still in the scanner (Yoshida et al., 2013; Yuan et al., 2009). It has even been claimed that sedation is required for younger children (Hermoye et al., 2006), but it is not common practice for ethical reasons (Copeland et al., 2021). Generally, to counteract the motion artifacts, single shot echo planar imaging (SS-EPI) sequences are preferred. They are fast and more importantly less sensitive to phase errors. However, SS-EPI sequences suffer from instrumental disadvantages including distortions caused by eddy currents and inhomogeneities in the magnetic field which become more problematic around the sinuses (Basser & Jones, 2002; Shen et al., 2004). Most of the artifacts can be corrected in postprocessing and in adult subjects, the artifacts do not usually hinder the DTI measurements. However, in pediatric imaging, it is more critical to be aware of these challenges and to evaluate the quality of the data. To make accurate interpretations of the DTI data, the technique needs to be applied carefully and the acquired images need to be processed with utmost caution (Jones & Cercignani, 2010).

In this study, we investigated the repeatability of DTI scalars in 5-year-old children by using tract-based spatial statistics (TBSS). We hypothesized that in the ROIs of the atlas and in the central WM (corpus callosum [CC] and WM around the subcortical nuclei) voxels in the TBSS skeleton DTI scalars would have a low intrascan test–retest variability (VAR), that is, less than 10% and a high test–retest reliability, that is, a good to excellent level (>0.75) of intraclass correlation coefficient (ICC(3,1)) values (Koo & Li, 2016; Shrout & Fleiss, 1979).

2 | MATERIALS AND METHODS

The data were acquired as a part of the FinnBrain Birth Cohort Study whose goal is to study the effects of genes and environment on the developmental and mental health of children (Karlsson et al., 2018). The cohort included 141 5-year-olds (mean age, SD [range] = 5.38, 0.11 [5.08–5.79]; 68 females, 73 males), who were scanned between 2018 and 2021. The criterion for inclusion in the imaging measurements was attending the standard neuropsychological assessment at five years of age.

The study exclusion criteria were 1) delivery before gestational week 35 (or before gestational week 32 in the cases who received routine prenatal synthetic glucocorticoid treatment) (Pulli et al., 2021); 2) developmental anomalies or abnormalities in the senses or communication skills (e.g., congenital heart disease, blindness, or deafness); 3) long-term diagnosis of epilepsy, autism, attention deficit hyperactivity disorder; 4) ongoing medical examinations or clinical follow-up in a hospital (i.e., a referral from primary care setting to specialists); 5) use of continuous, daily medication (including oral medications, topical creams, and inhalants with one exception of desmopressin); 6) a history of head trauma defined as concussion necessitating clinical follow-up in a health care setting; and 7) metallic ear tubes (to assure good quality scans), and routine MRI contraindications. The families were provided with both oral and written information about the study and gave their written informed consents in accordance with the Declaration of Helsinki. The wishes and needs of the children were considered at all times. The protocol was approved by the Ethics Committee of the South-Western Hospital District and the study was carried out in accordance with their recommendations.

2.1 | MRI scanning visits

The participants were recruited for the neuroimaging visits via telephone calls made by a research staff member. In the first call, the families were given general information about the study and the inclusion and exclusion criteria were checked. A follow-up call was used to confirm the families' participation and to answer any of their questions. Before the scanning session, each subject was met with in person and at the meeting they were given instructions to practice at home for the MRI.

We applied multiple methods to reduce anxiety and make the visit feel as safe as possible (Copeland et al., 2021; Pulli et al., 2021). The visit was conducted in a child-friendly manner with a flexible timetable for the preparation before the scan. The participants were either scanned awake or naturally asleep without sedation. During the structural and diffusion imaging, the participants were allowed to watch a cartoon or a movie of their choice. A parent and a research

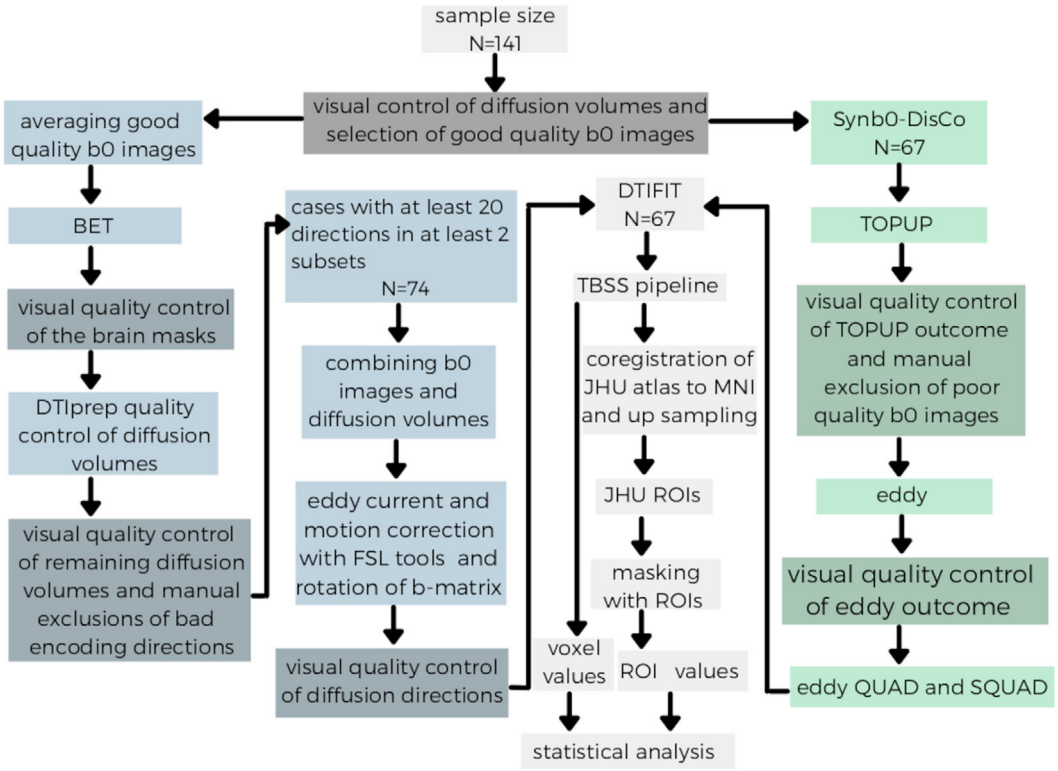

FIGURE 1 Schema of data analysis procedure for evaluation of intrascan test–retest repeatability of diffusion tensor imaging (DTI) scalars in 5-year-olds with two preprocessing pipelines: the volume censoring pipeline (blue) and the outlier replacement pipeline (green), after which common processing (gray) including a DTI model fitting and statistical analysis is applied. Fully automated processes are shown in a lighter shade of the color assigned to the pipeline process and manual visual control steps are shown in a darker shade of the color assigned to the pipeline process.

staff member were present in the scanner room throughout the scan. Everyone in the room had their hearing protected with earplugs and earmuffs. The maximum scan time was 60 min, and the subjects could stop the scan at any time.

