NeuroImage 197 (2019) 598–607

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Test-retest reliability of Diffusion Tensor Imaging metrics in neonates

Harri Merisaari [a,h,i,*,1], Jetro J. Tuulari [a,b,j,1], Linnea Karlsson [a,e], Noora M. Scheinin [a,b], Riitta Parkkola [d], Jani Saunavaara [c], Tuire Lähdesmäki [g], Satu J. Lehtola [a], Maria Keskinen [a], John D. Lewis [f], Alan C. Evans [f], Hasse Karlsson [a,b]

[a] FinnBrain Birth Cohort Study, Turku Brain and Mind Center, Institute of Clinical Medicine, University of Turku, Turku, Finland
[b] Department of Psychiatry, University of Turku and Turku University Hospital, Turku, Finland
[c] Department of Medical Physics, Turku University Hospital, Turku, Finland
[d] Department of Radiology, University of Turku and Turku University Hospital, Turku, Finland
[e] Department of Child Psychiatry, University of Turku and Turku University Hospital, Turku, Finland
[f] McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, McGill University, Montreal, Canada
[g] Department of Pediatric Neurology, Turku University Hospital and University of Turku, Finland
[h] Department of Future Technologies, University of Turku, Finland
[i] Center of Biomedical Engineering and Personalized Medicine, Case Western Reserve University, Cleveland, OH, USA
[j] Turku Collegium for Science and Medicine, University of Turku, Turku, Finland

ARTICLE INFO

Keywords:
DTI
Neonate
Test-retest reliability
TBSS
ICC

ABSTRACT

Diffusion tensor imaging (DTI) has been widely used in children and adults to study the microstructural features of the brain. Its use in neonate brains has been limited. Neonate brains are almost completely unmyelinated, and this together with the tendency for babies to move during a scanning session may affect the reliability of the measurements. Here we divided a 96 direction acquisition into three segments, and analysed the intra scan test-retest reliability for pairs of segments. Each segment was subjected to a rigorous quality control, and from the surviving data we chose 25 diffusion encoding directions from each segment, and assessed the pairwise reliability of the most common DTI metrics. This pairwise reliability was assessed for data from 86 infants. We used tract-based spatial statistics (TBSS), voxelwise and ROI analysis schemes, to see potential differential effects of analysis strategy and post processing on the obtained DTI metrics. We found that intra class correlation coefficient (ICC) values were generally high (ICC > 0.80). Residual motion in the data, after quality control, was not found to associate with the diffusion metrics. The results indicate that DTI metrics from neonate data can be reliable, even at relatively low angular resolution that are common for neonate scans. The results lend confidence to the use of neonate DTI data in cross sectional and longitudinal analyses in brain white matter skeleton. Future studies should assess the reliability of fiber tracking techniques in neonate data.

1. Introduction

Neonatal MR imaging provides measures for understanding the normally developing brain and the effects of prematurity and disease (Kelly et al., 2018). Diffusion Tensor Imaging (DTI) measures the diffusion of water molecules and, because that diffusion is restricted by cell membranes, myelin, etc., can be used to study the microstructural features of the brain. It is most often used to assess white matter (WM) structure. Common modelling includes calculation of scalar measures such as fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD). Controlling measurement error in the scalar measures is of high importance for studies of brain development. Infants frequently move even during evidently calm sleep, and head motion is known to cause bias in the parameter estimates of the DTI models (Roalf et al., 2016; Ling et al., 2012; Yendiki et al., 2014; Gumus et al., 2014; Graham et al., 2017). As DTI sequences are typically time consuming, it is not possible to repeat the entire sequence in case of excessive motion. A solution is to divide the acquisition scheme into several parts that enable redoing the individual parts, if excessive movement occurs. We have used such a scheme for neonates of 2–5 weeks of age. We divided a 96 direction sequence into three segments. An added benefit of this scheme is that it allows us to assess the test–retest reliability (TRTR) of the diffusion

* Corresponding author. Center of Computational Imaging and Personalized Diagnostics, Wickenden, 10900, Euclid Ave, Cleveland, OH, USA.
E-mail address: harri.merisaari@utu.fi (H. Merisaari).
[1] Both the authors have equal contribution.

https://doi.org/10.1016/j.neuroimage.2019.04.067
Received 20 September 2018; Received in revised form 17 April 2019; Accepted 24 April 2019
Available online 25 April 2019

scans. A high TRTR for pairs of the pieces of our 96 direction scan would lend confidence that the overall scan will provide true measures of microstructural characteristics.

To the best of our knowledge there has not yet been an assessment of the reliability of diffusion data acquired from neonates. Previous evaluations of the repeatability of diffusion data in developmental context have used data acquired from children from 5.5 years old and older (Bonekamp et al., 2007), and from adults. Extant studies on test retest reliability have typically used paired scans with one weeks interval (Duan et al., 2015), while some also included intra-session reliability measures (Madhyastha et al., 2014). Intra-class correlation coefficient (ICC) values vary but are often more than 0.60 for intra session comparisons. Even across-site comparisons show promisingly good reliabilities (Jovicich et al., 2014). But the reliability of neonatal diffusion imaging may be impacted by a number of factors such as greater head motion. For such data, quality control is of major importance, and automated quality control software is likely to prove helpful (Oguz et al., 2014). Head motion can be measured by post-processing tools aligning the dataset (Smith et al., 2004). Expectedly, techniques used for spatial manipulations affect the reliability measures (Cabeen et al., 2017). The infant brain has also very different white matter features than what is observed later in life, and this may pose additional methodological challenges (Li et al., 2016). The number of diffusion directions (Leemans and Jones, 2009), and Signal-to-Noise Ratio (SNR) (Landman et al., 2007; Farrell et al., 2007) have been reported to affect DTI parameters. Apart from actual diffusion direction orientations and slight difference in number of directions between parts of the 96 direction sequence (numbers of directions were 31, 32 and 33), we had the same acquisition schemes for all scans, the sequence set up was identical, and physiological noise is likely similar across sessions. If the motion and other artefacts are controlled for (Madhyastha et al., 2014), the TRTR should be high and here we aimed to describe this in more detail.

We focused on the effects that may occur due to variations in distribution of diffusion directions remaining after correction for motion and eddy currents; we standardized the comparisons, in terms of the number of directions, by choosing 25 spatially evenly distributed directions in each segment. In a DTI head motion study conducted on adults, it has been noted that the effect of motion can be present even after correction (Ling et al., 2012), and may yield false positive findings, if not controlled for (Yendiki et al., 2014). Extensive evaluation has not been done in neonates for potential factors that may have an effect on precision and accuracy, such as motion metrics, and metrics relating to the distribution of the diffusion directions used for tensor modelling. The purpose of this study was to evaluate the reliability of intra-session quantitative diffusion measures at both voxel and region of interest (ROI) levels.

2. Materials and methods

The FinnBrain Birth Cohort Study [www.finnbrain.fi] was established in 2011 to prospectively investigate the effects of ELS, including prenatal stress (PS) exposure, on child brain development and health (Karlsson et al., 2018). The aim is to identify biomarkers related to ELS exposures as well as eventual trajectories for common psychiatric and somatic illnesses. The MR scanning took place in 2012–2015.

For brain imaging, 180 infants at ages of 2–5 weeks were recruited by contacting the families via telephone (by study nurse or investigators). Exclusion criteria for infants were: occurrence of any perinatal complications with potential neurological consequences (e.g. hypoxia); less than 5 points in the 5 min Apgar score; previously diagnosed central nervous system anomaly or prior clinical MR scan at peripartum due to clinical indications; gestational age <32 weeks; or birth weight <1500 g. All of the children were term-born, with a normal >2500 g weight at birth. Families were provided oral and written information about the study, and the parents provided written consent to participate on behalf of their baby. The study was conducted in accordance with the Declaration of Helsinki, and it was approved by the Ethics Committee of the Hospital District of Southwest Finland (15.3.2011 §95, ETMK: 31/180/2011).

