MASiVar: Multisite, Multiscanner, and Multisubject Acquisitions for Studying Variability in Diffusion Weighted Magnetic Resonance Imaging

Leon Y. Cai[1], Qi Yang[2], Praitayini Kanakaraj[2], Vishwesh Nath[2], Allen T. Newton[3,4], Heidi A. Edmonson[5], Jeffrey Luci[6,7], Benjamin N. Conrad[8,9], Gavin R. Price[9], Colin B. Hansen[2], Cailey I. Kerley[2], Karthik Ramadass[2], Fang-Cheng Yeh[10], Hakmook Kang[11], Eleftherios Garyfallidis[12], Maxime Descoteaux[13], Francois Rheault[2,13], Kurt G. Schilling[3,4], and Bennett A. Landman[1,2,3,4]

[1] Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA
[2] Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA
[3] Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
[4] Vanderbilt University Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA
[5] Department of Radiology, Mayo Clinic, Rochester, MN, USA
[6] Department of Biomedical Engineering, The University of Texas at Austin, TX, USA
[7] Department of Psychiatry, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, New Jersey, USA
[8] Neuroscience Graduate Program, Vanderbilt Brain Institute, Vanderbilt University Medical Center, Nashville, TN, USA
[9] Department of Psychology and Human Development, Peabody College, Vanderbilt University, Nashville, TN, USA
[10] Department of Neurological Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[11] Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA
[12] Department of Intelligent Systems Engineering, Indiana University, Bloomington, IN, USA
[13] Department of Computer Science, Université de Sherbrooke, Sherbrooke, Québec, Canada

Abstract

Diffusion weighted imaging (DWI) allows investigators to identify microstructural differences between subjects, but variability due to session and scanner biases is still a challenge. To investigate DWI variability, we present MASiVar, a multisite dataset consisting of 319 diffusion scans acquired at 3T from b = 1000 to 3000 s/mm^2 across 97 different healthy subjects and four different scanners as a publicly available, preprocessed, and de-identified dataset. With these data we characterize variability on the intrasession intrascanner (N = 158), intersession intrascanner (N = 328), intersession interscanner (N = 53), and intersubject intrascanner (N = 80) levels. Our baseline analysis focuses on four common DWI processing approaches: (1) a tensor signal representation, (2) a multi-compartment neurite orientation dispersion and density model, (3) white matter bundle segmentation, and (4) structural connectomics. Respectively, we evaluate region-wise fractional anisotropy (FA), mean diffusivity, and principal eigenvector; region-wise cerebral spinal fluid volume fraction, intracellular volume fraction, and orientation dispersion index; bundle-wise shape, volume, length and FA; and connectome correlation and maximized modularity, global efficiency, and characteristic path length. We plot the scan/re-scan discrepancies in these measures at each level and find that variability generally increases with intrasession to intersession to interscanner to intersubject effects and that sometimes interscanner variability can approach intersubject variability. This baseline study suggests harmonization between scanners for multisite analyses is critical prior to inference of group differences on subjects and demonstrates the potential of MASiVar to investigate DWI variability across multiple levels and processing approaches simultaneously.

Keywords: Diffusion MRI, variability, reproducibility, tensor, microstructure, DTI, NODDI, bundle segmentation, connectome

Introduction

Diffusion weighted MRI imaging (DWI) is a noninvasive way of elucidating the brain's microstructural makeup (O'Donnell and Westin, 2011). Common modes of DWI analysis include representing the diffusion signal with tensors (Assaf and Pasternak, 2008), representing biological tissues with multi-compartment models (Jelescu et al., 2015; Zhang et al., 2012), identifying white matter bundles (Schilling et al., 2020b), and investigating the human structural connectome (Sotiropoulos and Zalesky, 2019). These approaches form the basis for many studies including those investigating a wide range of neurological disorders including autism (Di Martino et al., 2017; Travers et al., 2012), diabetes (Kodl et al., 2008; Repple et al., 2019), multiple sclerosis (De Santis et al., 2019), and schizophrenia (Cetin-Karayumak et al., 2019) as well as differences due to aging (Westlye et al., 2010) and sex (Menzler et al., 2011). These types of studies, however, rely on the identification of group differences with respect to an independent variable. Often this variable is whether the scanned subject has a particular disease, or the age or sex of the subject. Robust study design can control for additional subject-level confounders through age- and sex-matching and related approaches. However, one level of potential confounding in DWI studies that has not been thoroughly characterized is the variability of calculations due to differences within and between imaging sessions and scanners.

One particular reason for this is the difficulty in acquiring data configured to perform such a characterization. For instance, to quantify within session variation, imaging sessions with multiple scans are needed. To quantify between session and between scanner variation, multiple imaging sessions on at least one scanner and at least one imaging session on multiple scanners are required, respectively. Last, to assess the session and scanner effects relative to the subject effect size, multiple scanned subjects are needed as well.

Another reason for this is the low number of properly configured publicly available datasets. Some of the few that exist that allow for investigations of DWI variability are the MASSIVE dataset (Froeling et al., 2017), the Human Connectome Project (HCP) 3T dataset (Van Essen et al., 2013), the MICRA dataset (Koller et al., 2020), the SIMON dataset (Duchesne et al., 2019), and the multisite dataset published by Tong et al. (Tong et al., 2020). MASSIVE consists of one subject scanned repeatedly on one scanner (Froeling et al., 2017); HCP consists of multiple subjects with multiple acquisitions per session all on one scanner (Van Essen et al., 2013); MICRA consists of multiple subjects scanned repeatedly on one scanner (Koller et al., 2020); SIMON consists of one subject scanned at over 70 sites (Duchesne et al., 2019), and the Tong dataset consists of multiple subjects each scanned on multiple scanners (Tong et al., 2020).

Thus, the release of additional publicly available datasets specifically configured to study DWI variability on the session and scanner level in addition to the subject level would be an asset to the field. A baseline characterization of this variability is also needed to inform investigators about the degree to which session and scanner effects may be impacting their analyses. To fill the first need, we propose MASiVar, a multisite, multiscanner, and multisubject dataset designed for characterization of DWI variability due to session, scanner, and subject effects. To fill the latter, we present a baseline quantification of these effects on four different common diffusion approaches using MASiVar and hypothesize that intrasession variability will be less than intersession variability which will be less than interscanner and finally intersubject variability.

Methods

Data acquisition

We acquired data for MASiVar from 2016 to 2020 in three cohorts, designated I, II, and III (Figure 1).

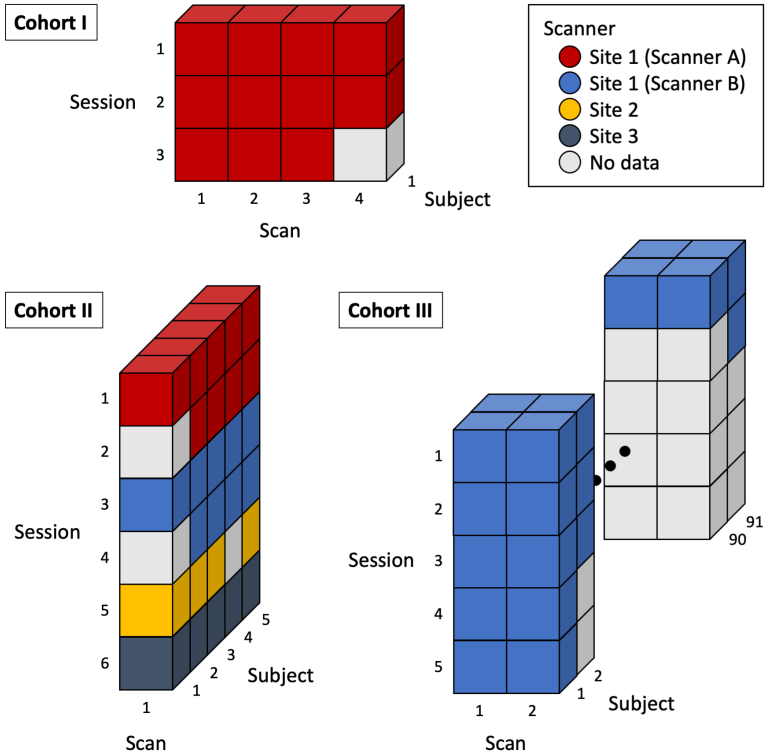

Figure 1. Overview of the MASiVar dataset. This dataset consists of three cohorts. Cohort I consists of one subject scanned repeatedly on one scanner. This subject underwent three separate imaging sessions and acquired 3-4 scans per session. Cohort II consists of 5 subjects each scanned on 3-4 different scanners across 3 institutions. Each subject underwent 1-2 sessions on each scanner and had one scan acquired per session. Cohort III consists of 91 subjects all scanned on one scanner. Each subject underwent 1-5 sessions on the scanner and had two scans acquired per session.

Cohort I consists of one healthy adult subject (male, age 25 years) with multiple imaging sessions on a 3T Philips Achieva scanner at site 1 (scanner A). This subject underwent three imaging sessions, one each consecutive day, and received two to three scans during each session. Each scan consisted of 96-direction acquisitions at b = 1000, 1500, 2000, 2500, and 3000 s/mm^2 (Table 1). These scans were acquired at 2.5mm isotropic resolution.

