

## A comprehensive reliability assessment of quantitative diffusion tensor tractography

Jun Yi Wang<sup>a</sup>, Hervé Abdi<sup>a,c</sup>, Khamid Bakhadirov<sup>a</sup>, Ramon Diaz-Arrastia<sup>b</sup>, Michael D. Devous, Sr.<sup>a,c,\*</sup>

<sup>a</sup> School of Brain and Behavioral Sciences, University of Texas at Dallas, Dallas, TX, USA

<sup>b</sup> Department of Neurology, University of Texas Southwestern Medical Center, Dallas, TX, USA

<sup>c</sup> Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX, USA

### ARTICLE INFO

#### Article history:

Received 18 August 2011

Revised 29 November 2011

Accepted 21 December 2011

Available online 29 December 2011

#### Keywords:

Reliability

Diffusion tensor imaging

Tractography

Variability

White matter

Fiber tracts

### ABSTRACT

Diffusion tensor tractography is increasingly used to examine structural connectivity in the brain in various conditions, but its test–retest reliability is understudied. The main purposes of this study were to evaluate 1) the reliability of quantitative measurements of diffusion tensor tractography and 2) the effect on reliability of the number of gradient sampling directions and scan repetition. Images were acquired from ten healthy participants. Ten fiber regions of nine major fiber tracts were reconstructed and quantified using six fiber variables. Intra- and inter-session reliabilities were estimated using intraclass correlation coefficient (ICC) and coefficient of variation (CV), and were compared to pinpoint major error sources. Additional pairwise comparisons were made between the reliability of images with 30 directions and NEX 2 (DTI30-2), 30 directions and NEX 1 (DTI30-1), and 15 directions and NEX 2 (DTI15-2) to determine whether increasing gradient directions and scan repetition improved reliability. Of the 60 tractography measurements, 43 showed inter-session  $CV \leq 10\%$ ,  $ICC \geq .70$ , or both for DTI30-2, 40 measurements for DTI30-1, and 37 for DTI15-2. Most of the reliable measurements were associated with the tracts corpus callosum, cingulum, cerebral peduncular fibers, uncinate fasciculus, and arcuate fasciculus. These reliable measurements included fractional anisotropy (FA) and mean diffusivity of all 10 fiber regions. Inter-session reliability was significantly worse than intra-session reliability for FA, mean length, and tract volume measurements from DTI15-2, indicating that the combination of MRI signal variation and physiological noise/change over time was the major error source for this sequence. Increasing the number of gradient directions from 15 to 30 while controlling the scan time, significantly affected values for all six variables and reduced inter-session variability for mean length and tract volume measurements. Additionally, while increasing scan repetition from 1 to 2 had no significant effect on the reliability for DTI with 30 directions, this significantly reduced the upward bias in FA values from all 10 fiber regions and fiber count, mean length, and tract volume measurements from 5 to 7 fiber regions. In conclusion, diffusion tensor tractography provided many measurements with high test–retest reliability across different fiber variables and various fiber tracts even for images with 15 directions (NEX 2). Increasing the number of gradient directions from 15 to 30 with equivalent scan time reduced variability whereas increasing repetition from 1 to 2 for 30-direction DTI improved the accuracy of tractography measurements.

© 2011 Elsevier Inc. All rights reserved.

### Introduction

Many brain functions are mediated by parallel distributed neural networks (Catani and Ffytche, 2005; McClelland and Rogers, 2003). For instance, language (Ben-Shachar et al., 2007; Duffau, 2008), attention, memory, working memory, consciousness (Naghavi and Nyberg, 2005), and executive function (Aron et al., 2007) all rely on coordinated activation of distinct cortical areas and efficient information transmission by long-range fiber pathways connecting these cortical areas. Recently, the non-invasive investigation of structural connectivity in the brain has been made possible by diffusion tensor imaging (DTI), a new MRI technique that estimates spatial distribution of water diffusion

and provides information on the predominant diffusion direction and diffusion directionality (Basser et al., 1994). Based on this information, white matter fiber tracts can be reconstructed *in vivo* by DTI tractography for visual examination of brain structural connectivity (Behrens et al., 2003; Conturo et al., 1999; Mori et al., 1999). DTI tractography also provides tract-derived measurements such as fractional anisotropy (FA)—which reflects structural alignment and mean diffusivity (MD)—which represents packing density (Basser and Pierpaoli, 1996; Beaulieu, 2002). DTI tractography is now a key area of research applied to a wide range of conditions.

As DTI tractography becomes a standard imaging tool, it is essential to evaluate its test–retest reliability. Recent studies have identified factors that affect DTI reliability through elevating variability of the measurements. These factors include choice of gradient sampling scheme (Gao et al., 2009; Jones et al., 1999; Skare et al., 2000), number of gradient sampling directions (Jones, 2004; Landman et al., 2007), diffusion

\* Corresponding author at: Department of Radiology, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, Texas 75390-9061, USA. Fax: +1 214 648 5641.  
E-mail address: [Michael.Devous@utsouthwestern.edu](mailto:Michael.Devous@utsouthwestern.edu) (M.D. Devous).

sensitivity  $b$ -value (Bisdas et al., 2008; Gao et al., 2009), number of scan repetitions (Farrell et al., 2007), image registration accuracy (Marenco et al., 2006; Pfefferbaum et al., 2003), rater reliability (Ciccarelli et al., 2003; Heiervang et al., 2006; Ozturk et al., 2008), fiber tracking method (Heiervang et al., 2006; Huang et al., 2004), and scanner model (Danielian et al., 2010; Pfefferbaum et al., 2003; Vollmar et al., 2010). Other factors affecting variability include imaging parameters, MRI signal variation, and subject physiological noise/change, motion, and positioning (Farrell et al., 2007; Gao et al., 2009; Landman et al., 2007; Marenco et al., 2006; Pfefferbaum et al., 2003). Some sources of variability are controllable by keeping consistency across all image acquisitions. Other sources that are not completely controllable include MRI signal variation, physiological noise/change, subject motion, and rater variability. These uncontrollable error sources are assumed to cause intra-session variability in tractography measurements. The same error sources are assumed to cause intersession variability, although MRI signal variation and physiological noise/change may contribute higher variability in measurement over longer periods of time than within the same day. Therefore, quantifying and comparing intra- and inter-session reliability reveal important error sources for tractography measurements.

Despite the importance of obtaining reliable tractography measurements, only four studies have provided limited information regarding the reliability of tractography measurements (Ciccarelli et al., 2003; Danielian et al., 2010; Heiervang et al., 2006; Vollmar et al., 2010). These studies—which used different fiber tracking algorithms—found that intersession reliability of tractography measurements depended on structure (e.g., the cingulum, corpus callosum) and variable (e.g., tract volume, FA). The structure- and variable-dependency of tractography measurements suggest that the reliability of a fiber measurement from a fiber tract cannot be directly inferred from the reliability of a fiber measurement from another fiber tract nor from the reliability of another fiber measurement from the same fiber tract. So, to fully understand the reliability of tractography measurements, intra-session reliability and the reliability of measurements from *all* major fiber tracts need to be determined.

One approach to improving the reliability of DTI measurements is to achieve high signal-to-noise ratio (SNR) through optimizing the DTI acquisition scheme. Studies have tested the effect of increasing the numbers of gradient directions and scan repetition on reliability. In adult brains where the diffusion directions in the white matter are least likely to be equally distributed, a large number of uniformly distributed gradient directions have been shown to reduce noise-induced rotational variance caused by the misalignment between the principal diffusion direction of the tensor and the gradient direction (Gao et al., 2009; Skare et al., 2000). Jones (2004) demonstrated through Monte-Carlo computer simulations that at least 30 gradient encoding directions were required to obtain robust diffusion measurements suitable for performing tractography. However, Heiervang et al. (2006) compared measurements from probabilistic fiber tracking performed on DTI with 60 versus 12 gradient directions with incongruent results. Even though the scan time was much shorter for DTI with 12 directions, the intersession variability was similar to the one obtained using 60 directions. This indicates that additional studies are needed to clarify the effect of number of gradient sampling directions on DTI reliability. Regarding the effect of scan repetition, Farrell et al. (2007) showed that scan repetition led to increased SNR, which reduced variability in voxel- and regions-of-interest (ROI)-based DTI measurements. To date, there are no available tractography data that determine whether scan repetition is an effective means of improving reliability.

The current study aimed to provide a comprehensive assessment of the reliability of quantitative DTI tractography for examining structural connectivity. Specifically, we reconstructed nine major white matter fiber tracts: the corpus callosum, cingulum, fornix, angular bundle (a.k.a. descending cingulum or cingulum at the temporal lobe), cerebral peduncular fibers, uncinate fasciculus, inferior longitudinal fasciculus (ILF), inferior frontooccipital fasciculus (IFO), and arcuate fasciculus.

These fiber tracts were quantified using six fiber variables: FA, MD, fiber count, mean length, tract volume, and fiber density. Intra- and inter-session reliabilities were estimated to identify reliable tractography measurements, and compared to evaluate the effect of MRI signal variation and subject physiological change over time. Additionally, the intra- and intersession reliabilities were compared between DTI acquired using both 30 and 15 gradient directions and with both NEX 1 (for 30-direction DTI only) and 2 to assess whether increasing the number of gradient directions and scan repetition improved the test–retest reliability of tractography measurements.

## Materials and methods

### Research participants and imaging sessions

Ten healthy subjects participated in the reliability study. These subjects had an average age of 28 years (standard deviation, SD, 5.5 years, range 19–38 years, 60% males). Inclusion criteria required that the subjects had no neurological impairments or psychiatric disorders and were under 40 years old. All healthy subjects participated in two imaging sessions. During the first session, two images were collected, with one using 30, and the other using 15 gradient directions and NEX 2 (DTI30-2A and DTI15-2A). During the second session about four weeks later (mean 41 days, SD 22 days, mode 30 days, range 23–88 days), four images of NEX 2, with two at 30 and two 15 directions, respectively, were collected (DTI30-2B, DTI30-2C, DTI15-2B, and DTI15-2C). As the images were not averaged at the scanner, we were able to select the first sets of the DTI30-2A, DTI30-2B, and DTI30-2C to form DTI30-1A, DTI30-1B, and DTI30-1C (i.e. NEX = 1) to determine the unique contribution to DTI reliability of increasing the number of gradient directions from 15 to 30. These images with 30 directions and NEX 1 were equivalent to DTI15-2 in terms of scan time and all other remaining imaging parameters.

### Imaging acquisition

A research dedicated Philips Achieva 3T MRI machine with multi-channel SENSE head coil (Best, the Netherlands) was utilized to collect DTI and T1-weighted images for the reliability study. DTI images for both Jones 15 and 30 gradient directions (gradient directions spread out evenly in the space as it was described in Jones, 2004) were obtained using a single-shot spin-echo, echo-planar imaging sequence in 65 axial slices of 2.2 mm thickness (no gap) with 224 mm FOV,  $112 \times 112$  acquisition matrix,  $256 \times 256$  image matrix, TR of 5630 ms, TE of 51 ms,  $90^\circ$  flip angle, and 2 NEX. The diffusion sensitizing gradients were applied at a  $b$ -value  $1000 \text{ s/mm}^2$ . One additional image with minimum diffusion weighting was also obtained. The acquisition times were ~8 min for DTI images with 30 directions and ~4 min for images with 15 directions. The image resolution was  $0.88 \times 0.88 \times 2.2 \text{ mm}^3$ . High-resolution T1-weighted 3D-Turbo Field Echo anatomical images were acquired in 160 axial slices of 1.0 mm thickness (no gap) with FOV 256 mm,  $256 \times 256$  matrix, TR of 8.1 ms, TE of 3.7 ms, and  $12^\circ$  flip angle.