2.1. MRI scanning visits

The scans were acquired at the Medical Imaging Centre of the Hospital District of Southwest Finland). The infants were fed with breastmilk or formula until they slept and subsequently gently swaddled into a vacuum mattress. No anesthetics were used.

All infants were provided with double hearing protection (wax plugs and custom-sized ear muffs). Standard ear muffs were given to parents, as they usually stayed in the scanning room throughout the scanning session. The personnel observed the scanning procedure through the window of the control room. A microphone and loudspeaker in the scanner room facilitated audio monitoring of the infant, and contact to the parent. If the baby woke up during the scan, the session was discontinued.

Each set of structural infant images was checked by an experienced neuroradiologist (author RP) to detect any possible incidental findings in the scans. In accordance with our study protocol, radiology reports were delivered to the researchers who then communicated them to the family within 1–4 weeks of the scans. In the case of an incidental finding, the parents were referred to a child neurologist (author TL) for a neurological check-up (N = 6, for this data set) as per study protocol. These incidental findings are relatively common in newborns (Rooks et al., 2008) and after visual inspection by experienced neuroradiologist (RP), they were deemed not to affect the imaging metrics in the regions of interest (the white matter), therefore we did not exclude participants based on these findings. None of the infants had any clinically identified, significant neurological symptoms or deficits in neurological development at the time of the neurological examination.

2.2. MRI acquisition

MRI scans were conducted on a Siemens Magnetom Verio 3T scanner (Siemens Medical Solutions, Erlangen, Germany) using a 12-element Head Matrix coil. Sequences were scanned in the following order: 1) Axial PD-T2, 2) Sagittal T1-MPRAGE, 3) field map, 4) DTI, and 5) task and rs-fMRI towards the end of the study (year 2015 only). The total duration of the complete scanning protocol did not exceed 60 min. Only DTI data were used in this study.

The DTI sequences were designed to maximize the success of data acquisition even with substantial motion. A standard twice-refocused Spin Echo-Echo Planar Imaging (SE-EPI) sequence with b-value of 1000 s/mm^2 and $2 \times 2x2$ mm^3 isotropic resolution (FOV 208 mm; 64 slices; TR 8500 ms; TE 90 ms) was used to acquire DTI data. Protocol containing 96 diffusion encoding directions was divided into three parts with 31, 32, and 33 directions, where diffusion encoding directions were uniformly distributed inside each part, and in composition of the three parts. This made it possible to repeat just one of the sequences in case of significant subject motion during the scan; this was done for 19 participants. In addition, each of the three sequences contained three volumes without diffusion encoding (b0 images, in the beginning, in the middle, and in the end). The diffusion vectors are provided in Appendix A.

2.3. DTI data preprocessing

Susceptibility artefacts commonly cause distortion in EPI images. We detected minimal distortion in our data set. We chose not to apply distortion correction, as we noted that some of the acquired field maps were corrupted by motion, and we considered that using them could compromise the data quality. The whole data analysis procedure is illustrated in Fig. 1.

First, the b0 volumes were visually checked by three raters, independently. The good quality b0 volumes were coregistered, averaged and moved to the front of the 4D series. A brain mask was then created from the selected b0 volume with FSL's (FMRIB Software Library v 5.0.9

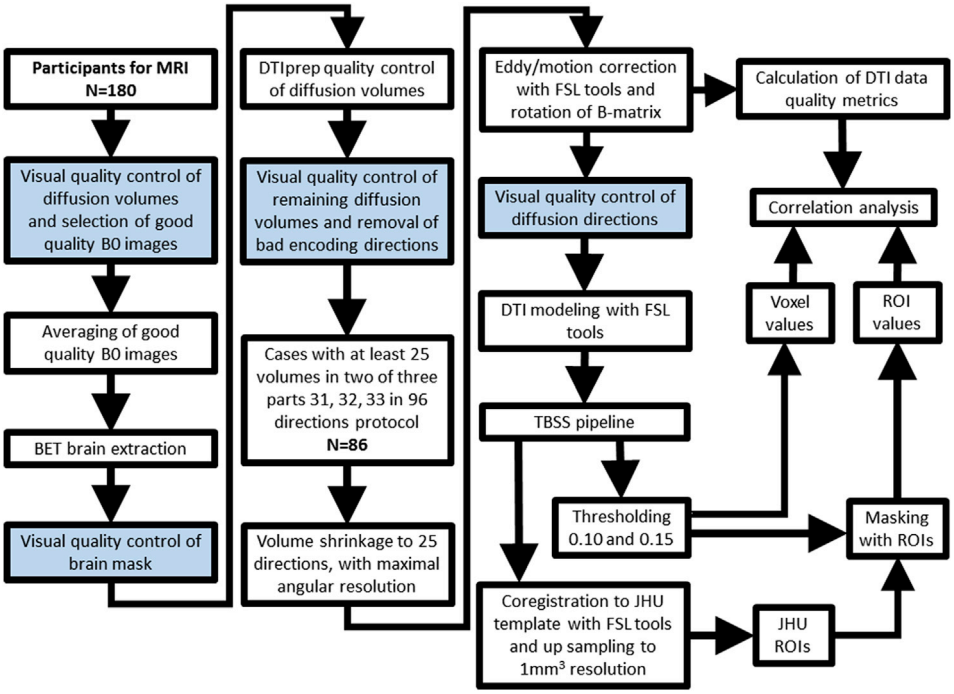

Fig. 1. Data analysis procedure of uniformized 25 directional dataset. Visual quality control steps were used to assure that automated steps worked as intended, and are highlighted in blue. After pre-processing of the DTI images, the TBSS (Tract Based Spatial Statistics) pipeline was applied. The infant JHU atlas (Oishi et al., 2011) was co-registered to the TBSS template volume to have white matter regions with and without TBSS masking in an automatic manner.

(Jenkinson et al., 2012), Brain Extraction Tool (BET) (Smith, 2002) with flags -R -f 0.3. The brain mask was visually inspected by two raters, independently, to assure that this step had been successful.

Second, the quality of diffusion datasets was quantitatively evaluated using DTIprep software, which identifies artefacts in 4D diffusion data (Oguz et al., 2014). Infant scans are commonly corrupted by artefacts that are clear and severe. From our experience a thorough quality control process is needed and strongly recommended for all studies. DTIprep was used inside our batch processing pipeline with default settings and used only to detect poor quality data before subsequent modelling with FSL tools to allow flexible modelling options and file formats for later steps (see details below). Directions/volumes that DTIprep suggested as being of bad quality were discarded. The remaining images were subject to by judgement of visual inspection by (by HM and JJT), and remaining artefactual images were discarded (1–3 diffusion images for 3 subjects). In the analysis presented here, we used only data for which there were two parts with each having at least 25 acceptable diffusion encoding directions. To eliminate the possibility of measurement bias from different numbers of diffusion directions, we kept only 25 gradient directions from data with more than 25 directions in such a way that maximized the final angular resolution. The angular resolution was calculated as the inverse of the largest empty region in the spherical representation of the diffusion directions. This was measured using cap angles of spherical segments between triplets of directions calculated with the spherical voronoi algorithm in Scipy python library (Millman and Aivazis, 2011; Caroli et al., 2010), and angles between pairs of directions. The final sample size available for TRTR analysis for this study was 86. The descriptive statistics are displayed in Table 1.

