TEST-RETEST RELIABILITY OF FRACTIONAL ANISOTROPY IN 5-YEAR-OLDS

Aylin ROSBERG

Master's Degree Program in Human Neuroscience

Faculty of Medicine, University of Turku

May 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

MRI	Magnetic Resonance Imaging
WM	White Matter
DTI	Diffusion Tensor Imaging
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
ROI	Region of Interest
sMRI	structural Magnetic Resonance Imaging
GM	Gray Matter
DWI	Diffusion Weighted Imaging
D	Diffusion Constant
CSF	Cerebrospinal Fluid
ADC	Apparent Diffusion Coefficient
MD	Mean Diffusivity
AD	Axial Diffusivity
RD	Radial Diffusivity
VBA	Voxel-Based Analysis
SE-EPI	Spin Echo-Echo Planar Imaging
FOV	Field of View
TR	Repetition Time
TE	Echo Time
FSL	FMRIB Software Library
BET	Brain Extraction Tool
TBSS	Tract-Based Spatial Statistics
ICBM	International Consortium of Brain Mapping
MNI	Montreal Neurological Institute
ICC	Intraclass Correlation Coefficient
CI	Confidence Interval
SEM	Standard Error of Measurement
LoA	Limits of Agreement
CC	Corpus Callosum
GCC	Genu of Corpus Callosum
SCC	Splenium of Corpus Callosum
BCC	Body of Corpus Callosum
ACR	Anterior Corona Radiata
PTR	Posterior Thalamic Radiation
SFOF	Superior Fronto-Occipital Fasciculus
PLIC	Posterior Limb of Internal Capsule
PCR	Posterior Corona Radiata

ABSTRACT

Diffusion tensor imaging (DTI) has provided great insights to the microstructural features of developing brain and has been shown to be reliable in infants. However, the repeatability of the DTI scalars for older pediatric age groups has not been thoroughly addressed. In this study, DTI scans of 5-year-olds were used to investigate the test-retest reliability of three different measurements with both voxel-wise and region of interest (ROI) analysis. Out of 96 diffusion encoding directions, divided into three parts, 20 unique diffusion encoding directions were chosen per measurement from 48 subjects. Tract based spatial analysis (TBSS) was used to extract fractional anisotropy (FA) values from those images and using the FA values the repeatability of the measurements was assessed by intraclass correlation coefficient (ICC) and standard error of measurement (SEM). Overall, FA values had high repeatability both in voxel-based analysis (ICC>0.73) and ROI analysis (for non-skeletonized ROI type 88% of the ROI labels: ICC>0.75, for skeletonized ROI type 87% of the ROI labels: ICC>0.75). Using a skeleton in the ROI analysis did not contribute to the repeatability and the volume size was found to be a contributing factor for repeatability. Interscanner reliability as well as reliability measured by using different atlases are yet to be investigated in 5-year-old data.

1 INTRODUCTION

Prior to magnetic resonance imaging (MRI) technology, the knowledge on brain anatomy was based on postmortem studies and animal studies. However, these studies have important limitations such as small sample size, sample preservation difficulties, no possibility to do follow up measurements and lack of quantification and thus lack of statistical comparison opportunities (Ferrer, Martinez, Boluda, Parchi, & Barrachina, 2008; Innocenti, Ansermet, & Parnas, 2003). MRI does not have these limitations and it provides functional and structural information about the brain in great detail. In addition, MRI is safe and non-invasive. Therefore, MRI was revolutionary when it was presented to the field of neuroscience in 1980's. Since then it has allowed researchers to quantitatively investigate the different types of tissues in brain (Cascio, Gerig, & Piven, 2007).

Diffusion tensor imaging (DTI) is a special MR imaging technique that allows unparalleled insights to the white matter microstructure (Paus et al., 2001). Most modern DTI sequences are based on echo planar imaging (EPI) since the images have relatively high spatial resolution and relatively low amount of motion artifacts (Poynton, Jenkinson, Whalen, Golby, & Wells III, 2008). However, EPI sequences are highly prone to distortions caused by eddy-currents and magnetic susceptibility of the cranial structures (brain, sinuses, bone). Eddy-currents occur due to the changing diffusion gradients of the scanner and even though they can be corrected to a certain level during acquisition, residual artifacts are common in the phase encoding direction (Shen et al., 2004). In addition to eddy-current artifacts, EPI images are prone to magnetic susceptibility effects which lead to signal loss and distortion in the images especially in the areas where the tissue meets the air (Basser & Jones, 2002).

Diffusion images are acquired based on the water motion in tissue, a small signal, and because of that they are also very sensitive to subject motion (Taylor et al., 2016). Subject movement during scanning results in blurring and ghosting in the image. During a DTI scan

multiple images are taken and the subject movement between image acquisitions result in coregisteration problems during the processing of the images (Mori & Tournier, 2014). In pediatric imaging, it is more critical to be aware of the challenges mentioned above and focus on the quality and the repeatability of the data. Previous studies, including experience from FinnBrain Birth Cohort Study (Copeland et al., in revision) suggest that children under 5 years old have difficulties in cooperating during the imaging process and have a hard time being still in the scanner (Yoshida et al., 2013). Considering the motion sensitivity of DTI, it is argued that sedation is required for younger children (Hermoye et al., 2006). Yet, sedation is neither recommended nor allowed in research setting (Copeland et al., in revision). Hence, as young children move during scanning, that motion shows up as ghosting or causes disturbances in signal intensities in the images. These artifacts could easily affect the interpretation of the DTI results in small children Therefore, to fully benefit from DTI in small children, the technique needs to be applied carefully and the acquired images need to be processed with utmost caution to assure generalizability of the results and accurate interpretation of the data (Jones & Cercignani, 2010).

In addition, collecting good quality data and repeatable results are critical also for practical sides of the research process. Repeating a scan, if a previous one fails, is not an easy task, and must be avoided as much as possible for the cost and effort goes into each scan for each subject. It is also crucial to have reliable baseline data for any longitudinal study for its benefits of allowing follow-up study possibilities, particularly when the inevitable decrease in number of subjects over time considered.

Taken together, it can be claimed that the quality and repeatability of the data, especially in pediatric population, is of utmost importance. Not only for the short term goal of getting accurate information from data of the current study but also for the long term goals of

understanding developing brain thoroughly and accumulating reliable literature for future research to refer back to and to take as a base level to build up on.

A previous study investigated the intrascan repeatability in neonates by analyzing and comparing the quantitative measures of DTI. The study included DTI scans of the neonates from 3 measurements with in total 96 different diffusion encoding directions. DTI scalars were extracted for all three measurements and test-retest repeatability of the scalars were analyzed in pairs by investigating the correlation between measurements and the measurement error (Merisaari et al., 2019). However, that study was only focusing on the neonate data.

Anatomically, age 5 is particularly critical in brain development. By age 5 the brain gets size-wise very close to the adult brain and any specific changes in the brain cannot be attributed to any age point after age 5 (Yoshida, Oishi, Faria, & Mori, 2013). Evidence suggest that there is a profound cortical growth and myelination affecting the white matter structures in the first years of life (Croteau-Chonka et al., 2016; Muircheartaigh et al., 2014). Therefore, the 5-year-old brain is quantitatively different than the small and unmyelinated neonate brain. In addition, compared to neonates, who slept through the scans, 5-year-olds potentially have more motion during scanning. Therefore, it is worth studying the test-retest reliability of the data acquired in this transitional stage of human brain development. The current study investigated the test-retest reliability of the 5-year-old children data using a similar approach as Merisaari and colleagues.

1.1 Magnetic Resonance Imaging and Neurodevelopment

Conventional structural MRI (sMRI) provides three dimensional, anatomical images of the brain (Mori & Zhang, 2006) and sMRI methods have improved our understanding of the anatomical neurodevelopment of humans starting from prenatal period with some studies

assessing even prenatal development (Dean et al., 2014; Girault et al., 2019; Tamnes et al., 2017; Tocchio, Kline-Fath, Kanal, Schmithorst, & Panigrahy, 2015).

sMRI studies show that the first 2 years in life are critical in neurodevelopment, as the brain goes through major changes and growth during this time (Knickmeyer et al., 2008). The whole brain volume doubles in size by the time the child reaches 2 years of age, reaching around 80% of an adult brain's size and weight (Yoshida, Oishi, Faria, & Mori, 2013). Even though the total brain volume gets close to the size of an adult brain quite early in life, the volumetric changes continues through childhood and adolescence. It is worth noting that, the growth does not happen linearly or with the same proportions throughout the brain. Whilst some structures such as the cerebellum almost triple in size, the size of caudate nucleus only increases by around 20% of its original size. The lateral ventricles seem to grow in size in the first year but then get smaller during the second year. Also, the overall growth seems to be more profound for gray matter (GM) at about 150% than it is for white matter (WM) at about 10% in the first year (Knickmeyer et al., 2008; Pfefferbaum et al., 1994).

Growth patterns of GM changes throughout the brain. For instance, temporal lobe volume shows periods of decrease and increase in different ages in childhood whilst frontal lobe volume peaks in adolescence. Another interesting finding about neurodevelopment of brain tissues is that while GM follows and inverted U shape growth curve, WM increase happens linearly even when the GM in cortex begins to decrease at around 10 years of age (Faria et al., 2010; Lenroot & Giedd, 2006; Richards & Xie, 2015). Neurodevelopment is a complex process and sMRI is a great tool to observe the neurodevelopmental changes on a large scale. Yet, it is not enough to explain the underlying neural mechanisms of the neurodevelopmental processes.

As diffusion weighted imaging (DWI) is now available to gain a better perspective of the microstructural features of the white matter (WM) and the changes WM goes through during neurodevelopment (Reynolds, Grohs, Dewey, & Lebel, 2019).

1.2 Diffusion Weighted Imaging

In its most basic form; diffusion is the movement of any given substance from high concentration to low concentration. In the context of diffusion imaging, diffusion refers to is the random motion of water (Brownian motion) led by the thermal energy caused by the microscopic movement of molecules even without the presence of factor (i.e. concentration, temperature or pressure) effecting the conditions (Baliyan, Das, Sharma, & Gupta, 2016; Rajagopalan et al., 2017). In theory, assuming no limits, water diffusion follows a Gaussian distribution. In this type of diffusion, the matter diffuses in a uniformly restricted, sphere shaped manner (isotropic diffusion; Figure 1). The diffusion constant (D) of that free motion of water at 37°C is calculated with 3.0×10^{-9} m^2/s. Based on this calculation most of the free water moves at least a couple dozen μm in 50 ms (Le Bihan & Iima, 2015). Though the movement is miniscule, this type of diffusion can be observed in the cerebrospinal fluid (CSF) for there are no restrictions to its flow.

In brain tissue, water diffusion is limited by for example fibers and cell membranes. Therefore, the distribution is not Gaussian and the motion of water remains limited to a few μm in 50 ms and random diffusion gains directionality (Le Bihan, 2007). This type of diffusion is called anisotropic diffusion (Figure 1). Anisotropic diffusion is observed in the WM. WM is comprised of the axon tracts and commissures that are covered with a mostly lipid layer called myelin sheath, not only giving the matter a lighter appearance compared to GM but also isolating the water content and giving direction to the water diffusivity (Purves et al., 2012).

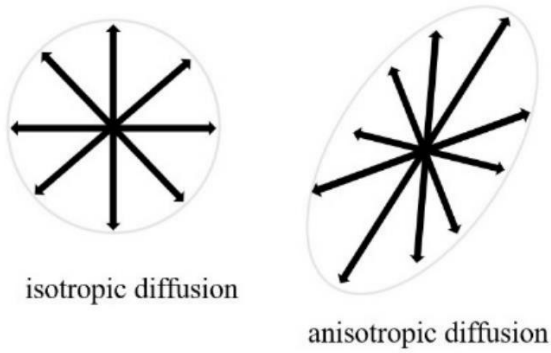

Figure 1. The comparison images for isotropic and anisotropic diffusion.

Diffusion weighted imaging (DWI) utilizes the same properties of the hydrogen nuclei as sMRI does to measure the diffusion of water. It is a technique where the differences in diffusion motion of the water creates the signal to provide information about the organization and orientation of the WM tracts (Baliyan et al., 2016). The main goal of DWI is creating magnetic field inhomogeneity along an axis so that the signal is sensitized to water diffusion and by doing so measuring the water motion in that axis. Magnetic field inhomogeneity is created by applying two different gradients. The gradient polarity changes depending on the sequence used. Regardless the polarity of the gradients when the first gradient is applied it disrupts the magnetic field homogeneity and after that the protons based on their locations resonate in different phases than before, hence this gradient is called a dephasing gradient, and this creates a signal loss. Following the dephasing gradient, based on the sequence used another gradient is applied. The second gradient rephases the protons, hence it is called a rephasing gradient (Mori & Tournier, 2014). The signal attenuation in this approach is calculated with a formula based on the diffusion coefficient. This process creates the diffusion weighted images. It is worthy of note that, the measured signal is not solely influenced by diffusion (Le Bihan & Johansen-Berg, 2012).

However, as explained above brain tissue enables a more complicated version of diffusion and the diffusion coefficient (D) is not enough to explain the non-Gaussian

distribution of the water motion. Therefore, D is replaced by another diffusion parameter called apparent diffusion coefficient (ADC). This parameter requires at least two diffusion weighting application to acquire quantitative diffusion images where signal attenuation is purely due to the diffusion sensitization (b). The common way to do this is to apply one gradient without any diffusion weighting (b=0 s/mm^2) and another one with high diffusion weighting (b=1000 s/mm^2). Based on the difference between the signal attenuation between different diffusion sensitizations (b-values) ADC enables direct physical interpretations of the brain tissues, especially the highly anisotropic WM tissue (Le Bihan, 2013).

Directionality of the anisotropic diffusion naturally turns the sphere shape of isotropic diffusion into an 3D ellipsoid shape. Since it is no longer possible to estimate the diffusivity with a diffusion constant only, a more detailed model called diffusion tensor model is presented to interpret the diffusivity accurately (Mori, 2007).

1.3 Diffusion Tensor Imaging

Diffusion tensor model consist of a 3×3 array where each number represent the diffusion rate in each combination of directions. DTI estimates the diffusivity in WM structure by utilizing the three gradient units (x, y, z) already existing as a given property of each MRI and their combinations to measure the diffusivity along a given WM tract. To have an accurate diffusion tensor model bare minimum is six parameters; three for the orientations of the three defining axes of the ellipsoid which are called eigenvectors and three for the lengths of those eigenvectors which are called eigenvalues (Drobyshevsky et al., 2004). The visualization of the six parameters defining the ellipsoid is shown in Figure 2. To be able to quantitatively estimate the diffusivity in WM tracts each DTI scanning must include an image without any direction gradient applied to set the baseline and at least six directions to cover all independent elements to acquire sufficient DTI data. The number of directions is increased depending on the analysis method and purpose of the (Catherine Lebel, Benner, & Beaulieu, 2012).

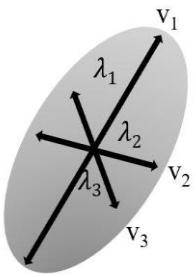

Figure 2. The six parameters, three eigenvectors (v_1, v_2, v_3) and three eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) needed to define the ellipsoid.

1.3.1 DTI scalars

The term DTI scalar is used to refer to the measures of magnitude and directionality of the diffusion. The DTI scalars can be used to quantitatively measure and compare the microstructural features of the brain tissues in a specific region or the whole brain to investigate a wide spectrum of cases including healthy neurodevelopment, aging and diseases caused by axonal and myelin damage (Soares, Marques, Alves, & Sousa, 2013). There are four diffusion scalars; mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA).

Mean Diffusivity

Mean diffusivity (MD) is a measure of diffusion magnitude. It refers to the average of the eigen values in all three directions. Because of the brain maturation processes such as increasing myelination and axonal changes MD tends to decrease throughout the childhood to adolescence (Snook, Plewes, & Beaulieu, 2007).

Axial Diffusivity

Axial diffusivity (AD) is a measure of diffusion magnitude parallel to the axon, it can be evaluated as a measure of axon alignment in fiber bundles (Billiet et al., 2015). It has been frequently used in research focusing on aging, where WM maturity and related axonal damage are studied. Evidence suggest that eventually AD increases in the whole brain with adult aging

yet this happens gradually in different time points and in different areas of the brain depending on the effects of aging (Kumar, Chavez, Macey, Woo, & Harper, 2013).

Radial Diffusivity

Radial diffusivity (RD) is a measure of diffusion magnitude perpendicular to the axon and it is associated with the myelin content in the fiber bundles. Therefore, it is taken as a reliable measure of myelin damage and/or demyelination (Winklewski et al., 2018). Similar to AD, RD also decreases during neurodevelopment because of the increased myelination as brain matures (Partridge et al., 2004).

Fractional Anisotropy

Fractional anisotropy (FA) value shows the directionality of the diffusivity and it is associated with axonal integrity of the fiber bundles. The FA value ranges from 0 to 1, with 0 indicating no directionality and 1 indicating highest amount of directionality. For instance, in an isotropic diffusion scenario, all the eigenvectors are equal thus the FA value would be 0. In an anisotropic diffusion scenario as the asymmetry in the diffusivity increases, the ellipsoid shape gets more prominent and FA value gets closer to 1. From infancy through adolescence FA increases as the brain matures due to the exact same factors causing MD to decrease over time (Moon et al., 2011).

$$\text{FA} = \sqrt{\frac{(\lambda_1-\lambda_2)^2 + (\lambda_2-\lambda_3)^2 + (\lambda_1-\lambda_3)^2}{2(\lambda_1^2+\lambda_2^2+\lambda_3^2)}}$$

In this study, FA was chosen particularly as it has been used as a robust and reliable measure of WM microstructure in test-retest studies as well as numerous other DTI studies conducted with both clinical and healthy population regardless the age group (Grieve, Williams, Paul, Clark, & Gordon, 2007; Heiervang, Behrens, Mackay, Robson, & Johansen-

Berg, 2006; Lewis et al., 2020; Mckinstry & Mathur, 2002; Siasios et al., 2016; Venkatraman et al., 2015; Voldsbekk et al., 2020).