2.2 | MRI acquisition

The exact imaging scheme used with neonates at the cohort's previous time point (Merisaari et al., 2019) was repeated at the 5-year-old time point.

MRI data were acquired using a Siemens Magnetom Skyra fit 3T scanner (Siemens Medical Solutions, Erlangen, Germany) with a Head/Neck 20 receiver coil. All participants went through the three following sequences: 1) sagittal T1-MPRAGE; 2) single shell DTI with *b*-value of 1000 s/mm^2 (for part of the participants, multishell DTI with additional b-values of 650 and 2000 s/mm^2); and 3) functional MRI (fMRI).

The single shell DTI data, used in the current study, were acquired by a standard twice-refocused spin echo-EPI sequence with a $2 \times 2 \times 2$ mm^3 isotropic resolution (FOV 208 mm; 64 slices; TR 9300 ms; TE 87 ms). There were three subsets in the multiscan DTI sequence. Each subset had three b0 images (images without diffusion coding) that were taken at the beginning of the scan. In addition, each subset included 31, 32, or 33 directions resulting in up to 96 unique uniformly distributed diffusion encoding directions per subject.

2.3 | DTI preprocessing

2.3.1 | The volume-censoring pipeline

The data analysis procedure from the previous neonate study on short-term repeatability (Merisaari et al., 2019) was adapted (see Figure 1) for the current study.

First, good quality b0 images were chosen, coregistered, averaged and moved in front of each 4D series. Brain masks based on the b0 volumes were created with FSL's (FMRIB Software Library v 5.0.9; (Jenkinson et al., 2012)) Brain Extraction Tool (Smith, 2002). DTIPrep software was used to quantitatively evaluate the quality of the diffusion volumes. DTIPrep is a quality control tool for diffusion MRI that removes diffusion volumes that are affected by three corrupting artifacts. The three artifacts are the interslice brightness artifact within the volume, venetian blind artifacts and motion within the volume and residual motion across all volumes (Oguz et al., 2014). Diffusion volumes considered to have poor quality were automatically discarded by the DTIPrep. According to the recommendations of the DTIPrep

creators, the remaining volumes were visually inspected in slice wise manner and more poor-quality directions were manually removed as necessary. In the manual exclusion, poor quality was identified as a visually detectable signal dropout in the diffusion encoding images. After the exclusion steps, there were 74 subjects remaining with at least 20 acceptable diffusion encoding directions in at least 2 out of the 3 subsets. The first and second subsets fulfilling the diffusion encoding direction criteria were taken in order of appearance to form test and retest measurements for the study and preprocessed separately. Eddy current and motion correction were done with FSL (Andersson & Sotiropoulos, 2016) and the b-matrix was corrected accordingly. A diffusion tensor model fit for each voxel was included in the brain mask with DTIFIT. As the DTIFIT failed in the case of 7 of the subjects, 67 subjects with at least 20 directions per test and retest were subsequently used for further analysis (see Figure 1).

The TBSS of FSL is expected to give robust scalar estimates through interscan coregistrations and the skeletonization process (Smith et al., 2006). Therefore, the TBSS pipeline (Smith et al., 2006) was applied to all DTI scalar parameter maps. In the TBSS procedure, in addition to the default 0.20 threshold value, we also tried a threshold value of 0.25 in order to reach a more robust exclusion of borderline voxels. This threshold was also tested due to previous tractography studies having used a thresholding of 0.25 to reduce the partial volume effect (Lebel et al., 2008; Lebel et al., 2012; Long et al., 2017; Tamnes et al., 2018).

2.3.2 | The outlier replacement pipeline

The data from the 134 subsets used in the volume-censoring pipeline were included in the analysis. The poor quality b0 images removed in the volume-censoring pipeline were again removed; however, this time, the b0 images were not averaged and all the diffusion-encoded images were kept in the dataset. Due to the lack of field-maps and reverse phase-encoding scans, synthesized images were used in the synthesized b0 for diffusion distortion correction (Synb0-DisCo) of the data. Synb0-DisCo (Schilling et al., 2019) utilizes a T1-weighted scan to create a synthetic undistorted b0 image which makes it possible to conduct the susceptibility and distortion correction with the FSL TOPUP tool (Andersson et al., 2003; Smith et al., 2004). The FSL eddy pipeline was run with the following parameters —repol—niter = 8—fwhm = 8,4,2,0,0,0,0,0—estimate_move_by_susceptibility and the TOPUP outcome (Andersson et al., 2016; Andersson et al., 2018; Andersson & Sotiropoulos, 2016). The FSL DTIFIT tool was used to fit the tensor model to each voxel. The TBSS analysis was run in exactly the same way as the volume-censoring pipeline (see Figure 1).

2.4 | Subject motion

Evidence suggests that both small of head motions and retrospective motion correction methods could quantitatively alter DTI scalars (Alhamud et al., 2015; Kreilkamp et al., 2016). Therefore, although the effect of motion was addressed by the volume wise data exclusion and coregistrations, we evaluated the potential association of residual head motion during the scan with the DTI scalars to verify this. Motion was assessed based on the joint eddy current and motion correction step used in the volume-censoring pipeline; this saves three translations across x, y, and z as well as three rotation parameters. Correlation between the DTI scalars were calculated with both the absolute value of maximum motion and the absolute value of total motion were calculated from framewise differences in the translational and rotational parameters.

2.5 | Statistical methods

All statistical analyses were conducted in the Rstudio (v 1.4.1103) (Rstudio, 2020). Both the R and python scripts used in the data analyses are available from the corresponding author upon request. To investigate the repeatability, the intraclass correlation coefficient for single measurement (ICC(3,1)) (Shrout & Fleiss, 1979), and test–retest variability (VAR) were used. ICC(3,1) values between 0.5 and 0.75 were considered as a moderate level, ICC(3,1) values over 0.75 were considered as a good level and ICC(3,1) values over 0.90 were considered as an excellent level of test–retest reliability (Koo & Li, 2016).