Third, the data were corrected for motion and eddy currents using FSL tools (Andersson and Sotiropoulos, 2016). The transform parameters were used to align the 4D diffusion volumes and the movement components were applied to the directional vectors. The b-matrix was rotated accordingly (Leemans and Jones, 2009). The resulting diffusion direction vectors were then examined to ensure that the remaining directions were evenly distributed (Roalf et al., 2016). Finally, the 4D diffusion dataset was processed with FSL's dtifit, using the brain mask to limit the modelling to brain tissue only.

Table 1
Descriptive statistics of the neonate datasets.

Description	Number of cases
Infant participants in the MRI	180
Datasets having more than 60 diffusion volumes after quality control	121
Datasets having more than 40 diffusion volumes after quality control	166
Datasets available for TRTR experiments having at least two parts of 96 the directional protocol, both having 25 directions after quality control	86

For further processing, we used the Tract-Based Spatial Statistics (TBSS) pipeline of FSL (Smith et al., 2007). It limits the analysis only to the skeletons of white matter tracts, estimated from individual images that are projected to a common skeleton space prior to statistical testing. Unlike conventional voxel-based analyses, TBSS does not require perfect brain-to-brain alignment or spatial smoothing and the estimation of tract center decreases partial volume effects.

We used the “tbss_2_reg -n” option that identifies the "most representative" image, which was then used as the target image for the linear and non-linear transformations for all the other data. The resulting study-specific infant template was reoriented to a straight position along the AC-PC line, co-registered to existing infant JHU template (Oishi et al., 2011) by using FSL (FLIRT, FNIRT) to facilitate later region of interest analysis, and upsampled to 1 mm^3 resolution.

We then ran a modified version of the “tbss_3_postreg —S” step to incorporate registrations to the study-specific infant template and upsample the data to 1 mm^3 resolution as per TBSS defaults. We used two FA threshold values of 0.15 and 0.10 in “tbss_4_prestats” module to create FA skeletons. Examples of axial slices of the TBSS template and skeleton are shown in Fig. 2. While various threshold values have been used with infants (some examples: 0.15 (Ball G, 2010), 0.18 (Ly, 2016), 0.2 (Lepomäki V, 2013)), we chose to use these values because smaller FA values than 0.1 were considered too noisy by visual inspection, and higher

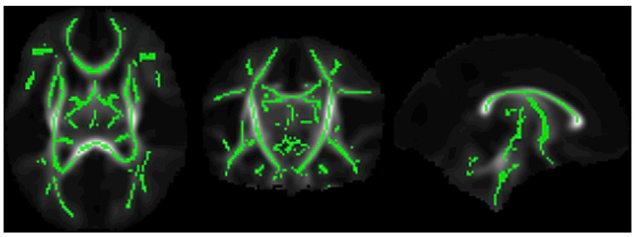

Fig. 2. The mean infant FA template (axial, coronal, and sagittal view) produced by the TBSS pipeline, overlayed with the skeleton generated using an FA threshold of 0.15. All the statistics are performed to the skeletonized data.

values than 0.15 were considered to produce too few remaining voxels for the remaining analyses. We then ran "tbss_non_FA" script with the MD, AD and RD maps (i.e. used the transformations estimated from FA maps to generate the coregistrations for the other tensor maps.

2.4. Data quality metrics to the DTI

Using spearman's rho test for non-linear associations, 32 measures that may potentially affect DTI reliability (Appendix B**)** were correlated with the DTI parameter mean values and test-retest variability (VAR), calculated as percentage of difference in proportion to average of test and retest acquisitions ($|scan_1\text{-}scan_2|/[[scan_1+scan_2]/2]$) of FA, MD, AD and RD parameter values in a voxelwise analysis. The data quality metrics contained estimates of residual motion from simultaneous eddy and motion correction step, angular resolution measures, and SNR for 32 ROI mean measurements.

Head motion is expected to increase RD and decrease AD, while MD as average diffusion is expected to be least affected (Yendiki et al., 2014). The amount of motion that occurred during acquisition may affect the diffusion measurements and was included in the models together with brain volume, which also may affect the diffusion measurements indirectly due to variations in relative resolution (individual voxels include less anatomical details in larger brains).

2.5. Statistical methods

We used intra class correlation coefficient for single measurement (ICC(3,1) (Shrout and Fleiss, 1979; Koo and Li, 2016)), and variability (VAR). For voxel-based statistical analysis, we calculated intra class correlation coefficients for single measurement (ICC(3,1)) and variability (VAR) at the voxel level within the TBSS skeleton with 0.10 and 0.15 thresholding, and for the 122 ROIs of JHU atlas. The overall ICC(3,1) and VAR measures over 86 subjects were compared between DTI parameter values for each analysis method, with 1-Way ANOVA test followed by a Tukey HSD test to see significant differences between reliability measures of ICC and VAR.

For region of interest analysis, we calculated ICC(3,1) and VAR for the mean values inside the 122 ROIs of the JHU atlas masked with the TBSS skeleton after thresholding at an FA of either 0.10 or 0.15. The JHU ROI's were thus calculated for each subject in skeleton space, which here corresponded to the skeletonized study-specific template space. The mean values of 122 ROIs of the infant JHU atlas within the TBSS skeletons were calculated. For the assessment of potential confounding factors we performed 32 comparisons per DTI scalar value (FA, MD, AD, RD), we applied Bonferroni corrected threshold per scalar (p-value threshold $0.001/32 = 3.125 \times 10^{-5}$). Here we report p-values that are below $p < 3.125 \times 10^{-5}$ preserve sensitivity, noting that these p-values are not corrected across the four scalars, 32 confounding factors, 122 ROIs, and the two statistics used. All statistical analyses were performed in Rstudio (v 1.1.456) (RStudio, 2016). The python and R programming codes that were used to analyze the data are available upon request from the corresponding author.

3. Results

3.1. Motion and data quality measurements

Residual (minor) motion in the data had negligible association to the diffusion metrics, as the image quality parameters significantly affected DTI parameters in only two of the 122 JHU ROIs. The mean absolute translation along the y axis was significantly associated with the mean AD in the left external capsule inside the skeleton region with both 0.15 and 0.10 threshold ($p = 1.51 \times 10^{-5}$ and $p = 9.36 \times 10^{-6}$). In addition, the absolute derivative of rotation around the y axis was significantly associated with the test-retest variance of FA in the left superior frontal gyrus ($p = 1.20 \times 10^{-5}$) with 0.15 threshold. In the other regions, no significant associations were found between the quality measures and either the DTI parameter values or test retest variance of the DTI parameter values. The mean (±SD) absolute motion estimated in eddy/motion correction step, in x, y, and z directions over 86 cases was: 1.229 ± 0.983 mm, 1.235 ± 0.978 mm and 1.224 ± 0.978 mm, and mean (±SD) rotations in degrees $0.70 \pm 0.67°$, $0.54 \pm 0.48°$ and $0.58 \pm 0.59°$.

3.2. Repeatability of voxelwise measurements

The ICC(3,1) values for voxelwise repeatability were generally high for all four DTI tensor measures (Figs. 3 and 4). For FA, using 0.15 thresholding was better than 0.10 in terms of VAR values ($p = 0.023$). FA had the higher ICC(3,1) and VAR in comparison to MD, AD and RD with 0.15 and 0.10 threshold values ($p < 3.125 \times 10^{-5}$). The distributions of mean ICC(3,1) and VAR values were generally symmetric apart from outliers towards low ICC(3,1) and higher VAR values. The ICC(3,1) performance across subjects was more squeezed in FA than other scalars, while in VAR, the FA was more spread than rest of the DTI parameters.