Cohort II consists of five healthy adult subjects (3 male, 2 female, age 27 to 47 years) scanned for one to two sessions on each of three to four different scanners. Each subject underwent all sessions within one year. The scanners included scanner A, another 3T Philips Achieva scanner at site 1 (scanner B), a 3T General Electric Discovery MR750 scanner at site 2, and a 3T Siemens Skyra scanner at site 3 (Figure 1). For each imaging session, each subject received one scan, consisting of 96-direction acquisitions at b = 1000, 1500, 2000, 2500 (or 2465 at site 3 due to hardware limitations) s/mm^2 and a 30- to 33-direction acquisition at b = 1000 s/mm^2 (Table 1). The scans acquired on scanner B, at site 2, and at site 3, and all the 30- to 33-direction scans were acquired at 2.5mm isotropic resolution. On scanner A, one subject's 96-direction acquisitions were also acquired at 2.5mm isotropic resolution while the remainder were acquired at 1.9mm by 1.9mm by 2.2mm (sagittal, coronal, and axial) resolution. For consistency, all sessions acquired on scanner A that contained scans of varying resolution were resampled to match the resolution of the 96-direction acquisitions prior to analysis.

Cohort III consists of 8 healthy adult subjects (4 male, 4 female, ages 21 to 31 years) scanned for one to six sessions on scanner B, and 83 healthy child subjects (48 male, 35 female, ages 5 to 8 years) scanned for one to two sessions on scanner B (Figure 1). Each adult subject underwent all sessions within one year, and for the 35 child subjects scanned twice, their sessions were spaced approximately one year apart. Each subject received one to two scans during each session, with each scan consisting of a 40-direction b = 1000 s/mm^2 and a 56-direction b = 2000 s/mm^2 acquisition (Table 1). These scans were acquired at 2.1mm by 2.1mm by 2.2mm (sagittal, coronal, and axial) resolution.

Table 1. Acquisitions acquired in each scan for the different MASiVar cohorts.

Acquisitions Per Scan		
Cohort	*Shell (b-value)*	*Number of Directions*
I	1000	96
	1500	96
	2000	96
	2500	96
	3000	96
II	1000	30 to 33
	1000	96
	1500	96
	2000	96
	2465 or 2500	96
III	1000	40
	2000	56

All images were phase encoded in the posterior to anterior direction (APP) and were acquired with b = 0 s/mm^2 volumes. Reverse phase encoded (APA) b = 0 s/mm^2 volumes were also acquired for all scans in all cohorts except for those from one subject in cohort II at site 3. Most sessions also included a T1-weighted image. All images were deidentified and all scans were acquired only after informed consent under supervision of the project Institutional Review Board.

Data preprocessing

Prior to analysis, all scans in MASiVar were preprocessed and quality checked with the PreQual pipeline (Cai et al., 2020). In brief, all acquisitions per scan were denoised with the Marchenko-Pastur technique (Cordero-Grande et al., 2019; Veraart et al., 2016b, 2016a), intensity normalized such that the average b = 0 s/mm^2 intensity distributions within the brain maximally intersected, and distortion corrected. Distortion correction included susceptibility-induced distortion correction (Andersson et al., 2003) using APA b = 0 s/mm^2 volumes when available and the Synb0-DisCo deep learning framework (Schilling et al., 2020a) when not, eddy current-induced distortion correction, intervolume motion correction, and slice-wise signal drop out imputation (Andersson et al., 2016; Andersson and Sotiropoulos, 2016).

Data processing

We sought to investigate the variability of four types of common diffusion processing approaches: the diffusion tensor imaging (DTI) signal representation, the multi-compartment neurite orientation dispersion and density imaging (NODDI) model (Zhang et al., 2012), the RecoBundles white matter bundle segmentation technique (Garyfallidis et al., 2018), and a connectomics representation with graph-based measures (Rubinov and Sporns, 2010) (Figure 2).

For the DTI approach, we extract the b = 1000 s/mm^2 acquisition from each scan with the largest number of directions. We then calculate the diffusion tensor for each scan using an iteratively reweighted least squares approach implemented in MRtrix3 (Veraart et al., 2013). The tensors are subsequently converted to fractional anisotropy (FA), mean diffusivity (MD), and principal eigenvector (V1) representations of the data (Westin et al., 1997). These images are then deformably registered to the Montreal Neurological Institute (MNI) image space with the ANTs software package (Avants et al., 2008; Tustison et al., 2014). From there, we identify the 48 regions of interest (ROIs) in each image defined by the Johns Hopkins white matter atlas (Hua et al., 2008; Mori et al., 2005; Wakana et al., 2007).

For the NODDI approach, we extract the b = 1000 and 2000 s/mm^2 acquisitions from each scan with the largest number of directions. We then fit the multicompartment model with the UCL NODDI Toolbox as implemented in MATLAB (Zhang et al., 2012). The models are subsequently converted to cerebrospinal fluid (CSF) volume fraction (cVF), intracellular volume fraction (iVF), and orientation dispersion index (ODI) representations. These images are then deformably registered to MNI space with the ANTs software package. From there, we identify the 48 regions of interest (ROIs) in each image defined by the Johns Hopkins white matter atlas.

For the white matter segmentation approach, we extract the b = 2000 s/mm^2 acquisition from each scan with the largest number of directions. We calculate a whole-brain tractogram with DIPY of 2 million streamlines (Garyfallidis et al., 2014). We use the constrained spherical deconvolution model (Tournier et al., 2007) with probabilistic local tracking with a maximum angle of 25°, a seeding criterion of FA > 0.3, and a stopping criterion of FA < 0.2. We extract 43 white matter bundles (Supplementary Table 1) from each tractogram using the RecoBundles algorithm as implemented in DIPY. In short, each tractogram is registered to an MNI tractogram template and streamlines from each tractogram are assigned to bundles within the template (Garyfallidis et al., 2018). The length, volume, and FA of each bundle are then calculated. We calculate bundle length by calculating the median streamline length. We calculate volume by first converting each bundle to a tract density image representation. From there, a binary bundle mask is calculated by thresholding the tract density image at 5% of the 99th percentile density. Volume is calculated by multiplying the number of voxels in the mask by the volume of each voxel. FA is calculated by first converting the image to a tensor representation (Veraart et al., 2013) and then to an FA representation (Westin et al., 1997). Each bundle's binary mask is then applied to obtain the voxel-wise median FA value per bundle.

For the connectomics approach, we extract the b = 2000 s/mm^2 acquisition from each scan with the largest number of directions. We then calculate a whole-brain tractogram with MRtrix3 (Tournier et al., 2019). We first use the constrained spherical deconvolution model with probabilistic tracking with a maximum angle of 25°, a seeding criterion of FA > 0.3 and a stopping criterion of FA < 0.2 to calculate a 10 million streamline tractogram. The tractogram is then filtered with the SIFT approach to 2 million streamlines (Smith et al., 2013). We parcellate the brain into 96 cortical regions using the Harvard-Oxford cortical atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) and compute a connectome where each edge represents the average streamline distance connecting the two nodes. The maximum modularity, global efficiency, and characteristic path length are then calculated from each connectome using the Brain Connectivity Toolbox as implemented in MATLAB (Rubinov and Sporns, 2010).

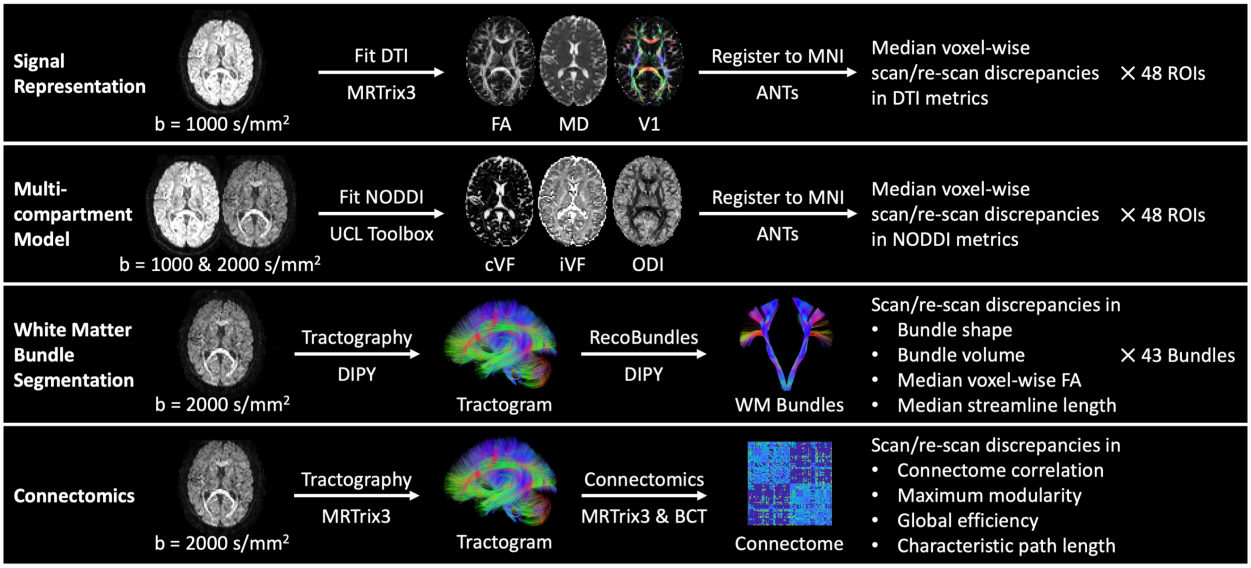

Figure 2. Outline of scan/re-scan discrepancy calculations in four types of common diffusion MRI processing.