The power analysis of ICC values can be found in the Supplementary Material 1. VAR in percentages was calculated by taking the absolute value of the difference between the two scans and dividing this value by the mean of the two scans and then multiplying by one hundred ($(|\text{scan1} - \text{scan2}|/[(\text{scan1} + \text{scan2})/2]) \times 100$).

For the ROI analysis, the template for the JHU ICBM-DTI-81 white-matter labels atlas (Mori et al., 2008) was coregistered with the images created by TBSS to extract the values of the DTI scalars. Of the 48 ROIs in the atlas, 36 were used and 12 were excluded due to their peripheral location (see Supplementary Material 2 for the list of ROIs included in the study). ICC(3,1) and VAR were calculated using the mean DTI scalar value of the ROIs.

Since the TBSS skeletons of the test and retest are not the same due to separate preprocessing, we created for the voxelwise analysis a new mask (overlap mask) from the overlapping voxels of the two TBSS skeleton masks (test mask and retest mask) per preprocessing pipeline. Dice similarity coefficients (DSCs) were calculated between the overlap mask and the original test TBSS skeleton mask and retest TBSS skeleton mask in order to assess the degree of similarity. Voxelwise test–retest reliability and variability were calculated for the voxels within the overlap masks.

The statistical significance level was corrected with the Bonferroni correction as regards the number of scalars, the number of subsets, the number of confounding factors and the number of ROIs. Raw p values have been reported before correction for multiple comparisons unless otherwise noted and p values <.05 after correction for multiple comparisons were considered statistically significant.

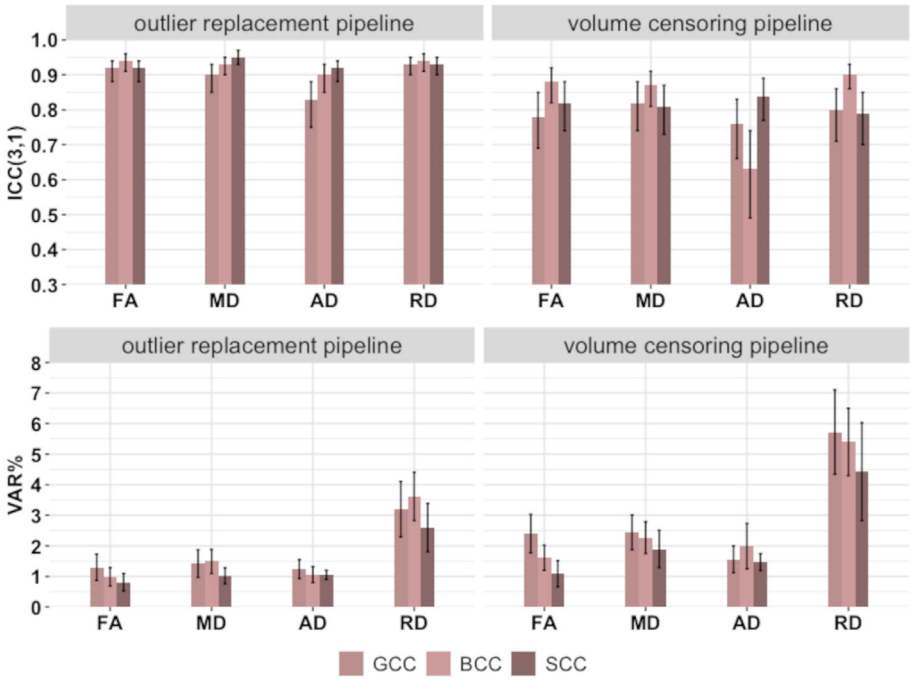

FIGURE 2 Test–retest reliability (ICC (3,1)) and variability (VAR%) of the diffusion tensor imaging (DTI) scalars shown with bar plots for three sections of the corpus callosum (the genu of corpus callosum [GCC], the body of corpus callosum [BCC], the splenium of corpus callosum [SCC]) as identified by the JHU atlas masked with the tract-based spatial statistics (TBSS) skeleton analyzed with the outlier replacement pipeline and the volume censoring pipeline.

3 | RESULTS

Regardless of the preprocessing pipeline chosen, using a higher threshold value (0.25) in the TBSS process created only negligible differences in repeatability. Therefore, only the results of the analysis conducted with the TBSS skeleton created with the default threshold value (0.2) are presented.

3.1 | Reliability and variability of ROI measures

With the outlier replacement pipeline, the test–retest repeatability was high. All scalars yielded good to excellent (ICC(3,1) > 0.75) reliability with the exception of MD values in the right uncinate fasciculus (ICC(3,1) = 0.73) (see Supplementary Material 2). Mean ICC(3,1) values for FA, MD, AD, and RD across the 36 ROIs of the atlas were 0.92, 0.92, 0.88, and 0.94, respectively. There was minimal variability between test and retest (see Supplementary Material 3). The highest variability was detected between the mean RD values in the left tapetum (VAR = 6.5%). The mean variability values for FA, MD, AD, and RD across the 36 ROIs of the atlas were 1.9, 1.2, 1.5, and 2%, respectively.

With the volume censoring pipeline, the FA and RD values showed good to excellent (ICC(3,1) > 0.75) reliability with the exceptions of the FA values in the right external capsule (ICC(3,1) = 0.73) and the right stria terminalis (ICC(3,1) = 0.74) and the RD values in the right external capsule (ICC(3,1) = 0.72), the right stria terminalis (ICC(3,1) = 0.62) and the right superior fronto-occipital fasciculus (ICC(3,1) = 0.71). The MD and AD values, however, had poorer test–retest reliability compared to FA and RD, with both showing less than a good level of test–retest reliability (ICC(3,1) < 0.75) in 13 of the ROIs. The mean ICC(3,1) values across the 36 evaluated ROIs for FA, MD, AD, and RD were 0.86, 0.81, 0.78, and 0.86, respectively (see Supplementary Material 2). Similar to the outlier replacement pipeline, the test–retest variability was low (<10%), while the highest variability was detected in the mean RD values for the left tapetum (VAR = 8.3%). The mean VAR values for FA, MD, AD, and RD were 2.3, 1.8, 1.8, and 2.9%, respectively (see Supplementary Material 3).

We evaluated the CC in three sections in accordance with the atlas parcellation: the genu of CC (GCC) which is the most anterior part of the CC, the splenium of the CC (SCC) which is the most posterior part of the CC and the body of the CC (BCC) which is the area between the GCC and SCC. With the outlier replacement pipeline, all scalars yielded good to excellent test–retest reliability (ICC(3,1) > 0.82). Similarly, with the volume censoring pipeline all scalars showed good test–retest reliability (ICC(3,1) > 0.76) with the exception of the AD values in the BCC (ICC(3,1) = 0.63). There was negligible test–retest variability in all three sections of the CC (volume censoring pipeline: VAR <6%, outlier replacement pipeline: VAR <4%) (see Figure 2). Regardless of the preprocessing pipeline chosen, using different threshold values in the TBSS process created only negligible differences in repeatability.