1.4 DTI in Pediatric Population

DTI has been commonly used to study the WM development in healthy and clinical pediatric population including perinatal period (Huber, Henriques, Owen, Rokem, & Yeatman, 2019; Jakab, Tuura, Kellenberger, & Scheer, 2017; C. Lebel et al., 2012; Li et al., 2018; Mckinstry & Mathur, 2002; Young et al., 2018). WM maturation is more profoundly observed in microstructural level than in a gross volume growth as it is observed in GM. In the first two years of life rapid increase rates in FA and decrease rates in MD due to myelination is observed. From late childhood to adolescence there is slower increase in FA and decrease in MD due to the increase in axonal density. More subtle changes continues to happen in specific tracts throughout the lifespan (Lebel, Walker, Leemans, Phillips, & Beaulieu, 2008; Long, Benischek, Dewey, & Lebel, 2017; Qiu, Mori, & Miller, 2015; Reynolds et al., 2019).

Two of the most used diffusion image analysis methods are voxel-based analysis (VBA) and region of interest (ROI) analysis. VBA involves image registration, either smoothing or the use of a skeleton to determine the voxels desired to be in the statistical analysis part. Tract based spatial statistics (TBSS) is a commonly used tool that extends VBA-style analysis. TBSS allows the investigators to create a white matter skeleton that estimates the individual tract centres before further analysis, and it provides robust results, reduces partial volume effect and does not necessitate spatial smoothing (Bach et al., 2014). VBA is a great analysis to use in group comparisons especially when answering a research question without a spatial hypothesis (Tamnes, Roalf, Goddings, & Lebel, 2018; Van Hecke, Emsell, & Sunaert, 2015).

When there is a spatial hypothesis, ROI analysis is more efficient. In the ROI analysis, same idea used for VBA applies within the predefined regions. ROIs can be defined by many factors depending on the research question. For instance, the anatomical boundaries, disease pathologies or results based on other studies can be used to determine the ROIs. This analysis is quite efficient for investigating the DTI scalars in the WM structures. ROIs can be defined manually or based on an atlas. Manual selection requires more effort and time compared to using an atlas, but it might be more useful in clinical populations for anatomical boundaries that may not coincide with the ones in the atlas. Using an atlas to define the ROIs, on the other hand, is useful in healthy populations. By using an atlas, reliable comparison between participants can be done and the time and effort that goes into each analysis are reduced (Faria et al., 2010; Oishi et al., 2008; Van Hecke et al., 2015).

Brain atlases are mostly based on adult brain and thus not well suited for the younger populations. Researchers claim that using adult templates and atlases in small children might be problematic (Machilsen et al., 2007). Due to high variability of age related anatomical changes and the possibility of misclassification of the brain tissues (Fonov et al., 2011; Yoon, Fonov, Perusse, & Evans, 2009).

1.5 Research Aims and Hypotheses

By the nature of magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) method there are artifacts and involuntary motion in the data. There are two main purposes of the current research: 1) to quantitatively investigate the intrascan test-retest reliability of the DTI data of 5-year-old children and 2) to critically reflect to previous test-retest reliability findings. The objectives of the study can be divided into three main parts:

1. to extract the DTI scalars to grasp the microstructural features of white matter (WM) in 5-year-olds a 96 direction DTI sequence in three parts

2. To quantitatively compare the fractional anisotropy (FA) values in between sessions across the whole brain white matter skeleton (voxel-wise approach) and,
3. to quantitatively compare the FA values in between sessions with a region of interest (ROI) based approach.

It is hypothesized that; the extracted FA values should show small amount of variability between measurements and they should have high repeatability in terms of intraclass correlation coefficient (ICC) and standard error of measurement (SEM) values.

2 METHODS

2.1 Participants

The DTI data analysed in this study was acquired by FinnBrain Birth Cohort Study (Karlsson et al., 2018). Data from 48 (23 boys and 25 girls) 5-year-olds (mean age on scan day in days:1945±24) were included in the analyses.

2.2 Data Acquisition

MRI data was acquired using a Siemens Magnetom Verio 3T scanner with 12-element Head Matrix coil in the same centre was used to collect the data. All participants went through 4 different sequences (sagittal T1-MPRAGE, field map acquisition, DTI with single shell and multishell parts, and task and resting state-fMRI during, which the participants viewed inscapes video). One scanning session lasted up to an hour.

The single shell DTI data, used in the current study, was acquired by standard twice-refocused Spin Echo-Echo Planar Imaging (SE-EPI) sequence. The resolution was $2\times2\times2$ mm^3 isotropic resolution (FOV 208 mm; 64 slices; TR 8500 ms; TE 90 ms) and the b-value 1000 s/mm^2. For each participant all three measurements in the DTI sequence included 3 b0 images and each measurement included uniformly distributed 31, 32 or 33 directions overall ended up to 96 unique diffusion encoding directions.

2.3 Ethics

This study was carried out in accordance with the recommendations of the Ethics Committee of the South-Western Hospital District with written informed consent from all mothers. All mothers gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the South-Western Hospital District.

2.4 Data Analysis

2.4.1 Preprocessing

The initial preprocessing of the data was done as per FinnBrain Neuroimaging Lab pipelines (Merisaari et al., 2019). First, good quality b0 images were chosen, coregistered and averaged. FSL's (FMRIB Software Library v 5.0.9; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) Brain Extraction Tool (BET) (Smith, 2002) using -R -f 0.3 flags was employed to create brain masks. Next, DTIPrep software (Oguz et al., 2014) was employed to assure the quantitative quality of the diffusion images. DTIPrep is an open source automatic quality control software developed to detect and correct a wide range of inevitable artifacts such as eddy-currents, motion related artifacts, gradient related artifacts and more (Oguz et al., 2014). Since DTI is susceptible to many artifacts it is highly recommended to use some procedure to assure the quality of the data for any artifact majorly effects the data and the interpretation of it. Based on suggestions of the DTIPrep quality control some directions and volumes were discarded (Merisaari et al., 2019). After the automatic exclusion, 48 subjects with at least 20 direction in each measurement were included in the further analyses. To maximize the number of subjects and number of directions included in the study, differently from the model study, there was not a manual quality control step in addition to DTIPrep. Finally, eddy current and motion correction were done on FSL (Andersson & Sotiropoulos, 2016).

After the preprocessing, the analyses were continued as follows; brain masks were used as the template for DTIFIT, which fits a diffusion tensor model to each voxel determined by the brain mask. The FA images created by DTIFIT were fed into the Tract-Based Spatial Statistics (TBSS) pipeline of FSL (Smith et al., 2006). TBSS pipeline created all participants' mean FA template and mean FA skeleton as a final output (Figure 3). The mean FA images were the averaged versions of the FA images that were aligned to a standard space. TBSS pipeline assumes that the highest FA values are located in the middle of the WM tracts and skeletons were created based on that assumption. The highest FA values were projected onto the skeleton and the skeleton was created (Van Hecke et al., 2015). More information about the TBSS steps / commands and the exact flags used in each step can be found in Table 1. This process prepared the data to further statistical analysis.

Table 1. The TBSS pipeline steps with flags used in the study and the goal task of each step.

TBSS pipeline step	flags	tasks
tbss_1_preproc	-	Copies all originals file to another folder to create a new folder to work on and creates an html for visual quality control (QC).
tbss_2_reg	-T	Aligns all FA images into FMRIB58_FA 1x1x1 MNI standard space through non-linear registration
tbss_3_postreg	-S	Brings all subjects into the standard-space and creates files including all FA files (all_FA), mean of all the FA files (mean_FA) and the skeletonized version of the mean FA file (mean_FA_skeleton).
tbss_4_prestats	0.25	Thresholds skeletonized mean FA image.

The skeletonized FA values were used for VBA. For the ROI analyses, the International Consortium of Brain Mapping (ICBM)-DTI-81 white-matter labels atlas was chosen among the two JHU DTI-based white-matter atlases available in FSL. The atlas includes 48 manually segmented WM tract labels. The standard space used for the segmentation was gathered from 81 healthy adults (mean age 39) (Mori et al., 2008). The atlas was co-registered to the TBSS template so that FA values could be extracted based on the 48 ROIs in the atlas from both skeletonized and non-skeletonized versions.

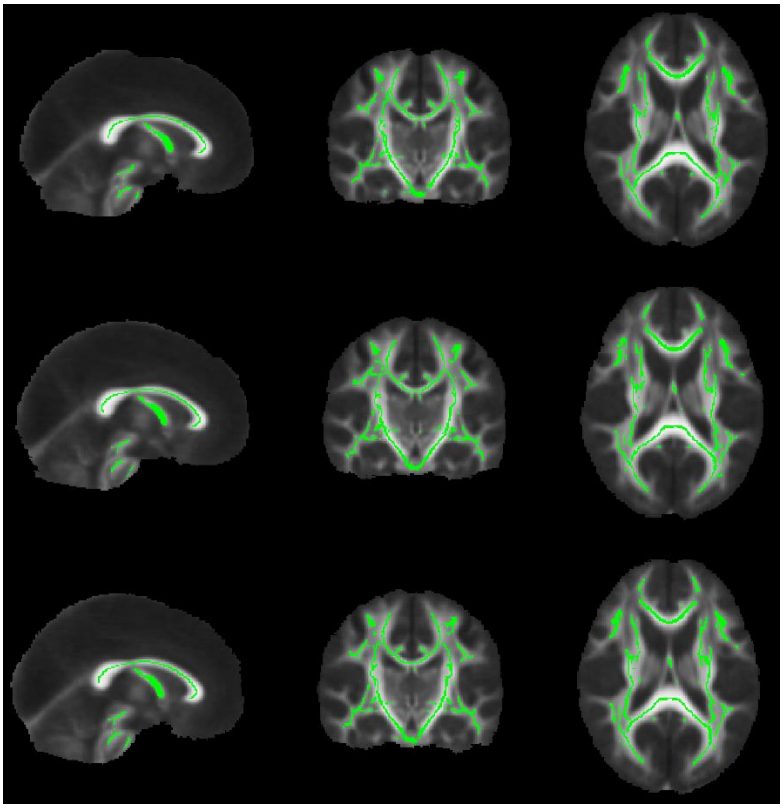

Figure 3. Skeletons (green) overlaid on mean FA images in sagittal, coronal and axial planes (respectively from left column to right column) created by TBSS for measurement repetitions 1, 2 and 3 (respectively from top row to bottom row in).

During the TBSS analysis in-house bash and python scripts were used to warp the atlas on skeletonized and non-skeletonized ROI types and to extract the FA values for whole WM and both ROI types. The statistical analyses with the extracted FA values for three datasets (whole WM, non-skeletonized ROI and skeletonized ROI) were completed in Rstudio (1.2.5033, Rstudio, 2015). Step by step information about the data analyses can be found in the data analysis procedure diagram below (Figure 4).

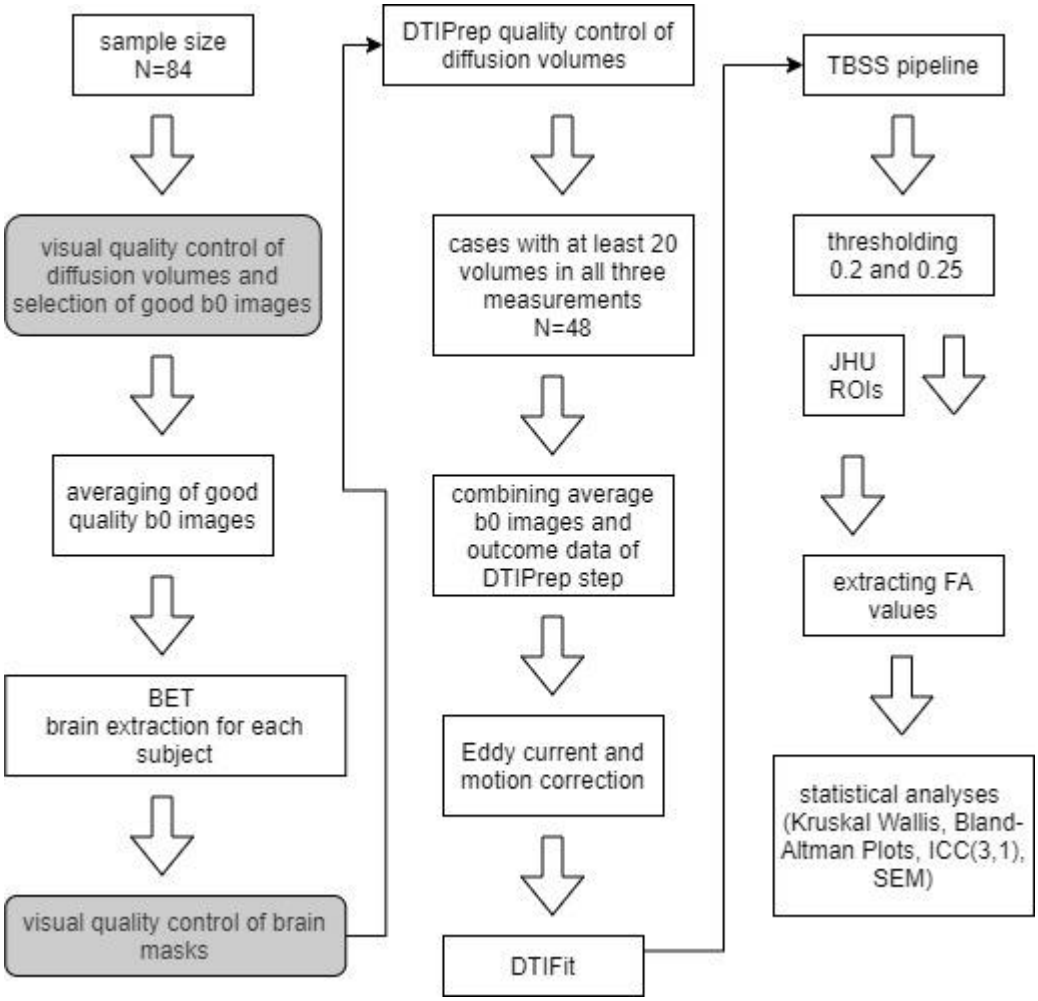

Figure 4. The data analysis procedure. Automated processes are shown in sharp-cornered rectangles and visual control steps are shown in round-cornered rectangles with gray shading.

2.4.2 Statistical analysis

Kruskal-Wallis Test

The Kruskal Wallis Test is a non-parametric method to investigate whether the samples come from the same population (Chan & Walmsley, 1997). The aim of using this testin the study was to determine whether there is a statistically significant difference in FA values extracted from three different repetitions, before moving on to the repeatability analyses.

Bland-Altman Plots

Bland and Altman suggested that since given two measurements are not identical there is bias and difference between them and they came up with a plotting method to quantitatively show

how much the different measures agree (Bland & Altman, 1999). Therefore, the idea of the Bland Altman plots is to check the agreement between the results acquired by two different paired methods, in the current study the agreement between DTI sequence parts, by plotting the paired measurements against the mean of the two measurements. (Giavarina, 2015). The Bland-Altman plot is a great visualization tool revealing a lot of quantitative information about the two given datasets.

When interpreting a Bland-Altman plot there are three main criteria to consider. When these criteria are met, it can be concluded that two methods agree well. The first criterium is, for a quantified predetermined significance level at 0.05, the 95% of the data must be in the data are in the range of ±2 SD of the mean difference. The second one is that the data creating a tube shape distribution and not a cone shape. This means that at any point data have same amount of mean difference and no specific mean difference can be attributed to a specific part of the data. The third criterium is to have the mean difference at 0, which is an indicator of the data not having a systematic bias.

Intraclass Correlation Coefficient (ICC)

ICC is a numerical value, providing information about how similar the given measurements are and hence how reproducible the measurements are. In other words, ICC involves both agreement and correlation between measures. Since one of the aims of the study is to investigate the possible differences between three different measurements acquired under same conditions from same participants, ICC analysis is a great tool achieve that goal. When ICC is calculated; a specific ICC value is generated based on a, preferably 95%, confidence interval (CI). To make a reliable assumption based on the ICC analysis, both CI and ICC value must be taken into account (Koo & Li, 2016). Reliability criteria of ICC value can be found in Table 2.

ICC has several forms that can be employed based on the study design. In the literature it is emphasized that the most appropriate form of ICC to study test-retest reliability is a two-way model (Shrout & Fleiss, 1979). Also, in this study for the reason that repeated measurements acquired were not randomized samples the chosen form of ICC for the analyses was a two-way fixed effects model for single measurement (ICC(3,1)).

Table2. The reliability criteria for ICC(3,1)

ICC(3,1)	reliability
<0.5	poor
0.5-0.75	moderate
0.75-0.9	good
>0.9	excellent

Standard Error of Measurement (SEM)

SEM is another metric that is used to provide quantitative information about the test reliability in measurement unit. It is an absolute number and it shows whether the acquired results are due to a random error or not. SEM is 0 when the test is completely reliable without error and it is equal to standard deviation (SD) when the test is completely not reliable (Harvill, 1991). Hence, the closer SEM value to 0 the more reliable the results are. SEM is calculated with the following formula:

$$\text{SEM} = \text{SD}\sqrt{1-r}$$

SEM =standard error of measurement
SD = standard deviation
r = Cronbach alpha

3 RESULTS

3.1 Whole White Matter Skeleton Analysis

Whole WM analysis was done by assuming the whole skeleton as the ROI and the FA values were calculated based on the skeletonized mask of the scalar globally. For each participant, FA values from each voxel were extracted, and mean FA values of each participant were used in the analyses. First, as a quality reassurance Kruskal Wallis Test was employed to investigate the possible differences within the following repetition combinations 1vs.2, 1vs.3, 2vs.3 and all three repetitions combined. Yet, no significant difference was found between the repetitions (in every case $p>=0.37$).

3.1.1 Agreement between measures

To evaluate whether having different repetitions affected the FA values Bland Altman plots were used. Bland Altman plots were created for repetition combinations 1vs.2, 1vs.3 and 2vs.3 (Figure 5a, 5b and 5c). Based on the agreement criteria mentioned earlier, it can be concluded that the repetition pairs in the study agreed well.