### Imaging preprocessing

The image preprocessing for DTI comprised brain extraction, which was carried out using the BET function (Smith, 2002) from FSL ([www.fmrib.ox.ac.uk/fsl/](http://www.fmrib.ox.ac.uk/fsl/), University of Oxford, UK), and eddy current and motion corrections, which were accomplished using the Automatic Image Registration program (Woods et al., 1998) implemented in DTI Studio ([www.mristudio.org](http://www.mristudio.org), Johns Hopkins Medical Institute, Baltimore, MD). For DTI30-2 and DTI15-2, the two sets of raw images were co-registered to one minimum diffusion weighted image using a 12-mode affine transformation. For DTI30-1, all diffusion weighted images were co-registered to a minimum diffusion weighted image using the same method. During the transformation process, the gradient tables were

modified according to the generated transformation matrix and subsequently utilized in tensor calculation and diffusion map generation.

### DTI tractography

Fiber tracking was performed in DTI Studio using the Fiber Assignment by Continuous Tracking algorithm (Mori et al., 1999). The FA seed threshold was set to 0.25, termination threshold 0.20, and angle threshold 60°. Fiber tracking adopted the multiple ROI approach which has been shown to have the potential of achieving higher inter-rater reliability in fiber quantification (Huang et al., 2004). ROI drawings were placed along the trajectory of the fiber tracts to extract fibers using ROI logic operations OR, AND, NOT, and CUT. Fiber tracking methods were carefully devised to achieve high accuracy and high inter-rater reliability. The ROIs slices were strategically determined based on anatomical knowledge (Augustinack et al., 2010; Crosby et al., 1962; Nolte, 2002) and referencing previous fiber tracking methods (Catani et al., 2002; Concha et al., 2005; Huang et al., 2005; Mori and van Zijl, 2002; Wakana et al., 2004, 2007).

Nine fiber tracts were chosen for reconstruction due to their clinical relevance and traceability under current DTI image resolution. As described in Wang et al. (2011), 28 fiber regions could be reconstructed from the 9 fiber tracts following the fiber tracking methods developed in the authors' laboratory. Appendix Fig. 1 depicts the ROI positions for reconstructing the 9 fiber tracts. In the current study, only one fiber region from each of the nine fiber tracts was chosen to simplify the analysis with the exception of the fornix. Both fornix body and crus were included in the analysis to account for the differences in inter-rater reliability of these two fiber regions (see next section) that might have affected the test–retest reliability. For fiber tracts containing two sides, the selections of the side were alternated left and right. The cerebral peduncular fibers contain five fiber regions. The superior frontal projections (CPSf) were chosen because of the prominence of the cortical spinal tract contained in this fiber region.

### Tractography measurements and inter-rater reliability

The 10 fiber regions were quantified using six fiber variables: FA, MD, fiber count (i.e. the number of DTI streamlines extracted for a fiber tract), mean length (i.e. the average length in mm for all streamlines belonging to a fiber tract), tract volume (number of voxels occupied by all streamlines for a particular fiber tract), and fiber density (fiber count/voxel). To correct for individual differences in head size, fiber count and tract volume were normalized by intracranial volume, which was calculated from T1-weighted images using the SIENAX program in FSL (Smith et al., 2002, 2004). The inter-rater reliability of the tractography measurements has been tested as a part of previous projects in our lab (Wang et al., 2008, 2011) using intraclass correlation coefficient (ICC) (McGraw and Wong, 1996). The ICC was  $\geq .90$  for all tractography measurements from the 10 fiber regions except for the tract volume of left fornix crus of which the ICC was .87.

### Statistical analysis

As one set each of DTI with 30 directions (NEX 2) and DTI with 15 directions (NEX 2) were acquired during the first session (DTI30-2A and DTI15-2A) while two sets of each were acquired during the second session (DTI30-2B, C and DTI15-2B, C), intra-session reliability was estimated by contrasting the images acquired within the second session (i.e. DTI30-2B vs. DTI30-2C), and inter-session reliability was quantified using the averaged values which contrasted the images acquired in the first session with the two images acquired in the second session (i.e. the average of DTI30-2A vs. DTI30-2B and DTI30-2A vs. DTI30-2C). The reliability indices were the intra class correlation coefficient (ICC) and the coefficient of variation (CV). ICC, which

estimates the amount of agreement in absolute value within the repeated measurements, is computed as:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where  $MS_R$  = mean square for rows of observations,  $MS_C$  = mean square for columns of subjects,  $MS_E$  = mean square error,  $k$  = number of observations, and  $n$  = number of subjects. ICC was computed using the MATLAB toolbox Intraclass correlation coefficients created by Arash Salarian ([www.mathworks.com/matlabcentral/fileexchange/22099](http://www.mathworks.com/matlabcentral/fileexchange/22099)). CV, which is defined as the ratio, expressed in percentage, of the standard deviation of a set of measurements by its mean (Abdi, 2010; Armitage, 1971; Bland and Altman, 1986), is computed as:

$$CV = \frac{SD}{\text{Mean}}.$$

Both between- and within-subject variability ( $CV_{\text{between}}$  and  $CV_{\text{within}}$ , respectively) were quantified, with  $CV_{\text{within}}$  being the SD of the differences in repeated measures divided by the mean. For biological measurements from MRI,  $CV_{\text{within}} \leq 10\%$  and  $ICC \geq .70$  are considered as acceptable, while an  $ICC \geq .80$  is considered ideal (Marengo et al., 2006). Therefore, tractography measurements with intersession  $CV_{\text{within}} \leq 10\%$ ,  $ICC \geq .70$ , or both, were classified as reliable. To compare different forms of reliability, we conducted paired *t*-tests to determine whether the changes in the ICC and  $CV_{\text{within}}$  for individual fiber variables (e.g., FA, MD) reached statistical significance. To correct for familywise errors, we used a 5% false discovery rate (FDR) (Benjamini and Hochberg, 1995) as computed by Thomas Nichols' matlab function, FDR.

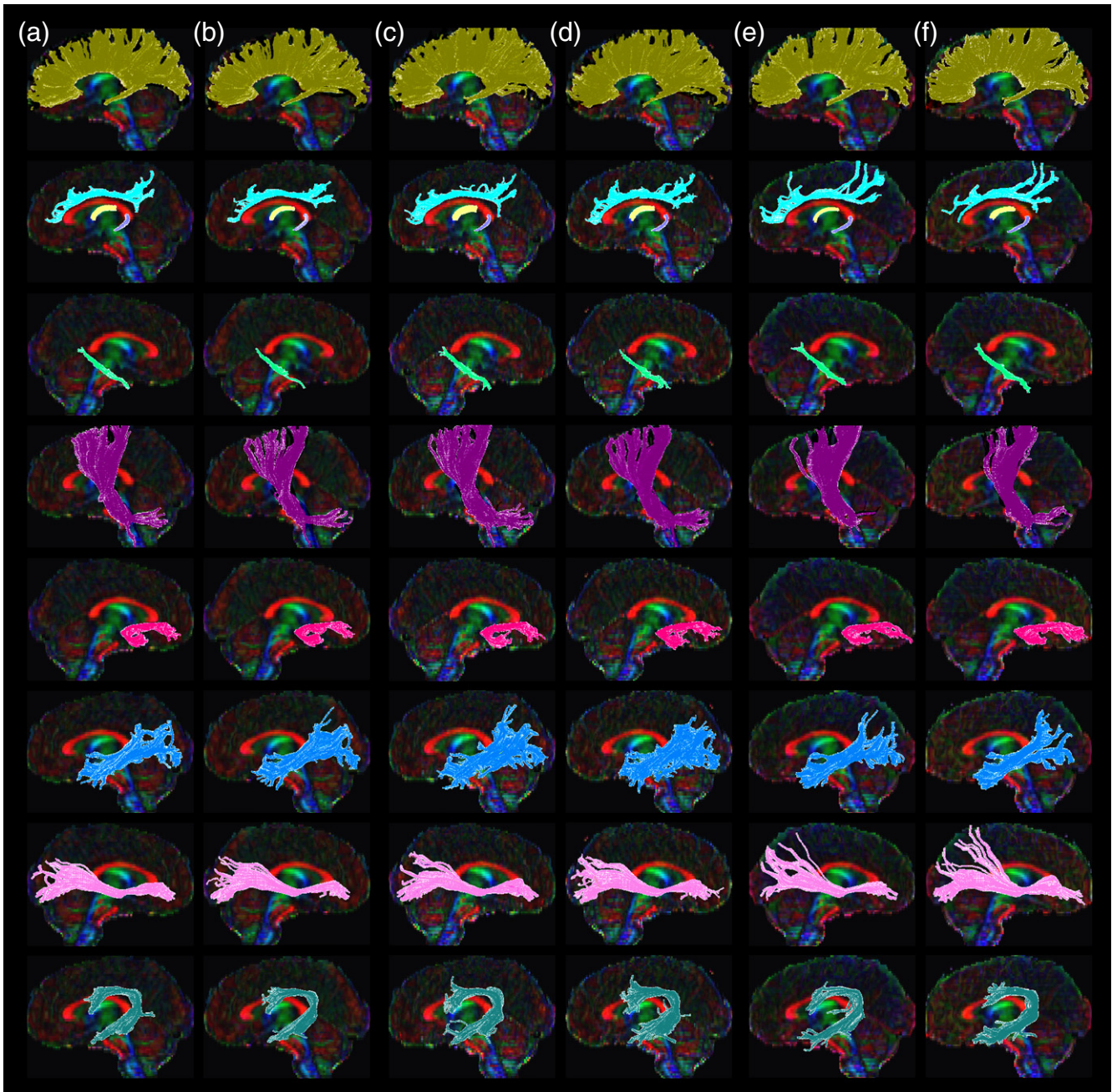
## Results

### Tractography results

Fig. 1 shows the representative fiber tracking results from one of the participants. Among the 10 fiber regions, left arcuate fasciculus was the only one that could not be reconstructed successfully from all 10 participants. This fiber region was missing from DTI30-2B, DTI30-1B, DTI30-2C, DTI30-1C, and DTI15-2B for a single subject even when the FA threshold for fiber tracking was lowered to .15. Consequently, left arcuate fasciculus for this subject was excluded from subsequent analyses. In addition, right IFO could not be reconstructed following the fiber tracking method for DTI30-2B, DTI30-1B, DTI30-2C, and DTI30-1C for one participant and DTI15-2A for another participant. Therefore, the ROI slices in the frontal lobe were moved posteriorly to extract the IFO from these scans.