3.2 | Reliability and variability of voxelwise measurements

In the evaluations with individual TBSS skeleton voxels, the overlap between the test and retest skeleton was high (with the outlier replacement pipeline both DSCs were 0.91, with the volume censoring pipeline the DSC was 0.87 and 0.83).

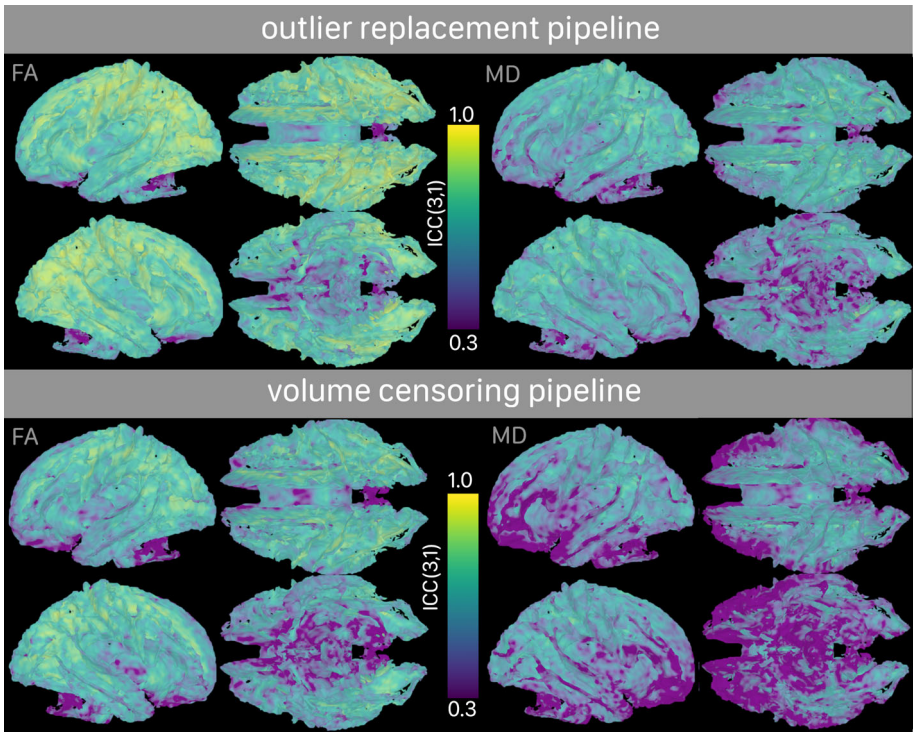

FIGURE 3 3D rendering of test–retest reliability (ICC(3,1)) maps of tract-based spatial statistics (TBSS) skeleton for fractional anisotropy (FA, left) and mean diffusivity (MD, right), calculated with 67 DTI images of 5-year-old subjects with two preprocessing pipelines: outlier replacement pipeline (top) and volume censoring pipeline (bottom)

In contrast to the high ICC(3,1) values found in the ROI analyses, the analysis in the TBSS skeleton voxels yielded a wider range of ICC (3,1) values (see Figure 3 and Supplementary Material 4). With the outlier replacement pipeline, the percentage of good to excellent voxel ICC values (ICC(3,1) > 0.75) in the skeleton were 50, 14, 26, and 37%; for moderate voxel ICC values (0.5 < ICC(3,1) < 0.75) they were 40, 59, 51, and 47% for FA, MD, AD, and RD, respectively (see Figure 5). With the volume censoring pipeline the percentage of good to excellent voxel ICC values (ICC(3,1) > 0.75) in skeleton were 18, 3, 4, and 14% and the moderate voxel ICC values (0.5 < ICC(3,1) < 0.75) were 52, 41, 43, and 49% for FA, MD, AD, and RD, respectively (see Figure 5). Test–retest variability remained low in both pipelines (see Figure 4 and Supplementary Material 5). Similar to the ROI analyses, most of the data points remained under 10% for all scalars both with the outlier replacement pipeline (percentage of data points in the skeleton were 68, 94, 94, and 84% for FA, MD, AD, and RD, respectively) and with the volume censoring pipeline (percentage of data points in the skeleton were 62, 85, 85, and 76% for FA, MD, AD, and RD, respectively) (see Figure 5).

3.3 | Subject motion and DTI scalars

The correlations were investigated with a nonparametric analysis between the motion observed during motion correction and the final DTI scalars preprocessed with the volume censoring pipeline for both test and retest. The nonlinear association was significant only between the absolute total derivative of the translation along the y axis and the mean of the AD values in the BCC ($R^2 = .22$) in the test subset. The mean (SD) absolute motion estimated in the eddy current and motion correction step in the x, y, and z directions across the 67 cases was 0.568 (0.746) mm, 0.308 (0.341) mm, and 0.311 (0.228) mm, respectively, and the mean (SD) rotations in degrees 0.957 (0.584)°, 1.097 (0.793)°, and 1.290 (1.331)°, respectively.

4 | DISCUSSION

Repeatability in DTI studies for specific age groups is of utmost importance because it provides a tool to make accurate interpretations about the changes that occur in neurodevelopmental processes (Madhyastha et al., 2014). To the best of our knowledge, this study is the first in-depth evaluation on DTI repeatability for 5-year-old subjects. We evaluated the intrascan repeatability of DTI scalars across voxels in the TBSS skeletons and in the ROIs created by masking the TBSS skeletons with a WM atlas in 5-year-olds. The two preprocessing pipelines representing the two approaches used to deal with the volumes of bad quality were evaluated: a volume censoring pipeline which discards the corrupted diffusion volumes, and an outlier replacement pipeline which replaces the corrupted volumes with non-parametric predictions. The data used after the volume censoring had, in total, at least 40 unique diffusion encoding directions per subject with at least 20 directions per test and a minimum of 20 directions per retest, however, the data used in the outlier replacement pipeline retained all the diffusion encoding directions. In addition, we applied two thresholding values to the TBSS skeleton and our findings support the results of a previous study (Madhyastha et al., 2014) that the small increase in the thresholding value in the TBSS process did not have an impact on the repeatability. Evidence suggest that an even more conservative thresholding value (0.4) must be applied to