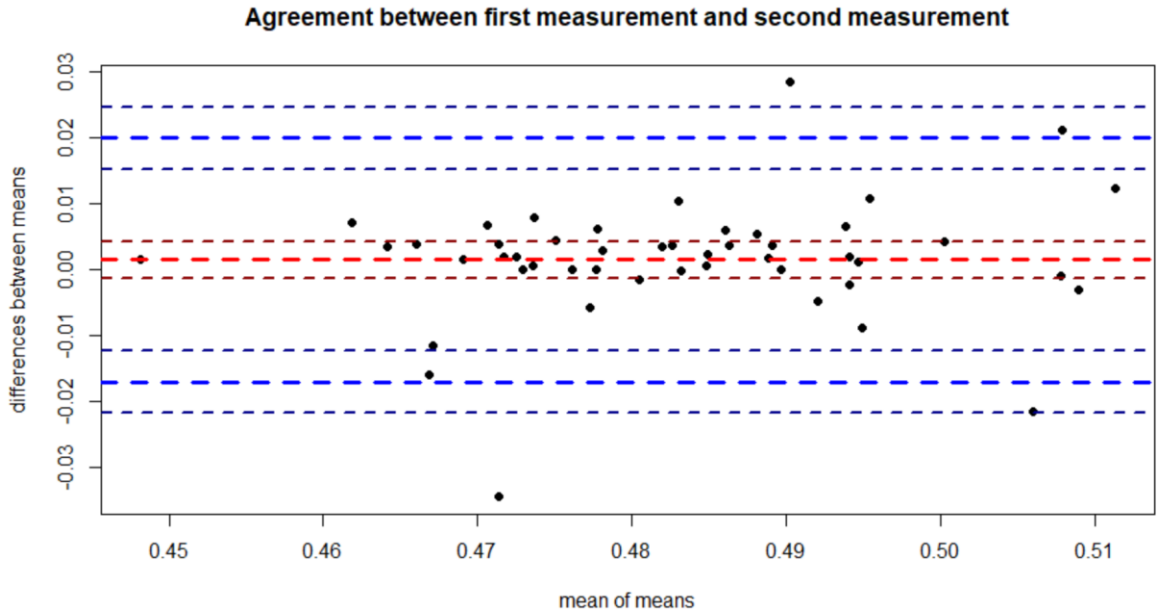

Figure 5a. Bland-Altman plot for the mean FA values from first and second repetitions out of three measurements. The thick red dash line at point 0 shows the mean difference and the thinner darker red dash lines above and below the mean difference shows the confidence interval of mean difference. The thick blue dash lines above and below the mean difference represents the minimum and maximum limits of agreement (LoA) and the thinner dash lines above and below the LoA show for the confidence intervals for maximum and minimum LoA.

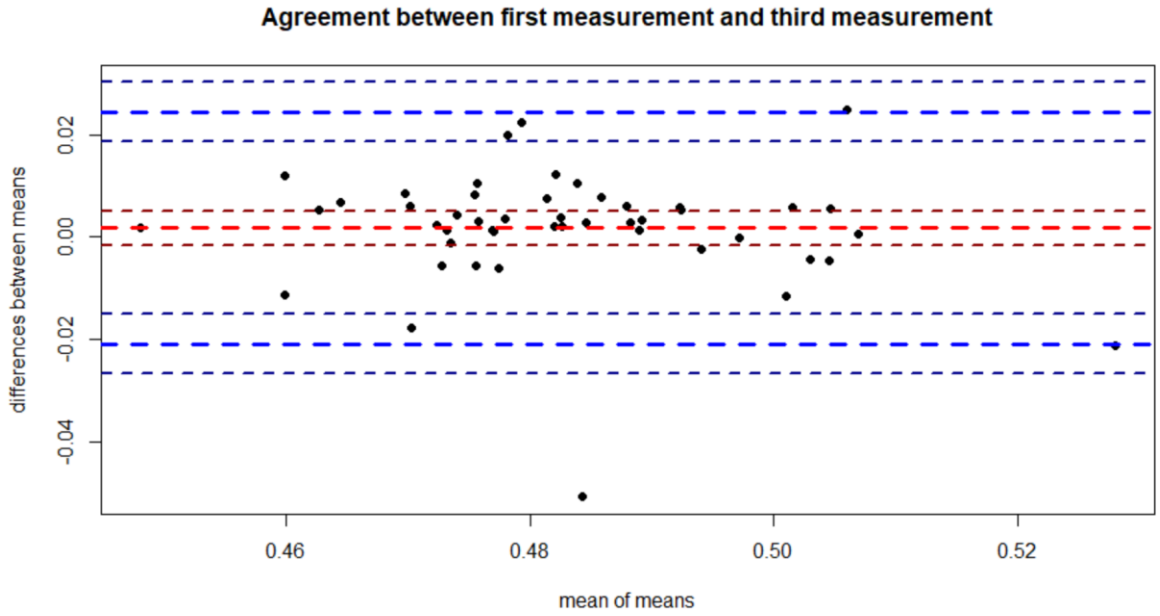

Figure 5b. Bland-Altman plot for the mean FA values from first and third repetitions out of three measurements. The thick red dash line at point 0 shows the mean difference and the thinner darker red dash lines above and below the mean difference shows the confidence interval of mean difference. The thick blue dash lines above and below the mean difference represents the minimum and maximum limits of agreement (LoA) and the thinner dash lines above and below the LoA show for the confidence intervals for maximum and minimum LoA.

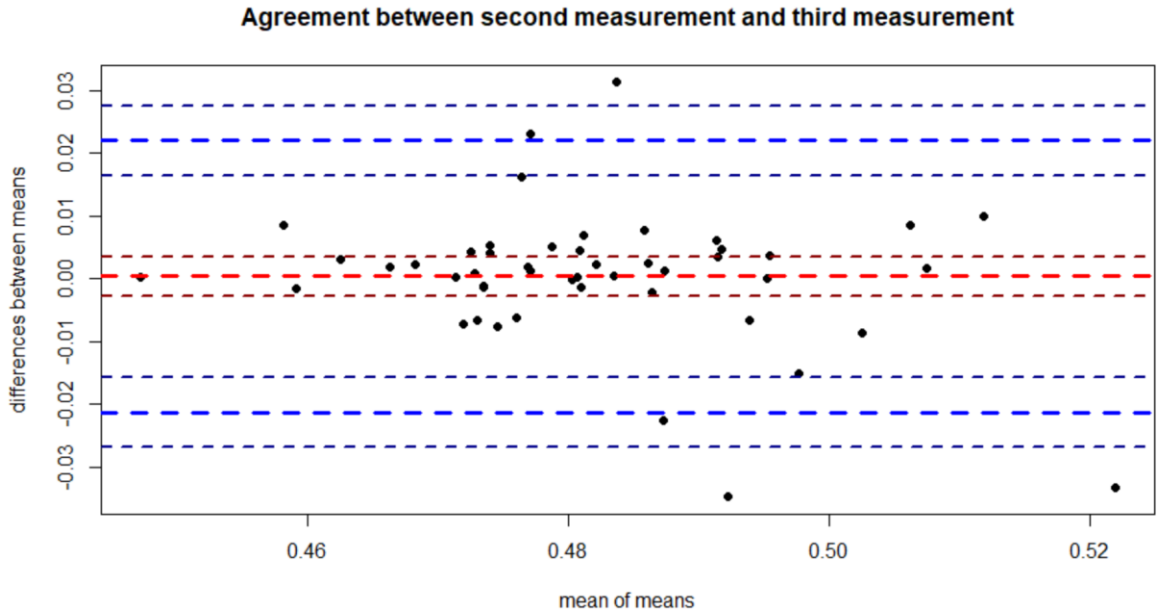

Figure 5c. Bland-Altman plot for the mean FA values from second and third repetitions out of three measurements. The thick red dash line at point 0 shows the mean difference and the thinner darker red dash lines above and below the mean difference shows the confidence interval of mean difference. The thick blue dash lines above and below the mean difference represents the minimum and maximum limits of agreement (LoA) and the thinner dash lines above and below the LoA show for the confidence intervals for maximum and minimum LoA.

3.1.2 Test-retest reliability analyses

ICC(3,1) values were calculated for each repetition combination, smallest ICC(3,1) value was 0.73. The ICC(3,1) value for each repetition combination with confidence intervals (CI) can be found in Figure 6. Based on the ICC values overall reliability of the repetitions is moderate to good level.

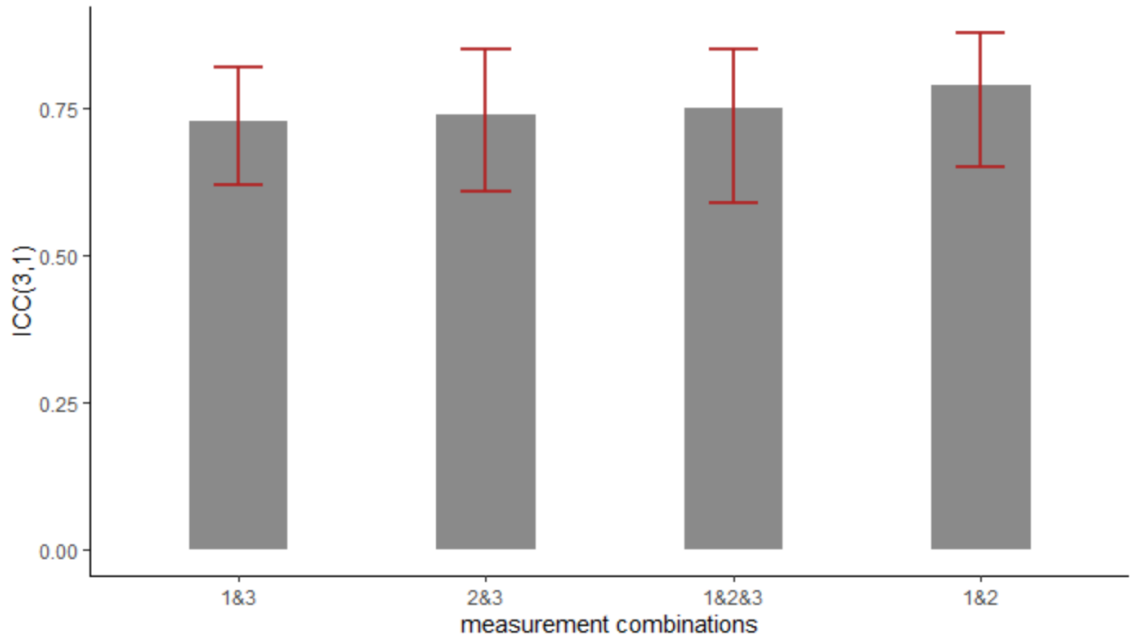

Figure 6. ICC(3,1) values with 95% confidence interval for each repetition combination. From left to right ICC(3,1) values calculated with the mean FA values from first and third repetitions, second and third repetitions, all three repetitions and first and second repetitions.

The SEM values were calculated for each repetition combination as well (Figure 7). All repetitions combined had the highest SEM value with 0.0037. Since that number is very close to zero it is safe to claim that the results were not due to random error and FA values were repeatable.

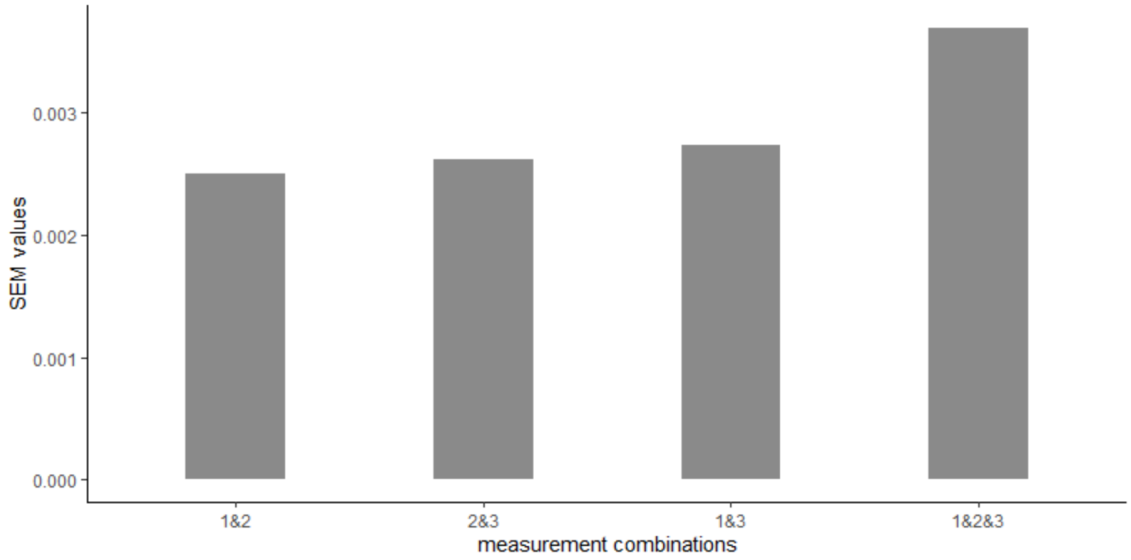

Figure 7. SEM values for each repetition combination. From left to right SEM values calculated with the mean FA values from first and second repetitions, second and third repetitions, first and third repetitions and all three repetitions.

3.2 ROI Analyses

For ROI analysis instead of extracting the FA values globally, they were calculated based on the atlas labels. From the non-skeleton mask, FA values were extracted for all 48 labels of the atlas whilst from the skeletonized mask FA values could only be calculated for 47 labels. The complete list of the ROI labels in the atlas and their mean FA values as well as the information about the ones excluded from the study can be found in Table 3a and 3b. As it was done with the global analysis, as a first step and a quality reassurance, possible differences between the three repetitions in each label for both ROI types were investigated with Kruskal Wallis Test. There were also no significant differences between measurements in any of the ROI labels (smallest p value for skeletonized ROI type $p=0.07$, smallest p value for non-skeletonized ROI type $p=0.33$).

Table 3a. ROI label names and numbers in JHU-ICBM DTI-81 white-matter labels atlas (ROI numbers from 1 to 24) and the mean FA values (mean ± SD) for non-skeletonized and skeletonized ROI types.

Label name	Label number	WM parcellation	Non-skeletonized			Skeletonized		
			Number of measurements			Number of measurements		
			1	2	3	1	2	3
Middle cerebellar peduncle*	1	Brainstem	0.51±0.02	0.51±0.02	0.50±0.02	0.58±0.02	0.59±02	0.58±0.02
Pontine crossing tract*	2	Brainstem	0.46±0.03	0.46±0.03	0.46±0.03	0.48±0.03	0.49±0.03	0.49±0.03
Genu of corpus callosum	3	Commissural	0.54±0.02	0.54±0.02	0.54±0.02	0.77±0.03	0.72±0.3	0.78±0.03
Body of corpus callosum	4	Commissural	0.58±0.02	0.59±0.03	0.59±0.03	0.71±0.03	0.71±0.03	0.71±0.4
Splenium of corpus callosum	5	Commissural	0.69±0.02	0.70±0.02	0.69±0.02	0.79±0.02	0.80±0.02	0.79±0.02
Fornix	6	Association	0.40±0.03	0.40±0.03	0.40±0.03	0.49±0.03	0.48±0.04	0.48±0.04
Corticospinal tract R*	7	Brainstem	0.49±0.02	0.49±0.02	0.49±0.03	0.52±0.02	0.52±0.03	0.51±0.03
Corticospinal tract L*	8	Brainstem	0.50±0.02	0.50±0.02	0.49±0.02	0.53±0.03	0.53±0.02	0.52±0.03
Medial lemniscus R*	9	Brainstem	0.51±0.03	0.50±0.02	0.51±0.02	0.56±0.03	0.59±0.03	0.56±0.02
Medial lemniscus L*	10	Brainstem	0.51±0.03	0.51±0.03	0.51±0.03	0.55±0.03	0.56±0.03	0.56±0.03
Inferior cerebellar peduncle R*	11	Brainstem	0.47±0.03	0.48±0.03	0.47±0.03	0.52±0.03	0.53±0.04	0.52±0.03
Inferior cerebellar peduncle L*	12	Brainstem	0.46±0.02	0.46±0.03	0.46±0.02	0.51±0.03	0.51±0.03	0.51±0.03
Superior cerebellar peduncle R*	13	Brainstem	0.52±0.02	0.52±0.02	0.52±0.02	0.61±0.03	0.61±0.03	0.62±0.03
Superior cerebellar peduncle L*	14	Brainstem	0.48±0.02	0.48±0.02	0.48±0.02	0.64±0.03	0.65±0.03	0.65±0.03
Cerebral peduncle R*	15	Projection	0.61±0.02	0.61±0.02	0.61±0.02	0.71±0.02	0.71±0.02	0.71±0.03
Cerebral peduncle L*	16	Projection	0.61±0.02	0.61±0.02	0.61±0.02	0.69±0.02	0.70±0.02	0.69±0.03
Anterior limb of internal capsule R	17	Projection	0.50±0.02	0.50±0.02	0.50±0.02	0.60±0.03	0.60±0.03	0.61±0.03
Anterior limb of internal capsule L	18	Projection	0.50±0.02	0.5±0.02	0.50±0.02	0.58±0.02	0.58±0.02	0.58±0.03
Posterior limb of internal capsule R	19	Projection	0.62±0.02	0.62±0.02	0.62±0.02	0.70±0.02	0.70±0.02	0.70±0.02
Posterior limb of internal capsule L	20	Projection	0.59±0.02	0.59±0.01	0.60±0.02	0.70±0.02	0.70±0.02	0.70±0.02
Retrolenticular part of internal capsule R	21	Projection	0.56±0.02	0.56±0.02	0.56±0.02	0.61±0.02	0.60±0.02	0.60±0.03
Retrolenticular part of internal capsule L	22	Projection	0.547±0.018	0.551±0.016	0.522±0.020	0.627±0.021	0.63±0.02	0.63±0.02
Anterior corona radiata R	23	Projection	0.440±0.026	0.441±0.027	0.443±0.027	0.499±0.030	0.50±0.03	0.50±0.03
Anterior corona radiata L	24	Projection	0.44±0.02	0.44±0.02	0.44±0.03	0.50±0.03	0.050±0.03	0.51±0.03

*excluded from the study because of their locations (brainstem)

Table 3b. ROI label names and numbers in JHU-ICBM DTI-81 white-matter labels atlas (ROI numbers from 25 to 48) and the mean FA values (mean ± SD) for non-skeletonized and skeletonized ROI types.