The distribution of ICC(3,1) and VAR across TBSS skeleton (threshold 0.15) is shown in Figs. 5 and 6. The ICC(3,1) varied across the brain, so that the peripheral regions and subcortical regions had lower repeatability in comparison to corpus callosum and main white matter pathways where the repeatability was high (see Fig. 5 and Supplemental Fig. S3). Test-retest variability (VAR) followed similar pattern, with overall variation between scans was generally higher in FA than the other three DTI scalars (see Fig. 6 and Supplemental Fig. S4).

3.3. Repeatability of ROI measurements

For 81 of the 122 ROI mean values of JHU atlas (with the 0.15 threshold, 4 of the ROIs did not have any voxels left for further analysis), ICC(3,1) was increased when using the TBSS skeleton with an FA

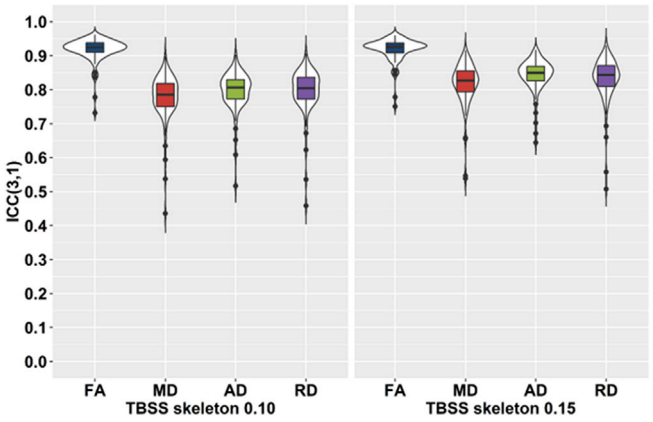

Fig. 3. Reliability of DTI parameters in 86 infant brains with violin plots of intra class correlation coefficient ICC(3,1) for voxels across the TBSS skeleton. The ICC(3,1) values are slightly lower for MD, AD, RD parameters and slightly higher for FA inside the skeleton mask.

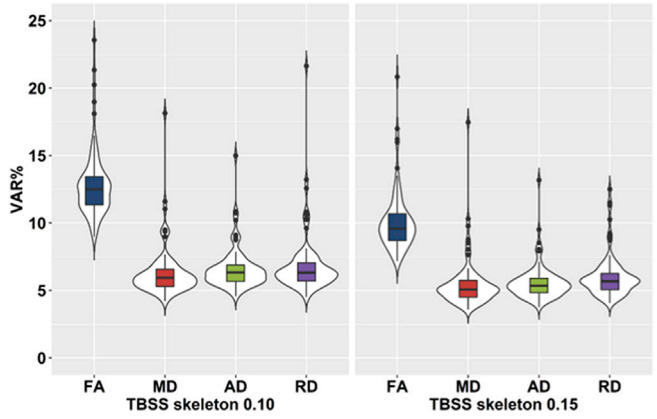

Fig. 4. Repeatability of DTI parameters in 86 infant brains with violin plots of variability (VAR) across voxels inside TBSS skeleton mask region. The VAR values are lower for MD, AD, RD parameters than for FA with both FA threshold values 0.10 and 0.15.

threshold of 0.15 instead of 0.10, and decreased in 37 ROIs. Test retest variance (VAR) was reduced when the 0.15 threshold was used for FA in 57 ROIs while in 61 ROIs, the 0.15 threshold affected increase in variance in comparison to a 0.10 threshold. The smaller ROIs are likely cause of increase in variance with 0.15. In conclusion, using 0.15 threshold resulted in better ICC(3,1) in majority of the ROIs. We selected 32 ROIs which overlap TBSS skeleton for further analysis (see supplemental Fig. S1 and Fig. S2). The SNR averaged over two acquisition segments was 3.6 ± 0.2 (see Appendix B for calculation). The biggest correlations over 86 subject between SNR and ROI mean value of FA was found in retrolenticular part of internal capsule ($p = 0.00021$ and $p = 0.00026$, for 0.10 and 0.15 threshold values), which were not considered significant ($p < 3.125 \times 10^{-5}$) after correction for multiple comparisons. In other DTI metrics, MD, AD and RD correlated with SNR in fornix, AD in corpus callosum, and cinculum cingular part ($p < 3.125 \times 10^{-5}$), while other correlations where not considered significant. Within the 32 ROIs and four DTI metrics, we found no major difference in ICC(3,1) between left and right hemisphere in any of the DTI metrics. Overall, 63 metrics had ICC(3,1) above 0.9. The highest ICC(3,1) values in all DTI metrics bilaterally were found in corona radiata (superior corona radiate FA ICC(3,1) = 0.913). Six out of total 128 values the ICC(3,1) was below 0.8. The lowest ICC(3,1) was found in internal capsule (retrolenticular part of internal capsule FA ICC(3,1) = 0.714). We also visually inspected overall intensity value patterns of 32 ROI mean values (see supplemental Fig. S5), where the individual cases were considered to be distinquishable.

ICC(3,1) values of the union of both sides of corpus callosum region (CC) are shown in Fig. 7. The DTI fitting models are expected to work well in the CC with the data due to predominantly single orientation of the fibers in left-to-right direction. In the corpus callosum, changing of threshold value between 0.10 and 0.15 did not result in improvement of reliability. To evaluate effect of individual voxels contribution to repeatability, we discarded 10% of voxels in corpus callosum that expressed highest variability. This resulted in increase of ICC(3,1) in FA. In corpus callosum, biggest difference between ROI level and voxel level analysis was in MD.

4. Discussion

We evaluated test retest reliability for a multi-part DTI acquisition scheme designed for use with neonates. The number of diffusion

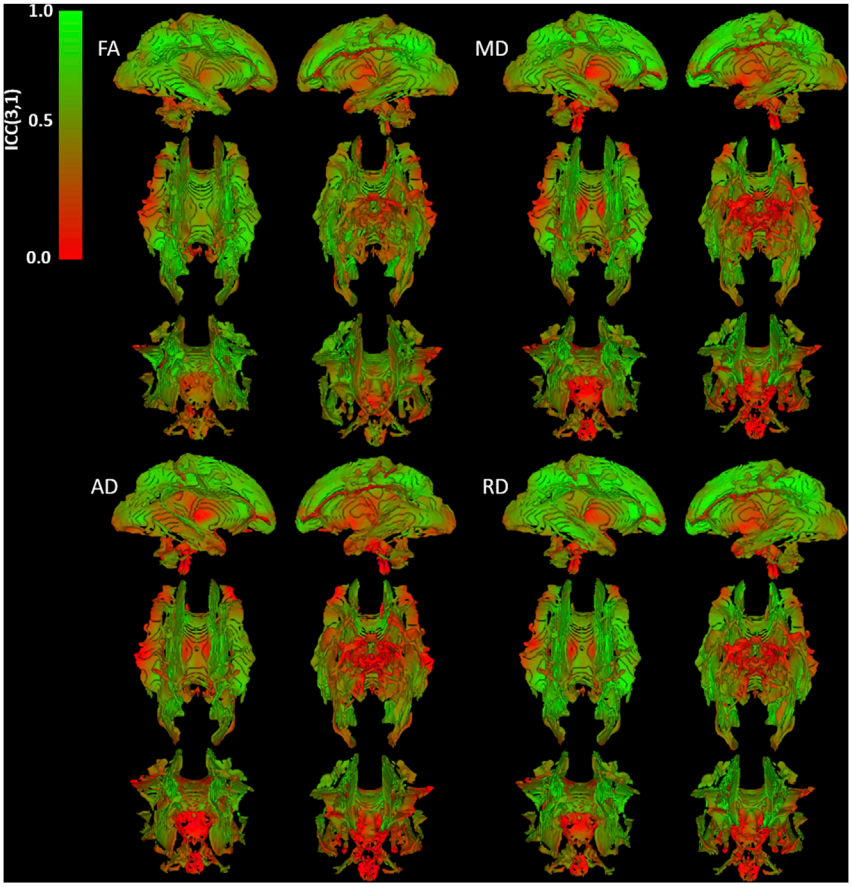

Fig. 5. Test-retest ICC(3,1) maps of TBSS skeleton calculated with 86 neonatal DTI images.