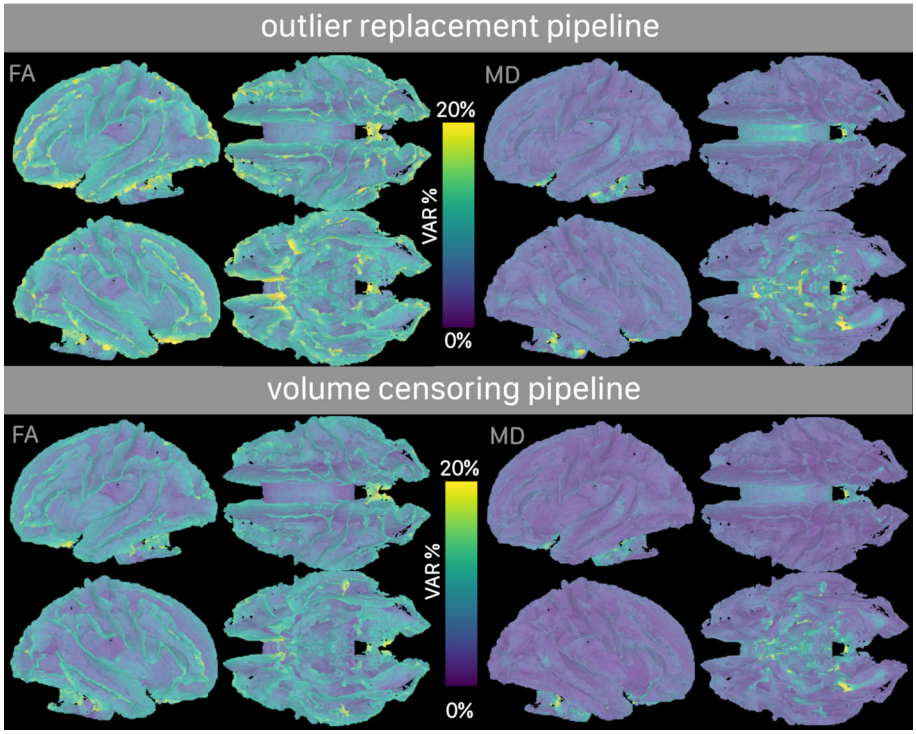

FIGURE 4 3D rendering of test–retest variability (VAR%) maps of tract-based spatial statistics (TBSS) skeleton for fractional anisotropy (FA, left) and mean diffusivity (MD, right), calculated with 67 DTI images of 5-year-old subjects with two preprocessing pipelines: outlier replacement pipeline (top) and volume censoring pipeline (bottom)

demonstrate improved ICC(3,1) values but at the risk of losing specific tracts all together (Boekel et al., 2017).

Despite the fact that 5-year-olds move more than adults in the scanner (Meissner et al., 2020; Theys et al., 2014) and that the rapid WM maturation period has just come to an end (Lebel et al., 2017), in the ROI analysis, we observed a generally low variability together with a high reliability, indicating a good test–retest repeatability of measurements. Our ICC(3,1) values gathered from the ROI analysis are similar to or higher than other studies using the same JHU atlas to measure intersession test–retest reliability in adults (Buimer et al., 2020; Duan et al., 2015).

The CC is a large and compact WM fiber bundle that has homogenous directionality, that is, without curves or crossing fibers, and consequently, has high anisotropy (Anand et al., 2019; Fabri, 2014; Hasan et al., 2005). For this reason, the CC can be considered as a good reference region in repeatability studies, and where achieving high repeatability is expected. We observed good to excellent test–retest reliability for all scalars in the CC except in one incidence where the AD values had moderate reliability in the BCC with the volume censoring pipeline. This moderate reliability of AD values in the BCC may be partially explained by the effect of motion residue since we found significant correlation between the mean AD values in the ROI and the motion metrics.

Voxelwise analysis with the TBSS skeleton gave additional insight into how the ICC(3,1) and the VAR varied across locations in the skeleton. Our voxelwise analysis complements another study that investigated the intersession reliability with adults in which the voxelwise analysis yielded poorer reliability than the ROI analysis (Luque Laguna et al., 2020). The ICC(3,1) values were lower and, even though the difference was negligible, the VAR was somewhat higher than in the ROI analysis. This was expected, since the analysis was done in the TBSS skeleton voxels which give a more in-detail view of the WM without averaging the voxels inside the ROI. We speculate that the thin structure and possible coregistration errors contribute to the low repeatability of individual voxels within the TBSS skeleton. However, as these locations are few in comparison to the total number of voxels (see Figure 5), we consider that in the future the overall voxel-level repeatability in the TBSS skeleton will be suitable for voxelwise statistics and other voxelwise analysis such as radiomics.

Overall repeatability was found to be higher when the outlier replacement pipeline was used compared to the approach that includes removal of bad quality volumes. To ensure the authenticity of the results, we used the eddy QC tools Quality Assessment for DMRI (QUAD) and Study-wise QUAD (Bastiani et al., 2019) to check the percentage of the slices classified as outliers in the data. The percentage of slices labeled as outliers in the data were below 6% and it was under 2% for most of the subjects. Hence, the repeatability is not expected to be due to a bias caused by the analysis process.

We found only one significant correlation between DTI scalars and motion metrics, which was the translation around the y axis (anterior–posterior direction). It has been speculated that the association between DTI scalars and the movement along the y axis is due to the motion induced susceptibility artifacts in the phase encoding direction used in the data acquisition (Merisaari et al., 2019). The particular sensitivity of AD as a scalar may further explain the correlation. Although the association between motion metrics and DTI scalars was trivial overall, it is advisable to consider residual motion after motion correction in DTI studies, especially when the study involves more sensitive metrics, such as AD.