### Descriptive statistics

Group means and standard deviations of the tractography measurements from the 10 fiber regions were averaged across the three sets of images for the same sequence (i.e. DTI30-2A, B, and C). Paired *t* tests were conducted to find group mean differences between the three sequences (Table 1). In general, DTI30-1 produced higher FA, mean length, fiber count, and tract volume measurements compared to DTI30-2, with FA showing the most consistent results, significantly higher for all 10 fiber regions. The remaining fiber variables showed higher values in DTI30-1 compared to DTI30-2 in 5–7 regions. DTI30-1 also produced higher FA, fiber count, mean length, and tract volume measurements and lower MD measurements compared to DTI15-2, with FA and MD showing the most consistent results in all fiber regions and the remaining variables showing consistency in 5–8 fiber regions. When comparing DTI30-2 to DTI15-2, fiber count, mean length, tract volume, and fiber density were significantly higher in 3–5 fiber regions and MD was lower in all fiber regions except fornix crus.



**Fig. 1.** Representative tractography results from a research participant. The fiber tracts were reconstructed from (a) DTI30-2A, (b) DTI30-2B, (c) DTI30-1A, (d) DTI30-1B, (e) DTI15-2A, and (f) DTI15-2B. The graph from top-down shows the corpus callosum (dark yellow), cingulum (light blue), fornix body (light yellow), fornix crus (light purple), angular bundle (light green), CPsf (purple), uncinate fasciculus (pink), ILF (blue), IFO (light pink), and arcuate fasciculus (dark teal).

### Inter-subject variability

The average  $CV_{\text{between}}$  was computed over the three sets of images for DTI30-2, DTI30-1, and DTI15-2. Among the six fiber variables, FA and MD generally had lower  $CV_{\text{between}}$  (2.4–7.6% for FA and 1.7–9.9% for MD) compared to mean length (4.3–22.4%), which in turn had lower  $CV_{\text{between}}$  compared to tract volume (9.5–42.9%), fiber density (11.8–41.0%), and fiber count (10.9–63.9%). Pairwise comparisons between DTI30-2, DTI30-1, and DTI15-2 revealed that DTI30-2 had significantly higher  $CV_{\text{between}}$  in fiber volume compared to DTI30-1, and lower  $CV_{\text{between}}$  in mean length and fiber density compared to DTI15-2. Additionally, DTI30-1 showed lower  $CV_{\text{between}}$  in mean length, tract

volume, fiber density compared to DTI15-2 (paired  $|t|(9) = 3.00\text{--}4.28$ ,  $p \leq .015$ , FDR  $q = 0.05$ ). Among the 10 fiber regions, the corpus callosum had relatively low  $CV_{\text{between}}$  for fiber count, tract volume and fiber density (9.5–17.2% vs. 9.3–63.9%), and fornix body and crus had relatively high  $CV_{\text{between}}$  for FA and MD (6.3–9.9% vs. 1.9–6.5%).

### Test-retest reliability

Intra- and inter-session  $CV_{\text{within}}$  and ICC were computed for all tractography measurements from DTI30-2, DTI30-1, and DTI15-2 (Table 2). Similar to  $CV_{\text{between}}$ , FA and MD had the lowest intra- and inter-session  $CV_{\text{within}}$ , which ranged from 0.7 to 8.1% for FA intra-session  $CV_{\text{within}}$ ,

**Table 1**  
Group mean and standard deviation of tractography measurements.

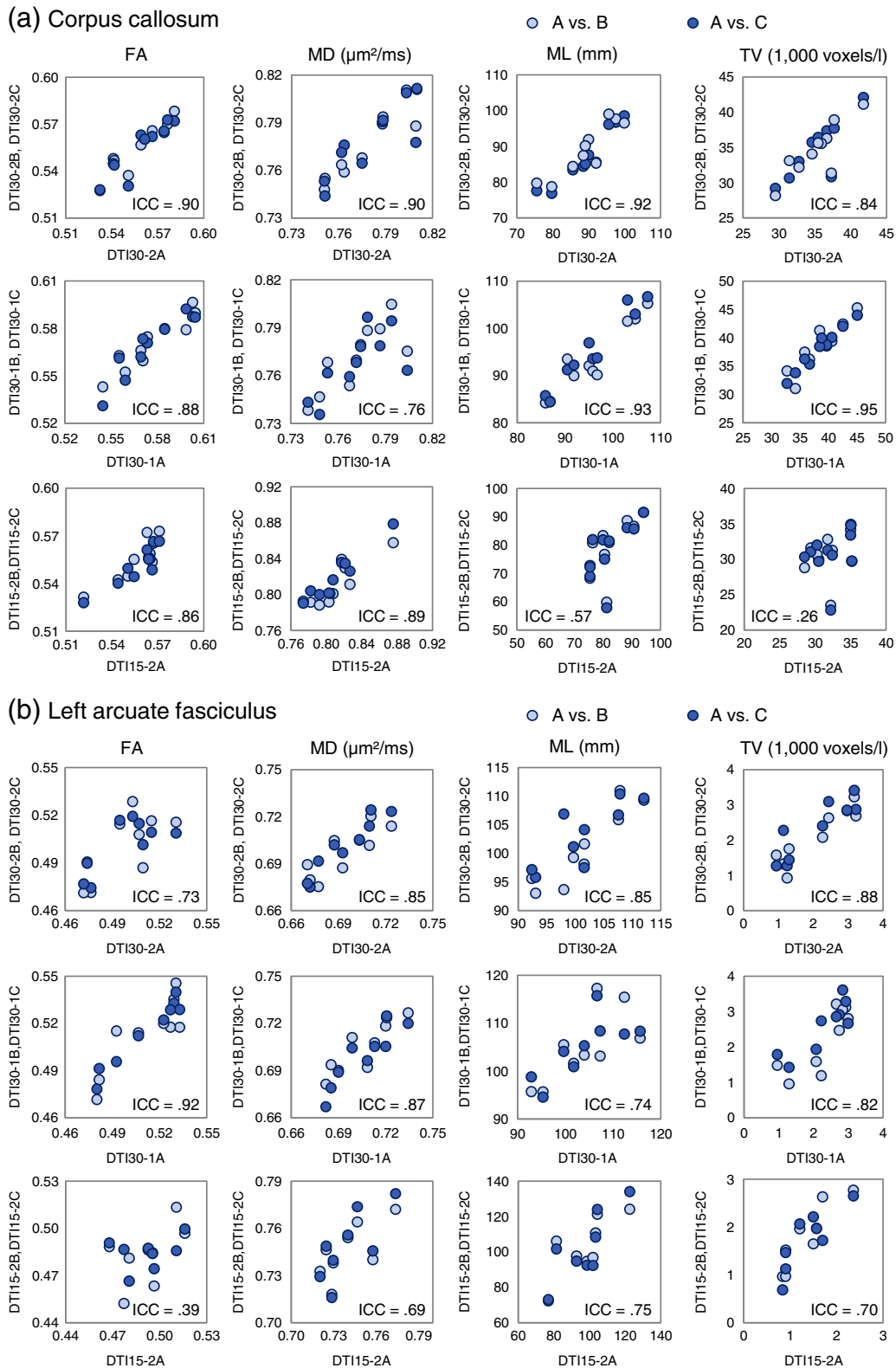
Measurements	DTI 30, NEX 2		DTI 30, NEX 1		DTI 15, NEX 2		DTI 30, NEX 2 Vs. DTI 30 NEX 1		DTI 30, NEX 1 Vs. DTI 15, NEX 2		DTI 30, NEX 2 Vs. DTI 15, NEX 2	
	Mean	SD	Mean	SD	Mean	SD	Paired t	p	Paired t	p	Paired t	p
	<i>Corpus callosum</i>											
FA	0.56	0.016	0.57	0.018	0.55	0.013	<b>-8.97</b>	<b>&lt;0.001</b>	<b>4.64</b>	<b>0.001</b>	0.19	0.550
MD (μm <sup>2</sup> /ms)	0.78	0.022	0.77	0.019	0.81	0.025	<b>4.44</b>	<b>0.001</b>	<b>-7.55</b>	<b>&lt;0.001</b>	<b>-5.88</b>	<b>&lt;0.001</b>
FC	16.4	2.8	16.9	2.8	14.7	1.4	<b>-4.60</b>	<b>0.001</b>	<b>3.40</b>	<b>0.008</b>	<b>2.63</b>	<b>0.027</b>
ML (mm)	88.6	7.4	94.9	7.4	79.9	8.1	<b>-13.24</b>	<b>&lt;0.001</b>	<b>5.85</b>	<b>&lt;0.001</b>	<b>3.31</b>	<b>0.009</b>
TV (voxels/l)	34.9	3.6	38.4	3.8	31.1	2.5	<b>-11.64</b>	<b>&lt;0.001</b>	<b>7.54</b>	<b>&lt;0.001</b>	<b>4.00</b>	<b>0.003</b>
FD (FC/voxel)	47.0	6.7	46.4	6.7	42.9	4.7	<b>2.82</b>	<b>0.020</b>	2.07	0.068	2.44	0.037
<i>Cingulum (left)</i>												
FA	0.50	0.017	0.53	0.016	0.51	0.020	<b>-12.53</b>	<b>&lt;0.001</b>	<b>3.44</b>	<b>0.008</b>	-0.32	0.779
MD (μm <sup>2</sup> /ms)	0.69	0.015	0.70	0.014	0.73	0.014	<b>-4.13</b>	<b>0.003</b>	<b>-9.22</b>	<b>&lt;0.001</b>	<b>-8.55</b>	<b>&lt;0.001</b>
FC	686	196	686	208	502	241	0.03	0.976	<b>2.89</b>	<b>0.018</b>	<b>2.92</b>	<b>0.017</b>
ML (mm)	83.5	9.8	89.0	9.0	73.8	12.1	<b>-5.70</b>	<b>&lt;0.001</b>	<b>4.41</b>	<b>0.002</b>	2.50	0.034
TV (voxels/l)	2251	551	2608	537	2045	742	<b>-7.51</b>	<b>&lt;0.001</b>	<b>3.04</b>	<b>0.014</b>	1.06	0.319
FD (FC/voxel)	30	4.6	27	4.8	21	7.4	<b>3.71</b>	<b>0.005</b>	<b>2.99</b>	<b>0.015</b>	<b>4.51</b>	<b>0.001</b>
<i>Angular bundle (right)</i>												
FA	0.41	0.016	0.44	0.021	0.41	0.025	<b>-12.50</b>	<b>&lt;0.001</b>	<b>4.99</b>	<b>0.001</b>	-0.45	0.677