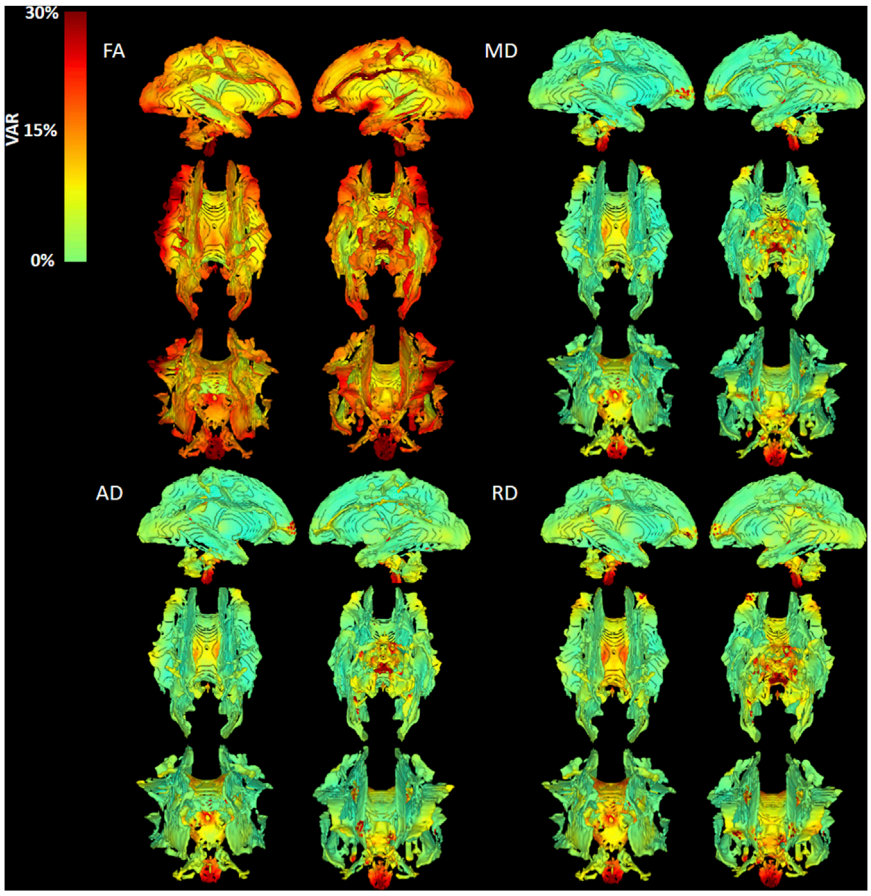

Fig. 6. Test-retest Variability maps of TBSS skeleton image calculated with 86 neonatal DTI images.

directions per part was standardized to 25 directions. The TBSS skeleton processing method was evaluated for reliability at the voxel level and the ROI level. The overall reliability was high (ICC(3,1) > 0.8). The repeatability of MD, AD, RD was robust to changing of TBSS threshold value, suggesting that the threshold parameter may be safely chosen between 0.10 and 0.15 with infant DTI data, and that 0.15 could be better for optimized processing for FA values. The performance of TBSS analysis in voxelwise analysis may be explained by smoothing which reduces the impact of Rician noise on the analysis, and results in increased effective ROI size (from one voxel to neighborhood), with the cost of losing spatial information (Jafari-Khouzani et al., 2019). We found that test-retest reliabilities were high in regions of interest, while voxel-level ICC values were spatially variable within the skeleton. Notably, most of the low voxelwise ICC values were located peripherally, which is expected. Additionally, lower ICC values were also found near the subcortical nuclei. This may be due to the narrowness of the white matter tracts within the infant brain that may cause minor registration errors. Admittedly, the DWI typically have lower SNR in deep brain regions but this effect was not pronounced with the used 12-channel head coil (as compared to using more channels). Similar plot for adults may thus have higher ICC.

We found only marginal associations with potential covariates including SNR, motion metrics and metrics related to distribution of diffusion directions, after rigorous quality control and standardization of number of directions to 25. Nevertheless, we recommend this to be tested for each study in question. Looking at the p-values not considered significant (data not shown), in alignment with (Ling et al., 2012), the most prominent association of data quality metric was found in translation along the y axis (anterior-posterior direction). We speculate that association of DTI to the effect of movement along y axis is to be caused by motion induced susceptibility artefacts in phase encoding direction that was used in the acquisition.

While the sample size was reasonable in the current study, the findings apply to infant scans only and it is left for future studies to address data quality metrics' effects on populations with older subjects. It is to be noted that only few data sets were perfect although the scans were deemed to be successful at the scanner (the scanner provided preview of the images with axial slices and an estimated FA map), which underlines the need for thorough quality control in infant DTI studies. The test retest scans were obtained here during the same session, and we acknowledge that it is hard to organize multiple scans for infants within narrow enough time margin where brain development would not hinder validity of test retest scans. Future studies may address reliability within a single day or paired days.

The number of directions was limited to only one setting with 25 directions to best reflect the test retest reliability between two scans/sequences. Importantly, the reliability metrics may be different with more or less available diffusion encoding directions (Landman et al., 2007; Leemans and Jones, 2009). However, for conciseness, we leave it for future to study the effects of a range of diffusion directions to the evaluated metrics. According to our experiments with various numbers of directions, 25 directions provided a good trade-off between number of cases left for analysis (as requirement for more directions forces exclusion of more cases), while having no association between data quality metrics and the mean DTI parameters. A simulation study (Jones, 2004) suggested the use of at least 30 directions for robust MD and tensor orientation estimation, and (Magnotta et al., 2012) found small within subject variability with 26 directions in an adult MR study. The fact that we used non-identical diffusion encoding directions may have result in decreased ICC values. However, even though the scanner directions would remain the same, head orientation/subject placement in the scanner makes every rescan to have a non-identical distribution of