Label name	Label number	WM parcellation	Non-skeletonized			Skeletonized		
			Number of measurements			Number of measurements		
			1	2	3	1	2	3
Superior corona radiata R	25	Projection	0.46±0.02	0.46±0.02	0.46±0.02	0.50±0.03	0.50±0.03	0.51-0.03
Superior corona radiata L	26	Projection	0.46±0.02	0.47±0.02	0.46±0.02	0.51±0.02	0.52±0.2	0.52±0.02
Posterior corona radiata R	27	Projection	0.44±0.02	0.45±0.02	0.45±0.02	0.48±0.03	0.49±0.03	0.49±0.03
Posterior corona radiata L	28	Projection	0.44±0.02	0.44±0.02	0.44±0.03	0.49±0.03	0.49±0.03	0.49±0.03
Posterior thalamic radiation R	29	Projection	0.57±0.03	0.58±0.03	0.58±0.03	0.61±0.03	0.62±0.03	0.62±0.03
Posterior thalamic radiation L	30	Projection	0.54±0.03	0.54±0.03	0.54±0.03	0.62±0.03	0.062±0.03	0.63±0.03
Sagittal stratum R	31	Association	0.49±0.02	0.50±0.02	0.50±0.02	0.55±0.03	0.55±0.03	0.55±0.03
Sagittal stratum L	32	Association	0.44±0.02	0.44±0.02	0.44±0.02	0.54±0.03	0.54±0.04	0.54±0.03
External capsule R	33	Association	0.38±0.01	0.38±0.02	0.38±0.02	0.45±0.02	0.45±0.02	0.45±0.02
External capsule L	34	Association	0.38±0.01	0.38±0.01	0.38±0.02	0.48±0.02	0.48±0.02	0.48±0.02
Cingulum (cingulate gyrus) R	35	Association	0.45±0.03	0.44±0.03	0.45±0.03	054±0.03	0.53±0.04	0.52±0.03
Cingulum (cingulate gyrus) L	36	Association	0.43±0.03	0.43±0.03	0.43±0.03	0.57±0.04	0.57±0.04	0.57±0.04
Cingulum (hippocampus) R	37	Association	0.36±0.02	0.36±0.02	0.36±0.02	0.41±0.02	0.41±0.02	0.40±0.03
Cingulum (hippocampus) L	38	Association	0.34±0.02	0.34±0.03	0.35±0.026	0.41±0.03	0.403±0.03	0.41±0.03
Fornix / Stria terminalis R	39	Association	0.43±0.02	0.44±0.02	0.44±0.02	0.56±0.03	0.563±0.03	0.57±0.03
Fornix / Stria terminalis L	40	Association	0.45±0.02	0.45±0.02	0.45±0.02	0.57±0.03	0.57±0.03	0.57±0.03
Superior longitudinal fasciculus R	41	Association	0.45±0.02	0.45±0.02	0.45±0.02	0.52±0.03	0.51±0.03	0.52±0.03
Superior longitudinal fasciculus L	42	Association	0.44±0.02	0.44±0.02	0.44±0.03	0.51±0.03	0.51±0.03	0.51±0.03
Superior fronto-occipital fasciculus R	43	Association	0.48±0.03	0.48±0.03	0.48±0.03	0.54±0.04	0.54±0.04	0.55±0.04
Superior fronto-occipital fasciculus L	44	Association	0.45±0.03	0.45±0.04	0.45±0.04	0.53±0.04	0.53±0.04	0.52±0.05
Uncinate fasciculus R	45	Association	0.44±0.03	0.44±0.03	0.44±0.03	0.51±0.03	0.50±0.04	0.51±0.04
Uncinate fasciculus L	46	Association	0.41±0.03	0.42±0.03	0.42±0.03	0.46±0.03	0.47±0.03	0.46±0.03
Tapetum R	47	Commissural	0.40±0.03	0.40±0.03	0.40±0.03	0.56±0.05	0.56±0.06	0.56±0.06
Tapetum L*	48	Commissural	0.30±0.03	0.31±0.03	0.31±0.03	-	-	-

*excluded from the skeletonized mask automatically due to its small volume

3.2.1 Repeatability of corpus callosum

Corpus callosum is a large size ROI mostly composed of myelinated WM tracts that have homogenous directionality (Anand et al., 2019; Fabri, 2014). For that reason, observing high repeatability of CC works as a reference region in repeatability studies. Therefore, one of the first ROIs to be analyzed must be corpus callosum (CC). The atlas used splits corpus callosum (CC) into three sections; genu of corpus callosum (GCC) the most anterior part of CC, splenium of corpus callosum (SCC) the most posterior part of the CC and body of corpus callosum (BCC) the area between the GCC and SCC. Even though the SEM values for the skeletonized ROI type seem higher in the plot, the difference between the skeletonized and non-skeletonized ROI types were quite small to create any significant difference (the largest difference between ROI types is 0.003). All three areas of CC found to have high ICC(3,1) values and low SEM values. Hence, it can be concluded that the repeatability is high in both ROI types (Figure 8a and 8b).

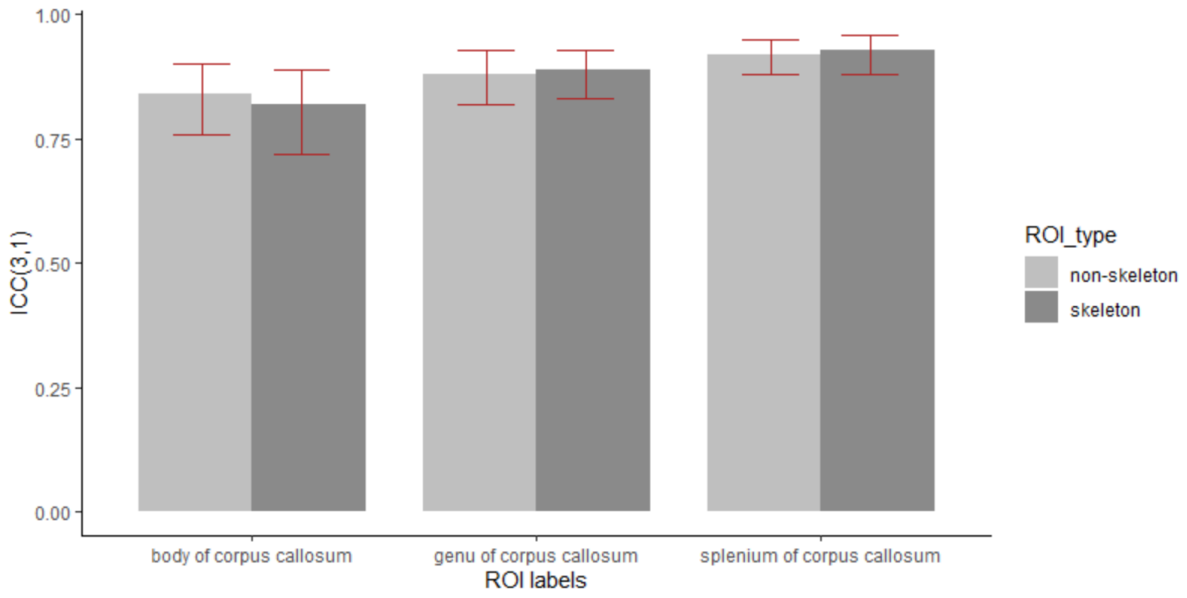

Figure 8a. The ICC(3,1) values for corpus callosum for non-skeletonized and skeletonized ROI types.

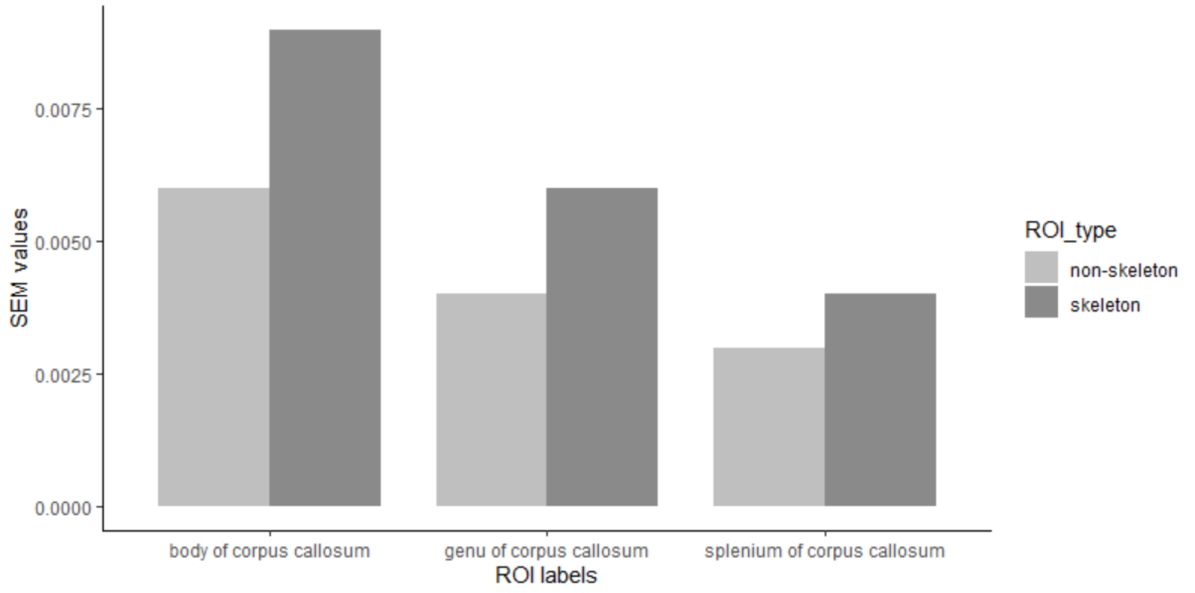

Figure 8b. The SEM values for corpus callosum for non-skeletonized and skeletonized ROI types.

3.2.2 Repeatability by ICC(3,1)

In terms of ICC(3,1) values, 87% of the skeletonized ROI type and 88% of the non-skeletonized ROI type had good to excellent repeatability. The ROI with the highest repeatability in the skeletonized ROI type that had excellent repeatability (ICC(3,1) = 0.94, CI: 0.90-0.96) was the anterior corona radiata (ACR) in the right hemisphere. ACR is a relatively large bundle of fibers connecting thalamus to the cortex i.e., prefrontal cortex (Karababa, Bayazıt, Kılıçaslan, & Celik, 2015). In the non-skeletonized ROI type, posterior thalamic radiation (PTR) and tapetum in the left hemisphere had the highest repeatability among ROI labels (ICC(3,1)=0.93, CI:0.88-0.95, ICC(3,1)=0.93, CI:089-0.96). PTR is a smaller fiber bundle compared to ACR, but it also connects the thalamus to cortex, to parietal and occipital lobes to be exact. Tapetum of corpus callosum is the temporal side of the CC yet it is located separately from the CC (Mori et al., 2008). Even though the areas mentioned with the highest ICC(3,1) values were not the same in both ROI types, both ROIs are still in the highest repeatability list on both ROI types. The ROI labels with excellent reliability for both ROI types (ICC(3,1) > 0.9) can be found in Figure 9a and 9b.

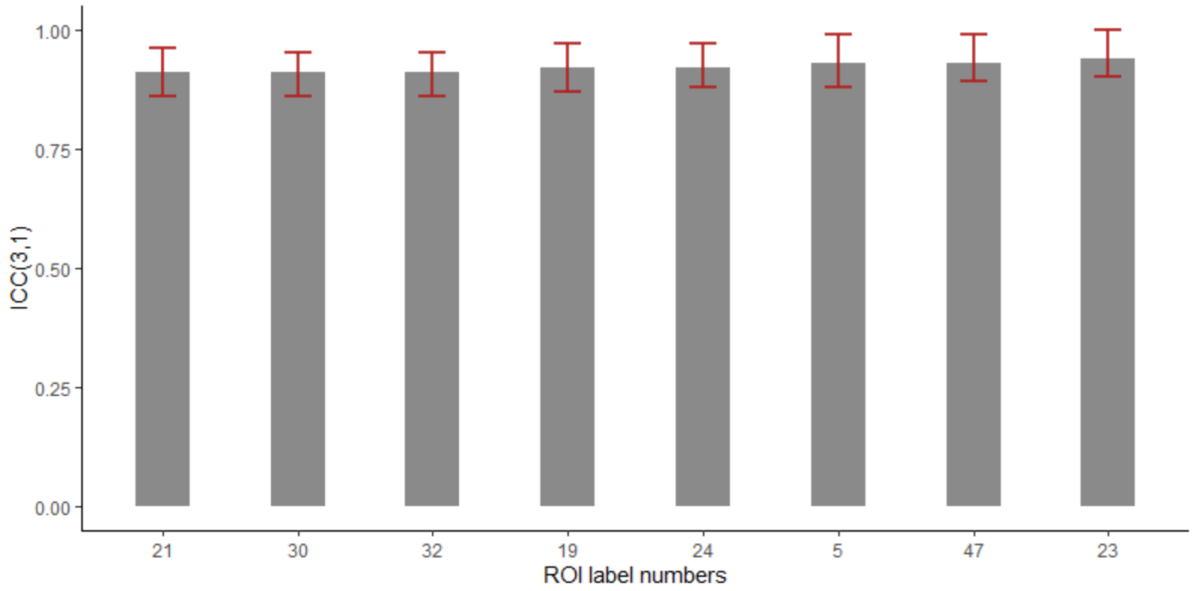

Figure 9a. The ICC(3,1) values for the ROI labels with excellent repeatability in the skeletonized ROI type. ROI labels respectively from left to right: (21)retrolenticular part of internal capsule R, (30)posterior thalamic radiation L, (32)sagittal stratum L, (19)posterior limb of internal capsule R, (24)anterior corona radiata L, (5)splenium of corpus callosum, (47)tapetum R and (23)anterior corona radiata R.

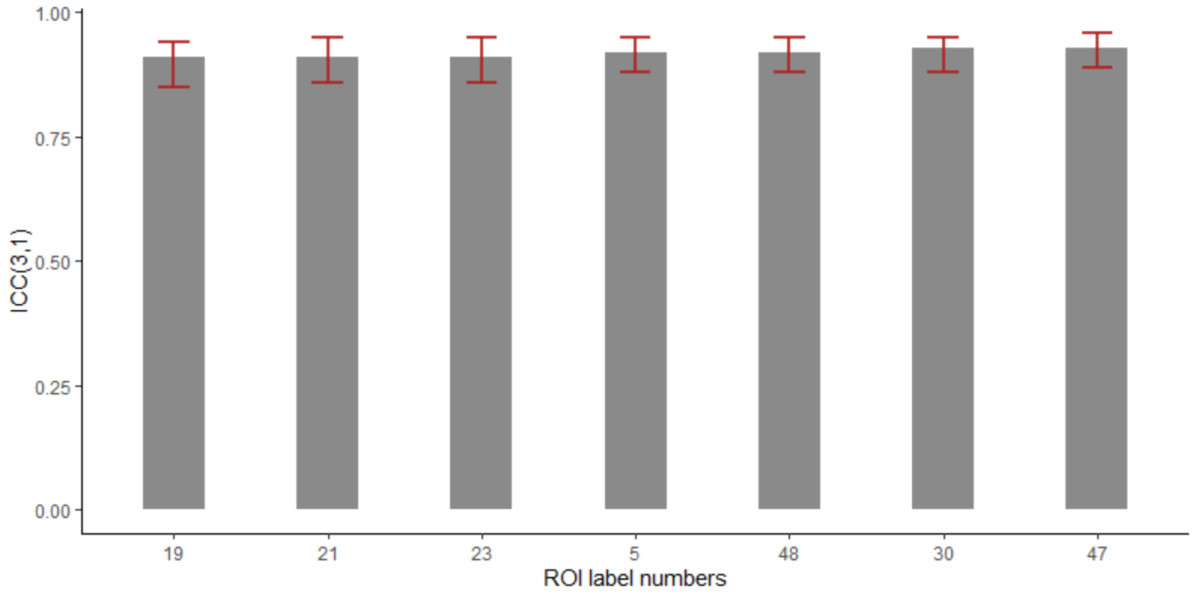

Figure 9b. The ICC(3,1) values for the ROI labels with excellent repeatability in the non-skeletonized ROI type. ROI labels respectively from left to right: (19)posterior limb of internal capsule R, (21) retrolenticular part of internal capsule R, (23) anterior corona radiata R, (5)splenium of corpus callosum, (48)tapetum L, (30) posterior thalamic radiation L, (47)tapetum R.

On the other hand, some ROIs had lower repeatability compared to others. The ROI label with the lowest repeatability in the skeletonized ROI type was the superior fronto-

occipital fasciculus (SFOF) in both hemispheres with poor to good reliability (ICC(3,1)=0.62, CI:0.47-0.75). SFOF is a rather small bundle of fibers assumed to be connecting frontal lobe to parietal lobe (Mori et al., 2008). In non-skeletonized ROI type, the ROI label with the smallest ICC(3,1) compared to others was again the SFOF in the right hemisphere with poor to moderate repeatability (ICC(3,1)=0.6, CI:0.44-0.73). Similar to plots of ROI labels with the the highest repeatability in terms of ICC(3,1), plots of ROI labels with the lowest repeatability in terms of ICC(3,1) include the almost same ROIs for both ROI types. The ROI labels with moderate reliability for both ROI types (ICC(3,1) < 0.75) can be found in Figure 10a and 10b.

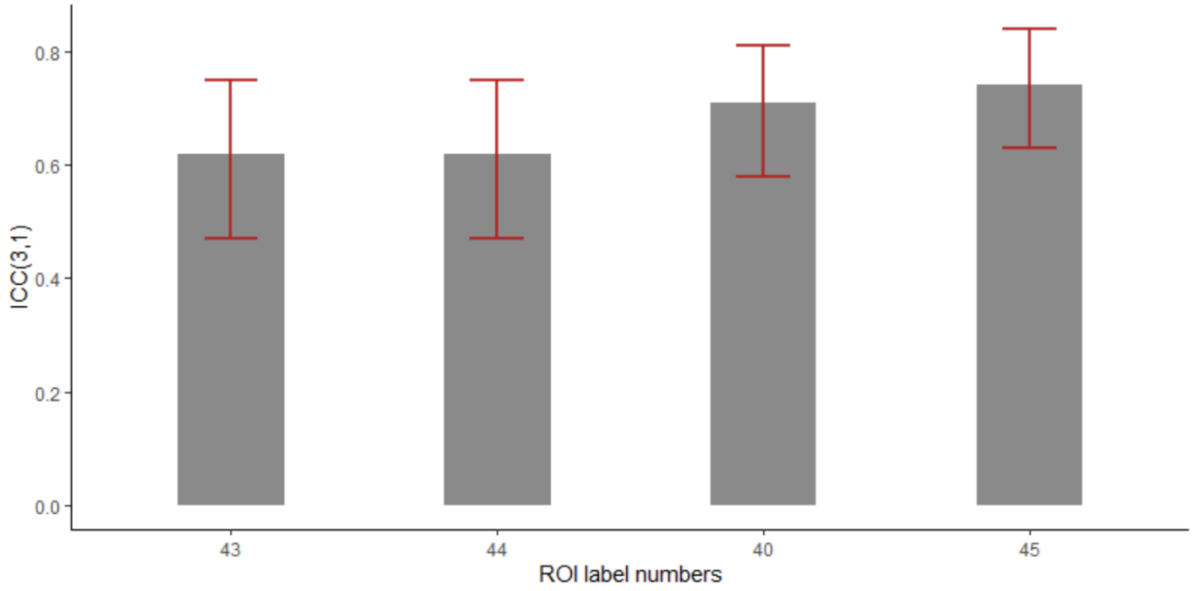

Figure 10a. The ICC(3,1) values for the ROI labels with moderate repeatability in the skeletonized ROI type. ROI labels respectively from left to right: (43)superior fronto-occipital fasciculus R, (44)superior fronto-occipital fasciculus L, (40)fornix (cres) / stria terminalis L, (45)uncinate fasciculus R.