MD (μm <sup>2</sup> /ms)	0.72	0.017	0.72	0.014	0.75	0.021	-1.39	0.208	<b>-4.82</b>	<b>0.001</b>	<b>-5.21</b>	<b>0.001</b>
FC	150	53	165	58	147	43	<b>-3.44</b>	<b>0.007</b>	1.01	0.339	0.19	0.855
ML (mm)	34.8	3.5	36.2	3.7	35.0	5.7	-2.08	0.067	0.61	0.556	-0.14	0.892
TV (voxels/l)	582	114	677	132	562	139	<b>-6.50</b>	<b>&lt;0.001</b>	2.29	0.048	0.41	0.695
FD (FC/voxel)	11	2.9	11	2.8	12	2.3	0.94	0.377	-0.63	0.545	-0.19	0.852
<i>Fornix body</i>												
FA	0.53	0.031	0.59	0.030	0.52	0.030	<b>-9.21</b>	<b>&lt;0.001</b>	<b>5.76</b>	<b>&lt;0.001</b>	0.98	0.362
MD (μm <sup>2</sup> /ms)	1.18	0.089	1.11	0.107	1.26	0.103	<b>5.93</b>	<b>&lt;0.001</b>	<b>-10.53</b>	<b>&lt;0.001</b>	<b>-8.19</b>	<b>&lt;0.001</b>
FC	334	121	300	81	281	113	1.75	0.115	0.81	0.440	1.52	0.162
ML (mm)	27.8	4.5	29.9	5.1	28.2	3.7	<b>-5.00</b>	<b>0.001</b>	<b>2.55</b>	<b>0.031</b>	-0.81	0.441
TV (voxels/l)	379	75	357	52	377	84	1.39	0.197	-0.58	0.574	0.06	0.954
FD (FC/voxel)	27	5.9	29	5.9	24	6.3	-0.87	0.409	<b>3.21</b>	<b>0.011</b>	1.95	0.083
<i>Fornix crus (left)</i>												
FA	0.53	0.036	0.56	0.027	0.50	0.036	<b>-5.44</b>	<b>0.001</b>	<b>5.49</b>	<b>&lt;0.001</b>	<b>2.81</b>	<b>0.015</b>
MD (μm <sup>2</sup> /ms)	0.89	0.075	0.88	0.058	0.96	0.069	1.26	0.244	<b>-3.30</b>	<b>0.009</b>	-2.41	0.039
FC	145	38	148	50	141	46	-0.16	0.875	0.43	0.680	0.25	0.807
ML (mm)	27.4	2.8	28.5	3.2	28.1	4.3	<b>-3.12</b>	<b>0.012</b>	0.47	0.651	-1.04	0.327
TV (voxels/l)	200	33	214	29	179	27	-1.23	0.248	<b>2.71</b>	<b>0.024</b>	<b>2.61</b>	<b>0.028</b>
FD (FC/voxel)	20	4.4	19	4.2	22	5.9	0.55	0.594	-1.32	0.220	-0.81	0.439
<i>Cerebral peduncular projections to superior frontal lobes</i>												
FA	0.53	0.013	0.55	0.012	0.54	0.014	<b>-14.34</b>	<b>&lt;0.001</b>	<b>4.52</b>	<b>0.001</b>	<b>-3.47</b>	<b>0.001</b>
MD (μm <sup>2</sup> /ms)	0.69	0.010	0.70	0.011	0.73	0.009	-1.36	0.134	<b>-8.15</b>	<b>&lt;0.001</b>	<b>-8.92</b>	<b>&lt;0.001</b>
FC	2246	830	2701	737	1871	949	<b>-3.69</b>	<b>0.005</b>	<b>3.19</b>	<b>0.011</b>	1.50	0.167
ML (mm)	120.9	5.8	126.3	5.1	114.0	8.9	<b>-3.64</b>	<b>0.005</b>	<b>5.56</b>	<b>&lt;0.001</b>	<b>2.63</b>	<b>0.027</b>
TV (voxels/l)	10,071	2155	13,185	1877	7926	2392	<b>-7.88</b>	<b>&lt;0.001</b>	<b>7.06</b>	<b>&lt;0.001</b>	<b>2.87</b>	<b>0.019</b>
FD (FC/voxel)	20	4.5	20	3.3	21	6.0	0.19	0.854	-0.53	0.607	-0.49	0.639
<i>Uncinate fasciculus (right)</i>												
FA	0.42	0.014	0.45	0.015	0.42	0.019	<b>-13.54</b>	<b>&lt;0.001</b>	<b>7.55</b>	<b>&lt;0.001</b>	0.80	0.446
MD (μm <sup>2</sup> /ms)	0.72	0.014	0.72	0.014	0.77	0.015	-1.06	0.372	<b>-15.78</b>	<b>&lt;0.001</b>	<b>-15.38</b>	<b>&lt;0.001</b>
FC	628	277	604	237	381	194	0.64	0.536	<b>2.79</b>	<b>0.021</b>	<b>3.12</b>	<b>0.012</b>
ML (mm)	76	12	76	13	67	14	0.03	0.979	2.34	0.044	2.29	0.048
TV (voxels/l)	2035	596	2240	551	1445	502	<b>-3.40</b>	<b>0.008</b>	<b>4.18</b>	<b>0.002</b>	<b>3.12</b>	<b>0.012</b>
FD (FC/voxel)	25	8.4	22	6.9	18	6.8	2.48	0.035	1.89	0.091	<b>3.10</b>	<b>0.013</b>
<i>Inferior longitudinal fasciculus (left)</i>												
FA	0.49	0.015	0.50	0.014	0.47	0.017	<b>-5.80</b>	<b>&lt;0.001</b>	<b>5.39</b>	<b>&lt;0.001</b>	<b>2.82</b>	<b>0.021</b>
MD (μm <sup>2</sup> /ms)	0.74	0.015	0.74	0.015	0.81	0.047	<b>-2.30</b>	<b>0.030</b>	<b>-5.94</b>	<b>&lt;0.001</b>	<b>-6.36</b>	<b>&lt;0.001</b>
FC	1782	444	2236	442	1313	364	<b>-4.59</b>	<b>0.001</b>	<b>7.70</b>	<b>&lt;0.001</b>	<b>3.83</b>	<b>0.004</b>
ML (mm)	81.5	6.8	85.0	8.2	79.2	8.0	<b>-4.24</b>	<b>0.002</b>	2.45	0.037	1.03	0.329
TV (voxels/l)	5913	1008	7907	1035	4975	779	<b>-9.26</b>	<b>&lt;0.001</b>	<b>11.21</b>	<b>&lt;0.001</b>	<b>3.38</b>	<b>0.008</b>
FD (FC/voxel)	27	4.1	26	3.4	24	5.0	1.29	0.232	2.11	0.065	2.20	0.055
<i>Inferior frontooccipital fasciculus (right)</i>												
FA	0.48	0.017	0.51	0.017	0.48	0.021	<b>-16.63</b>	<b>&lt;0.001</b>	<b>6.66</b>	<b>&lt;0.001</b>	-0.39	0.681
MD (μm <sup>2</sup> /ms)	0.74	0.021	0.73	0.021	0.76	0.021	2.50	0.034	<b>-5.09</b>	<b>0.001</b>	<b>-3.14</b>	<b>0.012</b>
FC	822	268	946	242	678	330	<b>-4.43</b>	<b>0.002</b>	<b>3.16</b>	<b>0.012</b>	1.99	0.078
ML (mm)	144.7	7.1	148.3	8.5	131.6	12.9	<b>-3.46</b>	<b>0.007</b>	<b>4.68</b>	<b>0.001</b>	<b>3.67</b>	<b>0.005</b>
TV (voxels/l)	4490	829	5545	822	3936	1477	<b>-7.83</b>	<b>&lt;0.001</b>	<b>4.75</b>	<b>0.001</b>	1.87	0.094
FD (FC/voxel)	31	6.5	30	4.0	25	8.6	1.18	0.268	2.38	0.041	<b>3.14</b>	<b>0.012</b>
<i>Arcuate fasciculus (left)</i>												
FA	0.500	0.018	0.513	0.022	0.487	0.012	<b>-4.70</b>	<b>0.002</b>	<b>4.40</b>	<b>0.003</b>	<b>2.90</b>	<b>0.024</b>
MD (μm <sup>2</sup> /ms)	0.698	0.017	0.703	0.018	0.745	0.018	<b>-6.71</b>	<b>&lt;0.001</b>	<b>-8.83</b>	<b>&lt;0.001</b>	<b>-10.16</b>	<b>&lt;0.001</b>
FC	529	346	520	277	340	169	0.52	0.617	2.16	0.067	1.91	0.098
ML (mm)	101.9	6.3	104.6	7.0	101.1	16.0	<b>-3.96</b>	<b>0.006</b>	0.77	0.466	0.22	0.829
TV (voxels/l)	2172	834	2362	782	1640	569	-2.01	0.085	<b>2.70</b>	<b>0.031</b>	1.90	0.099
FD (FC/voxel)	22	8.6	21	5.9	20	6.8	0.77	0.467	0.76	0.472	1.00	0.352