Overview of Variability Analysis

The unique configuration of the MASiVar dataset allows for the extraction of four different levels of scan/re-scan pairs of images for variability analysis (Figure 3). This includes intrasession intrascanner (N = 158), intersession intrascanner (N = 328), intersession interscanner (N = 53), and intersubject intrascanner (N = 80) scan/re-scan pairs. For the intersubject intrascanner pairings, we omit cohort III to prevent bias of results toward scanner B. For all pairings we do not consider direction and consider the pairing of scan X and scan Y to be the same as scan Y and scan X. Using these pairs, we calculate the scan/re-scan discrepancy at the four different levels for the four processing approaches.

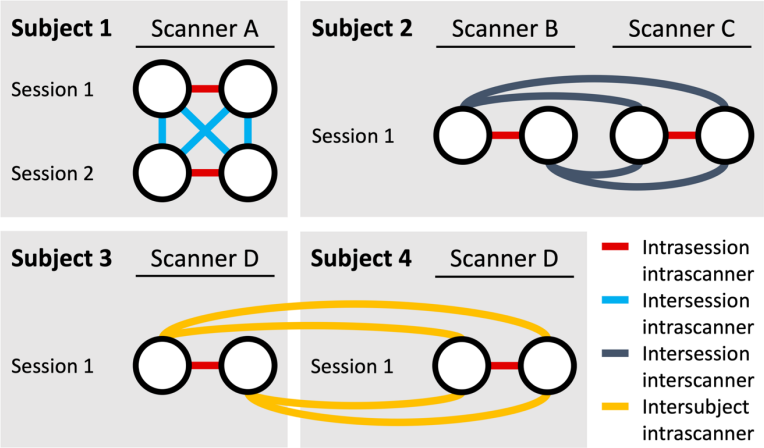

Figure 3. Example identification of scan/re-scan pairs at the four levels of variation. The MASiVar dataset consists of scan/re-scan pairs on four different levels: intrasession intrascanner, intersession intrascanner, intersession interscanner, and intersubject intrascanner. Identification of scan/re-scan pairs allows for quantification of variability at the four different levels.

Calculation of scan/re-scan discrepancy at a given level of variation

For a given level of variation, we identify all scan/re-scan pairs for that level in MASiVar. We then compute the scan/re-scan discrepancies within all pairs and take the median across pairs as the discrepancy at that level of variation, bootstrapping the image pairs used to obtain a 95% confidence interval (Efron, 1979). For the DTI, NODDI, and bundle segmentation analyses, we do this for each of multiple regions or bundles. An example of this is shown in Figure 4 for the FA DTI analysis. The remaining are shown in Supplementary Figures 1-9. In these cases, our final analysis is performed on the distribution of discrepancies across regions or bundles (Figures 5-7). Since we do not have multiple regions or bundles for the connectomics analysis, we use the bootstrapped distribution of discrepancies in their place (Figure 8).

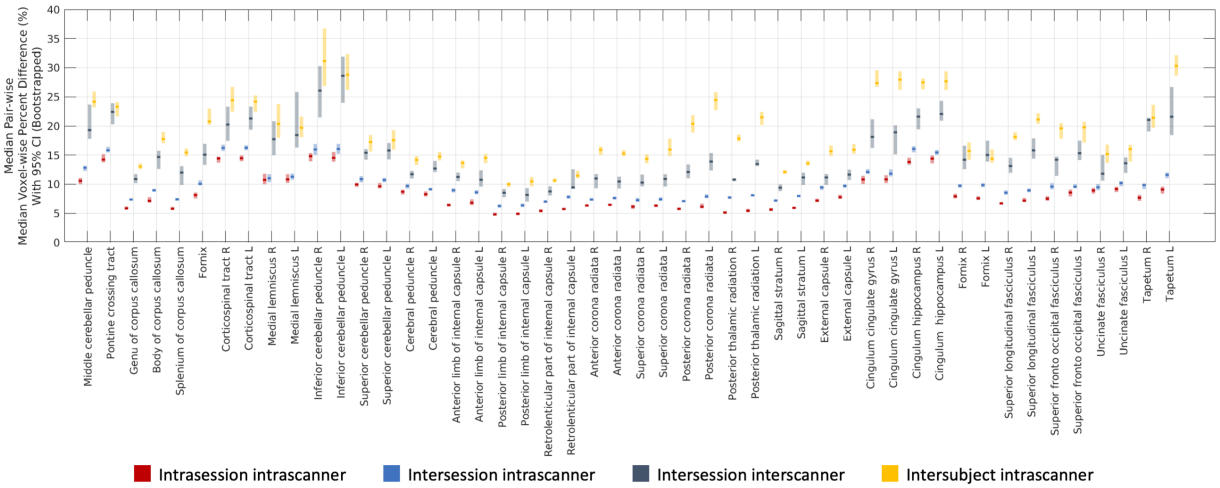

Figure 4. Bootstrapped scan/re-scan discrepancies of FA across 48 Johns Hopkins white matter atlas regions. The session and scanner effects across the white matter regions range from roughly 5% to 30%. For most regions, the scan/re-scan discrepancies increase from intrasession intrascanner to intersession intrascanner to intersession interscanner to intersubject intrascanner, suggesting the relative magnitude of session and scanner effects in white matter DWI analyses. The bootstrapped confidence intervals are generally tight and non-overlapping, supporting that these results are robust.

To compare the scan/re-scan discrepancies across levels, we use non-parametric statistical tests for all four processing approaches and report the uncorrected *p*-values. For the DTI, NODDI, and bundle segmentation analyses, we use the Wilcoxon signed-rank test for paired distributions, as all points in each distribution are measured from the same regions or bundles (Hollander et al., 2013). For the connectomics analysis, we use the Wilcoxon rank-sum test for unpaired distributions, as the distributions were constructed from bootstrapping and thus do not have correspondence (Hollander et al., 2013).

Calculation of scan/re-scan discrepancy within an image pair

We perform the region-based DTI and NODDI discrepancy analyses in MNI space with voxel-wise correspondence between images (Figure 2). For a given region and level of variation, we calculate the scan/re-scan discrepancy for FA, MD, cVF, iVF, and ODI as the median voxel-wise percent absolute difference. For V1, we define it as the median voxel-wise angular difference in degrees.

Because streamline-wise and subsequent voxel-wise correspondence cannot be achieved with tractography and bundle segmentation, we perform this analysis differently (Figure 2). For a given bundle and level of variation, we calculate the scan/re-scan discrepancy of bundle shape with the Dice similarity index (Dice, 1945) between the tract density images from the two paired images. We define the discrepancy of bundle volumes as the percent absolute

difference. We calculate the discrepancy of bundle FA as the percent absolute difference between the voxel-wise medians from each image and the discrepancy of bundle length as the percent absolute difference between the streamline-wise medians from each image.

To evaluate the scan/re-scan discrepancy of connectomics, we characterize each pair of connectomes as both (1) a whole and (2) through scalar measures (Figure 2). First, we calculate the Pearson correlation between the connectomes within each image pair as an estimate for connectome agreement within a pair. Second, we calculate the percent absolute difference in maximum modularity, global efficiency, and characteristic path length (Rubinov and Sporns, 2010) between the connectomes in the pair.

Analysis of Intrasubject to Intersubject Discrepancy Ratios

To understand how intrasubject variation compares to intersubject variation, we calculate the ratios of each of the intrasubject scan/re-scan discrepancy estimates to the corresponding intersubject intrascanner estimate. For the DTI, NODDI, and bundle segmentation analysis, we compute this ratio on a region- or bundle-wise basis, and for the connectome analysis, we compute this ratio on the bootstrapped distributions. For the Dice and correlation similarity measures, we first subtract them from 1 to obtain dissimilarity estimates. With this design, ratios of <1, 1, and >1 indicate the intrasubject variation is less than, equal to, or greater than the intersubject variation, respectively.

Results

Variability in tensor signal representations

We find that intrasession intrascanner FA measurements can vary as much as 7.6%, that intersession intrascanner measurements can vary as much as 9.5%, that intersession interscanner measurements can vary as much as 13.7%, and that intersubject intrascanner measurements can vary as much as 17.7%. We find the corresponding measurements in the MD case to be 2.8%, 4.8%, 8.6%, and 9.1% and for the V1 case to be 9.7°, 12.4°, 14.5°, and 20.2°, respectively. All these differences were statistically significant ($p < 0.0005$, Wilcoxon signed-rank test), except for the intersession interscanner and intersubject intrascanner comparison for MD (Figure 5).