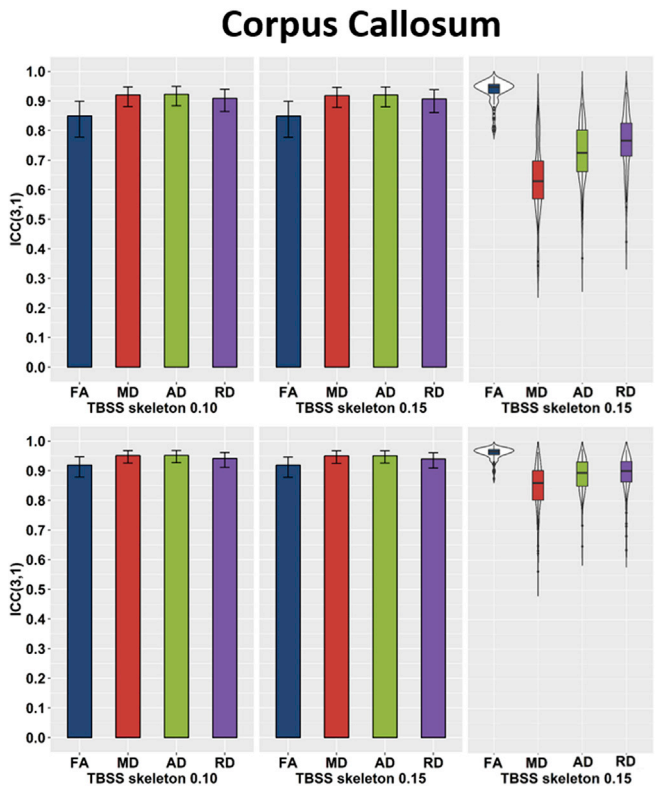

Fig. 7. Reliability of the DTI parameters in 86 infant brains assessed with intraclass correlation coefficient ICC(3,1) values in the corpus callosum of the JHU atlas masked with the TBSS skeleton thresholded at either 0.10 or 0.15. On the left and in the middle: ICC(3,1) of ROI mean values with two FA threshold values. On the right: Distributions of voxelwise ICC(3,1) values for the same region with 0.15 threshold. Top row: All voxels included in analysis. Bottom row: ICC(3,1) values when 10% of voxels with highest VAR were excluded.

diffusion encoding directions, which supports that similar, and high, test retest reliability likely apply in other studies as well.

The evaluated diffusion metrics may be sensitive to other analyses of the data, such as fiber tracking algorithms (Baum et al., 2018) and tract-specific analysis (TSA). As there are various fiber tracking and tract segmenting methods available, which may have different reliability also depending on parameters used in their execution, it is left for future studies to address data quality metrics' effect on the reliability of the resolved fibers and in TSA, together with test-retest repeatability of the principal tensor orientation that was not analysed here. However, we speculate that fiber tracking may be sensitive to inhomogeneities in the distribution of diffusion directions, since lower angular resolution results in higher angular error (Prckovska et al., 2013; Landman et al., 2007). We did not assess the effects of including minor artefact corrupted data, and how much that may affect the reliability. However, some novel approaches (Bastiani M, 2018) may help in partial data recovery for corrupted scans and such approaches may increase the amount of useable scans (Andersson et al., 2016). In addition, we speculate that other acquisition schemes such as multi-shell DTI procedures may benefit from using the presented data quality metrics (Appendix B) for controlling of potential limitations to accuracy and precision that they may pose to the results.

5. Conclusion

We have presented an assessment of the reliability of neonate DTI measurements using a test-retest analysis of pieces of a 96 direction acquisition divided into three segments. The results showed that the DTI measurements within the WM skeleton were reliable in voxelwise analyses and in most of the ROIs of the JHU atlas. We did not find any major dependency between data quality metrics, including residual motion, and final DTI parameters at the voxel or the ROI level. The results indicate that DTI metrics from neonate data can be reliable, even at relatively low angular resolution, and thus lend confidence to the use of neonate DTI data in cross sectional and longitudinal analyses.

Declarations of interest

None.

Acknowledgements

We want to thank all the participating families, the staff of the Medical Imaging Centre of Turku University Hospital as well as the FinnBrain Birth Cohort Study research personnel. This research has received funding from the Jane and Aatos Erkko Foundation (HK), Academy of Finland, the Sigrid Jusélius Foundation (HM), Hospital District of Southwest Finland State Research Grants (HM, JJT, NMS, HK), Alfred Kordelin Foundation, Turku University Foundation (JJT), Emil Aaltonen Foundation (JJT), The Gyllenberg Foundation (NMS) and Cultural Foundation of Finland (HM).

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.04.067.

Appendix A. Table of 96 diffusion encoding directions

Table of diffusion encoding vectors used in DTI data acquisition. Acquisition was divided in three parts with 31, 32 and 33 diffusion encoding directions. Each contained three volumes without diffusion encoding (b0 images, in the beginning, in the middle, and in the end), otherwise b-value of 1000 s/mm^2 was used.

DTI part 1 31 directions 3 b0 images		DTI part 2 32 directions 3 b0 images		DTI part 3 33 directions 3 b0 images	
[0]	(0.000, 0.000, 0.000)	[0]	(0.000, 0.000, 0.000)	[0]	(0.000, 0.000, 0.000)
[1]	(0.049, −0.919, −0.391)	[1]	(0.212, −0.754, −0.622)	[1]	(-0.809, 0.324, −0.491)
[2]	(0.726, 0.301, −0.618)	[2]	(0.912, −0.104, 0.398)	[2]	(-0.281, −0.955, 0.096)

(*continued on next page*)

(*continued*)

[3]	(-0.683, 0.255, −0.684)	[3]	(-0.311, 0.947, −0.077)	[3]	(0.592, 0.483, −0.645)
[4]	(0.845, −0.502, −0.186)	[4]	(0.679, 0.632, −0.374)	[4]	(0.382, 0.593, 0.709)
[5]	(-0.730, −0.619, −0.288)	[5]	(0.135, −0.286, 0.949)	[5]	(0.237, −0.506, −0.829)
[6]	(-0.051, 0.039, 0.998)	[6]	(-0.647, 0.230, 0.727)	[6]	(-0.711, 0.625, −0.323)
[7]	(-0.018, 0.871, −0.491)	[7]	(0.904, 0.397, 0.158)	[7]	(-0.146, 0.519, −0.842)
[8]	(-0.444, 0.494, 0.747)	[8]	(-0.757, 0.647, −0.087)	[8]	(0.927, −0.243, −0.286)
[9]	(-0.989, −0.086, −0.116)	[9]	(0.143, 0.284, 0.948)	[9]	(-0.809, −0.587, −0.004)
[10]	(-0.470, −0.855, 0.221)	[10]	(0.233, 0.894, −0.382)	[10]	(-0.976, 0.142, −0.165)
[11]	(0.412, 0.400, 0.819)	[11]	(0.664, −0.531, 0.527)	[11]	(-0.135, −0.870, −0.474)
[12]	(-0.552, 0.790, −0.267)	[12]	(0.157, 0.710, 0.686)	[12]	(0.274, −0.048, −0.960)
[13]	(-0.123, −0.477, 0.871)	[13]	(-0.895, −0.214, 0.392)	[13]	(0.518, −0.855, 0.037)
[14]	(-0.848, 0.141, 0.510)	[14]	(0.594, 0.080, 0.801)	[14]	(-0.482, −0.767, 0.423)
[15]	(0.000, 0.000, 0.000)	[15]	(-0.006, 0.688, −0.726)	[15]	(0.750, −0.039, 0.660)
[16]	(-0.341, −0.788, −0.512)	[16]	(0.000, 0.000, 0.000)	[16]	(0.000, 0.000, 0.000)
[17]	(0.361, −0.529, 0.768)	[17]	(0.665, 0.746, 0.024)	[17]	(0.143, −0.048, 0.989)
[18]	(-0.472, 0.850, 0.234)	[18]	(-0.277, 0.276, 0.920)	[18]	(0.632, −0.434, −0.642)
[19]	(-0.856, −0.481, 0.189)	[19]	(-0.962, 0.268, 0.046)	[19]	(0.799, 0.429, −0.422)
[20]	(0.797, 0.162, 0.582)	[20]	(-0.133, −0.970, −0.202)	[20]	(0.480, −0.632, 0.608)
[21]	(0.467, −0.009, −0.884)	[21]	(0.790, −0.405, −0.461)	[21]	(-0.547, −0.799, −0.250)
[22]	(0.013, 0.998, −0.056)	[22]	(-0.194, −0.193, 0.962)	[22]	(-0.616, −0.310, −0.724)
[23]	(0.882, −0.387, 0.267)	[23]	(-0.236, 0.952, 0.195)	[23]	(-0.028, −0.962, 0.270)
[24]	(0.017, −0.536, −0.844)	[24]	(-0.884, −0.272, −0.379)	[24]	(-0.324, −0.506, 0.799)
[25]	(-0.442, −0.651, 0.617)	[25]	(0.463, −0.307, 0.831)	[25]	(0.268, −0.875, −0.403)
[26]	(0.365, −0.058, 0.929)	[26]	(0.700, 0.066, −0.711)	[26]	(0.843, 0.059, −0.534)
[27]	(0.977, −0.004, −0.213)	[27]	(-0.200, 0.928, −0.314)	[27]	(-0.886, 0.462, −0.044)
[28]	(-0.406, −0.902, −0.145)	[28]	(0.550, 0.705, 0.449)	[28]	(0.187, 0.489, 0.852)
[29]	(-0.627, 0.614, 0.479)	[29]	(-0.670, 0.727, 0.153)	[29]	(0.184, −0.746, 0.640)
[30]	(-0.354, 0.772, −0.528)	[30]	(0.237, 0.722, −0.650)	[30]	(0.793, 0.450, 0.412)
[31]	(-0.658, −0.472, −0.586)	[31]	(0.960, 0.260, −0.100)	[31]	(0.475, −0.248, −0.844)
[32]	(0.423, 0.322, −0.847)	[32]	(0.407, −0.756, −0.512)	[32]	(0.662, 0.728, −0.181)
[33]	(0.000, 0.000, 0.000)	[33]	(-0.355, −0.180, −0.917)	[33]	(0.943, 0.083, 0.323)
		[34]	(0.000, 0.000, 0.000)	[34]	(-0.045, 0.720, 0.692)
				[35]	(0.000, 0.000, 0.000)