Note: only one fiber region from each fiber tract was chosen for the reliability assessment to simply the analysis except for fornix. **Bold**, significant at 5% FDR (p=.031). Abbreviations: FC, fiber count; FD, fiber density; ML, mean length; TV, tract volume.

**Table 2**  
Test–retest reliability of representative tractography measurements.

Measurements	Intrasection CV			Interseccion CV			Intraseccion ICC			Interseccion ICC		
	DTI30-2	DTI30-1	DTI15-2	DTI30-2	DTI30-1	DTI15-2	DTI30-2	DTI30-1	DTI15-2	DTI30-2	DTI30-1	DTI15-2
<i>Corpus callosum</i>												
FA	<b>0.74</b>	<b>1.48</b>	<b>0.89</b>	<b>1.21</b>	<b>1.88</b>	<b>1.25</b>	<b>0.97</b>	<b>0.90</b>	<b>0.89</b>	<b>0.90</b>	<b>0.88</b>	<b>0.86</b>
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>1.07</b>	<b>1.01</b>	<b>1.01</b>	<b>1.34</b>	<b>1.45</b>	<b>1.40</b>	<b>0.94</b>	<b>0.92</b>	<b>0.90</b>	<b>0.90</b>	<b>0.76</b>	<b>0.89</b>
FC	<b>5.22</b>	<b>7.56</b>	<b>4.59</b>	<b>7.44</b>	<b>5.99</b>	11.28	<b>0.95</b>	<b>0.90</b>	<b>0.93</b>	<b>0.91</b>	<b>0.93</b>	0.39
ML (mm)	<b>2.37</b>	<b>2.28</b>	<b>1.60</b>	<b>3.20</b>	<b>3.22</b>	<b>9.23</b>	<b>0.93</b>	<b>0.93</b>	<b>0.99</b>	<b>0.92</b>	<b>0.93</b>	0.57
TV (voxels/l)	<b>3.65</b>	<b>4.13</b>	<b>3.11</b>	<b>6.07</b>	<b>5.00</b>	10.90	<b>0.95</b>	<b>0.92</b>	<b>0.96</b>	<b>0.84</b>	<b>0.95</b>	0.26
FD (FC/voxel)	<b>5.26</b>	<b>6.71</b>	<b>3.32</b>	<b>5.83</b>	<b>6.88</b>	<b>8.93</b>	<b>0.86</b>	<b>0.84</b>	<b>0.97</b>	<b>0.91</b>	<b>0.92</b>	0.62
<i>Cingulum (left)</i>												
FA	<b>2.07</b>	<b>1.54</b>	<b>1.85</b>	<b>2.26</b>	<b>3.29</b>	<b>2.88</b>	<b>0.77</b>	<b>0.85</b>	<b>0.92</b>	<b>0.76</b>	0.67	<b>0.77</b>
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>1.57</b>	<b>1.17</b>	<b>1.51</b>	<b>2.02</b>	<b>1.83</b>	<b>1.73</b>	<b>0.77</b>	<b>0.89</b>	0.65	0.65	<b>0.71</b>	0.55
FC	14.23	15.12	17.96	22.04	21.63	29.38	<b>0.86</b>	<b>0.88</b>	<b>0.94</b>	<b>0.70</b>	<b>0.76</b>	<b>0.82</b>
ML (mm)	<b>7.48</b>	<b>8.26</b>	11.26	<b>9.50</b>	<b>8.34</b>	18.75	<b>0.84</b>	<b>0.77</b>	<b>0.76</b>	<b>0.73</b>	<b>0.76</b>	0.50
TV (voxels/l)	<b>9.82</b>	10.87	13.56	12.32	14.44	19.74	<b>0.92</b>	<b>0.88</b>	<b>0.91</b>	<b>0.88</b>	<b>0.79</b>	<b>0.83</b>
FD (FC/voxel)	<b>8.09</b>	11.30	12.54	16.03	16.42	29.69	<b>0.83</b>	<b>0.80</b>	<b>0.93</b>	0.50	0.53	<b>0.71</b>
<i>Angular bundle (right)</i>												
FA	<b>3.70</b>	<b>5.21</b>	<b>2.74</b>	<b>3.96</b>	<b>2.97</b>	<b>3.71</b>	<b>0.74</b>	0.59	<b>0.90</b>	0.55	<b>0.75</b>	<b>0.80</b>
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>3.45</b>	<b>3.25</b>	<b>2.32</b>	<b>3.17</b>	<b>3.19</b>	<b>2.99</b>	0.29	0.32	0.64	0.48	0.25	0.58
FC	13.71	47.61	37.50	38.17	31.67	47.98	<b>0.96</b>	0.55	0.52	0.48	0.54	0.28
ML (mm)	<b>8.19</b>	12.60	<b>9.79</b>	10.49	14.07	13.88	<b>0.82</b>	0.47	<b>0.84</b>	0.50	0.37	<b>0.72</b>
TV (voxels/l)	<b>8.73</b>	29.48	30.23	19.62	17.85	30.32	<b>0.93</b>	0.42	0.56	0.54	0.49	0.47
FD (FC/voxel)	15.21	29.93	19.12	28.97	28.19	32.55	<b>0.90</b>	0.62	0.58	0.49	0.48	0.30
<i>Fornix body</i>												
FA	<b>3.58</b>	<b>8.09</b>	<b>3.03</b>	<b>6.52</b>	<b>7.38</b>	<b>5.28</b>	<b>0.85</b>	0.13	<b>0.86</b>	0.57	0.41	0.65
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>2.65</b>	<b>4.54</b>	<b>3.04</b>	<b>3.85</b>	<b>5.09</b>	<b>4.24</b>	<b>0.92</b>	<b>0.88</b>	<b>0.93</b>	<b>0.89</b>	<b>0.84</b>	<b>0.87</b>
FC	22.99	40.37	40.09	26.38	43.50	40.96	<b>0.81</b>	0.39	0.66	<b>0.77</b>	0.28	0.59
ML (mm)	<b>6.80</b>	<b>7.73</b>	<b>6.36</b>	<b>9.01</b>	<b>5.23</b>	14.83	<b>0.91</b>	<b>0.91</b>	<b>0.89</b>	<b>0.87</b>	<b>0.96</b>	0.41
TV (voxels/l)	21.99	17.89	23.19	17.09	21.24	25.10	0.53	0.47	0.61	0.69	0.43	0.51
FD (FC/voxel)	16.00	42.25	33.42	20.87	51.74	33.94	<b>0.74</b>	0.27	0.56	0.57	0.08	0.47
<i>Fornix crus (left)</i>												
FA	<b>5.36</b>	<b>4.58</b>	<b>3.81</b>	<b>6.18</b>	<b>5.50</b>	<b>5.20</b>	<b>0.71</b>	0.65	<b>0.81</b>	0.64	0.47	0.63
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>7.83</b>	<b>6.32</b>	<b>9.71</b>	<b>7.58</b>	<b>3.46</b>	<b>6.36</b>	<b>0.70</b>	0.52	0.48	0.69	<b>0.73</b>	0.68
FC	27.49	44.31	48.95	31.17	60.12	36.14	0.51	0.44	0.35	0.49	0.35	0.51
ML (mm)	11.16	10.34	10.27	10.69	<b>9.52</b>	13.43	0.55	<b>0.73</b>	<b>0.79</b>	0.65	0.55	<b>0.70</b>
TV (voxels/l)	24.34	23.59	22.25	21.91	27.11	28.16	0.47	0.12	0.27	0.32	0.13	0.20
FD (FC/voxel)	20.97	28.07	32.34	27.73	38.48	37.25	<b>0.71</b>	0.32	0.56	0.42	0.53	0.39
<i>Cerebral peduncular projections to superior frontal lobes</i>												
FA	<b>2.21</b>	<b>1.94</b>	<b>1.26</b>	<b>2.92</b>	<b>2.19</b>	<b>3.25</b>	0.64	0.59	<b>0.88</b>	0.51	0.58	0.38
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>1.58</b>	<b>1.62</b>	<b>0.98</b>	<b>1.45</b>	<b>1.61</b>	<b>2.92</b>	0.51	0.49	<b>0.86</b>	0.57	0.53	−0.16
FC	23.06	19.08	24.96	34.66	28.62	31.35	<b>0.84</b>	<b>0.72</b>	<b>0.88</b>	0.63	0.57	<b>0.81</b>
ML (mm)	<b>2.49</b>	<b>2.95</b>	<b>6.45</b>	<b>4.17</b>	<b>4.50</b>	10.07	<b>0.86</b>	<b>0.81</b>	<b>0.81</b>	0.68	0.65	0.38
TV (voxels/l)	<b>8.61</b>	11.36	12.10	17.61	16.68	25.33	<b>0.93</b>	0.53	<b>0.92</b>	<b>0.71</b>	0.56	<b>0.73</b>
FD (FC/voxel)	20.08	<b>9.26</b>	15.63	18.86	14.66	19.06	0.67	<b>0.86</b>	<b>0.89</b>	<b>0.72</b>	<b>0.73</b>	<b>0.80</b>
<i>Uncinate fasciculus (right)</i>												
FA	<b>2.28</b>	<b>2.60</b>	<b>2.65</b>	<b>3.06</b>	<b>3.21</b>	<b>5.28</b>	<b>0.86</b>	<b>0.81</b>	<b>0.87</b>	0.63	0.66	0.53
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>0.49</b>	<b>1.11</b>	<b>1.18</b>	<b>1.14</b>	<b>1.65</b>	<b>1.83</b>	<b>0.92</b>	<b>0.86</b>	<b>0.85</b>	<b>0.84</b>	<b>0.77</b>	0.58
FC	16.60	21.13	31.13	29.42	30.58	38.03	<b>0.95</b>	<b>0.87</b>	<b>0.86</b>	<b>0.80</b>	<b>0.71</b>	<b>0.76</b>
ML (mm)	<b>3.95</b>	<b>6.46</b>	10.81	<b>8.04</b>	<b>7.33</b>	12.91	<b>0.97</b>	<b>0.93</b>	<b>0.88</b>	<b>0.89</b>	<b>0.93</b>	<b>0.85</b>
TV (voxels/l)	12.34	10.02	17.78	16.12	12.45	26.39	<b>0.93</b>	<b>0.93</b>	<b>0.88</b>	<b>0.86</b>	<b>0.82</b>	<b>0.75</b>
FD (FC/voxel)	11.74	16.90	27.61	16.42	22.82	27.97	<b>0.95</b>	<b>0.87</b>	<b>0.81</b>	<b>0.89</b>	<b>0.74</b>	<b>0.72</b>
<i>Inferior longitudinal fasciculus (left)</i>												
FA	<b>2.62</b>	<b>2.01</b>	<b>2.20</b>	<b>2.94</b>	<b>3.32</b>	<b>3.41</b>	<b>0.71</b>	<b>0.78</b>	<b>0.81</b>	0.64	0.52	0.65
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>1.11</b>	<b>0.89</b>	<b>0.65</b>	<b>1.60</b>	<b>1.23</b>	<b>5.72</b>	<b>0.90</b>	<b>0.89</b>	<b>0.99</b>	0.69	<b>0.70</b>	0.46
FC	17.13	16.26	20.74	21.55	23.32	26.83	<b>0.76</b>	<b>0.78</b>	<b>0.71</b>	0.69	0.46	0.58
ML (mm)	<b>5.13</b>	<b>4.43</b>	<b>3.36</b>	<b>4.50</b>	<b>5.46</b>	<b>7.69</b>	<b>0.85</b>	<b>0.88</b>	<b>0.95</b>	<b>0.87</b>	<b>0.84</b>	<b>0.78</b>
TV (voxels/l)	<b>7.33</b>	11.79	15.62	11.44	14.02	23.84	<b>0.90</b>	<b>0.75</b>	<b>0.70</b>	<b>0.80</b>	0.44	0.21
FD (FC/voxel)	13.63	10.22	<b>9.87</b>	16.28	17.86	33.46	0.65	<b>0.76</b>	<b>0.93</b>	0.54	0.49	0.21
<i>Inferior frontooccipital fasciculus (right)</i>												
FA	<b>3.51</b>	<b>4.01</b>	<b>2.35</b>	<b>3.60</b>	<b>4.37</b>	<b>6.96</b>	0.65	0.47	<b>0.90</b>	0.58	0.47	0.25
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>1.08</b>	<b>1.49</b>	<b>1.09</b>	<b>1.72</b>	<b>1.97</b>	<b>3.26</b>	<b>0.93</b>	<b>0.89</b>	<b>0.95</b>	<b>0.83</b>	<b>0.75</b>	0.41
FC	51.11	61.60	28.87	38.59	35.30	83.43	0.39	−0.15	<b>0.91</b>	0.38	0.38	0.13
ML (mm)	<b>2.32</b>	<b>4.78</b>	<b>4.75</b>	<b>3.99</b>	<b>4.73</b>	17.65	<b>0.82</b>	<b>0.70</b>	<b>0.95</b>	0.67	<b>0.77</b>	−0.04
TV (voxels/l)	16.11	24.60	10.66	17.70	17.09	44.53	<b>0.80</b>	0.34	<b>0.96</b>	0.43	0.48	0.42
FD (FC/voxel)	42.12	45.27	28.58	30.68	35.50	60.74	0.04	−0.49	<b>0.83</b>	0.36	0.19	0.10
<i>Arcuate fasciculus (left)</i>												
FA	<b>1.47</b>	<b>1.80</b>	<b>3.29</b>	<b>2.83</b>	<b>2.51</b>	<b>3.11</b>	<b>0.92</b>	<b>0.92</b>	0.32	<b>0.73</b>	<b>0.92</b>	0.39
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>1.24</b>	<b>0.97</b>	<b>0.61</b>	<b>1.15</b>	<b>1.10</b>	<b>1.75</b>	<b>0.84</b>	<b>0.88</b>	<b>0.96</b>	<b>0.85</b>	<b>0.87</b>	0.69
FC	21.03	28.17	59.48	20.17	29.42	43.27	<b>0.93</b>	<b>0.75</b>	0.38	<b>0.94</b>	<b>0.82</b>	0.63
ML (mm)	<b>3.99</b>	<b>3.31</b>	<b>4.19</b>	<b>3.06</b>	<b>4.46</b>	<b>9.99</b>	<b>0.72</b>	<b>0.87</b>	<b>0.97</b>	<b>0.85</b>	<b>0.74</b>	<b>0.75</b>
TV (voxels/l)	16.85	21.58	23.95	17.85	19.83	21.29	<b>0.86</b>	<b>0.73</b>	<b>0.75</b>	<b>0.88</b>	<b>0.82</b>	<b>0.70</b>
FD (FC/voxel)	14.26	18.76	41.63	21.03	26.67	33.77	<b>0.93</b>	0.65	0.31	<b>0.83</b>	0.67	0.58

**Bold.** CV  $\leq 10\%$  or ICC  $\geq .70$ . Abbreviations: FC, fiber count; FD, fiber density; ML, mean length; TV, tract volume.