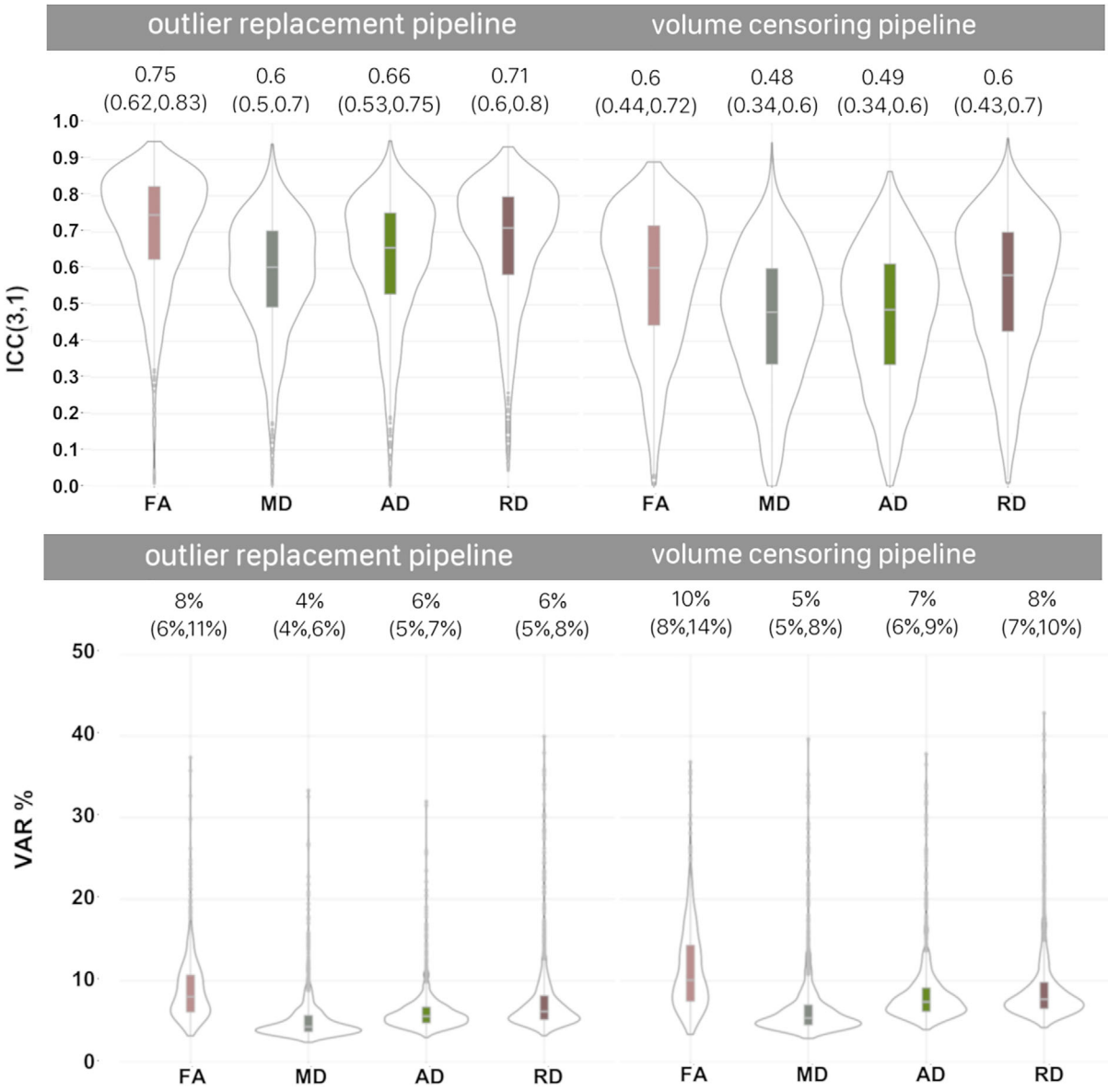

FIGURE 5 Test–retest reliability (ICC(3,1)) and variability (VAR%) of diffusion tensor imaging (DTI) scalars for all voxels inside the tract-based spatial statistics (TBSS) skeleton preprocessed with the outlier replacement pipeline and the volume censoring pipeline. Median (interquartile range) per scalar are presented the top of each violin plot

This study had a number of limitations. The number of evaluated cases was somewhat limited due to the requirement of at least 20 directions in at least two subsets in the volume censoring pipeline. This requirement was considered to simulate a data acquisition compatible with majority of both contemporary and earlier DTI studies around this age. Investigating how the number of directions affects the repeatability of the DTI scalars will be the responsibility of future studies. It remains important to assess similar measurements using different preprocessing pipelines in other populations and other sequences in order to define how well they generalize. In addition, other metrics, such as tractography metrics, were not evaluated here and may express different repeatability depending on how robust the metrics are that they provide.

Using an atlas standardizes and automizes the analysis especially for healthy population (Faria et al., 2010). Since the anatomical restrictions to drawing the boundaries for ROIs are not always available (Mori et al., 2008), the WM atlases include both anatomical and hypothetical boundaries and this causes the differences between the atlases. We leave investigating the effect of different atlases on test–retest repeatability to a future study.

It is important to note that the results obtained from the current study only apply to intrascan repeatability, and not interscan repeatability, nor repeatability across scanner models or imaging sites. However, with very few studies on the topic, we feel the current study has provided a promising first step for future studies that specifically address the repeatability between scan sessions. In future studies, for

instance, ICC(2,1) for interscanner/intersite reproducibility could be investigated. We only evaluated the DTI scalars; we hope that the repeatability of other analysis techniques with DTI data will be the subject of future studies. While differences in the pipelines regarding susceptibility correction and b0 averaging hinder direct method comparison, the evaluation of two different pipelines provided a more exhaustive coverage of repeatability, which was the focus of this study. Direct methodological comparison between two pipelines and tools is a crucially important issue for future studies.

5 | CONCLUSION

We conclude that the DTI scalars in the data of 5-year-olds, analyzed with TBSS, regardless of the preprocessing pipeline have high intrascan repeatability at the ROI-level and moderate to high intrascan repeatability at the voxel-level. The outlier replacement pipeline yielded a more robust test–retest reliability and a lower test–retest variability compared to the volume censoring pipeline. Conservative thresholding in TBSS did not impact repeatability. It is recommended that the effect of motion during acquisition on the final DTI scalars is considered as a confounding effect especially when using AD. The results acquired in the study can be used as a baseline assessment of reliability for future DTI studies with a similar age group.

ACKNOWLEDGMENTS

The authors would like to thank the families and the research staff in FinnBrain Birth Cohort Study and Turku University Hospital. This research was funded by Jane and Aatos Erkko Foundation (H. K.); Hospital District of Southwest Finland State Research Grants (A. R., L. K., H. K., H. M., J. T.); Finnish Medical Foundation (J. T.); Emil Aaltonen Foundation (J. T.); Sigrid Jusélius Foundation (J. T., A. R.); Cultural Foundation of Finland (V. K., H. M.); Signe and Ane Gyllenberg Foundation (L. K., H. K., M. L.); Academy of Finland (L. K. [#325292 Profi 5, #308176], H. K. [#134950, #253270]), H. M. (#26080983); Juho Vainio Foundation (E. S.); Finnish Brain Foundation (E. S.); and Turunmaa Duodecim Society (E. S.).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The current Finnish legislation and our Ethical Board approval do not permit the open data sharing of the imaging data or derived measures. The analysis code can be made available upon request to the corresponding author.