Appendix B. Data quality metrics correlated with DTI parameter estimates

Metrics 1.-24. relate to motion and eddy current effects measured from the data during correction process, and are calculated using 4×4 transformation matrix obtained in coregistration of volumes for eddy and motion correction.

1. Absolute derivative of rotation around x axis
2. Absolute derivative of rotation around y axis
3. Absolute derivative of rotation around z axis
4. Mean of absolute rotation around x axis
5. Mean of absolute rotation around y axis
6. Mean of absolute rotation around z axis
7. Absolute derivative of scaling along x axis
8. Absolute derivative of scaling along y axis
9. Absolute derivative of scaling along z axis
10. Mean of absolute scaling align x axis
11. Mean of absolute scaling align y axis
12. Mean of absolute scaling align z axis
13. Absolute derivative of skew along x axis
14. Absolute derivative of skew along y axis
15. Absolute derivative of skew along z axis
16. Mean absolute skew along x axis
17. Mean absolute skew along y axis
18. Mean absolute skew along z axis
19. Absolute derivative of translation along x axis
20. Absolute derivative of translation along y axis
21. Absolute derivative of translation along z axis
22. Mean absolute translation along x axis
23. Mean absolute translation along y axis
24. Mean absolute translation along z axis

All absolute derivate metrics are calculated as absolute difference divided by time between acquisitions which was considered the same for all gradient directions:

$$Metric_{Derivative} = \sum_{Gradient\ Directions} \frac{\Delta metric}{\Delta acquisition\ time}$$

All mean absolute metrics were calculated as average of absolute metric values:

$$Metric_{Mean} = \frac{\sum_{Gradient\ Directions} \Delta metric}{Number\ of\ Gradient\ Directions}$$

Metric 25. (brain volume) estimates voxel size and scanner FOV relative to the size of the brain, and is calculated from BET brain extraction tool mask.

25. Brain volume in mL

Metrics 26.-32. address distribution of the measurement directions after preprocessing. The 27.-28. are calculated by applying spherical voronoi tesselation algorithm (Caroli et al., 2010) to the diffusion direction vectors to find largest non-covered region in the sphere together with pairwise comparisons between diffusion direction vectors.

26. Maximum of neighboring tensor directions after preprocessing

$$AngularGap_{Max\ pairs} = \max_{Gradients\ i} \left\{ \min_{Neighbors\ j\ of\ Gradient\ i} \left| angle\left(V_i - V_j\right) \right| \right\}$$

27. Theta angle of orientation in center direction of biggest gap in the coverage of diffusion directions.
28. Phi angle of orientation in center direction of biggest gap in the coverage of diffusion directions.

The biggest gap is the center orientation of either two neighboring gradients *i* and *j*, or center of three gradient directions if the cap angle is greater than angle between neighboring gradients. The cap angle is calculated with center orientation vectors from spherical voronoi tesselation as:

$$AngularGap_{Max\ triplets} = \max_{Voronoi\ Centers\ K_i} \left\{ \min_{Neighbors\ j\ of\ K_i} \left| angle\left(K_i - V_j\right) \right| \right\}$$

The metrics 29.-32. Are estimated from distribution of angles, where one angle is resolved for each diffusion vector by finding largest non-covered region from $AngularGap_{Max\ triplets}$ values, which is bordered by evaluated diffusion vector.

29. Average angle between neighboring diffusion directions after preprocessing
30. Median angle between neighboring diffusion directions after preprocessing
31. Standard deviation of angles between neighboring diffusion directions after preprocessing
32. Range of angles between neighboring diffusion directions after preprocessing
33. Signal-to-noise ratio of DWI imaging data

The estimated signal-to-noise ratio of the DTI acquisition before processing was calculated for ROI from two B_0 images inside each of two used acquisition segments, as in [1]:

$$B_{Avg} = \left(B_{0,1st} + B_{0,2nd}\right)$$

$$B_{Diff} = B_{0,1st} - B_{0,2nd}$$

$$SNR_{ROI} = 1 \Big/ \sqrt{2} \times Mean\left(ROI\left(B_{Avg}\right)\right) \Big/ SD\left(ROI\left(B_{Diff}\right)\right)$$

where $B_{0,1st}$ and $B_{0,2nd}$ are first and second accepted B_0 in their order of occurrence in the acquisition segment, and *ROI* is the brain mask with three voxels erosion. *Mean* and *SD* refer to average and standard deviations of all voxels inside *ROI*.

[1] National Electrical Manufacturers Association (NEMA). 2001.'Determination of signal-to-noise ratio (SNR) in diagnostic magnetic resonance imaging. NEMA Standards Publication MS 1–2001'. Rosslyn: National Electrical Manufacturers Association.

References

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.

Ball G, C.S., 2010. An optimised tract-based spatial statistics protocol for neonates: applications to prematurity and chronic lung disease. Neuroimage 94–102.

Bastiani, M.,C.M.-A., 2018. 'Automated quality control for within and between diffusion MRI studies using a non-parametric framework for movement and distortion correction. Neuroimage. https://doi.org/10.1016/j.neuroimage.2018.09.073. Epub 2018 Sep. 26.

Baum, G.L., Roalf, D.R., Cook, P.A., Ciric, R., Rosen, A.F.G., Xia, C., Elliott, M.A., Ruparel, K., Verma, R., Tunc, B., Gur, R.C., Gur, R.E., Bassett, D.S., Satterthwaite, T.D., 2018. The impact of in-scanner head motion on structural connectivity derived from diffusion MRI. Neuroimage 173, 275–286.

Bonekamp, D.N., 2007. Diffusion Tensor Imaging in Children and Adolescents: Reproducibility, Hemispheric, and Age-Related Differences. NeuroImage, pp. 733–742.