**Fig. 2.** The scatter plots of FA, MD, mean length (ML), and tract volume (TV) of (a) the corpus callosum and (b) the left arcuate fasciculus reconstructed from the DTI images acquired from the first sessions (x-axis) versus the second sessions (y-axis). Each row shows the data from a different DTI sequence, with the first row showing the data from DTI30-2, the second row DTI30-1, and the third row DTI15-2A. Light blue dots contrast the data from the first session (DTI30-2A, DTI30-1A, or DTI15-2A) with the data of the first DTI sequences acquired from the second sessions (DTI30-2B, DTI30-1B, or DTI15-2B) and the dark blue dots contrasting the data from the first session (DTI30-2A, DTI30-1A, or DTI15-2A) with the data of the second DTI sequences acquired from the second sessions (DTI30-2C, DTI30-1C, or DTI15-2C). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 3**  
Tractography measurements from the 10 fiber regions showing either  $CV \leq 10\%$ ,  $ICC \geq .70$ , or both.

Fiber regions	DTI30-2	DTI30-1	DTI15-2	Total
Corpus callosum	FA, MD, FC, ML, TV, FD	FA, MD, FC, ML, TV, FD	FA, MD, ML, FD	16
Cingulum	FA, MD, FC, ML, TV	FA, MD, FC, ML, TV	FA, MD, FC, TV, FD	15
Angular bundle	FA, MD	FA, MD	FA, MD, ML	7
Fornix body	FA, MD, FC, ML	FA, MD, ML	FA, MD	9
Fornix crus	FA, MD	FA, MD, ML	FA, MD, ML	8
CPSf	FA, MD, ML, TV, FD	FA, MD, ML, FD	FA, MD, FC, TV, FD	14
Uncinate fasciculus	FA, MD, FC, ML, TV, FD	FA, MD, FC, ML, TV, FD	FA, MD, FC, ML, TV, FD	18
ILF	FA, MD, ML, TV	FA, MD, ML	FA, MD, ML	10
IFO	FA, MD, ML	FA, MD, ML	FA, MD	8
Arcuate fasciculus	FA, MD, FC, ML, TV, FD	FA, MD, FC, ML, TV	FA, MD, ML, TV	15
Total	43	40	37	120

Abbreviations: FC, fiber count; FD, fiber density; ML, mean length; TV, tract volume.

0.5–9.7% for MD intra-session  $CV_{within}$ , 1.2–7.4% for FA intersession  $CV_{within}$ , and 1.1–7.6% for MD intersession  $CV_{within}$ . Mean length had the next lowest intra- and inter-session  $CV_{within}$  (1.6–12.6% and 3.2–18.8%, respectively), followed by tract volume (3.1–30.2% and 5.0–44.5%, respectively), fiber density (3.2–45.3% and 5.8–60.7%, respectively), and fiber count (4.6–61.6% and 6.0–83.4%, respectively). Among the 10 fiber regions, the corpus callosum had the lowest intra- and inter-session  $CV_{within}$  (0.7–7.6% and 1.2–11.3%, respectively) and produced the most number of acceptable  $CV_{within}$  (34) of the 36 values (6 fiber variables times intra- and inter-session  $CV_{within}$  for DTI30-2, DTI30-1, and DTI15-2). The number of acceptable  $CV_{within}$  for the remaining nine fiber regions ranged from 13 (fornix crus) to 20 (ILF).

In contrast, intra- and inter-session ICCs were more uniform across all six fiber variables. Mean length provided more tractography measurements (46/60) with acceptable ICC values compared to other fiber variables (28–35/60). Of the 10 fiber regions, the corpus callosum and uncinate fasciculus provided the most number of acceptable ICC values (32 out of 36), followed in turn by cingulum (28), arcuate fasciculus (26), ILF (20), CPSf (18), fornix body (16), IFO (14), angular bundle (10), and fornix crus (7) (see Fig. 2).

Table 3 summarizes the tractography measurements from the ten fiber regions with acceptable intersession CV, ICC, or both for the three DTI sequences. The corpus callosum, cingulum, CPSf, uncinate fasciculus, and arcuate fasciculus provided 14–18/18 reliability measurements across the three DTI sequences. Two thirds (120/180) of the tested tractography measurements were identified as reliable among which 43 were from DTI30-2, 40 from DTI30-1, and 37 from DTI15-2.

#### Intra- versus inter-session reliability

To compare intra- and inter-session  $CV_{within}$  and ICC, paired  $t$  tests were performed on individual fiber variables across the 10 fiber regions and FDR was utilized to control for familywise errors (Table 4). For  $CV_{within}$ , only the DTI15-2 showed significantly higher intersession than intra-session values in FA and mean length (paired  $t(9) = -3.82$  and  $-5.83$ ,  $p = .004$  and  $.0002$ , respectively). For ICC, both DTI30-2 and DTI15-2 showed significant differences in some of the fiber variables. In DTI30-2, intra-session ICC for FA was significantly higher than intersession ICC (paired  $t(9) = 4.59$ ,  $p = .001$ ). In DTI15-2, three fiber variables showed significantly higher intra-session ICC compared to inter-session ICC. These were FA (paired  $t(9) = 3.30$ ,  $p = .009$ ), mean length (paired  $t(9) = 3.62$ ,  $p = .006$ ), and tract volume (paired  $t(9) = 3.24$ ,  $p = .010$ ).

#### The effect of gradient directions and scan repetitions

Next, pairwise comparisons of  $CV_{within}$  and ICC were performed between DTI30-2, DTI30-1, and DTI15-2 to assess the effect of scan repetition and increasing the number of gradient directions from 15 to 30 on intra- and inter-session reliability (Table 5). For intra- and inter-session  $CV_{within}$ , while the values from DTI30-2 were not significantly different from the values from DTI30-1, mean length (paired  $t(9) = -4.98$ ,  $p = .001$ ) and tract volume (paired  $t(9) = -3.65$ ,  $p = .005$ ) showed significantly lower intersession  $CV_{within}$  in DTI30-1 relative to DTI15-2. When comparing DTI30-2 to DTI 15-2, mean length (paired  $t(9) = -5.97$ ,  $p = .0002$ ), tract volume (paired  $t(9) = -4.71$ ,  $p = .001$ ), and

**Table 4**  
Comparisons of intra- and inter-session reliability.

Reliability Indices	DTI30-2		DTI30-1		DTI15-2	
	Paired $t$	$P$	Paired $t$	$P$	Paired $t$	$P$
CV						
FA	-2.98	0.02	-0.95	0.37	-3.85*	0.004
MD	-1.94	0.09	-0.06	0.95	-1.53	0.16
Fiber count	-1.87	0.09	-0.22	0.83	-1.24	0.25
Mean length	-2.56	0.03	-0.94	0.37	-5.83*	< 0.001
Tract volume	-1.89	0.09	-0.02	0.98	-2.60	0.03
Fiber density	-1.67	0.13	-2.07	0.07	-2.43	0.04
ICC						
FA	4.59*	0.001	0.66	0.53	3.30*	0.009
MD	0.94	0.37	1.57	0.15	2.42	0.04
Fiber count	2.51	0.03	0.45	0.66	1.72	0.12
Mean length	1.49	0.17	1.78	0.11	3.62*	0.006
Tract volume	2.36	0.04	0.49	0.63	3.24*	0.010
Fiber density	1.61	0.14	0.15	0.88	2.59	0.029

\*significant at FDR .05,  $p \leq .004$  for CV, and  $p \leq .010$  for ICC.

**Table 5**  
Reliability comparisons between different DTI sequences.

Reliability Indices	Intra-session reliability						Intersession reliability					
	DTI30-2 Vs. DTI30-1		DTI30-1 Vs. DTI15-2		DTI30-2 Vs. DTI15-2		DTI30-2 Vs. DTI30-1		DTI30-1 Vs. DTI15-2		DTI30-2 Vs. DTI15-2	
	Paired <i>t</i>	<i>P</i>	Paired <i>t</i>	<i>P</i>	Paired <i>t</i>	<i>P</i>	Paired <i>t</i>	<i>P</i>	Paired <i>t</i>	<i>P</i>	Paired <i>t</i>	<i>P</i>
<b>CV</b>												
FA	−1.17	0.27	1.61	0.14	1.14	0.29	−0.49	0.63	−0.86	0.41	−1.11	0.30
MD	−0.11	0.91	0.07	0.95	−0.01	0.99	0.54	0.61	−1.87	0.10	−1.56	0.15
Fiber count	−2.48	0.04	−0.26	0.80	−1.93	0.09	−1.14	0.29	−1.38	0.20	−2.79	0.02
Mean length	−1.71	0.12	−0.77	0.46	−1.72	0.12	−0.03	0.97	−4.98*	0.001	−5.97*	<0.001
Tract volume	−1.59	0.15	−0.39	0.71	−1.83	0.10	−0.89	0.40	−3.65*	0.005	−4.71*	0.001
Fiber density	−1.63	0.14	−0.15	0.88	−1.45	0.18	−1.82	0.10	−1.62	0.14	−4.27*	0.002
<b>ICC</b>												
FA	1.57	0.15	−1.35	0.21	−0.44	0.67	0.42	0.68	0.57	0.58	1.06	0.32
MD	0.79	0.45	−1.26	0.24	−0.85	0.42	1.69	0.12	1.57	0.15	2.39	0.04
Fiber count	2.88	0.02	−0.86	0.41	0.87	0.40	1.88	0.09	0.35	0.74	1.92	0.09
Mean length	0.57	0.58	−2.01	0.08	−1.45	0.18	0.49	0.64	1.75	0.11	2.37	0.04
Tract volume	3.43	0.008	−2.13	0.06	1.44	0.18	2.42	0.04	1.07	0.31	2.71	0.02
Fiber density	2.31	0.05	−1.34	0.21	−0.05	0.96	1.59	0.15	0.69	0.51	2.41	0.04

\*significant at FDR .05,  $p \leq .005$  for CV.

fiber density (paired  $t(9) = -4.27$ ,  $p = .002$ ) showed significantly lower intersession  $CV_{within}$ . In contrast, none of the comparisons of intra-session  $CV_{within}$  or intra- and inter-session ICC survived an FDR of 5%.

## Discussion

The current study provided an explicit evaluation of the reliability of tractography measurements of nine major fiber tracts in the brain. The examined fiber variables included FA and MD—which are the most widely used DTI parameters—together with the less commonly used volumetric measurements such as fiber count, mean length, tract volume, and fiber density. Different types of intra- and intersession reliabilities were compared to find important sources of variability and to evaluate whether increasing the number of gradient directions from 15 to 30 and scan repetition improved reliability.