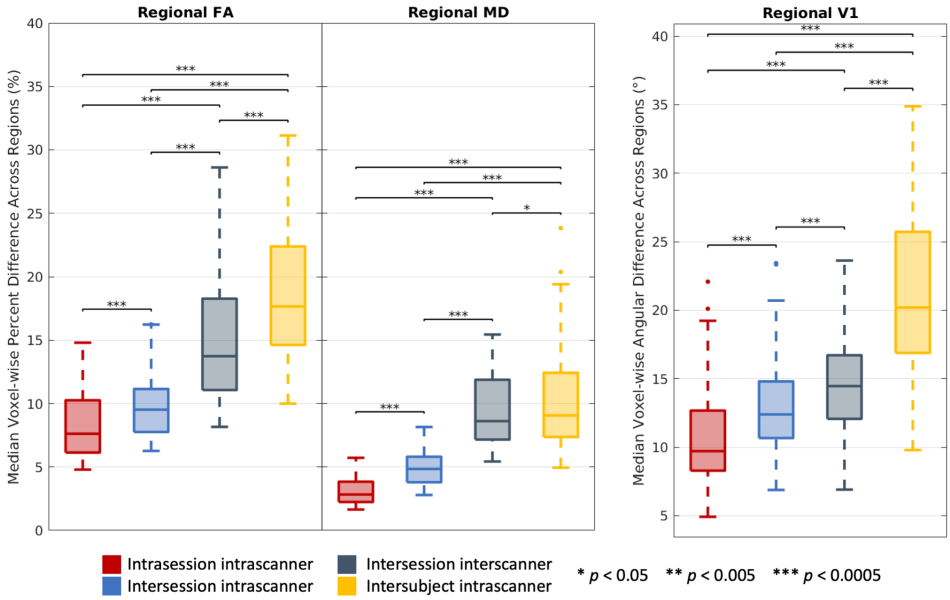

Figure 5. Scan/re-scan discrepancies of tensor signal representations at four levels of variability across 48 Johns Hopkins white matter atlas regions. Statistical significance was determined with the Wilcoxon signed-rank test.

Variability in multi-compartment modeling

We find that the intrasession intrascanner cVF measurements can vary as much as 50.1%, that intersession intrascanner measurements can vary as much as 56.6%, that intersession interscanner measurements can vary as much as 57.5%, and that intersubject intrascanner measurements can vary as much as 61.9%. We find the corresponding measurements in the iVF case to be 6.2%, 8.2%, 11.8%, and 12.1% and for the ODI case to be 10.2%, 14.0%, 19.1%, and 26.5%, respectively. All these differences were statistically significant ($p < 0.0005$, Wilcoxon signed-rank test), except for the intersession intrascanner and intersession interscanner cVF comparison ($p < 0.005$) and the intersession interscanner and intersubject intrascanner iVF comparison (Figure 6). We evaluated cVF only in white matter regions defined by the Johns Hopkins atlas and thus dealt with very low cVF values when calculating percent difference.

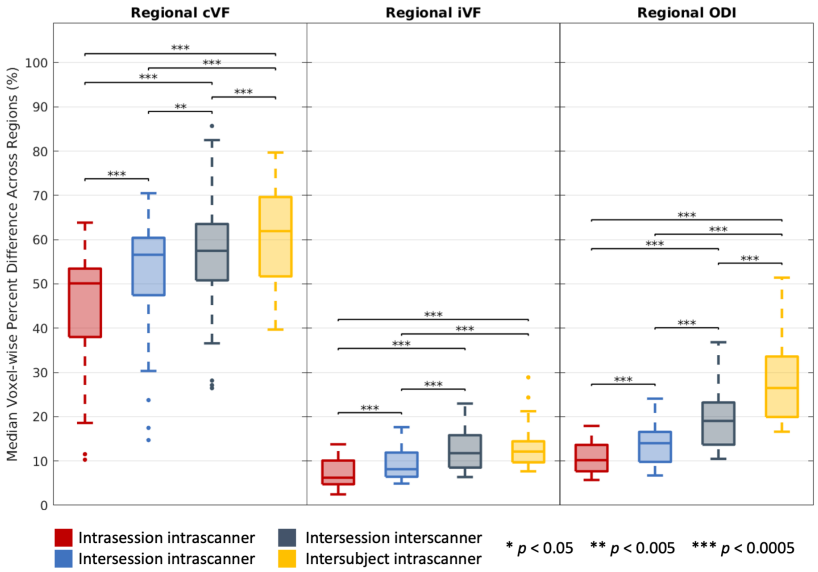

Figure 6. Scan/re-scan discrepancies of multi-compartment NODDI models at four levels of variability across 48 Johns Hopkins white matter atlas regions. Statistical significance was determined with the Wilcoxon signed-rank test.

Variability in white matter bundle segmentation

We find that intrasession intrascanner bundles overlap at 0.60 Dice, that intersession intrascanner bundles overlap at 0.59 Dice, that intersession interscanner bundles overlap at 0.57 Dice, and that intersubject intrascanner bundles overlap at 0.53 Dice. We find the measurements for the corresponding levels of variation in the bundle volume comparisons to be 11.0%, 10.8%, 12.8%, and 17.9%, in the FA case to be 1.6%, 2.0%, 5.0%, and 4.9%, and in the bundle length case to be 3.7%, 4.0%, 4.0%, and 8.0%, respectively. All these differences were statistically significant ($p < 0.0005$, Wilcoxon signed-rank test), except for the intrasession intrascanner and intersession intrascanner shape ($p < 0.05$), volume, and length ($p < 0.005$) comparisons as well as the intersession interscanner and intersubject intrascanner FA comparison and the intersession intrascanner and intersession interscanner length comparison ($p < 0.05$) (Figure 7).

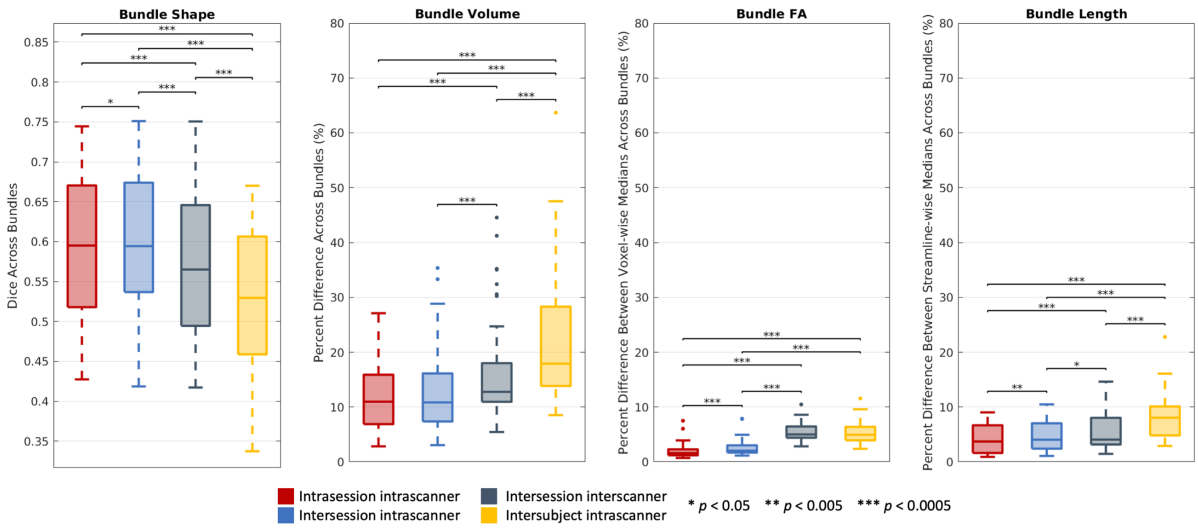

Figure 7. Scan/re-scan discrepancies of white matter bundle segmentation at four levels of variability across 43 bundles. Statistical significance was determined with the Wilcoxon signed-rank test.

Variability in connectomics

We find that the intrasession intrascanner connectomes correlate at 0.61, that the intersession intrascanner connectomes correlate at 0.59, that the intersession interscanner connectomes correlate at 0.59 and that the intersubject intrascanner connectomes correlate at 0.54. We find the measurements for the corresponding levels of variation in the maximum modularity case to be 4.7%, 6.4%, 9.5%, and 13.1%, in the global efficiency case to be 1.2%, 1.9%, 6.1%, and 6.2%, and in the characteristic path length case to be 1.1%, 1.9%, 6.2%, and 6.1%, respectively. All these differences were statistically significant ($p < 0.0005$, Wilcoxon rank-sum test) except for the intersession interscanner and intersubject intrascanner comparison of characteristic path length (Figure 8).