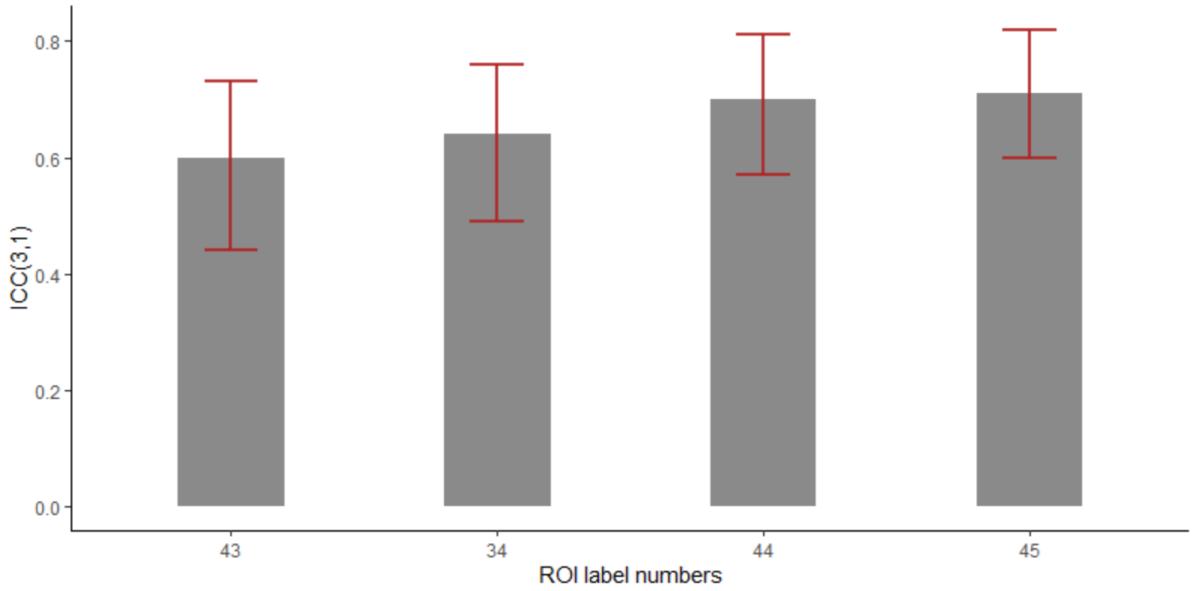

Figure 10b. The ICC(3,1) values for the ROI labels with moderate repeatability in the non-skeletonized ROI type. ROI labels respectively from left to right: (43)superior fronto-occipital fasciculus R, (34)external capsule L, (44)superior fronto-occipital fasciculus L, (45)uncinate fasciculus R.

3.2.3 Repeatability by SEM

All ROIs in both ROI types had really close to zero numbers for SEM which suggests that the results are repeatable and did not occur by chance or a random error. The lowest SEM value was observed in posterior limb of internal capsule (PLIC) in right hemisphere in the skeletonized ROI type (SEM=0.003). PLIC is the part of the internal capsule lies between thalamus and globus pallidus (Mori, Oishi, & Faria, 2009). In non-skeletonized ROI type, again, the lowest SEM value was observed in PLIC and additionally in SCC (SEM=0.003). As expected, same ROIs with high ICC repeatability measures were also observed in the low SEM plots. The ROI labels with lowest SEM values for both ROI types can be found in Figure 11a and 11b (SEM<0.005).

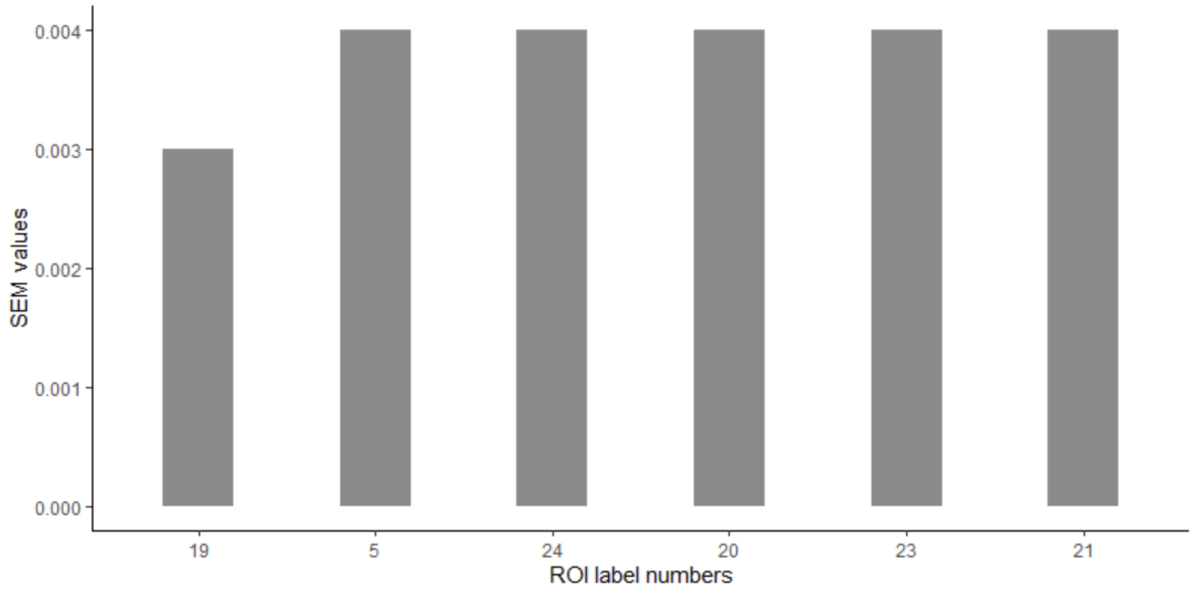

Figure 11a. The lowest SEM values for the ROI labels in the skeletonized ROI type. ROI labels respectively from left to right: (19)posterior limb of internal capsule R, (5)splenium of corpus callosum, (24)anterior corona radiata L, (20)posterior limb of internal capsule L, (23)anterior corona radiata R, (21)retrolenticular part of internal capsule R.

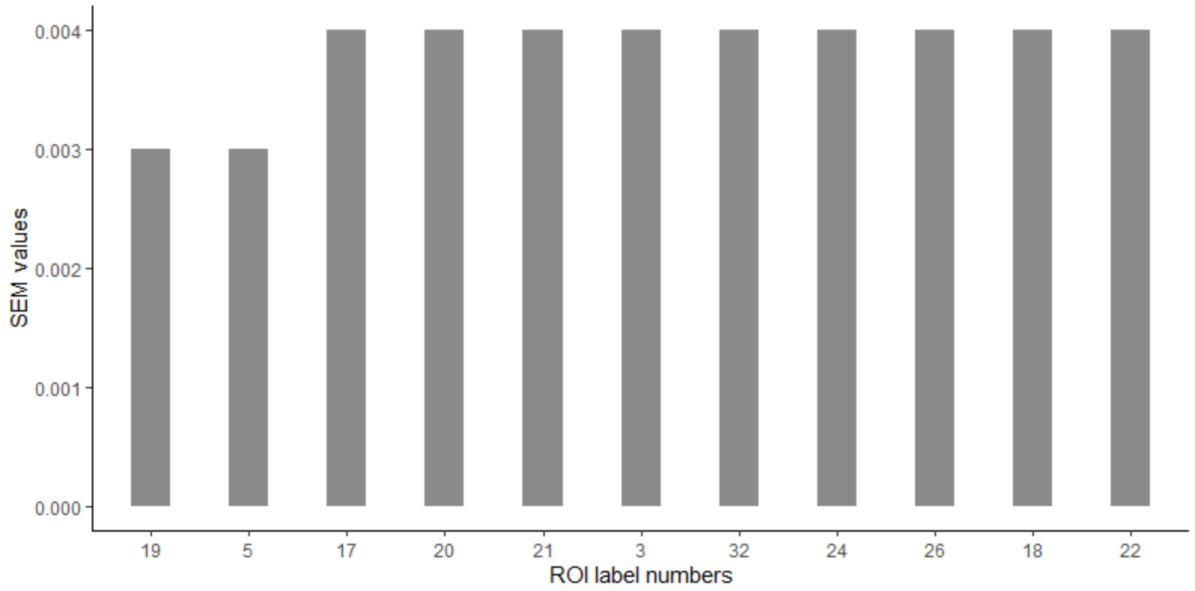

Figure 11b. The lowest SEM values for the ROI labels in the non-skeletonized ROI type. ROI labels respectively from left to right: (19)posterior limb of internal capsule R, (5)splenium of corpus callosum, (17)anterior limb of internal capsule R, (20)posterior limb of internal capsule L, (21)retrolenticular part of internal capsule R, (3)genu of corpus callosum, (32)sagittal stratum L, (24)anterior corona radiata L, (26)superior corona radiata L, (18)anterior limb of internal capsule L, (22)retrolenticular part of internal capsule L.

In accordance with the lowest ICC(3,1) value, the highest SEM value was observed in posterior SFOF in both hemispheres in the skeletonized ROI type (SEM=0.02). In non-skeletonized ROI type one more label joined SFOF in lowest repeatability group based on SEM value; posterior corona radiata (PCR) in right hemisphere (SEM=0.01). Corona radiata is

located in the frontal motor cortex and the ROI label refers to the most posterior part of those projections. In the atlas used the anterior, superior, posterior parcellation was made based on the CC (Mori et al., 2008). The ROI labels with highest SEM values for both ROI types can be found in Figure 12a and 12b (SEM>0.007).

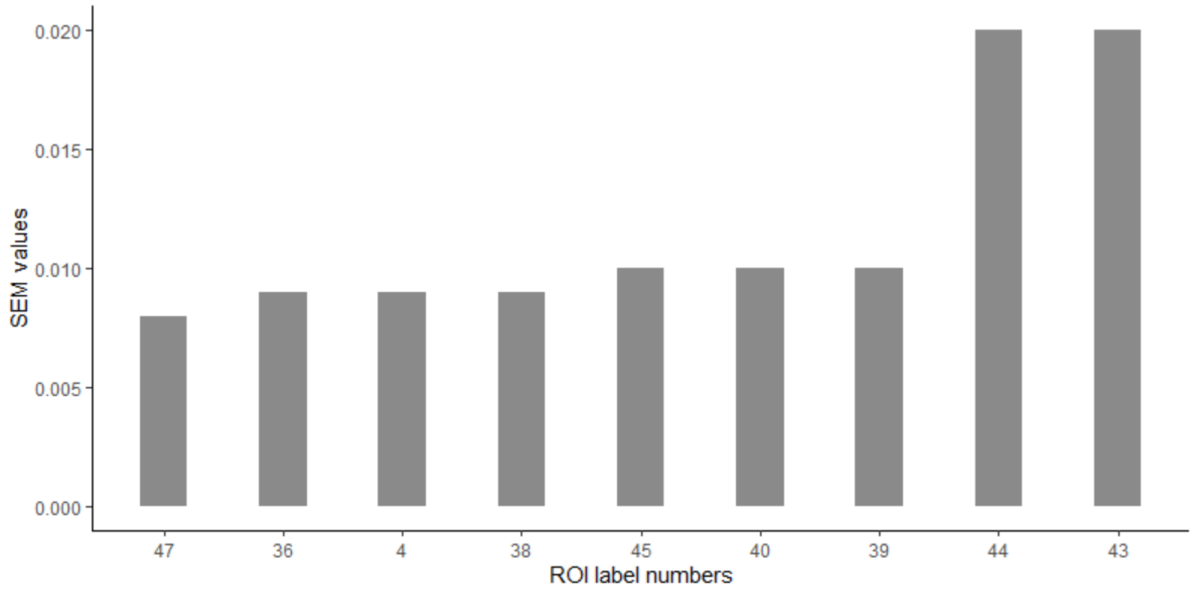

Figure 12a. The highest SEM values for the ROI labels in the skeletonized ROI type. ROI labels respectively from left to right: (47)tapetum R, (36)cingulum (cingulate gyrus) L, (4)body of corpus callosum, (38)cingulum (hippocampus) L, (45)uncinate fasciculus R, (40)fornix (cres) / stria terminalis L, (39) fornix (cres) / stria terminalis R, (44)superior fronto-occipital fasciculus L, (43)superior fronto-occipital fasciculus R.

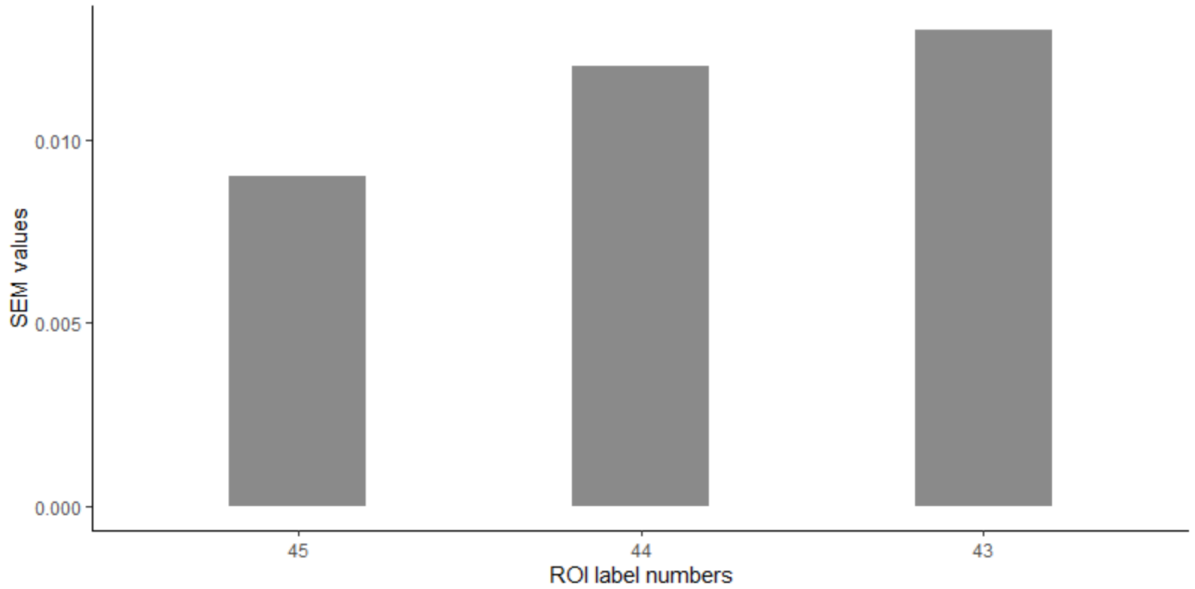

Figure 12b. The highest SEM values for the ROI labels in the non-skeletonized ROI type. ROI labels respectively from left to right: (45)uncinate fasciculus R, (44)superior fronto-occipital fasciculus L, (43)superior fronto-occipital fasciculus R.

3.2.4 Volume size and repeatability

Volume sizes were calculated by taking the average of all three measurements for all label sizes for both non-skeletonized ROI and skeletonized ROI types and multiplying by the volume of a single voxel (8ml).

The average volume sizes and both ICC(3,1) scores and SEM scores were analyzed to reveal any possible relationship between the measures. In skeletonized set after around 5 ml in volume size, ICC values tend to get higher and SEM values tend to get lower, same trend was observed in non-skeletonized set around the 20 ml mark in volume size. To investigate the relationship between the volume size and repeatability further different regression models i.e. linear regression, exponential regression and polynomial regression with a quadratic model were applied to find how much the trend can be statistically explained, in other words how well the volume size can predict the repeatability. Even though the results were significant for almost all of the relationships between the volume and SEM and volume and ICC(3,1) in both ROI types, the explanatory power of the exponential and linear models were quite low (highest R^2 =0.22). Interestingly, the magnitude of the relationship between SEM and volume size (R^2 =0.55) and ICC(3,1) and volume size (R^2 =0.34) for skeletonized ROI type and ICC(3,1) and volume size for non-skeletonized ROI type (R^2 =0.31) seem to increase considerably with polynomial model. However, since the ending trend of the polynomial model was depending on one ROI label point only, though the magnitude of the relationship seemed relatively large, it was still weak evidence for the effect of the volume size on repeatability. Plots for the regression models for skeletonized ROI type can be found in Figure 13a and 13b and for non-skeletonized ROI type in Figure 14a and 14b. The detailed information about the statistical measures of the regression models can be found in Table 4.