### Reliability of tractography measurements

The major findings of the current study were the identification of fiber tracts that provide tractography measurements with high test–retest reliability, and fiber variables that could show acceptable test–retest reliability across different fiber tracts. In line with previous reports (Ciccarelli et al., 2003; Danielian et al., 2010; Heiervang et al., 2006; Vollmar et al., 2010), the reliability of tractography measurements is shown to be tract and variable specific. Across the three different DTI sequences (i.e. DTI30-2, DTI30-1, and DTI15-2), measurements from the corpus callosum tended to show lower intra- and inter-session CV ( $\leq 10\%$  except 11.3% for fiber count and 10.1% for tract volume from DTI15-2) than those from other fiber tracts. Furthermore, the measurements from the corpus callosum, uncinate fasciculus, cingulum, and arcuate fasciculus provided relatively high intra- and inter-session ICC ( $ICC \geq .70$  for 26–32/60 measurements) compared to other fiber tracts ( $ICC \geq .70$  for 7–20/60 measurements). Considering both intersession CV and ICC, 120/180 tractography measurements have shown either low test–retest variability, high consistency, or both across the three DTI sequences. Five fiber regions, namely the corpus callosum, cingulum, CPsf, uncinate fasciculus, and arcuate fasciculus, provided reliable measurements (i.e. intersession  $CV \leq 10\%$  or  $ICC \geq .70$ ) beyond FA, MD, and mean length, while the reliable tractography measurements from the remaining five fiber regions were more restricted to FA and MD. Of the six fiber variables, FA and MD measurements showed the lowest intra- and inter-session CV, followed in turn by mean length, fiber volume, fiber density, and fiber count. Whereas all 60 FA and MD measurements

had  $CV \leq 10\%$ , only 44 mean length, 9 tract volume, 9 fiber density, and 5 fiber count measurements attained such levels. In contrast, the six fiber variables exhibited similar intra- and inter-session ICC levels, with mean length showing the highest number of measurements with acceptable ICC values ( $ICC \geq .70$  for 46/60 measurements for mean length, and 28–35/60 for the remaining fiber variables).

Among previous studies, only Danielian et al. (2010) performed deterministic fiber tracking. They acquired DTI with 32 non-collinear directions (NEX 4) and assessed the reliability of tractography measurements from the complete, genu, motor, and splenium of the corpus callosum, and bilateral uncinate fasciculus. Despite the differences in scan repetition, image acquisition scheme, experimental design, image preprocessing, and fiber tracking methods, the previously reported reliability was at a similar level to that of the current results, notably the DTI30-2 images for both complete corpus callosum and uncinate fasciculus and also the DTI30-1 images for the corpus callosum (see Appendix Table 1). The findings from Danielian et al. and ours indicate that the choices of the number of directions and scan repetition are both more important than the choice of MR scanner for the reliability of tractography measurements (see below for more detailed discussion regarding the effect of the number of gradient encoding directions on reliability). We can compare our results with those of two other previous studies, one using fast marching, the other using a probabilistic fiber tracking method (Ciccarelli et al., 2003; Heiervang et al., 2006). Despite the fact that both previous studies acquired DTI using 60 directions (NEX 1) and thus had equivalent or more scan time than the current three sequences, our data showed lower variability (i.e. the intersession CVs for all three sequences from our study were higher than the intersession CVs from Ciccarelli et al. and Heiervang et al.) (see Appendix Table 1).

### Error sources in DTI tractography measurements

We estimated and then compared intra- and inter-session reliabilities in order to find major error sources. For DTI30-2, only FA showed significantly lower intersession ICC compared to intra-session ICC. For DTI15-2, both FA and mean length exhibited significantly higher intersession CV compared to intra-session CV, and FA as well as mean length and tract volume exhibited significantly lower intersession ICC compared to intra-session ICC. No significant differences were detected between intra- and inter-session CV and ICC for DTI30-1 even though many measurements showed higher intersession CV than intra-session CV (44/60).

**Appendix Table 1**

Comparisons of intersession CV and ICC to previous reports.

Tractography measurements	Ciccarelli et al., 2003	Heiervang et al., 2006 <sup>a</sup>	Danielian et al., 2010		DTI30-2		DTI30-1		DTI15-2	
	CV (%)	CV (%)	CV (%)	ICC	CV (%)	ICC	CV (%)	ICC	CV (%)	ICC
<b>CC</b>										
FA	6.2	5.9	1.4	0.87	1.2	0.90	1.9	0.88	1.3	0.86
MD	–	4.7	1.2	0.83	1.3	0.90	1.5	0.76	1.4	0.89
Tract volume	7.8	9.0	4.1	0.92	6.1	0.84	5.0	0.95	10.9	0.26
<b>Left CB</b>										
FA	–	7.3	–	–	2.3	0.76	3.3	0.67	2.9	0.77
MD	–	2.2	–	–	2.0	0.65	1.8	0.71	1.7	0.55
Tract volume	–	28.9	–	–	12.3	0.88	14.4	0.79	19.7	0.83
<b>Right UF</b>										
FA	–	7.3	1.2	0.89	3.1	0.63	3.2	0.66	5.3	0.53
MD	–	2.2	1.4	0.91	1.1	0.84	1.7	0.77	1.8	0.58
Tract volume	–	28.9	12.3	0.68	16.1	0.86	12.5	0.82	26.4	0.75

<sup>a</sup> The values were from the two ROI approach and the genu instead of the complete corpus callosum was reported.

The analysis indicated for the first time that MRI signal variation and physiological noise/change over time were important sources of error for test–retest reliability of tractography measurements especially FA, which has been used in a variety of studies. Although FA showed low inter-subject variability across the three DTI sequences—a pattern that helped detecting group differences between samples, the low intersession ICC for many FA measurements could potentially affect the capabilities of FA to reveal longitudinal changes and to find associations with other types of measurement. Additionally, we showed that MRI signal variation and physiological noise/change became more prominent for images acquired with 15 gradient directions, affecting both CV and ICC for several fiber variables. However, the results did not imply that the intra-session reliabilities for all tractography measurements were acceptable. In fact, 37/180 (21%) tractography measurements showed both high intra-session  $CV_{within}$  and low ICC. These measurements were mostly fiber count, tract volume, and fiber density measurements from fiber tracts such as the fornix, angular bundle, and IFO. The low reliability in some tractography measurements, even for images acquired from the same scan session, indicated that the combination of MRI signal variation within a scan session, subject motion, physiological noise, and rater variability was also an important error source in addition to MRI signal variation and physiological noise/change over time. Future studies should concentrate on improving image quality by stabilizing the MRI signal within a scan session and across sessions, and minimizing subject motion. Our data suggests that utilizing DTI acquisition schemes with 30 directions may help to reduce MRI signal variation over time. In addition, using phantoms to calibrate water diffusivity and directionality routinely (Lorenz et al., 2008) may be a useful procedure to avoid the drifting of DTI measurements.

#### The effect of number of gradient directions and scan repetitions

To determine whether increasing the number of gradient directions from 15 to 30 and scan repetition improves reliability, intra- and intersession CV and ICC were compared between DTI30-2, DTI30-1, and DTI15-2. While the number of gradient directions and scan repetition displayed non-significant effect on intra-session reliability, we found that increasing the number of gradient directions significantly reduced intersession CV for mean length and tract volume when comparing DTI15-2 to DTI30-1. This finding indicates that using 30 instead of 15 uniformly distributed gradient directions could significantly reduce intersession variability for mean length and tract volume measurements even when scan times are equivalent. However, our result is at odds with the findings of Heiervang et al. (2006) who acquired images with a 1.5T Siemens Sonata MR scanner. In Heiervang et al. (2006), probabilistic fiber tracking was performed to obtain FA, MD, and tract volume measurements of the cingulum, pyramidal tracts, optic radiations, and genu of the corpus callosum. The intersession CV for 12-direction images was reported as

being similar to that for 60-direction images. One major difference between the Heiervang's study and the present one, however, was the number of fiber tracts examined (Heiervang et al. used 4 when we used 9), and this difference might have affected the sensitivity in detecting differences in tract volume. Other factors contributing to the discrepancy could be the differences in scanner performance, image acquisition scheme, tracts-of-interest, and fiber tracking algorithms employed in the analyses. Further analysis is needed to investigate these discrepancies.

Unexpectedly, although increasing scan repetition from 1 to 2 for DTI with 30 directions did not significantly influence intra- or intersession reliability when combining tractography measurements from the 10 fiber regions, the values obtained were significantly different for many measurements between the two datasets. The reduced SNR in DTI30-1 relative to DTI30-2 led to significantly elevated FA measurements from all 10 fiber regions (2.6–10% increase) as well as significantly elevated fiber count, mean length, and tract volume measurements from many fiber regions. In contrast, the effect on MD was mixed, with either increased, reduced, or no changes in values. The results were consistent with Farrell et al. (2007)'s findings of an upward bias in FA, but no bias in MD when reducing SNR for ROI- and voxel-based measurements. A practical implication is that DTI with different numbers of scan repetitions should not be mixed in the same analysis. Mixing DTI with different numbers of scan repetitions could occur in situations when multiple scans are acquired to improve SNR. When subject motion is detected in one of the scans for a subset of the subjects, we may opt to discard the scans with substantial subject motion. However, this may significantly affect the FA and other DTI measurements for these subjects.

In addition, increasing the number of gradient directions from 15 to 30 with either equivalent or doubled scan time had an effect on both the values and  $CV_{between}$  of tractography measurements. All FA measurements had significantly higher values in DTI30-1 compared to DTI15-2 (1.8–13% increase) as did many measurements of fiber count, mean length, and tract volume. MD was also affected, with significantly lower values in DTI30-1 compared to DTI15-2 (3.7–11% decrease). Comparing DTI15-2 with DTI30-2, the increased gradient directions had a significant effect on all MD measurements except fornix crus (3.1–15% decrease) and on some of the measurements of the remaining fiber variables (lower in DTI15-2 except for CPsf FA). Lower fiber volumes in DTI with 12 directions compared to those in DTI with 60 directions have been reported by Heiervang et al. (2006). Furthermore, we found that the values of the  $CV_{between}$  were also higher for mean length, tract volume, fiber density in DTI15-2 compared to DTI30-1, and for mean length and fiber density when comparing DTI15-2 with DTI30-2.

In summary, the comparisons of the three DTI sequences suggested that each had its own advantages. While DTI30-1 provided tractography measurements with lower variability in mean length, tract volume, and fiber density than DTI15-2, the values of tractography

measurements from DTI15-2 were closer to DTI30-2 compared to those of DTI30-1. Among the three sequences, DTI30-2 provided most reliable measurements. However, it is important to keep in mind that the current reliability study was conducted on healthy subjects. The choice of DTI sequence could be different in clinical populations who may experience difficulties in staying still in the scanner. In subject populations where motion is likely, DTI15-2 or DTI30-1 may produce more reliable data than DTI30-2. Indeed, the intersession reliability was not significantly different between DTI30-1 and DTI30-2, and 15 directions were sufficient for obtaining FA and MD measurements from all 10 fiber regions, as were mean length and tract volume measurements from some of the fiber regions (Table 3).

#### *The reliability of DTI tractography*

While the reconstruction of the 10 fiber regions was successful for most images, it was unexpected that left arcuate fasciculus could not be reconstructed from some of the images for one participant. The reconstruction failure could be caused by biologically small tract size since the tract volumes for successfully reconstructed arcuate from other images of the same participant were also small. A second issue encountered during fiber reconstruction was the occurrences of so called 'red voxels.' These 'red voxels' have abnormally high FA values and were located along the longitudinal fissure where water diffusion directionality is normally low. Consequently, some corpus callosum fibers propagated through the longitudinal fissure and formed loops. These fibers were excluded from the result. However, the deletion might have affected the reliability. A third issue was the truncation of the corpus callosum fibers reconstructed from DTI15-2. The corpus callosum fibers failed to propagate to the cortex in some of the images, affecting the inter-session reliability of tractography measurements other than FA and MD.