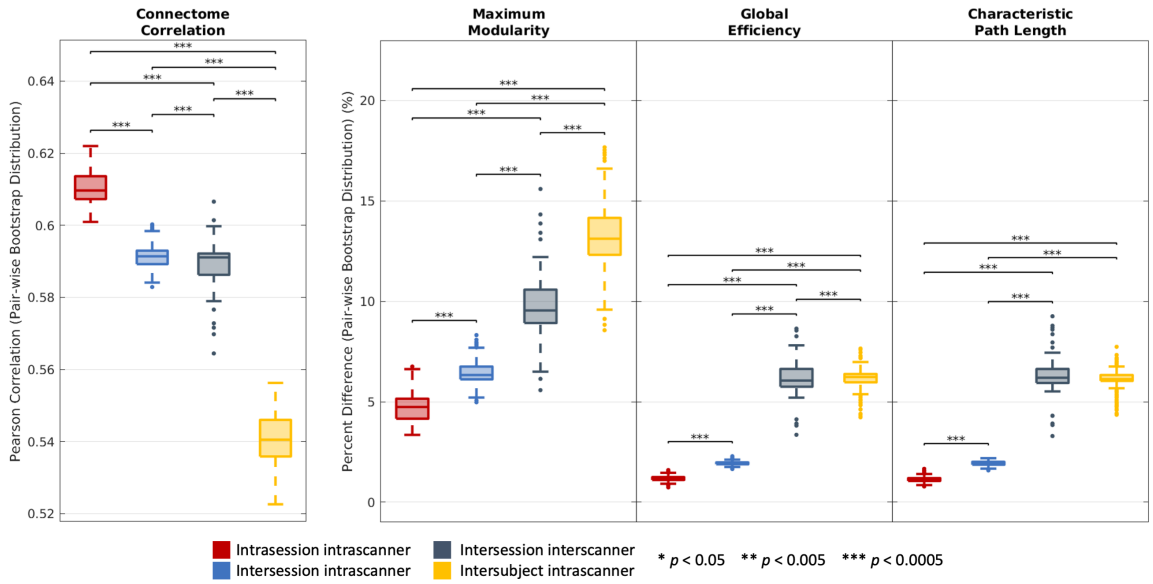

Figure 8. Scan/re-scan discrepancies of connectomics analysis at four levels of variability with bootstrapped distributions. Statistical significance was determined with the Wilcoxon rank-sum test.

Ratios of Intrasubject Discrepancies to Intersubject Discrepancies

We plot the intrasubject to intersubject discrepancy ratios in Figure 9. We consider a ratio of <0.6 to represent good reproducibility, 0.6-0.8 to represent moderate, and >0.8 to represent poor reproducibility. With this model, we find good to moderate intrasession intrascanner and intersession intrascanner reproducibility for most DTI- and NODDI-based microstructural measures. The bundle volume, FA, length, and the connectomics graph estimates also exhibit this behavior, with the graph measures even demonstrating ratios <0.4. Notably, the white matter cVF, bundle shape, and connectome correlation investigations exhibit moderate to poor intrasession intrascanner and intersession intrascanner reproducibility. Last, across the board we find that intersession interscanner effects can introduce variability that is on par with variability from intersubject intrascanner effects. We find this trend is most striking for the global efficiency and characteristic path length analyses that demonstrated intrasession intrascanner ratios of around 0.2 but intersession interscanner ratios near 1. In summary, we find that across all four DWI analysis approaches, intrasubject variability is within an order of magnitude of intersubject variability, and the size of interscanner effects can approach the size of intersubject effects.

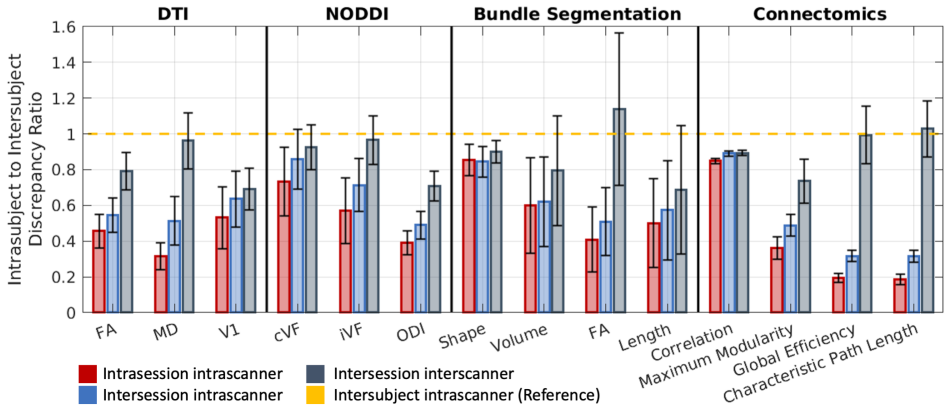

Figure 9. The ratio of intrasubject to intersubject discrepancies across all four DWI analysis approaches. Average ratios are reported with error bars denoting standard deviation.

Discussion and Conclusions

Here, we present, MASiVar, a dataset designed for investigation of DWI variability. We characterize intrasession intrascanner, intersession intrascanner, intersession interscanner, and intersubject intrascanner variability in four common diffusion processing approaches. In support of our hypothesis, we consistently find across all approaches that variability increases with consideration of session, scanner, and subject effects. We also find that intrasubject variability accounts for a non-negligible portion of intersubject variability and that at times intersession interscanner variability can approach intersubject intrascanner variability. We interpret these results to mean that harmonization between scanners for multisite analyses is critical prior to inference of group differences on subjects. We do not perform harmonization here. Instead we hope that our baseline estimates for variability will provide a benchmark for future investigations into harmonization and allow investigators looking to make group inferences from DWI data to optimize the study designs needed to resolve certain effects in their research.

The reproducibility of DWI analysis has received significant attention in the field, including the analysis of tensor representations (Andica et al., 2020; Magnotta et al., 2012; Palacios et al., 2017; Vollmar et al., 2010), multi-compartment models (Andica et al., 2020; Tariq et al., 2013), tractography and bundle segmentation (Besseling et al., 2012; Nath et al., 2020), and connectomics (Prčkovska et al., 2016; Roine et al., 2019). However, these studies have each primarily focused on one type of approach and one or two levels of variation. Thus, to the best of our knowledge, this study and dataset represent the first attempt and first dataset configured to characterize all four types of diffusion processing and all four levels of variation simultaneously.

For instance, regarding DTI, previous studies have estimated within and between scanner coefficients of variation (CVs) for regional FA at roughly 1% and up to 3% (Magnotta et al., 2012), at 1-2% and 2-4% (Vollmar et al., 2010), and at roughly 3% and 8% (Palacios et al., 2017), respectively. The same estimates of variability for regional MD have been estimated at roughly 1% and up to 2% (Magnotta et al., 2012) and 2% and 6% (Palacios et al., 2017). Another study put white matter FA between session and between scanner CVs at 0.5% and 2% and at 0.2% and 3% for MD, respectively (Andica et al., 2020). Intra- and intersession variability of FA for white matter regions have also been estimated with a CV of 2% and 3%, respectively, and at 1% for both for MD (Farrell et al., 2010). Between session variability of V1 has been estimated at roughly 7° to 10° (Farrell et al., 2010). We find the same general trends in our results that variation in DTI measures increases from consideration of session effects to consideration of scanner effects. Our results, however, are larger overall with session and scanner effects ranging from roughly 5 to 15% difference in scan/re-scan pairs for FA and MD. Our between session V1 error at 10-15° is also larger than the prior estimate.

Previously, one study estimated between session and between scanner CVs for NODDI measures at roughly 1-4% and 2-14%, respectively (Andica et al., 2020). Another study estimated the between session CV to be around 4-5% (Tariq et al., 2013). Our NODDI results place the percent difference for iVF and ODI at 5-15% for between session effects and 10-20% for between scanner effects. As with the DTI analysis, we observe consistent increases in variation from session to scanner effects, but larger estimates of variation overall.

Another study quantified estimates of bundle shape reproducibility at around 0.65 to 0.92 Dice for between sessions comparisons (Besseling et al., 2012). Another study similarly demonstrated between session bundle shape Dice greater than 0.88 was achievable (Nath et al., 2020). Between session comparisons of bundle-based volume and FA estimates with CV were calculated around 3-22% and 1-4%, respectively (Besseling et al., 2012). We observe similar bundle shape consistency around 0.6 Dice. We also observe similar estimates of variation for bundle volume and FA due to between session effects.

Between session connectome correlations have previously been calculated to range between 0.6 and 0.95 (Prčkovska et al., 2016). The between session CV of global efficiency and characteristic path length has been calculated as roughly 31% and 2% (Roine et al., 2019), respectively. Our intersession connectome correlations were roughly just under 0.6. In addition, our intersession estimates for global efficiency and characteristic path length were both around 2%.