Cabeen, R.P., Bastin, M.E., Laidlaw, D.H., 2017. A Comparative evaluation of voxel-based spatial mapping in diffusion tensor imaging. Neuroimage 146, 100–112.

Caroli, Manuel, Pedro, M., de Castro, M., Loriot, Sébastien, Rouiller, Olivier, Teillaud, Monique, Wormser, Camille, 2010. Robust and efficient delaunay triangulations of points on or close to a sphere. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 462–473.

Duan, F., Zhao, T.D., He, Y., Shu, N., 2015. 'Test-Retest reliability of diffusion measures in cerebral white matter: a multiband diffusion MRI study. J. Magn. Reson. Imaging 42, 1106–1116.

Farrell, J.A., Landman, B.A., Jones, C.K., Smith, S.A., Prince, J.L., van Zijl, P.C., Mori, S., 2007. 'Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. J. Magn. Reson. Imaging 26, 756–767.

Graham, M.S., Drobnjak, I., Jenkinson, M., Zhang, H., 2017. 'Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI. PLoS One 12, e0185647.

Gumus, K., Keating, B., Poser, B.A., Armstrong, B., Chang, L., Maclaren, J., Prieto, T., Speck, O., Zaitsev, M., Ernst, T., 2014. Prevention of motion-induced signal loss in

diffusion-weighted echo-planar imaging by dynamic restoration of gradient moments. Magn. Reson. Med. 71, 2006–2013.
Jafari-Khouzani, K., Paynabar, K., Hajighasemi, F., Rosen, B., 2019. The effect of region of interest size on the repeatability of quantitative brain imaging biomarkers. IEEE Trans. Biomed. Eng. 66 (3), 864–872.
Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782–790.
Jones, D.K., 2004. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. Magn. Reson. Med. 51, 807–815.
Jovicich, J., Marizzoni, M., Bosch, B., Bartres-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Picco, A., Nobili, F., Blini, O., Bombois, S., Lopes, R., Bordet, R., Chanoine, V., Ranjeva, J.P., Didic, M., Gros-Dagnac, H., Payoux, P., Zoccatelli, G., Alessandrini, F., Beltramello, A., Bargallo, N., Ferretti, A., Caulo, M., Aiellou, M., Ragucci, M., Soricelli, A., Salvadori, N., Tarducci, R., Floridi, P., Tsolaki, M., Constantinidis, M., Drevelegas, A., Rossini, P.M., Marra, C., Otto, J., Reiss-Zimmermann, M., Hoffmann, K.T., Galluzzi, S., Frisoni, G.B., Consortium, PharmaCog, 2014. Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. Neuroimage 101, 390–403.
Karlsson, L., Tolvanen, M., Scheinin, N.M., Uusitupa, H.M., Korja, R., Ekholm, E., Tuulari, J.J., Pajulo, M., Huotilainen, M., Paunio, T., Karlsson, H., Group FinnBrain Birth Cohort Study, 2018. 'Cohort profile: the FinnBrain birth cohort study (FinnBrain)'. Int. J. Epidemiol. 47, 15–16j.
Kelly, C.J., Hughes, E.J., Rutherford, M.A., Counsell, S.J., 2018. Advances in Neonatal MRI of the Brain: from Research to Practice. Arch Dis Child Educ Pract.
Koo, T.K., Li, M.Y., 2016. 'A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine 15, 155–163.
Landman, B.A., Farrell, J.A., Jones, C.K., Smith, S.A., Prince, J.L., Mori, S., 2007. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. Neuroimage 36, 1123-38.
Leemans, A., Jones, D.K., 2009. 'The B-matrix must be rotated when correcting for subject motion in DTI data. Magn. Reson. Med. 61, 1336–1349.
Lepomäki, V.,L.M., 2013. Preterm infants' early growth and brain white matter maturation at term age. Pediatr. Radiol. 1357–1364.
Li, X., Gao, J., Wang, M., Wan, M., Yang, J., 2016. Rapid and reliable tract-based spatial statistics pipeline for diffusion tensor imaging in the neonatal brain: applications to the white matter development and lesions. Magn. Reson. Imaging 34, 1314–1321.
Ling, J., Merideth, F., Caprihan, A., Pena, A., Teshiba, T., Mayer, A.R., 2012. 'Head injury or head motion? Assessment and quantification of motion artifacts in diffusion tensor imaging studies. Hum. Brain Mapp. 33, 50–62.
Ly, M.T., 2016. Comparing tract-based spatial statistics and manual region-of-interest labeling as diffusion analysis methods to detect white matter abnormalities in infants with hypoxic-ischemic encephalopathy. J. Magn. Reson. Imaging 42, 1689–1697.
Madhyastha, T., Merillat, S., Hirsiger, S., Bezzola, L., Liem, F., Grabowski, T., Jancke, L., 2014. Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. Hum. Brain Mapp. 35, 4544–4555.
Magnotta, V.A., Matsui, J.T., Liu, D., Johnson, H.J., Long, J.D., Bolster Jr., B.D., Mueller, B.A., Lim, K., Mori, S., Helmer, K.G., Turner, J.A., Reading, S., Lowe, M.J., Aylward, E., Flashman, L.A., Bonett, G., Paulsen, J.S., 2012. Multicenter reliability of diffusion tensor imaging. Brain Connect. 2, 345–355.
Millman, K Jarrod, Aivazis, Michael, 2011. Python for scientists and engineers. Comput. Sci. Eng. 13, 9–12.
Oguz, I., Farzinfar, M., Matsui, J., Budin, F., Liu, Z., Gerig, G., Johnson, H.J., Styner, M., 2014. DTIPrep: quality control of diffusion-weighted images. Front. Neuroinf. 8, 4.
Oishi, K., Mori, S., Donohue, P.K., Ernst, T., Anderson, L., Buchthal, S., Faria, A., Jiang, H., Li, X., Miller, M.I., van Zijl, P.C., Chang, L., 2011. Multi-contrast human neonatal brain atlas: application to normal neonate development analysis. Neuroimage 56, 8–20.
Prckovska, V., Achterberg, H.C., Bastiani, M., Pullens, P., Balmashnova, E., Ter Haar Romeny, B.M., Vilanova, A., Roebroeck, A., 2013. Optimal short-time acquisition schemes in high angular resolution diffusion-weighted imaging. Int. J. Biomed. Imaging 2013, 658583.
Roalf, D.R., Quarmley, M., Elliott, M.A., Satterthwaite, T.D., Vandekar, S.N., Ruparel, K., Gennatas, E.D., Calkins, M.E., Moore, T.M., Hopson, R., Prabhakaran, K., Jackson, C.T., Verma, R., Hakonarson, H., Gur, R.C., Gur, R.E., 2016. The impact of quality assurance assessment on diffusion tensor imaging outcomes in a large-scale population-based cohort. Neuroimage 125, 903–919.
Rooks, V.J., Eaton, J.P., Ruess, L., Petermann, G.W., Keck-Wherley, J., Pedersen, R.C., 2008. 'Prevalence and evolution of intracranial hemorrhage in asymptomatic term infants'. AJNR Am J Neuroradiol 29, 1082–1089.
RStudio, 2016. RStudio: Integrated Development Environment for R. RStudio, Inc, Boston, MA.
Shrout, P.E., Fleiss, J.L., 1979. 'Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, 208–219.
Smith, S.M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T.E., Miller, K.L., Robson, M.D., Jones, D.K., Klein, J.C., Bartsch, A.J., Behrens, T.E., 2007. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. Nat. Protoc. 2, 499–503.
Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., Fischl, B., 2014. 'Spurious group differences due to head motion in a diffusion MRI study. Neuroimage 88, 79–90.