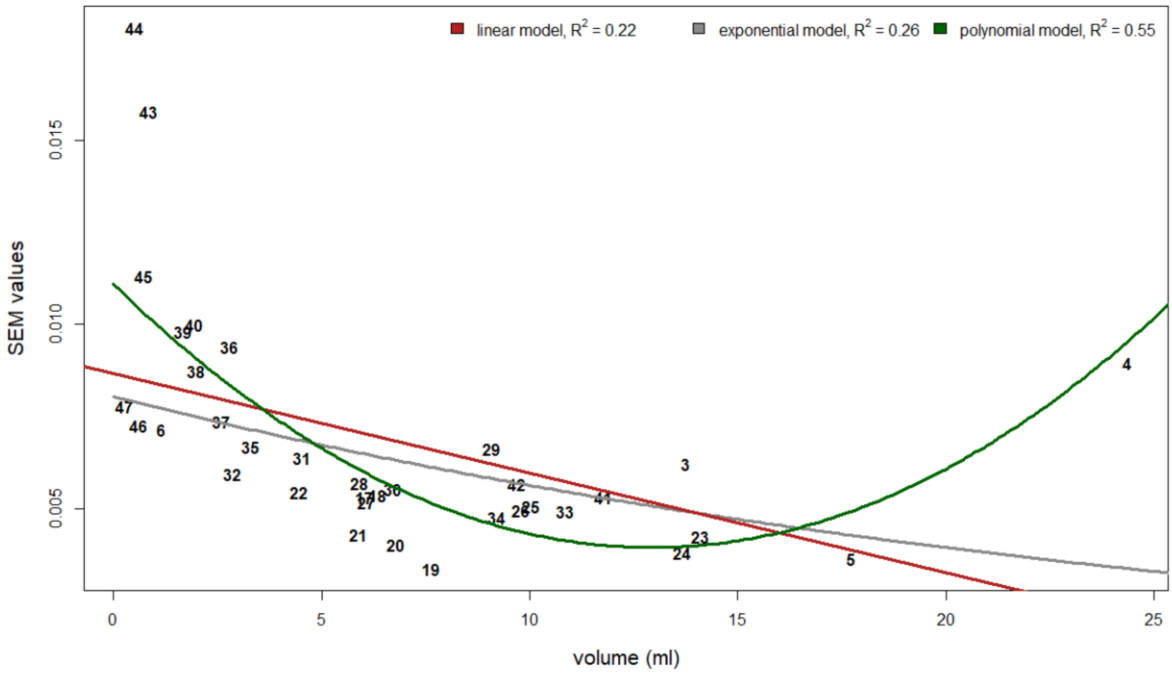

Figure 13a. Relationship between the volume sizes of the labels and the corresponding SEM values for three repetitions in skeletonized ROI type.

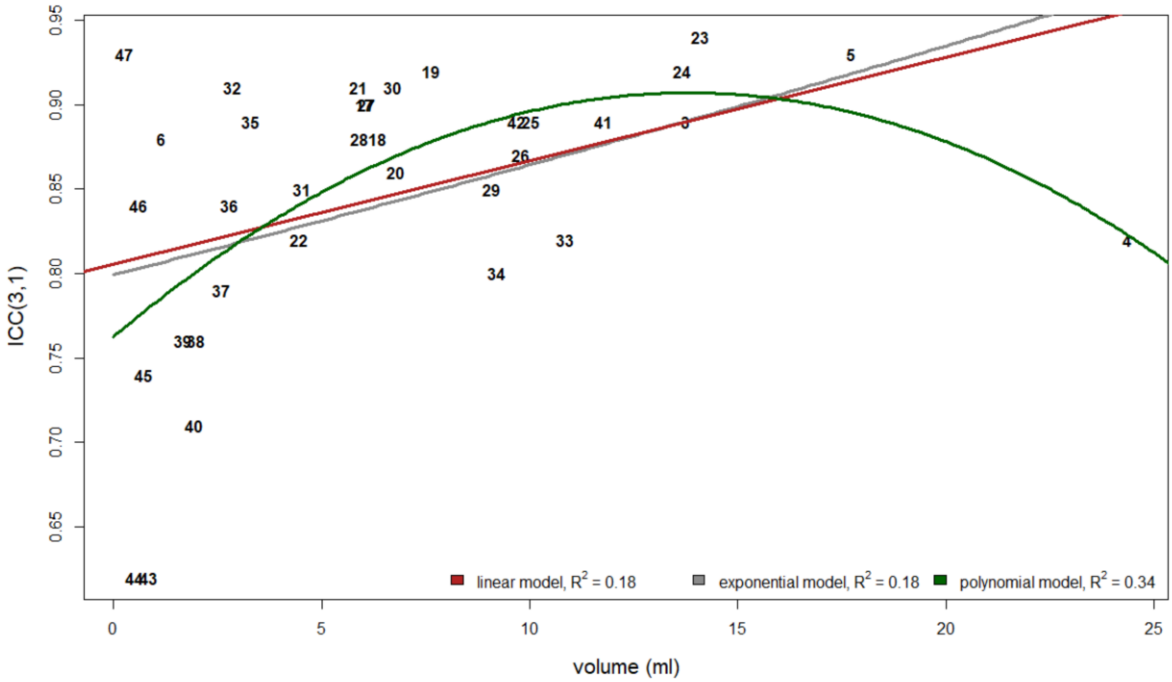

Figure 13b. Relationship between the volume sizes of the labels and the corresponding ICC(3,1) values for three repetitions in skeletonized ROI type. Commissural tracts (3-5,47) had higher ICC(3,1) values compared to others, association fibers (6,31-46) show more variability in terms of ICC(3,1) compared to projection fibers (17-30).

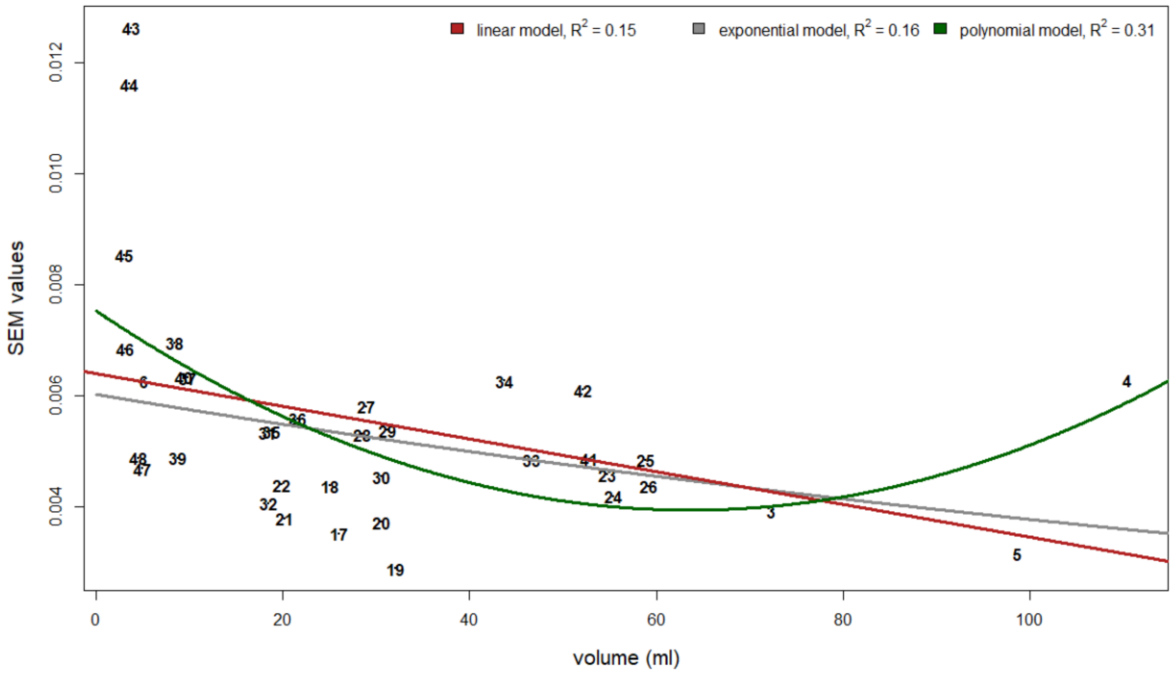

Figure 14a. Relationship between the volume sizes of the labels and the corresponding SEM values for three repetitions in non-skeletonized ROI type.

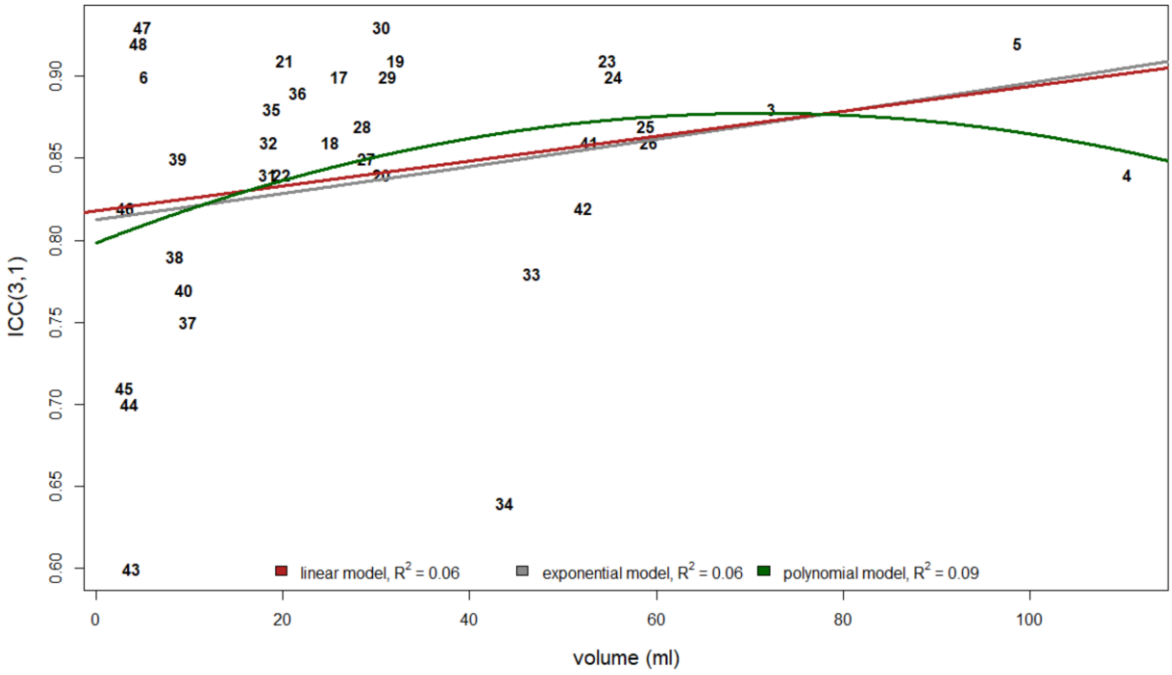

Figure 14b. Relationship between the volume sizes of the labels and the corresponding ICC(3,1) values for three repetitions in non-skeletonized ROI type. Commissural tracts (3-5,47,48) had higher ICC(3,1) values compared to others, association fibers (6,31-46) show more variability in terms of ICC(3,1) compared to projection fibers (17-30).

Table 4. Different types of regression analyses for the relationship between repeatability and volume size in skeletonized ROI and non-skeletonized ROI types.

regression type	data	p value	R^2
linear model	skeletonized SEM vs volume	<0.01	0.22
	skeletonized ICC vs volume	<0.05	0.18
	non-skeletonized SEM vs volume	<0.05	0.15
	non-skeletonized ICC vs volume	>0.05	0.06
exponential model	skeletonized SEM vs volume	<0.01	0.26
	skeletonized ICC vs volume	<0.05	0.18
	non-skeletonized SEM vs volume	<0.05	0.16
	non-skeletonized ICC vs volume	>0.05	0.06
polynomial regression (quadratic model)	skeletonized SEM vs volume	<0.01	0.55
	skeletonized ICC vs volume	<0.01	0.34
	non-skeletonized SEM vs volume	<0.01	0.31
	non-skeletonized ICC vs volume	>0.05	0.09

3.2.5 Lateralization and repeatability

To investigate whether there are any hemispheric differences, both repeatability measures were compared for the ROI labels located in both hemispheres in both skeletonized ROI and non-skeletonized ROI types. Based on the results it can be concluded that structures did not have any hemispheric differences in terms of repeatability.

3.2.6 The effect of skeleton and repeatability

ICC(3,1) and SEM values for the ROI labels were plotted to visualize any possible difference might be caused by the use of the skeleton in right hemisphere (Figure 15a and 15b) and in left hemisphere (Figure 16a and 16b). Evaluating relevant plots (Figure 8a and 8b, Figure 15a and 15b and Figure 16a and 16b), it can be concluded that none of them were suggestive of a significant effect of the use of the skeleton neither for whole WM nor for individual ROIs.

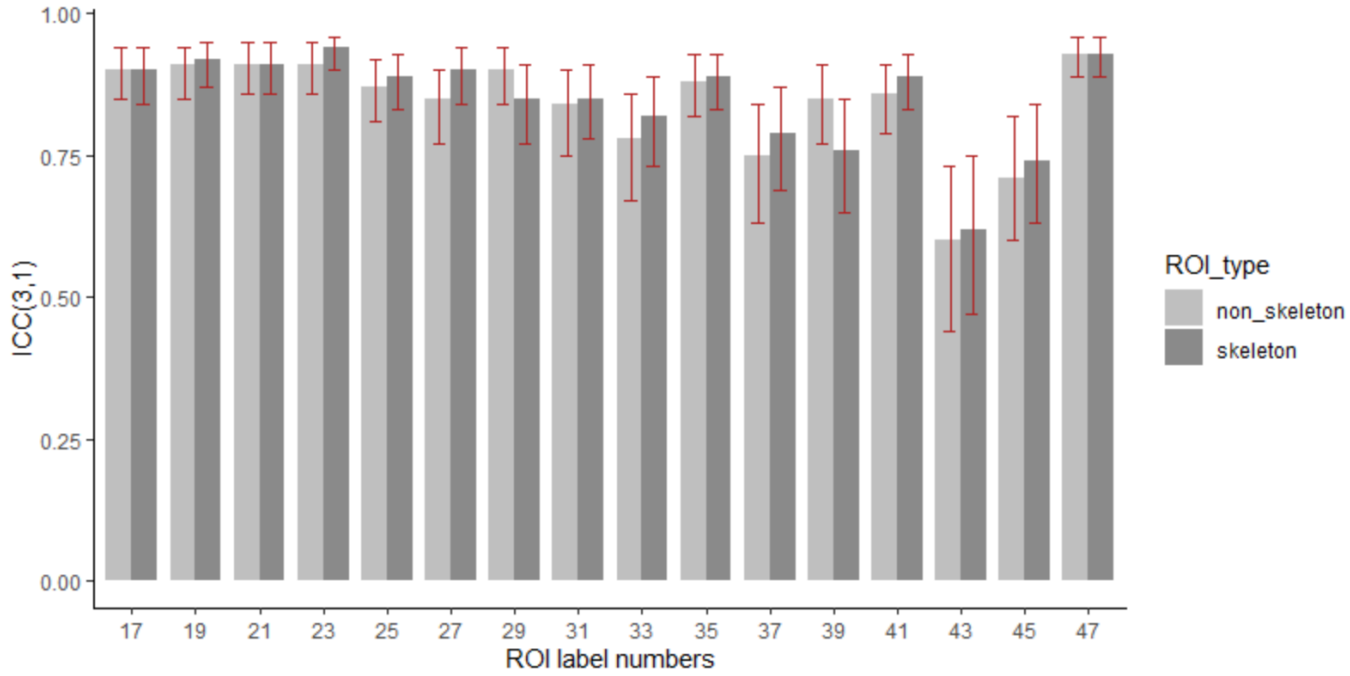

Figure 15a. The ICC(3,1) values of the ROI labels located in the right hemisphere for skeletonized ROI and non-skeletonized ROI types. ROI labels respectively from left to right: (17)anterior limb of internal capsule R, posterior limb of internal capsule R, retrolenticular part of internal capsule R, anterior corona radiata R, superior corona radiata R, posterior corona radiata R, posterior thalamic radiation R, sagittal stratum R, external capsule R, cingulum (cingulate gyrus) R, cingulum (hippocampus) R, fornix (cres) / stria terminalis R, superior longitudinal fasciculus R, superior fronto-occipital fasciculus R, uncinate fasciculus R, tapetum R.

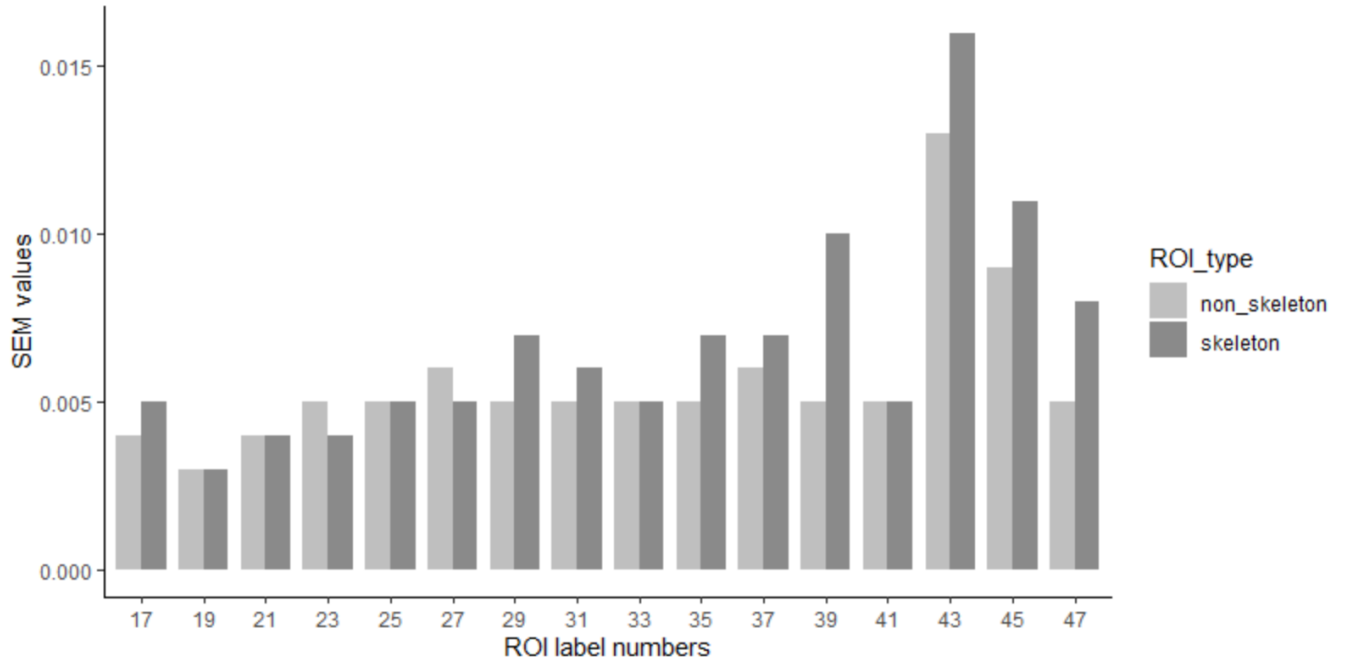

Figure 15b. The SEM values of the ROI labels located in the right hemisphere for skeletonized ROI and non-skeletonized ROI types. ROI labels respectively from left to right: (17)anterior limb of internal capsule R, (19)posterior limb of internal capsule R, (21)retrolenticular part of internal capsule R, (23)anterior corona radiata R, (25)superior corona radiata R, (27)posterior corona radiata R, (29)posterior thalamic radiation R, (31)sagittal stratum R, (33)external capsule R, (35)cingulum (cingulate gyrus) R, (37)cingulum (hippocampus) R, (39)fornix (cres) / stria terminalis R, (41)superior longitudinal fasciculus R, (43)superior fronto-occipital fasciculus R, (45) uncinate fasciculus R, (47)tapetum R.