Additionally, the within and between scanner intraclass correlation coefficient (ICC) for regional FA have previously been estimated at 0.90-0.99 and 0.82-0.99, respectively (Vollmar et al., 2010), and at 0.74-1.00 and 0.54-0.97, respectively (Andica et al., 2020). Other studies have placed the ICC for the cVF, iVF, and ODI NODDI measures in white matter regions at 0.133-0.997, 0.773-0.989, and 0.789-0.998, respectively, for between session analysis and at 0.013-0.545, 0.300-0.935, and 0.181–0.962 for between scanner analysis (Andica et al., 2020). Between session ICC for bundle volume and FA have been estimated at 0.53-0.96 and 0.65-0.94, respectively (Besseling et al., 2012), and between session ICC for global efficiency and characteristic path length have been calculated at 0.78 and 0.77, respectively. We observe similar trends in our discrepancy ratios that variability between subjects can be more largely attributed to scanner effects than session effects.

In summary, when comparing to previous variability studies, we find similar trends across all approaches that variation increases with session and scanner biases. We also find that our estimates of variation are generally larger than previously reported, especially for the DTI approach. These existing studies were also limited in their ability to do a full characterization of DWI variability. Whether due to limited subject number, missing acquisitions, or missing multiple sessions or scanners, they all fell short of being able to characterize all four levels of variability at once for multiple different diffusion approaches. As a result, it was difficult to obtain a full picture of DWI variability. Thus, we hope that MASiVar and the baseline characterization presented here will promote further

investigation into a wide spectrum of DWI variability issues from a large pool of models to push the field toward a global understanding of the effects of session, scanner, and subject biases on DWI measurements.

Of note, the bulk of existing studies used CV to estimate variation and ICC to characterize the proportion of between subjects variation attributable to session or scanner effects. In contrast, we used a paired intrasubject and intersubject percent difference approach for the former and quantified their ratios for the latter. We chose these approaches to improve isolation of session and scanner effects. For instance, by pairing images, we remove other potential confounders like subject or acquisition effects. This approach also has the benefit of characterizing the relative sizes of within and between subjects variation without the necessary assumptions of traditional ICC, including not having additional confounders within each class, an assumption that is inherently violated when looking at sessions nested within scanners nested with subjects.

For this study, we chose popular software toolboxes to do all the analyses, parameter configurations that we were familiar with, and common similarity assessments that we found to be interpretable. However, we recognize that there are many other software options available to do similar tasks, each with a large number of different configurations, and a large number of ways to assess variability. For instance, there are different methods for fitting tensors (Chang et al., 2005; Cook et al., 2006; Hernandez-Fernandez et al., 2019), for identifying regions (Desikan et al., 2006; Figley et al., 2017; Hansen et al., 2020; Volz et al., 2018) and bundles (Warrington et al., 2020; Yeh, 2020; Yeh et al., 2013; Yendiki et al., 2011), for comparing bundles (Rheault et al., 2020), and for configuring and representing connectomes (Hagmann et al., 2008; Roine et al., 2019; Rubinov and Sporns, 2010; Sporns et al., 2005). Additionally, there are a number of other microstructural measures that can be characterized as well (Koller et al., 2020). Thus, the goal of the present study was not to provide an analysis between different processing toolboxes, parameters, or analysis approaches. Specifically, since each approach was not necessarily optimized, we do not recommend thorough comparisons of reproducibility between the four different approaches. Instead, we aimed to establish a baseline understanding of DWI variability and its trends across sessions, scanners, and subjects in four common processing approaches and DWI characterizations in a generally interpretable way that demonstrated the potential of the dataset. As such, we hope that the release of MASiVar will prompt other investigators in the field to further characterize differences between software tools and their parameters, different DWI processing and variability measures, and other potential confounders in DWI analysis.

In addition to the ability of MASiVar to serve as a utility for variability analysis, we note that the pediatric subjects in cohort III present another unique resource for the field. The majority of the existing DWI datasets and studies for variability use adult subjects. Of existing pediatric datasets, many have focused on older age ranges. For example, the Adolescent Brain Cognitive Development project (Casey et al., 2018) and the Lifespan Human Connectome Project in Development (Somerville et al., 2018) contain longitudinal DWI data acquired from children starting at age 9 and 10 through adolescence. Thus, to the best of our knowledge, MASiVar represents one of the first publicly available longitudinal DWI datasets of children prior to adolescence aged 5-8 years old and is further distinguished by its inclusion of repeated scans within each session. We hope that investigators in developmental neuroscience and pediatric neurology will be able to take advantage of this resource for their work.

One limitation of the variability study is the differences in number of gradient directions between the different cohorts. Cohort III consists of a 40-direction b = 1000 s/mm^2 acquisition and a 56-direction b = 2000 s/mm^2 acquisition in contrast to the 96 directions for cohorts I and II. There is a potential effect that could be biasing the results. However, our study design focuses on scan/re-scan pairs and thus all variability measurements are between acquisitions of the same configuration, which should minimize this effect. In a similar vein, due to hardware limitations, the data collected at site 3 in cohort II was collected at a maximum shell of 2465 s/mm^2 as opposed to the 2500 s/mm^2 across the rest of MASiVar. This shell was not used for the present variability analysis, but this discrepancy should be noted on future studies using the dataset. Additionally, considerations for the pediatric subjects in cohort III that contributed to this study must be taken into account. One aspect is that pediatric patients often move more while being scanned. All images were preprocessed, motion corrected, and quality checked prior to analysis (Cai et al., 2020), but small residual artifacts may remain thus potentially increasing variability estimates.

Another aspect that must be considered is that for some of the intersession measurements, the one-year separation between sessions may result in increased variability from inherent white matter changes in the developing brain. Due to the paired approach and the omission of cohort III from the intersubject analysis, we do not expect this to impact the other variability levels, and thus do not expect it to significantly impact the overall conclusions of this study.

Last, we have made the MASiVar dataset publicly available at https://openneuro.org/datasets/ds003416 in Brain Imaging Data Structure (BIDS) format (Gorgolewski et al., 2016) with deidentified metadata and defaced images.

Acknowledgements

The authors thank E. Brian Welch for his help with image acquisition and study design and Zachary J. Williams for his insight into bootstrapping. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. This work was supported by the National Institutes of Health (NIH) under award numbers 5R01EB017230, 5T32EB001628, 5T32GM007347, and 1UL1RR024975. This work was also supported by the National Science Foundation (NSF) under award numbers 1452485, 1660816, and 1750213. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

References

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

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

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

Andica, C., Kamagata, K., Hayashi, T., Hagiwara, A., Uchida, W., Saito, Y., Kamiya, K., Fujita, S., Akashi, T., Wada, A., Abe, M., Kusahara, H., Hori, M., Aoki, S., 2020. Scan–rescan and inter-vendor reproducibility of neurite orientation dispersion and density imaging metrics. Neuroradiology 62, 483–494. https://doi.org/10.1007/s00234-019-02350-6

Assaf, Y., Pasternak, O., 2008. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review. J. Mol. Neurosci. 34, 51–61. https://doi.org/10.1007/s12031-007-0029-0

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41. https://doi.org/10.1016/j.media.2007.06.004

Besseling, R.M.H., Jansen, J.F.A., Overvliet, G.M., Vaessen, M.J., Braakman, H.M.H., Hofman, P.A.M., Aldenkamp, A.P., Backes, W.H., 2012. Tract specific reproducibility of tractography based morphology and diffusion metrics. PLoS One 7. https://doi.org/10.1371/journal.pone.0034125

Cai, L.Y., Yang, Q., Hansen, C.B., Nath, V., Ramadass, K., Johnson, G.W., Conrad, B.N., Boyd, B.D., Begnoche, J.P., Beason-Held, L.L., Shafer, A.T., Resnick, S.M., Taylor, W.D., Price, G.R., Morgan, V.L., Rogers, B.P., Schilling, K.G., Landman, B.A., 2020. PreQual: An automated pipeline for integrated preprocessing and quality assurance of diffusion weighted MRI images. bioRxiv 2020.09.14.260240. https://doi.org/10.1101/2020.09.14.260240

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Daniela Cornejo, M., Sicat, C.S., Harms, M.P., Dosenbach, N.U.F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni,

J.R., Kuperman, J.M., Fair, D.A., Dale, A.M., 2018. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Cetin-Karayumak, S., Di Biase, M.A., Chunga, N., Reid, B., Somes, N., Lyall, A.E., Kelly, S., Solgun, B., Pasternak, O., Vangel, M., Pearlson, G., Tamminga, C., Sweeney, J.A., Clementz, B., Schretlen, D., Viher, P.V., Stegmayer, K., Walther, S., Lee, J., Crow, T., James, A., Voineskos, A., Buchanan, R.W., Szeszko, P.R., Malhotra, A.K., Hegde, R., McCarley, R., Keshavan, M., Shenton, M., Rathi, Y., Kubicki, M., 2019. White matter abnormalities across the lifespan of schizophrenia: a harmonized multi-site diffusion MRI study. Mol. Psychiatry. https://doi.org/10.1038/s41380-019-0509-y

Chang, L.C., Jones, D.K., Pierpaoli, C., 2005. RESTORE: Robust estimation of tensors by outlier rejection. Magn. Reson. Med. 53, 1088–1095. https://doi.org/10.1002/mrm.20426

Cook, P. a, Bai, Y., Seunarine, K.K., Hall, M.G., Parker, G.J., Alexander, D.C., 2006. Camino: Open-Source Diffusion-MRI Reconstruction and Processing. 14th Sci. Meet. Int. Soc. Magn. Reson. Med. 14, 2759.

Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A.N., Hajnal, J. V., 2019. Complex diffusion-weighted image estimation via matrix recovery under general noise models. Neuroimage 200, 391–404. https://doi.org/10.1016/j.neuroimage.2019.06.039

De Santis, S., Bastiani, M., Droby, A., Kolber, P., Zipp, F., Pracht, E., Stoecker, T., Groppa, S., Roebroeck, A., 2019. Characterizing Microstructural Tissue Properties in Multiple Sclerosis with Diffusion MRI at 7 T and 3 T: The Impact of the Experimental Design. Neuroscience 403, 17–26. https://doi.org/10.1016/j.neuroscience.2018.03.048

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Di Martino, A., O'Connor, D., Chen, B., Alaerts, K., Anderson, J.S., Assaf, M., Balsters, J.H., Baxter, L., Beggiato, A., Bernaerts, S., Blanken, L.M.E., Bookheimer, S.Y., Braden, B.B., Byrge, L., Castellanos, F.X., Dapretto, M., Delorme, R., Fair, D.A., Fishman, I., Fitzgerald, J., Gallagher, L., Keehn, R.J.J., Kennedy, D.P., Lainhart, J.E., Luna, B., Mostofsky, S.H., Müller, R.A., Nebel, M.B., Nigg, J.T., O'Hearn, K., Solomon, M., Toro, R., Vaidya, C.J., Wenderoth, N., White, T., Craddock, R.C., Lord, C., Leventhal, B., Milham, M.P., 2017. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Sci. Data 4, 170010. https://doi.org/10.1038/sdata.2017.10

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Duchesne, S., Chouinard, I., Potvin, O., Fonov, V.S., Khademi, A., Bartha, R., Bellec, P., Collins, D.L., Descoteaux, M., Hoge, R., McCreary, C.R., Ramirez, J., Scott, C.J.M., Smith, E.E., Strother, S.C., Black, S.E., 2019. The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts. J. Magn. Reson. Imaging 49, 456–465. https://doi.org/10.1002/jmri.26197

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. Ann. Stat. 7, 1–26. https://doi.org/10.1214/aos/1176344552

Farrell, J.A.D., Landman, B.A., Jones, C.K., Smith, A., Prince, J.L., Zijl, P.C.M. Van, Mori, S., 2010. Effects of SNR on the Accuracy and Reproducibility of DTI-derived Fractional Anisotropy, Mean Diffusivity, and Principal Eigenvector Measurements at 1.5T. J. Magn. Reson. 26, 756–767. https://doi.org/10.1002/jmri.21053.Effects

Figley, T.D., Mortazavi Moghadam, B., Bhullar, N., Kornelsen, J., Courtney, S.M., Figley, C.R., 2017. Probabilistic white matter atlases of human auditory, basal ganglia, language, precuneus, sensorimotor, visual and visuospatial networks. Front. Hum. Neurosci. 11, 1–12. https://doi.org/10.3389/fnhum.2017.00306

Frazier, J.A., Chiu, S., Breeze, J.L., Makris, N., Lange, N., Kennedy, D.N., Herbert, M.R., Bent, E.K., Koneru, V.K., Dieterich, M.E., Hodge, S.M., Rauch, S.L., Grant, P.E., Cohen, B.M., Seidman, L.J., Caviness, V.S., Biederman, J., 2005. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. Am. J. Psychiatry 162, 1256–1265. https://doi.org/10.1176/appi.ajp.162.7.1256

Froeling, M., Tax, C.M.W., Vos, S.B., Luijten, P.R., Leemans, A., 2017. "MASSIVE" brain dataset: Multiple

acquisitions for standardization of structural imaging validation and evaluation. Magn. Reson. Med. 77, 1797–1809. https://doi.org/10.1002/mrm.26259
Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., Nimmo-Smith, I., 2014. Dipy, a library for the analysis of diffusion MRI data. Front. Neuroinform. 8, 1–17. https://doi.org/10.3389/fninf.2014.00008
Garyfallidis, E., Côté, M.A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., Descoteaux, M., 2018. Recognition of white matter bundles using local and global streamline-based registration and clustering. Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.07.015
Goldstein, J.M., Seidman, L.J., Makris, N., Ahern, T., O'Brien, L.M., Caviness, V.S., Kennedy, D.N., Faraone, S. V., Tsuang, M.T., 2007. Hypothalamic Abnormalities in Schizophrenia: Sex Effects and Genetic Vulnerability. Biol. Psychiatry 61, 935–945. https://doi.org/10.1016/j.biopsych.2006.06.027
Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3, 160044. https://doi.org/10.1038/sdata.2016.44
Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Van Wedeen, J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol. 6, 1479–1493. https://doi.org/10.1371/journal.pbio.0060159
Hansen, C.B., Yang, Q., Lyu, I., Rheault, F., Kerley, C., Chandio, B.Q., Fadnavis, S., Williams, O., Shafer, A.T., Resnick, S.M., Zald, D.H., Cutting, L., Taylor, W.D., Boyd, B., Garyfallidis, E., Anderson, A.W., Descoteaux, M., Landman, B.A., Schilling, K.G., 2020. Pandora: 4-D white matter bundle population-based atlases derived from diffusion MRI fiber tractography. bioRxiv 2020.06.12.148999. https://doi.org/10.1101/2020.06.12.148999
Hernandez-Fernandez, M., Reguly, I., Jbabdi, S., Giles, M., Smith, S., Sotiropoulos, S.N., 2019. Using GPUs to accelerate computational diffusion MRI: From microstructure estimation to tractography and connectomes. Neuroimage 188, 598–615. https://doi.org/10.1016/j.neuroimage.2018.12.015
Hollander, M., Wolfe, D.A., Chicken, E., 2013. Nonparametric statistical methods. John Wiley & Sons.
Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P.A., Pekar, J.J., van Zijl, P.C.., Mori, S., 2008. Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. Neuroimage 39, 336–347. https://doi.org/10.1016/j.neuroimage.2007.07.053
Jelescu, I.O., Veraart, J., Adisetiyo, V., Milla, S.S., Novikov, D.S., Fieremans, E., 2015. One diffusion acquisition and different white matter models: How does microstructure change in human early development based on WMTI and NODDI? Neuroimage 107, 242–256. https://doi.org/10.1016/j.neuroimage.2014.12.009
Kodl, C.T., Franc, D.T., Rao, J.P., Anderson, F.S., Thomas, W., Mueller, B.A., Lim, K.O., Seaquist, E.R., 2008. Diffusion tensor imaging identifies deficits in white matter microstructure in subjects with type 1 diabetes that correlate with reduced neurocognitive function. Diabetes 57, 3083–3089. https://doi.org/10.2337/db08-0724
Koller, K., Rudrapatna, S.U., Chamberland, M., Raven, E.P., Parker, G.D., Tax, C.M.W., Drakesmith, M., Fasan, F., Owen, D., Hughes, G., Charron, C., Evans, J.C., Jones, D.K., 2020. MICRA: Microstructural Image Compilation with Repeated Acquisitions. Neuroimage 225, 117406. https://doi.org/10.1016/j.neuroimage.2020.117406
Magnotta, V.A., Matsui, J.T., Liu, D., Johnson, H.J., Long, J.D., Bolster, 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. https://doi.org/10.1089/brain.2012.0112
Makris, N., Goldstein, J.M., Kennedy, D., Hodge, S.M., Caviness, V.S., Faraone, S. V., Tsuang, M.T., Seidman, L.J., 2006. Decreased volume of left and total anterior insular lobule in schizophrenia. Schizophr. Res. 83, 155–171. https://doi.org/10.1016/j.schres.2005.11.020
Menzler, K., Belke, M., Wehrmann, E., Krakow, K., Lengler, U., Jansen, A., Hamer, H.M., Oertel, W.H.,

Rosenow, F., Knake, S., 2011. Men and women are different: Diffusion tensor imaging reveals sexual dimorphism in the microstructure of the thalamus, corpus callosum and cingulum. Neuroimage 54, 2557–2562. https://doi.org/10.1016/j.neuroimage.2010.11.029

Mori, S., Wakana, S., Van Zijl, P.C.M., Nagae-Poetscher, L.M., 2005. MRI atlas of human white matter. Elsevier.