#### *Limitations and future directions*

Several limitations of the present study need to be mentioned. The first is the well-known 'crossing fiber' issue that can occur with the deterministic fiber tracking algorithm such as those implemented in DTI Studio. This algorithm is considered advantageous for clinical applications compared to many alternative computationally intensive probabilistic fiber tracking methods (Behrens et al., 2003). This advantage notwithstanding, the deterministic fiber tracking algorithm fails to resolve complex fiber organization when fiber tracts are crossing or "kissing" and this can cause either early termination of fiber propagation because of the sub-threshold FA value, or erroneous fibers due to the incorrect principal diffusion direction. It is unknown how the 'crossing fiber' problem has affected the reliability of our tractography measurements. Additional studies are needed to evaluate the performance of different fiber tracking algorithms and compare the reliability of the resulting quantifications.

The second limitation is in the recognition of important error sources. The current investigation estimated the effect of the combination of MRI signal variation, subject motion, and subject physiological noise/change, scan repetition, and number of gradient directions on reliability. However, the study neither exhaustively differentiated all error sources, nor did it estimate reliability as a function of number of gradient sampling directions. Fulfilling these requirements would need further experimentation to be carried out. For example, the contribution of MRI signal variation to variability in tractography measurements can be tested using a phantom specifically designed for calibrating water diffusion properties and fiber tracking results. Nevertheless, the present study has identified important error sources in test–rest reliability of tractography measurements which should form a guide for future improvement.

Finally, we performed tractography in subjects' native space instead of automatic fiber tracking that required the transformation

all images to a common space, followed by the application of a common set of ROIs for tractography. Our intent was to provide a reliability assessment of tractography measurements obtained using the most popular method in order to provide insight regarding their potential efficacy in clinical research and practice. Although we acquired the images with AC-PC alignment and defined the ROI slices rigorously, the small misalignment between images and minor variances in intra-rater reliability might have had an effect on the reliability assessment. However, automatic fiber tracking may also add variability to the reliability assessment because individual differences in brain morphology make image registration and transformation less than perfect (Klein et al., 2009). Further, the need to perform tensor transformations coupled with the low image resolution of DTI make the process even more challenging (Zhang et al., 2008). Future experiments could beneficially explore whether automatic fiber tracking via spatially normalized data can further improve the reliability of tractography measurements.

#### *Conclusions*

The present study assessed the test–retest reliability of quantitative diffusion tensor tractography. Through a comprehensive evaluation of tractography measurements from nine major white matter fiber tracts, 37–43/60 measurements were identified as reliable, depending on the number of gradient directions and scan repetition. Many of these came from the corpus callosum, cingulum, CPsf, uncinate fasciculus, and arcuate fasciculus, and from FA and MD values of all 10 fiber regions. Notably, the identification of reliable tractography measurements was verified in a separate study (Wang et al., 2011), which showed improvement in the validity of tractography measurements for detecting traumatic axonal injury using the reliable measurements.

Different types of intra- and inter-session reliability of the tractography measurements were compared. While intersession reliability was lower than intra-session reliability for DTI with 15 directions, even intra-session reliability was low for some tractography measurements from DTI with 15 as well as 30 directions. The results indicate that the combination of MRI signal variation, subject motion, and physiological noise/change is an important error source. To improve reliability, increasing the number of gradient directions can be an effective method especially for volumetric tractography measurements, a conclusion supported by reduced variability in 30- versus 15-direction DTI for mean length and tract volume even when the scan time is equivalent. In contrast, the accuracy of the values can be improved through scan repetition, evident by comparing group means of tractography measurements from DTI30-1 with DTI30-2. These findings have important implications for future experiments. As long as subject motion is under control, scan repetition and increasing number of gradient directions will likely lead to tractography measurements with high test–retest reliability and accuracy. Nevertheless, DTI images with 15 directions with two scan repetitions are capable of providing reliable FA and MD measurements from all nine major fiber tracts, and volumetric tractography measurements from some of the fiber tracts.

Supplementary materials related to this article can be found online at [doi:10.1016/j.neuroimage.2011.12.062](https://doi.org/10.1016/j.neuroimage.2011.12.062).

#### **Acknowledgments**

The authors would like to express gratitude to volunteers and to researchers and staff at the Meadow's Imaging Center at UT Southwestern Medical Center. We would also like to acknowledge Bennett Landman from Johns Hopkins University who kindly provided MATLAB scripts for generating 15 gradient directions with minimal potential energy, and to Tom Porteous for editorial comments. This study is supported by National Institute on Disability and Rehabilitation Research (H133A020526 to R.D.-A.) and National Institutes of Health (U01 HD042652, R01 HD48179 to R.D.-A.).

## References

- Abdi, H., 2010. Coefficient of variation. In: Salkind, N.J., Dougherty, D.M., Frey, B. (Eds.), *Encyclopedia of Research Design*. SAGE Publications, Inc., Thousand Oaks, CA, pp. 169–171.
- Armitage, P., 1971. *Statistical Methods in Medical Research*. Blackwell Scientific Publications, Oxford.
- Aron, A.R., Behrens, T.E., Smith, S., Frank, M.J., Poldrack, R.A., 2007. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* 27, 3743–3752.
- Augustinack, J.C., Helmer, K., Huber, K.E., Kakunoori, S., Zollei, L., Fischl, B., 2010. Direct visualization of the perforant pathway in the human brain with ex vivo diffusion tensor imaging. *Front. Hum. Neurosci.* 4, 42.
- Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* 111, 209–219.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B* 103, 247–254.
- Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed.* 15, 435–455.
- Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Ben-Shachar, M., Dougherty, R.F., Wandell, B.A., 2007. White matter pathways in reading. *Curr. Opin. Neurobiol.* 17, 258–270.
- Bisdas, S., Bohning, D.E., Besenski, N., Nicholas, J.S., Rumboldt, Z., 2008. Reproducibility, interrater agreement, and age-related changes of fractional anisotropy measures at 3T in healthy subjects: effect of the applied b-value. *AJNR Am. J. Neuroradiol.* 29, 1128–1133.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310.
- Catani, M., Ffytche, D.H., 2005. The rises and falls of disconnection syndromes. *Brain* 128, 2224–2239.
- Catani, M., Howard, R.J., Pajevic, S., Jones, D.K., 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17, 77–94.
- Ciccarelli, O., Parker, G.J., Toosy, A.T., Wheeler-Kingshott, C.A., Barker, G.J., Boulby, P.A., Miller, D.H., Thompson, A.J., 2003. From diffusion tractography to quantitative white matter tract measures: a reproducibility study. *Neuroimage* 18, 348–359.
- Concha, L., Gross, D.W., Beaulieu, C., 2005. Diffusion tensor tractography of the limbic system. *AJNR Am. J. Neuroradiol.* 26, 2267–2274.
- Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E., 1999. Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10422–10427.
- Crosby, E.C., Humphrey, T., Lauer, E.W., 1962. *Correlative Anatomy of the Nervous System*. The MacMillan Company, New York.
- Danielian, L.E., Iwata, N.K., Thomasson, D.M., Floeter, M.K., 2010. Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. *Neuroimage* 49, 1572–1580.
- Duffau, H., 2008. The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography. *Neuropsychologia* 46, 927–934.
- 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.
- Gao, W., Zhu, H., Lin, W., 2009. A unified optimization approach for diffusion tensor imaging technique. *Neuroimage* 44, 729–741.
- Heiervang, E., Behrens, T.E., Mackay, C.E., Robson, M.D., Johansen-Berg, H., 2006. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage* 33, 867–877.
- Huang, H., Zhang, J., van Zijl, P.C., Mori, S., 2004. Analysis of noise effects on DTI-based tractography using the brute-force and multi-ROI approach. *Magn. Reson. Med.* 52, 559–565.
- Huang, H., Zhang, J., Jiang, H., Wakana, S., Poetscher, L., Miller, M.J., van Zijl, P.C., Hillis, A.E., Wytik, R., Mori, S., 2005. DTI tractography based parcellation of white matter: application to the mid-sagittal morphology of corpus callosum. *Neuroimage* 26, 195–205.
- 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.
- Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.* 42, 515–525.
- Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802.
- 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, 1123–1138.
- Lorenz, R., Bellemann, M.E., Hennig, J., Il'yasov, K.A., 2008. Anisotropic phantoms for quantitative diffusion tensor imaging and fiber-tracking validation. *Appl. Magn. Reson.* 33, 419–429.
- Marenco, S., Rawlings, R., Rohde, G.K., Barnett, A.S., Honea, R.A., Pierpaoli, C., Weinberger, D.R., 2006. Regional distribution of measurement error in diffusion tensor imaging. *Psychiatry Res.* 147, 69–78.
- McClelland, J.L., Rogers, T.T., 2003. The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.* 4, 310–322.
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46.
- Mori, S., van Zijl, P.C., 2002. Fiber tracking: principles and strategies – a technical review. *NMR Biomed.* 15, 468–480.
- Mori, S., Crain, B.J., Chacko, V.P., van Zijl, P.C., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45, 265–269.
- Naghavi, H.R., Nyberg, L., 2005. Common fronto-parietal activity in attention, memory, and consciousness: shared demands on integration? *Conscious. Cogn.* 14, 390–425.
- Nolte, J., 2002. *The Human Brain: An Introduction to Its Functional Anatomy*, 5th ed. Mosby, St. Louis, MO.
- Ozturk, A., Sasson, A.D., Farrell, J.A., Landman, B.A., da Motta, A.C., Aralasmak, A., Yousem, D.M., 2008. Regional differences in diffusion tensor imaging measurements: assessment of intrarater and interrater variability. *AJNR Am. J. Neuroradiol.* 29, 1124–1127.
- Pfefferbaum, A., Adalsteinsson, E., Sullivan, E.V., 2003. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J. Magn. Reson. Imaging* 18, 427–433.
- Skare, S., Hedehus, M., Moseley, M.E., Li, T.Q., 2000. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. *J. Magn. Reson.* 147, 340–352.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17, 479–489.
- 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., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 (Suppl. 1), S208–S219.
- 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.
- Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C., Mori, S., 2004. Fiber tract-based atlas of human white matter anatomy. *Radiology* 230, 77–87.
- 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.
- Wang, J.Y., Bakhadirov, K., Devous Sr., M.D., Abdi, H., McColl, R., Moore, C., Marquez de la Plata, C.D., Ding, K., Whittemore, A., Babcock, E., Rickbeil, T., Dobervich, J., Kroll, D., Dao, B., Mohindra, N., Madden, C.J., Diaz-Arrastia, R., 2008. Diffusion tensor tractography of traumatic diffuse axonal injury. *Arch. Neurol.* 65, 619–626.
- Wang, J.Y., Bakhadirov, K., Abdi, H., Devous Sr., M.D., Marquez de la Plata, C.D., Moore, C., Madden, C.J., Diaz-Arrastia, R., 2011. Longitudinal changes of structural connectivity in traumatic axonal injury. *Neurology* 77, 818–826.
- Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C., 1998. Automated image registration: I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* 22, 139–152.
- Zhang, W., Olivi, A., Hertig, S.J., van Zijl, P., Mori, S., 2008. Automated fiber tracking of human brain white matter using diffusion tensor imaging. *Neuroimage* 42, 771–777.