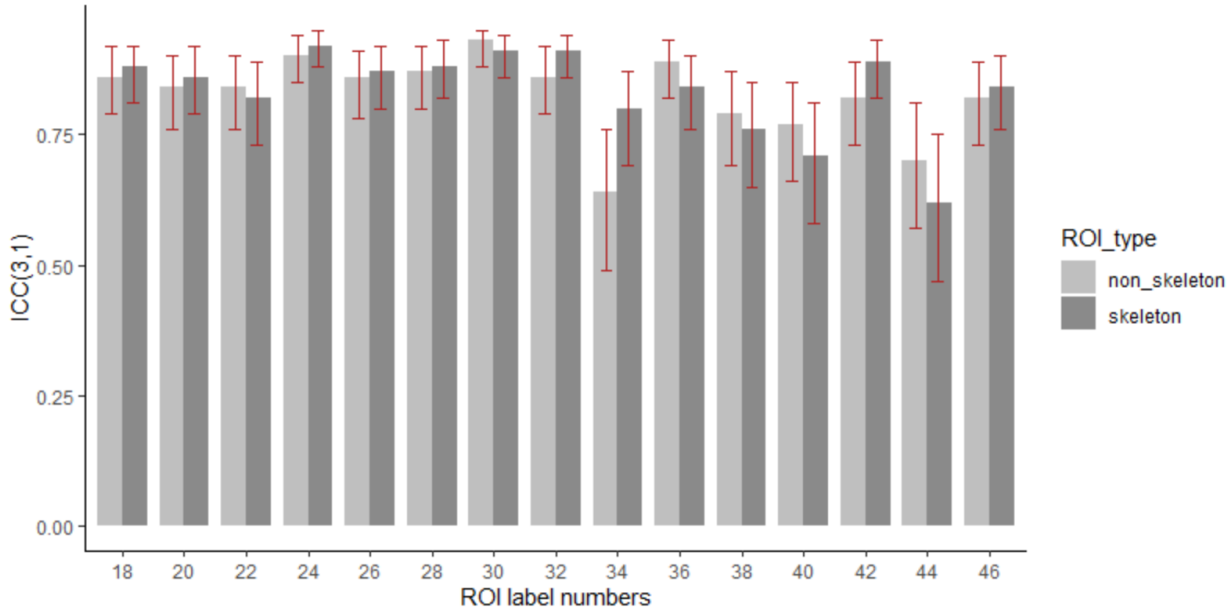

Figure 16a. The ICC(3,1) values of the ROI labels located in the left hemisphere for ROI skeletonized ROI and non-skeletonized ROI types. ROI labels respectively from left to right: (18)anterior limb of internal capsule L, (20)posterior limb of internal capsule L, (22)retrolenticular part of internal capsule L, (24)anterior corona radiata L, (26)superior corona radiata L, (28)posterior corona radiata L, (30)posterior thalamic radiation L, (32)sagittal stratum L, (34)external capsule L, (36)cingulum (cingulate gyrus) L, (38)cingulum (hippocampus) L, (40)fornix (cres) / stria terminalis L, (42)superior longitudinal fasciculus L, (44)superior fronto-occipital fasciculus L, (46)uncinate fasciculus L.

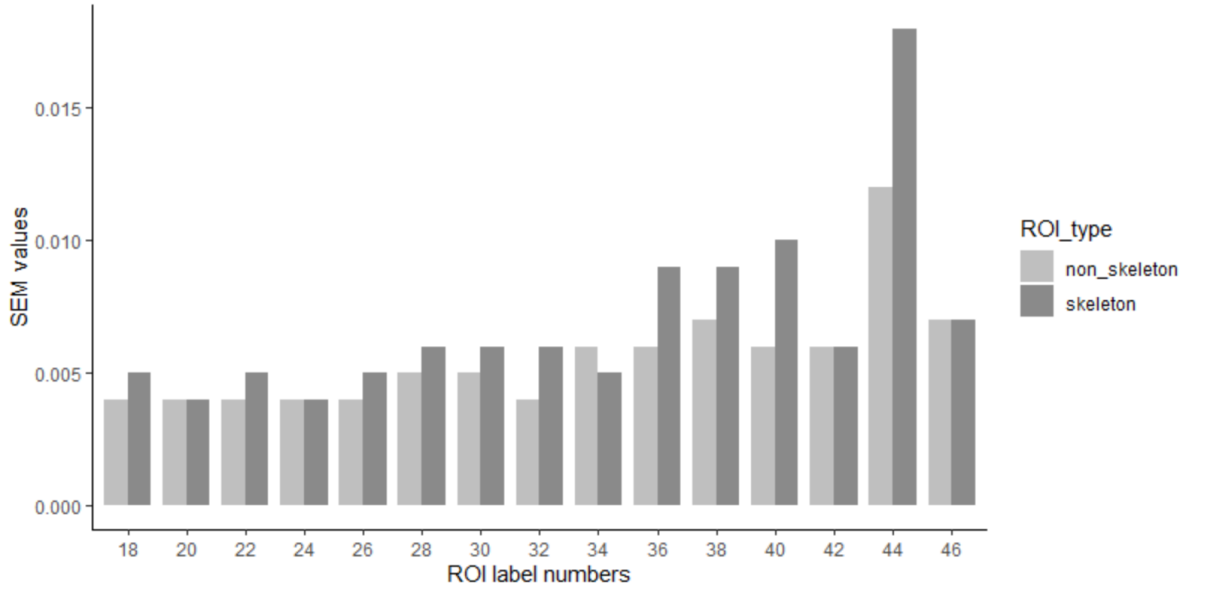

Figure 16b. The SEM values of the ROI labels located in the left hemisphere for skeletonized ROI and non-skeletonized ROI types. ROI labels respectively from left to right: (18)anterior limb of internal capsule L, (20)posterior limb of internal capsule L, (22)retrolenticular part of internal capsule L, (24)anterior corona radiata L, (26)superior corona radiata L, (28)posterior corona radiata L, (30)posterior thalamic radiation L, (32)sagittal stratum L, (34)external capsule L, (36)cingulum (cingulate gyrus) L, (38)cingulum (hippocampus) L, (40)fornix (cres) / stria terminalis L, (42)superior longitudinal fasciculus L, (44)superior fronto-occipital fasciculus L, (46)uncinate fasciculus L.

4 DISCUSSION

The current study followed the footsteps of a previous test-retest reliability study done with neonates to assure the same quality and reliability level with 5 year-olds (Merisaari et al., 2019). Adopting the data acquisition and preprocessing methods stated in the model study, two hypotheses were formulated. First; FA values should have really small differences, if any at all, between the three repetitions. Second, FA values should have high repeatability, i.e. close to zero SEM values and good to excellent level of ICC(3,1) values.

The first two methods, Kruskal Wallis test and Bland Altman plots, used to investigate the possible differences in all measurement combinations showed that, even though there were slight variations in the data they were not significant. Kruskal-Wallis test revealed non-significant results and the three requirements Bland-Altman plots stipulate for the agreement between measures were met. 95% percent of the data remained in the range of ±2 SD of the mean difference, none of the mean differences could be attributed to any point in the data specifically and the mean difference was zero.

The results of the repeatability measures showed that globally, overall three repetitions had ICC(3,1) values referring to moderate to good level of repeatability (ICC(3,1)=0.75, CI:0.64-0.84) and locally almost 90% of the datasets showed good to excellent level of repeatability (ICC(3,1)>0.75). Considering the time difference between the first repetition and the third repetition, having high ICC(3,1) values and low SEM values for each repetition and those values not being significantly different from the each other are supportive evidence for the repeatability of the data. Also, even though an adult atlas was used, the high repeatability provides one justification for using and adult template and atlas in 5-year-old population. Future studies should confirm this by additionally using a study-specific template.

In accordance with the previous studies conducted in different populations and age groups there was quite little variability in FA between measurements (Boekel, Forstmann, & Keuken, 2017; Carlson et al., 2014; Hakulinen et al., 2012; Merisaari et al., 2019; Shahim, Holleran, Kim, & Brody, 2017; Zhou et al., 2018). The ICC(3,1) values, in this study they were a bit smaller than the ones acquired from neonate data (Merisaari et al., 2019). The reason for this is thought to be the subject motion, since the neonate data in collected during sleep, there was considerably less effect of subject motion on the data in those cases.

The ROIs in the study is manipulated through 2 different processes. The first one is exclusion of the ROIs classified as brainstem ROIs (Mori et al., 2008) and cerebral peduncle in both hemispheres. In total of 12 ROIs were excluded for they were of no interest to the study and their locations were not consistently included in the field of view as per scanning instructions. The second one is the conservative thresholding (0.25) used for FA values. The threshold value is different than the default value in FSL and this approach has disadvantage of losing WM unnecessarily. In tract base analysis studies it was shown that, especially in specific ROIs, having a more conservative FA threshold can lead to problems due to its effect on the DTI metrics, for instance having a lesser number of tracts or not being able to draw the WM tracts at all (Domin, Langner, Hosten, & Lotze, 2014; Taoka et al., 2009). In this study there were uneven numbers of ROI labels in skeletonized (47 labels before exclusions) and non-skeletonized (48 labels before exclusions) ROI types. Tapetum of corpus callosum in the left hemisphere was not represented in the ROIs of the skeletonized ROI type and tapetum of the corpus callosum in the right hemisphere which was represented in the skeletonized ROI type had the smallest volume among others. Therefore, it is likely that tapetum in the left hemisphere was just ruled out during the skeletonization process due to conservative thresholding. Yet, in this study after evaluating the images with thresholding of recommended 0.2, a higher threshold was needed to assure accuracy in WM selection. These manipulations

made on ROIs in the study are good examples of how research question and through data quality assessment must be considered as a guide in ROI definition and selection.

In ROI based repeatability results, a strong indicator of good test-retest reliability of the data is having higher ICC(3,1) values for corpus callosum (CC). CC is a large and directionally coherent i.e. without curves or crossing fibers, compact WM fiber bundle and consequently, has high anisotropy (Hasan, Gupta, Santos, Wolinsky, & Narayana, 2005). For that reason, CC is a good reference ROI and having high repeatability in the region is crucial.

Besides CC, almost all the ROIs had high ICC(3,1) values but when ROIs were investigated according to the classification of tracts provided with the atlas a variety was observed in terms of repeatability level. In accordance with literature, association fibers show more variability in terms of ICC(3,1) values compared to commissural and projection fibers and commissural tracts had higher ICC(3,1) values in both ROI types (Duan, Zhao, He, & Shu, 2015).

Investigating the repeatability of ROIs separately also revealed the agreement between ICC(3,1) and SEM values. When the ROIs with highest repeatability and lowest repeatability were plotted, most of the ROIs in both low and high repeatability plots were the same in both repeatability by ICC(3,1) and repeatability by SEM plots without major conflicts. One of the ROIs classified as an association tract in the atlas had specifically lower repeatability on all measures for both ROI types, Superior fronto-occipital fasciculus (SFOF). Yet, it is rather controversial whether it is an association tract, and whether it even exists at all (Schmahmann & Pandya, 2007). Whilst some studies could not locate a WM tract where the SFOF was thought to be, some studies did find WM tracts in the location of interest but they also found that SFOF was not a tract on its own connecting frontal lobe to any specific areas, but might belong to other WM tracts like superior thalamic peduncle or anterior thalamic radiation (Bao,

Wang, Wang, & Wang, 2017; Ture, Yaşargil, & Pait, 1997). The low repeatability of FA values in the ROI can be explained by how established the ROI itself is and the challenges of such interconnected the area with many different crossing fibers bring.

All volume sizes for all the labels in three repetitions were equal in non-skeletonized version, and there were small differences in volume size in skeletonized version. However, since the biggest difference between the volume size and the average was 29, it was safe to take the average value as a representative volume size for the label. Based on the results, higher repeatability values could be associated with smaller volume sizes. However, the magnitude was low, none of the regression models used could not predict the data well. Therefore, it can be concluded that small volume size was not a standalone predictor of lower repeatability and the measurements were repeatable regardless the size of the label. As literature suggests, repeatability cannot only be explained by volume alone (Vos, Jones, Viergever, & Leemans, 2011). The results of the test-retest reliability of the neonate study supported this finding. Even though neonate brains were much smaller than the 5-year-old brains, neonate data showed relatively higher ICC(3,1) values compared to the 5-year-olds data (Merisaari et al., 2019). Based on this, it can be argued that motion is a much more profound factor in repeatability than the volume size. However, caution is required when it comes to interpreting the repeatability of the smaller ROIs since the volume size is still one of the contributing factors and FA seems to be sensitive to volume changes (Vos et al., 2011). One viable additional analysis would be a detailed probing of different motion profiles and their influences on repeatability.

Lastly, to investigate the differences between the ROI types, the SEM and ICC(3,1) values were plotted next to each other for each ROI label of both ROI types, for better visuality purposes the values were plotted separately for each hemisphere. No profound effect of using a skeletonized ROI type was observed based on the SEM values and the ICC(3,1) values and the CIs.

4.1 Limitations of the Study

There are some limitations in this study. The first one is the small number of subjects, which is partly due to ongoing data collection. Sample size affects the statistical power of the analyses, confidence intervals and overall conclusions drawn based on those analyses. The effort was made to keep the number of subjects as high as possible by only employing an automated exclusion step in the analysis pipeline and eliminating the manual exclusion step to avoid further reduction of the number of subjects. That is also affecting the quality of data and is therefore a limitation. Yet, considering the robust repeatability of the results, it can be claimed that FA measures had high test-retest reliability.

Previous studies suggest that in adult data FA measures tend to be more similar in intrascanner studies compared to interscanner studies (Pfefferbaum, Adalsteinsson, & Sullivan, 2003; Vollmar et al., 2010). Hence, it would be worth conducting interscanner repeatability studies in 5-year-olds as well. Such approaches are in use in state-of-the-art multi-site studies (Casey et al., 2018).

The atlas used in the study, as per other WM atlases, inevitably includes both anatomical boundaries for the ROIs and hypothetical ones for anatomical restrictions to draw the boundaries are not always available (Mori et al., 2008). Because of that reason even though using an atlas standardize and automize the analysis especially for healthy population (Faria et al., 2010), there are differences between the atlases and that is another aspect of the test-retest reliability analyses a future study should investigate further.

5 CONCLUSION

The current study conducted a test-retest analyses of the 96 direction DTI data of 5-year-old children gathered in three repetitions based on two different measures. In the light of the results, it can be claimed that in 5-year-old children data FA values gathered from VBA and ROI

analysis using adult MNI space ICBM atlas with and without a skeleton were in line with the literature and had good repeatability. High repeatability was achieved even in the data that has less than highest possible quality. The use of skeleton in ROI analysis did not have a significant effect on repeatability and despite the fact that volume of the ROI could not predict the repeatability, it was found to be a contributing factor.

REFERENCES

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 and Function*, *224*(9), 3373–3385. https://doi.org/10.1007/s00429-019-01981-y

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

Bach, M., Laun, F. B., Leemans, A., Tax, C. M. W., Biessels, G. J., Stieltjes, B., & Maier-Hein, K. H. (2014). Methodological considerations on tract-based spatial statistics (TBSS). *NeuroImage*, *100*, 358–369. https://doi.org/10.1016/j.neuroimage.2014.06.021

Baliyan, V., Das, C. J., Sharma, R., & Gupta, A. K. (2016). Diffusion weighted imaging: Technique and applications. *World Journal of Radiology*, *8*(9), 785–799. https://doi.org/10.4329/wjr.v8.i9.785

Bao, Y., Wang, Y., Wang, W., & Wang, Y. (2017). The Superior Fronto-Occipital Fasciculus in the Human Brain Revealed by Diffusion Spectrum Imaging Tractography : An Anatomical Reality or a Methodological Artifact ? *Frontiers in Neuroanatomy* |, *11*, 1–10. https://doi.org/10.3389/fnana.2017.00119

Basser, P. J., & Jones, D. K. (2002). Diffusion-tensor MRI: Theory, experimental design and data analysis - A technical review. *NMR in Biomedicine*, *15*(7–8), 456–467. https://doi.org/10.1002/nbm.783

Billiet, T., Vandenbulcke, M., Mädler, B., Peeters, R., Sunaert, S., & Emsell, L. (2015). Neurobiology of Aging Age-related microstructural differences quantified using myelin water imaging and advanced diffusion MRI. *Neurobiology of Aging*, *36*(6), 2107–2121. https://doi.org/10.1016/j.neurobiolaging.2015.02.029

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135–160.