ORCID

Aylin Rosberg https://orcid.org/0000-0002-4587-4882
Anni Copeland https://orcid.org/0000-0002-6482-9008
John D. Lewis https://orcid.org/0000-0002-0832-7396
Harri Merisaari https://orcid.org/0000-0002-8515-5399

REFERENCES

Alhamud, A., Taylor, P. A., Laughton, B., Van Der Kouwe, A. J. W., & Meintjes, E. M. (2015). Motion artifact reduction in pediatric diffusion tensor imaging using fast prospective correction. *Journal of Magnetic Resonance Imaging*, *41*, 1353–1364. https://doi.org/10.1002/jmri.24678

Anand, C., Brandmaier, A. M., Arshad, M., Lynn, J., Stanley, J. A., & Raz, N. (2019). White-matter microstructural properties of the corpus callosum: Test–retest and repositioning effects in two parcellation schemes. *Brain Structure & Function*, *224*, 3373–3385. https://doi.org/10.1007/s00429-019-01981-y

Andersson, J. L. R., Graham, M. S., Drobnjak, I., Zhang, H., & Campbell, J. (2018). Susceptibility-induced distortion that varies due to motion: Correction in diffusion MR without acquiring additional data. *NeuroImage*, *171*, 277–295. https://doi.org/10.1016/j.neuroimage.2017.12.040

Andersson, J. L. R., Graham, M. S., Zsoldos, E., & Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, *141*, 556–572. https://doi.org/10.1016/j.neuroimage.2016.06.058

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*, 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078. https://doi.org/10.1016/j.neuroimage.2015.10.019

Basser, P. J., & Jones, D. K. (2002). Diffusion-tensor MRI: Theory, experimental design and data analysis—A technical review. *NMR in Biomedicine*, *15*, 456–467. https://doi.org/10.1002/nbm.783

Bastiani, M., Cottaar, M., Fitzgibbon, S. P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S. N., Jbabdi, S., & Andersson, J. L. R. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage*, *184*, 801–812. https://doi.org/10.1016/j.neuroimage.2018.09.073

Boekel, W., Forstmann, B. U., & Keuken, M. C. (2017). A test-retest reliability analysis of diffusion measures of white matter tracts relevant for cognitive control. *Psychophysiology*, *54*, 24–33. https://doi.org/10.1111/psyp.12769

Buimer, E. E. L., Pas, P., Brouwer, R. M., Froeling, M., Hoogduin, H., Leemans, A., Luijten, P., van Nierop, B. J., Raemaekers, M., Schnack, H. G., Teeuw, J., Vink, M., Visser, F., Hulshoff Pol, H. E., & Mandl, R. C. W. (2020). The YOUth cohort study: MRI protocol and test-retest reliability in adults. *Developmental Cognitive Neuroscience*, *45*, 100816. https://doi.org/10.1016/j.dcn.2020.100816

Copeland, A., Silver, E., Korja, R., Lehtola, S. J., Merisaari, H., Saukko, E., Sinisalo, S., Saunavaara, J., Lähdesmäki, T., Parkkola, R., Nolvi, S., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2021). Infant and child MRI: A review of scanning procedures. *Frontiers in Neuroscience*, *15*, 1–16. https://doi.org/10.3389/fnins.2021.666020

Croteau-Chonka, E. C., Dean, D. C., Remer, J., Dirks, H., O'Muircheartaigh, J., Deoni, S. C. L. L., Muircheartaigh, J. O., & Deoni, S. C. L. L. (2016). Examining the relationships between cortical maturation and white matter myelination throughout early childhood. *NeuroImage*, *125*, 413–421. https://doi.org/10.1016/j.neuroimage.2015.10.038

Dekaban, A. S., & Sadowsky, D. (1978). Changes in brain weights during the span of human life: Relation of brain weights to body heights and body weights. *Annals of Neurology*, *4*, 345–356. https://doi.org/10.1002/ana.410040410

Duan, F., Zhao, T., He, Y., & Shu, N. (2015). Test-retest reliability of diffusion measures in cerebral white matter: A multiband diffusion MRI

study. *Journal of Magnetic Resonance Imaging*, *42*, 1106–1116. https://doi.org/10.1002/jmri.24859

Fabri, M. (2014). Functional topography of the corpus callosum investigated by DTI and fMRI. *World Journal of Radiology*, *6*, 895–906. https://doi.org/10.4329/wjr.v6.i12.895

Faria, A. V., Zhang, J., Oishi, K., Li, X., Jiang, H., Akhter, K., Hermoye, L., Lee, S., Hoon, A., Stashinko, E., Miller, M. I., van Zijl, P. C. M., & Mori, S. (2010). Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection. *NeuroImage*, *52*, 415–428. https://doi.org/10.1016/j.neuroimage.2010.04.238

Hasan, K. M., Gupta, R. K., Santos, R. M., Wolinsky, J. S., & Narayana, P. A. (2005). Diffusion tensor fractional anisotropy of the normal-appearing seven segments of the corpus callosum in healthy adults and relapsing-remitting multiple sclerosis patients. *Journal of Magnetic Resonance Imaging*, *21*, 735–743. https://doi.org/10.1002/jmri.20296

Hermoye, L., Saint-Martin, C., Cosnard, G., Lee, S. K., Kim, J., Nassogne, M. C., Menten, R., Clapuyt, P., Donohue, P. K., Hua, K., Wakana, S., Jiang, H., Van Zijl, P. C. M., & Mori, S. (2006). Pediatric diffusion tensor imaging: Normal database and observation of the white matter maturation in early childhood. *NeuroImage*, *29*, 493–504. https://doi.org/10.1016/j.neuroimage.2005.08.017

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jones, D. K., & Cercignani, M. (2010). Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR in Biomedicine*, *23*, 803–820. https://doi.org/10.1002/nbm.1543

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H. M., Korja, R., Ekholm, E., Tuulari, J. J., Pajulo, M., Huotilainen, M., Paunio, T., & Karlsson, H. (2018). Cohort profile: The FinnBrain Birth Cohort Study (FinnBrain). *International Journal of Epidemiology*, *47*, 15–16j. https://doi.org/10.1093/ije/dyx173

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kreilkamp, B. A. K., Zacà, D., Papinutto, N., & Jovicich, J. (2016). Retrospective head motion correction approaches for diffusion tensor imaging: Effects of preprocessing choices on biases and reproducibility of scalar diffusion metrics. *Journal of Magnetic Resonance Imaging*, *43*, 99–106. https://doi.org/10.1002/jmri.24965