Nath, V., Schilling, K.G., Parvathaneni, P., Huo, Y., Blaber, J.A., Hainline, A.E., Barakovic, M., Romascano, D., Rafael-Patino, J., Frigo, M., Girard, G., Thiran, J.P., Daducci, A., Rowe, M., Rodrigues, P., Prčkovska, V., Aydogan, D.B., Sun, W., Shi, Y., Parker, W.A., Ould Ismail, A.A., Verma, R., Cabeen, R.P., Toga, A.W., Newton, A.T., Wasserthal, J., Neher, P., Maier-Hein, K., Savini, G., Palesi, F., Kaden, E., Wu, Y., He, J., Feng, Y., Paquette, M., Rheault, F., Sidhu, J., Lebel, C., Leemans, A., Descoteaux, M., Dyrby, T.B., Kang, H., Landman, B.A., 2020. Tractography reproducibility challenge with empirical data (TraCED): The 2017 ISMRM diffusion study group challenge. J. Magn. Reson. Imaging 51, 234–249. https://doi.org/10.1002/jmri.26794

O'Donnell, L.J., Westin, C.F., 2011. An introduction to diffusion tensor image analysis. Neurosurg. Clin. N. Am. https://doi.org/10.1016/j.nec.2010.12.004

Palacios, E.M., Martin, A.J., Boss, M.A., Ezekiel, F., Chang, S., Yuh, E.L., Vassar, M.J., Schnyer, D.M., Donald, C.L. Mac, Crawford, K.L., Irimia, A., Toga, A.W., 2017. Towards Precision and Reproducibility of DTI. AJNR Am J Neuroradiol. 38, 537–545. https://doi.org/10.3174/ajnr.A5025.Towards

Prčkovska, V., Rodrigues, P., Puigdellivol Sanchez, A., Ramos, M., Andorra, M., Martinez-Heras, E., Falcon, C., Prats-Galino, A., Villoslada, P., 2016. Reproducibility of the Structural Connectome Reconstruction across Diffusion Methods. J. Neuroimaging 26, 46–57. https://doi.org/10.1111/jon.12298

Repple, J., Karliczek, G., Meinert, S., Förster, K., Grotegerd, D., Goltermann, J., Redlich, R., Arolt, V., Baune, B.T., Dannlowski, U., Opel, N., 2019. Variation of HbA1c affects cognition and white matter microstructure in healthy, young adults. Mol. Psychiatry 1–10. https://doi.org/10.1038/s41380-019-0504-3

Rheault, F., De Benedictis, A., Daducci, A., Maffei, C., Tax, C.M.W., Romascano, D., Caverzasi, E., Morency, F.C., Corrivetti, F., Pestilli, F., Girard, G., Theaud, G., Zemmoura, I., Hau, J., Glavin, K., Jordan, K.M., Pomiecko, K., Chamberland, M., Barakovic, M., Goyette, N., Poulin, P., Chenot, Q., Panesar, S.S., Sarubbo, S., Petit, L., Descoteaux, M., 2020. Tractostorm: The what, why, and how of tractography dissection reproducibility. Hum. Brain Mapp. 41, 1859–1874. https://doi.org/10.1002/hbm.24917

Roine, T., Jeurissen, B., Perrone, D., Aelterman, J., Philips, W., Sijbers, J., Leemans, A., 2019. Reproducibility and intercorrelation of graph theoretical measures in structural brain connectivity networks. Med. Image Anal. 52, 56–67. https://doi.org/10.1016/j.media.2018.10.009

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: Uses and interpretations. Neuroimage 52, 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Schilling, K.G., Blaber, J., Hansen, C., Cai, L., Rogers, B., Anderson, A.W., Smith, S., Kanakaraj, P., Rex, T., Resnick, S.M., Shafer, A.T., Cutting, L.E., Woodward, N., Zald, D., Landman, B.A., 2020a. Distortion correction of diffusion weighted MRI without reverse phase-encoding scans or field-maps. PLoS One 15, e0236418.

Schilling, K.G., Petit, L., Rheault, F., Remedios, S., Pierpaoli, C., Anderson, A.W., Landman, B.A., Descoteaux, M., 2020b. Brain connections derived from diffusion MRI tractography can be highly anatomically accurate—if we know where white matter pathways start, where they end, and where they do not go. Brain Struct. Funct. 225, 2387–2402. https://doi.org/10.1007/s00429-020-02129-z

Smith, R.E., Tournier, J.D., Calamante, F., Connelly, A., 2013. SIFT: Spherical-deconvolution informed filtering of tractograms. Neuroimage 67, 298–312. https://doi.org/10.1016/j.neuroimage.2012.11.049

Somerville, L.H., Bookheimer, S.Y., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Dapretto, M., Elam, J.S., Gaffrey, M.S., Harms, M.P., Hodge, C., Kandala, S., Kastman, E.K., Nichols, T.E., Schlaggar, B.L., Smith, S.M., Thomas, K.M., Yacoub, E., Van Essen, D.C., Barch, D.M., 2018. The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. Neuroimage 183, 456–468. https://doi.org/10.1016/j.neuroimage.2018.08.050

Sotiropoulos, S.N., Zalesky, A., 2019. Building connectomes using diffusion MRI: why, how and but. NMR Biomed. 32, 1–23. https://doi.org/10.1002/nbm.3752

Sporns, O., Tononi, G., Kötter, R., 2005. The human connectome: A structural description of the human brain.

PLoS Comput. Biol. 1, 0245–0251. https://doi.org/10.1371/journal.pcbi.0010042
Tariq, M., Schneider, T., Alexander, D.C., Wheeler-Kingshott, C.A., Zhang, H., 2013. Assessing scan-rescan reproducibility of the parameter estimates from NODDI. Proc. Intl. Soc. Mag. Reson. Med. 21 21, 3187.
Tong, Q., He, H., Gong, T., Li, C., Liang, P., Qian, T., Sun, Y., Ding, Q., Li, K., Zhong, J., 2020. Multicenter dataset of multi-shell diffusion MRI in healthy traveling adults with identical settings. Sci. Data 7. https://doi.org/10.1038/s41597-020-0493-8
Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A., 2019. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. https://doi.org/10.1016/j.neuroimage.2019.116137
Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35, 1459–1472. https://doi.org/10.1016/j.neuroimage.2007.02.016
Travers, B.G., Adluru, N., Ennis, C., Tromp, D.P.M., Destiche, D., Doran, S., Bigler, E.D., Lange, N., Lainhart, J.E., Alexander, A.L., 2012. Diffusion Tensor Imaging in Autism Spectrum Disorder: A Review. Autism Res. https://doi.org/10.1002/aur.1243
Tustison, N.J., Cook, P.A., Klein, A., Song, G., Das, S.R., Duda, J.T., Kandel, B.M., van Strien, N., Stone, J.R., Gee, J.C., Avants, B.B., 2014. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 99, 166–179. https://doi.org/10.1016/j.neuroimage.2014.05.044
Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn Human Connectome Project: An overview. Neuroimage 80, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041
Veraart, J., Fieremans, E., Novikov, D.S., 2016a. Diffusion MRI noise mapping using random matrix theory. Magn. Reson. Med. 76, 1582–1593. https://doi.org/10.1002/mrm.26059
Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016b. Denoising of diffusion MRI using random matrix theory. Neuroimage 142, 394–406. https://doi.org/10.1016/j.neuroimage.2016.08.016
Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B., 2013. Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls. Neuroimage 81, 335–346. https://doi.org/10.1016/j.neuroimage.2013.05.028
Vollmar, C., O'Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V., Duncan, J.S., Richardson, M.P., Koepp, M.J., 2010. Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. Neuroimage 51, 1384–1394. https://doi.org/10.1016/j.neuroimage.2010.03.046
Volz, L.J., Cieslak, M., Grafton, S.T., 2018. A probabilistic atlas of fiber crossings for variability reduction of anisotropy measures. Brain Struct. Funct. 223, 635–651. https://doi.org/10.1007/s00429-017-1508-x
Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., Mori, S., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage 36, 630–644. https://doi.org/10.1016/j.neuroimage.2007.02.049
Warrington, S., Bryant, K.L., Khrapitchev, A.A., Sallet, J., Charquero-Ballester, M., Douaud, G., Jbabdi, S., Mars, R.B., Sotiropoulos, S.N., 2020. XTRACT - Standardised protocols for automated tractography in the human and macaque brain. Neuroimage 217, 1–15. https://doi.org/10.1016/j.neuroimage.2020.116923
Westin, C.F., Peled, S., Gudbjartsson, H., Kikinis, R., Jolesz, F.A., 1997. Geometrical Diffusion Measures for MRI from Tensor Basis Analysis. Proc. 5th Annu. Meet. ISMRM 1742.
Westlye, L.T., Walhovd, K.B., Dale, A.M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., Grydeland, H., Tamnes, C.K., Østby, Y., Fjell, A.M., 2010. Life-span changes of the human brain white matter: Diffusion tensor imaging (DTI) and volumetry. Cereb. Cortex 20, 2055–2068. https://doi.org/10.1093/cercor/bhp280
Yeh, F.C., 2020. Shape analysis of the human association pathways. Neuroimage 223. https://doi.org/10.1016/j.neuroimage.2020.117329
Yeh, F.C., Verstynen, T.D., Wang, Y., Fernández-Miranda, J.C., Tseng, W.Y.I., 2013. Deterministic diffusion fiber tracking improved by quantitative anisotropy. PLoS One 8, 1–16.

https://doi.org/10.1371/journal.pone.0080713
Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., Jbabdi, S., Gollub, R., Fischl, B., 2011. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. Front. Neuroinform. 5, 1–12. https://doi.org/10.3389/fninf.2011.00023
Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C., 2012. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61, 1000–1016. https://doi.org/10.1016/j.neuroimage.2012.03.072