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

Carlson, H. L., Laliberté, C., Brooks, B. L., Hodge, J., Kirton, A., Bello-espinosa, L., … Sherman, E. M. S. (2014). Epilepsy & Behavior Reliability and variability of diffusion tensor imaging (DTI) tractography in pediatric epilepsy. *Epilepsy & Behavior*, *37*, 116–122. https://doi.org/10.1016/j.yebeh.2014.06.020

Cascio, C. J., Gerig, G., & Piven, J. (2007). Diffusion Tensor Imaging : Application to the Study of the Developing Brain. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*(2), 213–223. https://doi.org/10.1097/01.chi.0000246064.93200.e8

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., … Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study : Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*(May 2017), 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Chan, Y., & Walmsley, R. P. (1997). Learning and Understanding the Kruskal-Wallis One-

Way Analysis-of-Variance-by-Ranks Test for Differences Among Three or More Independent Groups. *Physical Therapy*, *77*(12), 1755–1761. https://doi.org/10.1093/ptj/77.12.1755

Croteau-Chonka, E. C., Dean, D. C., Remer, J., Dirks, H., O'Muircheartaigh, J., & Deoni, S. C. 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

Dean, D. C., Dirks, H., O'Muircheartaigh, J., Walker, L., Jerskey, B. A., Lehman, K., … Deoni, S. C. L. (2014). Pediatric neuroimaging using magnetic resonance imaging during non-sedated sleep. *Pediatric Radiology*, *44*, 64–72. https://doi.org/10.1007/s00247-013-2752-8

Domin, M., Langner, S., Hosten, N., & Lotze, M. (2014). Comparison of Parameter Threshold Combinations for Diffusion Tensor Tractography in Chronic Stroke Patients and Healthy Subjects. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0098211

Drobyshevsky, A., Back, S., Wyrwicz, A. M., Li, L., Derrick, M., Ji, X., … Tan, S. (2004). Diffusion tensor imaging of the developing rabbit brain. *Magnetic Resonance in Medicine*, *11*(July), 2004–2004. Retrieved from https://www.sciencedirect.com/science/article/pii/S1933721307000955

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 M*, *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*(12), 895. https://doi.org/10.4329/wjr.v6.i12.895

Faria, A. V, Zhang, J., Oishi, K., Li, X., Jiang, H., Akhter, K., … Mori, S. (2010). Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection. *NeuroImage*, *52*(2), 415–428. https://doi.org/10.1016/j.neuroimage.2010.04.238

Ferrer, I., Martinez, A., Boluda, S., Parchi, P., & Barrachina, M. (2008). Brain banks : benefits , limitations and cautions concerning the use of post-mortem brain tissue for molecular studies. *Cell Tissue Banking*, *9*, 181–194. https://doi.org/10.1007/s10561-008-9077-0

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased Average Age-Appropriate Atlases for Pediatric Studies. *NeuroImage*, *54*(1), 313–327. https://doi.org/10.1161/CIRCULATIONAHA.110.956839

Girault, J. B., Cornea, E., Goldman, B. D., Jha, S. C., Murphy, V. A., Li, G., … Gilmore, J. H. (2019). Cortical Structure and Cognition in Infants and Toddlers. *Cerebral Cortex*, *i*(July), 1–15. https://doi.org/10.1093/cercor/bhz126

Grieve, S. M., Williams, L. M., Paul, R. H., Clark, C. R., & Gordon, E. (2007). Cognitive aging, executive function, and fractional anisotropy: A diffusion tensor MR imaging study. *American Journal of Neuroradiology*, *28*(2), 226–235.

Hakulinen, U., Brander, A., Ryymin, P., Öhman, J., Soimakallio, S., Helminen, M., … Eskola, H. (2012). Repeatability and variation of region-of-interest methods using quantitative diffusion tensor MR imaging of the brain. *BMC Medical Imaging*, *12*.

https://doi.org/10.1186/1471-2342-12-30

Harvill, L. M. (1991). ITEMS lnstructional Topics in Educational Measurement Standard Error of Measurement.

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

Heiervang, E., Behrens, T. E. J., Mackay, C. E., Robson, M. D., & Johansen-Berg, H. (2006). Between session reproducibility and between subject variability of diffusion MR and tractography measures. *NeuroImage*, *33*(3), 867–877. https://doi.org/10.1016/j.neuroimage.2006.07.037

Huber, E., Henriques, R. N., Owen, J. P., Rokem, A., & Yeatman, J. D. (2019). Applying microstructural models to understand the role of white matter in cognitive development. *Developmental Cognitive Neuroscience*, *36*(January), 100624. https://doi.org/10.1016/j.dcn.2019.100624

Innocenti, G. M., Ansermet, F., & Parnas, J. (2003). Schizophrenia, neurodevelopment and corpus callosum. *Molecular Psychiatry*, 261–274. https://doi.org/10.1038/sj.mp.4001205

Jakab, A., Tuura, R., Kellenberger, C., & Scheer, I. (2017). In utero diffusion tensor imaging of the fetal brain: A reproducibility study. *NeuroImage: Clinical*, *15*, 601–612. https://doi.org/10.1016/j.nicl.2017.06.013

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

Karababa, I. F., Bayazıt, H., Kılıçaslan, N., & Celik, M. (2015). Microstructural Changes of Anterior Corona Radiata in Bipolar Depression. *Psychiatry Investig*, *12*(3), 367–371.

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H. M., Korja, R., Ekholm, E., … Karlsson, H. (2018). Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain). *International Journal of Epidemiology*, *47*(1), 15-16j. https://doi.org/10.1093/ije/dyx173

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., … Gilmore, J. H. (2008). A structural MRI study of human brain development from birth to 2 years. *Journal of Neuroscience*, *28*(47), 12176–12182. https://doi.org/10.1523/JNEUROSCI.3479-08.2008

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kumar, R., Chavez, A. S., Macey, P. M., Woo, M. A., & Harper, R. M. (2013). Brain Axial and Radial Diffusivity Changes with Age and Gender in Healthy Adults. *Brain Research*, (1512), 22–36. https://doi.org/10.1016/j.brainres.2013.03.028.Brain

Le Bihan, D. (2007). The “wet mind”: Water and functional neuroimaging. *Physics in Medicine and Biology*, *52*(7), 57–90. https://doi.org/10.1088/0031-9155/52/7/R02

Le Bihan, D. (2013). Apparent diffusion coefficient and beyond : What diffusion mr imaging can tell us about tissue structure. *Radiology*, *268*(2), 318–322.

https://doi.org/10.1148/radiol.13130420

Le Bihan, D., & Iima, M. (2015). Diffusion magnetic resonance imaging: What water tells us about biological tissues. *PLoS Biology*, *13*(7), 1–13. https://doi.org/10.1371/journal.pbio.1002203

Le Bihan, D., & Johansen-Berg, H. (2012). Diffusion MRI at 25: Exploring brain tissue structure and function. *Neuroimage.*, *6*(12), 324–341. https://doi.org/doi:10.1016/j.neuroimage.2011.11.006. Diffusion

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

Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, *40*(3), 1044–1055. https://doi.org/10.1016/j.neuroimage.2007.12.053

Lebel, Catherine, Benner, T., & Beaulieu, C. (2012). Six is enough? Comparison of diffusion parameters measured using six or more diffusion-encoding gradient directions with deterministic tractography. *Magnetic Resonance in Medicine*, *68*(2), 474–483. https://doi.org/10.1002/mrm.23254

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

Lewis, A. F., Myers, M., Heiser, J., Kolar, M., Baird, J. F., & Stewart, J. C. (2020). Test–retest reliability and minimal detectable change of corticospinal tract integrity in chronic stroke. *Human Brain Mapping*, (December 2019), 1–13. https://doi.org/10.1002/hbm.24961

Li, S. J., Wang, Y., Qian, L., Liu, G., Liu, S. F., Zou, L. P., … Ma, L. (2018). Alterations of white matter connectivity in preschool children with autism spectrum disorder. *Radiology*, *288*(1), 209–217. https://doi.org/10.1148/radiol.2018170059

Long, X., Benischek, A., Dewey, D., & Lebel, C. (2017). Age-related functional brain changes in young children. *NeuroImage*, *155*(November 2016), 322–330. https://doi.org/10.1016/j.neuroimage.2017.04.059

Machilsen, B., Maes, F., Vandermeulen, D., Hahn, H. K., Lagae, L., & Stiers, P. (2007). Linear normalization of MR brain images in pediatric patients with periventricular leukomalacia. *Human Brain Mapping Journal*, *35*(2), 686–697. https://doi.org/10.1016/j.neuroimage.2006.12.037

Mckinstry, R. C., & Mathur, A. (2002). Radial organization of developing human cerebral cortex revealed by non- invasive water diffusion anisotropy MRI Radial Organization of Developing Preterm Human Cerebral Cortex Revealed by Non-invasive Water Diffusion Anisotropy. *Cerebral Cortex*, (12), 1237–1243. https://doi.org/10.1093/cercor/12.12.1237

Merisaari, H., Tuulari, J. J., Karlsson, L., Scheinin, N. M., Parkkola, R., Saunavaara, J., … Karlsson, H. (2019). Test-retest reliability of Diffusion Tensor Imaging metrics in neonates. *NeuroImage*, *197*(September 2018), 598–607. https://doi.org/10.1016/j.neuroimage.2019.04.067

Moon, W. J., Provenzale, J. M., Sarikaya, B., Ihn, Y. K., Morlese, J., Chen, S., & DeBellis, M. D. (2011). Diffusion-tensor imaging assessment of white matter maturation in childhood and adolescence. *American Journal of Roentgenology*, *197*(3), 704–712. https://doi.org/10.2214/AJR.10.6382

Mori, S., Oishi, K., & Faria, A. V. (2009). White matter atlases based on diffusion tensor imaging. *Curr Opin Neurol.*, *22*(4), 362–369. https://doi.org/10.1097/WCO.0b013e32832d954b.White

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., … Mazziotta, J. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *NeuroImage*, *40*(2), 570–582. https://doi.org/10.1016/j.neuroimage.2007.12.035

Mori, S., & Tournier, J. D. (2014). Introduction to Diffusion Tensor Imaging and Higher Order Models. In *Introduction to Diffusion Tensor Imaging and Higher Order Models* (2nd ed., pp. 1–11). Elsevier B. V.

Mori, S., & Zhang, J. (2006). Principles of Diffusion Tensor Imaging and Its Applications to Basic Neuroscience Research. *Neuron*, *51*(5), 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., … Deoni, S. C. L. (2014). White Matter Development and Early Cognition in Babies and Toddlers. *Human Brain Mapping*, *4487*(February), 4475–4487. https://doi.org/10.1002/hbm.22488

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

Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., … Mori, S. (2008). Human Brain White Matter Atlas: Identification and Assignment of Common Anatomical Structures in Superficial White Matter. *NeuroImage*, *43*(3), 447–457. https://doi.org/doi:10.1016/j.neuroimage.2008.07.009. Human

Partridge, S. C., Mukherjee, P., Henry, R. G., Miller, S. P., Berman, J. I., Jin, H., … Vigneron, D. B. (2004). Diffusion tensor imaging: Serial quantitation of white matter tract maturity in premature newborns. *NeuroImage*, *22*(3), 1302–1314. https://doi.org/10.1016/j.neuroimage.2004.02.038

Paus, T., Collins, D. L., Evans, A. C., Leonard, G., Pike, B., & Zijdenbos, A. (2001). Maturation of white matter in the human brain : A review of magnetic resonance studies. *Brain Research Bulletin*, *54*(3), 255–266.

Pfefferbaum, A., Adalsteinsson, E., & Sullivan, E. V. (2003). Replicability of Diffusion Tensor Imaging Measurements of Fractional Anisotropy and Trace in Brain. *Journal of Magnetic Resonance Imaging*, *18*, 427–433. https://doi.org/10.1002/jmri.10377

Pfefferbaum, A., Mathalon, D. H., Sullivan, E. V., Rawles, J. M., Zipursky, R. B., & Lim, K. O. (1994). A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Arch Neurol*, *51*, 874–887.

Poynton, C., Jenkinson, M., Whalen, S., Golby, A. J., & Wells III, W. (2008). *Medical Image Computing and Computer-Assisted Intervention*. (D. Metaxas, L. Axel, G. Fichtinger, & G. Székely, Eds.) (11th ed.). Springer-Verlag Berlin Heidelberg.

Qiu, A., Mori, S., & Miller, M. I. (2015). Diffusion Tensor Imaging for Understanding Brain Development in Early Life. *Annu Rev Psychol*, *3*(66), 853–876. https://doi.org/10.1146/annurev-psych-010814-015340

Rajagopalan, V., Radic, J. S., Pioro, E. P., Wylie, G. R., Jiang, Z., Das, A., & Yue, G. H. (2017). A Basic Introduction to Diffusion Tensor Imaging Mathematics and Image Processing Steps. *Brain Disorders & Therapy*, *06*(02). https://doi.org/10.4172/2168-975x.1000229

Reynolds, J. E., Grohs, M. N., Dewey, D., & Lebel, C. (2019). Global and regional white matter development in early childhood. *NeuroImage*, *196*(October 2018), 49–58. https://doi.org/10.1016/j.neuroimage.2019.04.004

Richards, J. E., & Xie, W. (2015). *Brains for All the Ages: Structural Neurodevelopment in Infants and Children from a Life-Span Perspective*. *Advances in Child Development and Behavior* (Vol. 48). https://doi.org/10.1016/bs.acdb.2014.11.001

Schmahmann, J. D., & Pandya, D. N. (2007). The Complex History of the Fronto-Occipital Fasciculus. *Journal of the History of the Neuroscience*, *16*, 362–377. https://doi.org/10.1080/09647040600620468

Shahim, P., Holleran, L., Kim, J. H., & Brody, D. L. (2017). Test-retest reliability of high spatial resolution diffusion tensor and diffusion kurtosis imaging. *Scientific Reports*, *7*, 1–14. https://doi.org/10.1038/s41598-017-11747-3

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*(5), 1184–1189. https://doi.org/10.1002/mrm.20267

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability.1. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979, 86:420–8. *Psychological Bulletin*, *86*(2), 420–428. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18839484

Siasios, I., Kapsalaki, E. Z., Fountas, K. N., Fotiadou, A., Dorsch, A., Vakharia, K., … Dimopoulos, V. G. (2016). The role of diffusion tensor imaging and fractional anisotropy in the evaluation of patients with idiopathic normal pressure hydrocephalus: A literature review. *Neurosurgical Focus*, *41*(3), 1–9. https://doi.org/10.3171/2016.6.FOCUS16192

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 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., … Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*(4), 1487–1505. https://doi.org/10.1016/j.neuroimage.2006.02.024

Snook, L., Plewes, C., & Beaulieu, C. (2007). Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment. *NeuroImage*, *34*(1), 243–252. https://doi.org/10.1016/j.neuroimage.2006.07.021

Tamnes, C. K., Herting, M. M., Goddings, A. L., Meuwese, R., Blakemore, S. J., Dahl, R. E., … Mills, K. L. (2017). Development of the cerebral cortex across adolescence: A

multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *Journal of Neuroscience*, *37*(12), 3402–3412. https://doi.org/10.1523/JNEUROSCI.3302-16.2017

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*(May 2017), 161–175. https://doi.org/10.1016/j.dcn.2017.12.002

Taoka, T., Morikawa, M., Akashi, T., Miyasaka, T., Nakagawa, H., Kiuchi, K., … Kichikawa, K. (2009). Fractional Anisotropy - Threshold Dependence in Tract-Based Diffusion Tensor Analysis : Evaluation of the Uncinate Fasciculus in Alzheimer Disease. *American Journal of Neuroradiology*, *30*, 1700–1703. https://doi.org/10.3174/ajnr.A1698

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*(12), 4405–4424. https://doi.org/10.1002/hbm.23318.Assessing

Tocchio, S., Kline-Fath, B., Kanal, E., Schmithorst, V. J., & Panigrahy, A. (2015). MRI Evaluation and Safety in the Developing Brain. *Semin Perinatol*, *39*(2), 73–104. https://doi.org/10.1016/j.physbeh.2017.03.040

Ture, U., Yaşargil, M. G., & Pait, T. G. (1997). Is There a Superior Occipitofrontal Fasciculus? A Microsurgical Anatomic Study. *Neurosurgery*, *40*(6), 1226–1232. https://doi.org/10.1097/0006123-199706000-00022

Van Hecke, W., Emsell, L., & Sunaert, S. (2015). Diffusion Tensor Imaging: A Practical Handbook. In *Diffusion Tensor Imaging: A Practical Handbook* (pp. 1–440). https://doi.org/10.1007/978-1-4939-3118-7

Venkatraman, V. K., Christopher, G. E., Landman, B., Goh, J., Reiter, D. A., An, Y., & Resnick, S. M. (2015). Region of Interest Correction Factors Improve Reliability of Diffusion Imaging Measures Within and Across Scanners and Field Strengths. *NeuroImage*, *01*(119), 406–416. https://doi.org/10.1016/j.neuroimage.2015.06.078

Voldsbekk, I., Maximov, I. I., Zak, N., Roelfs, D., Geier, O., Due-Tønnessen, P., … Groote, I. (2020). Evidence for wakefulness-related changes to extracellular space in human brain white matter from diffusion-weighted MRI. *NeuroImage*, *212*(October 2019). https://doi.org/10.1016/j.neuroimage.2020.116682

Vollmar, C., Muircheartaigh, J. O., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., … Koepp, M. J. (2010). Identical , but not the same : Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3 . 0 T scanners. *NeuroImage*, *51*(4), 1384–1394. https://doi.org/10.1016/j.neuroimage.2010.03.046

Vos, S. B., Jones, D. K., Viergever, M. A., & Leemans, A. (2011). NeuroImage Partial volume effect as a hidden covariate in DTI analyses. *NeuroImage*, *55*(4), 1566–1576. https://doi.org/10.1016/j.neuroimage.2011.01.048

Winklewski, P. J., Sabisz, A., Naumczyk, P., Jodzio, K., Szurowska, E., & Szarmach, A. (2018). Understanding the physiopathology behind axial and radial diffusivity changes-what do we Know? *Frontiers in Neurology*, *9*(FEB). https://doi.org/10.3389/fneur.2018.00092

Yoon, U., Fonov, V. S., Perusse, D., & Evans, A. C. (2009). The effect of template choice on morphometric analysis of pediatric brain data. *NeuroImage*, *45*(3), 769–777. https://doi.org/10.1016/j.neuroimage.2008.12.046

Yoshida, S., Oishi, K., Faria, A. V., & Mori, S. (2013). Diffusion tensor imaging of normal brain development. *Pediatric Radiology*, *43*(1), 15–27. https://doi.org/10.1007/s00247-012-2496-x

Young, J. M., Vandewouw, M. M., Morgan, B. R., Smith, M. Lou, Sled, J. G., & Taylor, M. J. (2018). Altered white matter development in children born very preterm. *Brain Structure and Function*, *223*(5), 2129–2141. https://doi.org/10.1007/s00429-018-1614-4

Zhou, X., Sakaie, K. E., Debbins, J. P., Narayanan, S., Fox, R. J., & J., L. M. (2018). Scan-rescan repeatability and cross-scanner comparability of DTI metrics in healthy subjects in the SPRINT-MS multicenter trial. *Magn Reson Imaging*, *53*, 105–111. https://doi.org/10.1016/j.physbeh.2017.03.040