Le Bihan, D., & Johansen-Berg, H. (2012). Diffusion MRI at 25: Exploring brain tissue structure and function. *NeuroImage*, *6*, 324–341. https://doi.org/10.1016/j.neuroimage.2011.11.006

Lebel, C., & Deoni, S. (2018). The development of brain white matter microstructure. *NeuroImage*, *182*, 207–218. https://doi.org/10.1016/j.neuroimage.2017.12.097

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*, 340–352. https://doi.org/10.1016/j.neuroimage.2011.11.094

Lebel, C., Treit, S., & Beaulieu, C. (2017). A review of diffusion MRI of typical white matter development from early childhood to young adulthood. *NMR in Biomedicine*, *23*, e3778. https://doi.org/10.1002/nbm.3778

Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, *40*, 1044–1055. https://doi.org/10.1016/j.neuroimage.2007.12.053

Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews*, *30*, 718–729. https://doi.org/10.1016/j.neubiorev.2006.06.001

Long, X., Benischek, A., Dewey, D., & Lebel, C. (2017). Age-related functional brain changes in young children. *NeuroImage*, *155*, 322–330. https://doi.org/10.1016/j.neuroimage.2017.04.059

Luque Laguna, P. A., Combes, A. J. E., Streffer, J., Einstein, S., Timmers, M., Williams, S. C. R., & Dell'Acqua, F. (2020). Reproducibility, reliability and variability of FA and MD in the older healthy population: A test-retest multiparametric analysis. *NeuroImage: Clinical*, *26*, 102168. https://doi.org/10.1016/j.nicl.2020.102168

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

Meissner, T. W., Walbrin, J., Nordt, M., Koldewyn, K., & Weigelt, S. (2020). Head motion during fMRI tasks is reduced in children and adults if participants take breaks. *Developmental Cognitive Neuroscience*, *44*, 100803. https://doi.org/10.1016/j.dcn.2020.100803

Merisaari, H., Tuulari, J. J., Karlsson, L., Scheinin, N. M., Parkkola, R., Saunavaara, J., Lähdesmäki, T., Lehtola, S. J., Keskinen, M., Lewis, J. D., Evans, A. C., & Karlsson, H. (2019). Test-retest reliability of diffusion tensor imaging metrics in neonates. *NeuroImage*, *197*, 598–607. https://doi.org/10.1016/j.neuroimage.2019.04.067

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A. V., Mahmood, A., Woods, R., Toga, A. W., Pike, G. B., Neto, P. R., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P., & Mazziotta, J. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *NeuroImage*, *40*, 570–582. https://doi.org/10.1016/j.neuroimage.2007.12.035

Mori, S., & Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, *51*, 527–539. https://doi.org/10.1016/j.neuron.2006.08.012

Muircheartaigh, J. O., Iii, D. C. D., Ginestet, C. E., Walker, L., Waskiewicz, N., Lehman, K., Dirks, H., Piryatinsky, I., & Deoni, S. C. L. (2014). White matter development and early cognition in babies and toddlers. *Human Brain Mapping*, *4487*, 4475–4487. https://doi.org/10.1002/hbm.22488

Oguz, I., Farzinfar, M., Matsui, J., Budin, F., Liu, Z., Gerig, G., Johnson, H. J., & Styner, M. (2014). DTIPrep: Quality control of diffusion-weighted images. *Frontiers in Neuroinformatics*, *8*, 1–11. https://doi.org/10.3389/fninf.2014.00004

Pulli, E. P., Silver, E., Kumpulainen, V., Copeland, A., Merisaari, H., Saunavaara, J., Parkkola, R., Lähdesmäki, T., Saukko, E., Nolvi, S., Kataja, E.-L., Korja, R., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2021). Feasibility of FreeSurfer processing for T1-weighted brain images of 5-year-olds: Semiautomated protocol of FinnBrain neuroimaging lab. bioRxiv, 20212021.05.25.445419.

Schilling, K. G., Blaber, J., Huo, Y., Newton, A., Hansen, C., Shafer, A. T., Williams, O., Resnick, S. M., Rogers, B., & Adam, W. (2019). Synthesized b0 for diffusion distortion correction (Synb0-DisCo). *Magnetic Resonance Imaging*, *64*, 62–70. https://doi.org/10.1016/j.mri.2019.05.008

Shen, Y., Larkman, D. J., Counsell, S., Pu, I. M., Edwards, D., & Hajnal, J. V. (2004). Correction of high-order eddy current induced geometric distortion in diffusion-weighted echo-planar images. *Magnetic Resonance in Medicine*, *52*, 1184–1189. https://doi.org/10.1002/mrm.20267

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

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

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*, 1487–1505. https://doi.org/10.1016/j.neuroimage.2006.02.024

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M.,

Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, 208–219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Tamnes, C. K., Roalf, D. R., Goddings, A. L., & Lebel, C. (2018). Diffusion MRI of white matter microstructure development in childhood and adolescence: Methods, challenges and progress. *Developmental Cognitive Neuroscience*, *33*, 161–175. https://doi.org/10.1016/j.dcn.2017.12.002

Taylor, P. A., Alhamud, A., van der Kouwe, A., Saleh, M. G., Laughton, B., & Meintjes, E. (2016). Assessing the performance of different DTI motion correction strategies in the presence of EPI distortion correction. *Human Brain Mapping*, *37*, 4405–4424. https://doi.org/10.1002/hbm.23318.Assessing

Theys, C., Wouters, J., & Ghesquière, P. (2014). Diffusion tensor imaging and resting-state functional MRI-scanning in 5- and 6-year-old children: Training protocol and motion assessment. *PLoS One*, *9*, 1–7. https://doi.org/10.1371/journal.pone.0094019

Yoshida, S., Oishi, K., Faria, A. V., & Mori, S. (2013). Diffusion tensor imaging of normal brain development. *Pediatric Radiology*, *43*, 15–27. https://doi.org/10.1007/s00247-012-2496-x

Yuan, W., Altaye, M., Ret, J., Schmithorst, V., Byars, A. W., Plante, E., & Holland, S. K. (2009). Quantification of head motion in children during various fMRI language tasks. *Human Brain Mapping*, *30*, 1481–1489. https://doi.org/10.1002/hbm.20616

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rosberg, A., Tuulari, J. J., Kumpulainen, V., Lukkarinen, M., Pulli, E. P., Silver, E., Copeland, A., Saukko, E., Saunavaara, J., Lewis, J. D., Karlsson, L., Karlsson, H., & Merisaari, H. (2022). Test–retest reliability of diffusion tensor imaging scalars in 5-year-olds. *Human Brain Mapping*, 1–11. https://doi.org/10.1002/hbm.26064