

Test–Retest Reliability of Diffusion Measures in Cerebral White Matter: A Multiband Diffusion MRI Study

Fei Duan, BS,^{1,2} Tengda Zhao, MS,^{1,2} Yong He, PhD,^{1,2} and Ni Shu, PhD^{1,2*}

Background: To investigate the test–retest (TRT) reliability of the diffusion measures in cerebral white matter obtained from the diffusion MRI dataset acquired with multiband acceleration.

Methods: With the multiband diffusion MRI dataset with two repeated scanning sessions, the TRT reliability of diffusion measures (fractional anisotropy [FA], mean diffusivity [MD], primary diffusivity [PD] and transverse diffusivity [TD]) was investigated through several fully automated analysis methods, including two voxel-level analyses (voxel-based analysis [VBA] and tract-based spatial statistics [TBSS]) and an atlas ROI-based analysis. The reproducibility was assessed by the intra-class correlation coefficient (ICC).

Results: Our results demonstrated moderate to high reproducibility ($ICC > 0.4$) of diffusion measures from the multiband EPI sequence with different analysis approaches. Across different measures, FA exhibited the highest reproducibility (mean $ICC = 0.70$), while MD showed the lowest reliability (mean $ICC = 0.55$) ($P = 0.006$). Additionally, ICCs varied across different tract ROIs: Commissural tracts showed higher reproducibility than other categories of tracts (projection, association and brainstem), while the brainstem tracts exhibited the poorest reliability ($P = 0.004$).

Conclusion: Our findings suggest a potential utility of the multiband EPI sequence for exploring individual differences of cerebral white matter and provide reference for future white matter studies.

J. MAGN. RESON. IMAGING 2015;42:1106–1116.

Diffusion Tensor Imaging (DTI), an imaging technique that measures diffusion movement of water molecules in biological tissues,^{1,2} has been widely used to study the microstructural properties of cerebral white matter (WM).^{3,4} DTI can generate several scalar diffusion measures, such as fractional anisotropy (FA), mean diffusivity (MD), primary diffusivity (PD), and transverse diffusivity (TD), with each measure providing specific quantitative information on the WM architecture at a voxel level.^{2,5} Many application studies have suggested these measures are sensitive to the changes in WM microstructure with normal ageing⁶ and development⁷ as well as in neuropsychiatric diseases,^{8,9} making diffusion metrics potential biomarkers for clinical applications.

Until now, numerous DTI studies have used the conventional echo planar imaging (EPI) sequence to acquire DTI data. However, conventional DTI suffers from prolonged scan times (tens of minutes for the scan parameters

used in this current study), resulting in increased risk of motion induced artifacts.^{10–12} Some promising fast-collecting imaging techniques, such as multiband EPI, have been proposed recently.¹³ This sequence can accelerate acquisition by simultaneously imaging multiple slices in the human brain, while not significantly sacrificing spatial resolution or the signal-to-noise ratio (SNR). Recently, this sequence is being applied in the human connectome project to acquire a large sample of healthy subjects with the purpose of uncovering individual differences in brain circuitry related to behavior.¹⁴ However, to reveal the true individual variations rather than the false-positive differences induced by iterative measurements, the test–retest (TRT) reliability of the diffusion measures is critical for the future application of group comparisons and longitudinal evaluation of this sequence.

Previous studies have evaluated the reliability of diffusion measures from both the voxel-level and region of

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.24859

Received Nov 23, 2014, Accepted for publication Jan 9, 2015.

*Address reprint requests to: N.S., State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China. E-mail: nshu@bnu.edu.cn

From the ¹State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China; and ²Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China

TABLE 1. Summary of Phenotype Information of Subjects^a

ID	Sex	Age(y)	Current diagnosis	Lifetime diagnosis
2475376	M	21	NO	NO
2799329	M	30	NO	NO
2842950	M	27	NO	NO
3201815	M	48	NO	NO
3315657	M	19	NO	NO
3795193	M	57	NO	NO
3808535	M	25	NO	NO
4176156	M	46	NO	NO
7055197	F	22	NO	NO
8735778	F	31	NO	NO
9630905	F	36	NO	NO

^aThe diagnostic information for each subject was collected using a structured clinical interview for the DSM Disorder (SCID) by trained professionals, and the numbers are DSM-IV codes. ‘No’ indicates no psychiatric disorder was identified during the interview.

interest (ROI) level perspectives of the conventional EPI sequence.^{15–19} Although moderate to high reproducibility of diffusion metrics was found, these studies identified the reliability variations across measures and WM tracts or regions¹⁹ that would be useful for clinical applications by generating reference data for inter-subject variability and intra-subject reproducibility. There are many possible factors affecting reliability, mainly including acquisition schemes,^{20–22} imaging parameters,^{20,23–25} image registration accuracy,^{18,26} subject physiological noise and head motion.^{18,22,24,26} For the multiband EPI sequence, whether the multiband DTI scans can effectively quantify the WM architecture and the normal values of diffusion measures in the human brain remain unknown. Moreover, whether the diffusion metrics from this new sequence can exhibit good TRT reliability is still unclear.

The purpose of this study is to evaluate the following: (i) the reliability of quantitative diffusion measures from both voxel and ROI levels and (ii) the reliability differences among diffusion measures and among WM tracts.

Materials and Methods

Test–retest Datasets

We used the multiband imaging test–retest pilot dataset that is publicly available from INDI (http://fcon_1000.projects.nitrc.org/indi/pro/eNKI_RS_TRT/FrontPage.html), which consists of 24 subjects. Among these subjects, 11 healthy participants (age: 32.9 ± 12.5 years; 3 females) without a history of neurologic or psychiatric diseases were included for the current analysis, whose phenotype information is presented in Table 1. All individuals included in the sample underwent semi-structured diagnostic psychiatric interviews and completed a battery of psychiatric, cogni-

tive, and behavioral assessments. Written informed consent was obtained from all participants. The Nathan Kline Institute (NKI) Institutional Review Board approved the research protocol to collect and share the data. Recently, the test–retest functional MRI data in this dataset has been used to examine the TRT reliability of regional functional homogeneity in the human brain,²⁷ identify global hubs in the human voxel-wise functional brain networks and examine TRT reliability over scanning time.²⁸

Data Acquisition

Each participant received test–retest diffusion MRI scans (at least one week apart) using a Siemens Trio 3.0 Tesla (T) scanner. Diffusion MRI data were acquired using a recently developed multiband EPI (mEPI) sequence^{13,29} with the following parameters: repetition time (TR) = 2400 ms, echo time (TE) = 85 ms, 64 slices, slice thickness of 2 mm, field of view (FOV) = 212×180 mm², voxel size of 2 mm isotropic, b value = 1500 s/mm², 128 gradient directions with 9 b = 0 images, multiband acceleration factor = 4, averages = 1, and total acquisition time = 5:58 min. All the test–retest diffusion MRI data were acquired on a single scanner. For each diffusion scan, the data quality was checked by visual inspection to avoid the major artifacts and/or distortions caused by the head motion. The head motions are within 2 mm for both test–retest diffusion scans of all participants. Additionally, MPRAGE T1-weighted images and the test–retest resting-state functional MRI images were also acquired but were not used in the present study.

Data Preprocessing

The preprocessing of DTI data included eddy current and motion correction, estimation of the diffusion tensor and calculation of the diffusion measures. Briefly, the eddy current distortions and motion artifacts in the DTI data were corrected by applying an affine alignment of each diffusion-weighted image to the b = 0

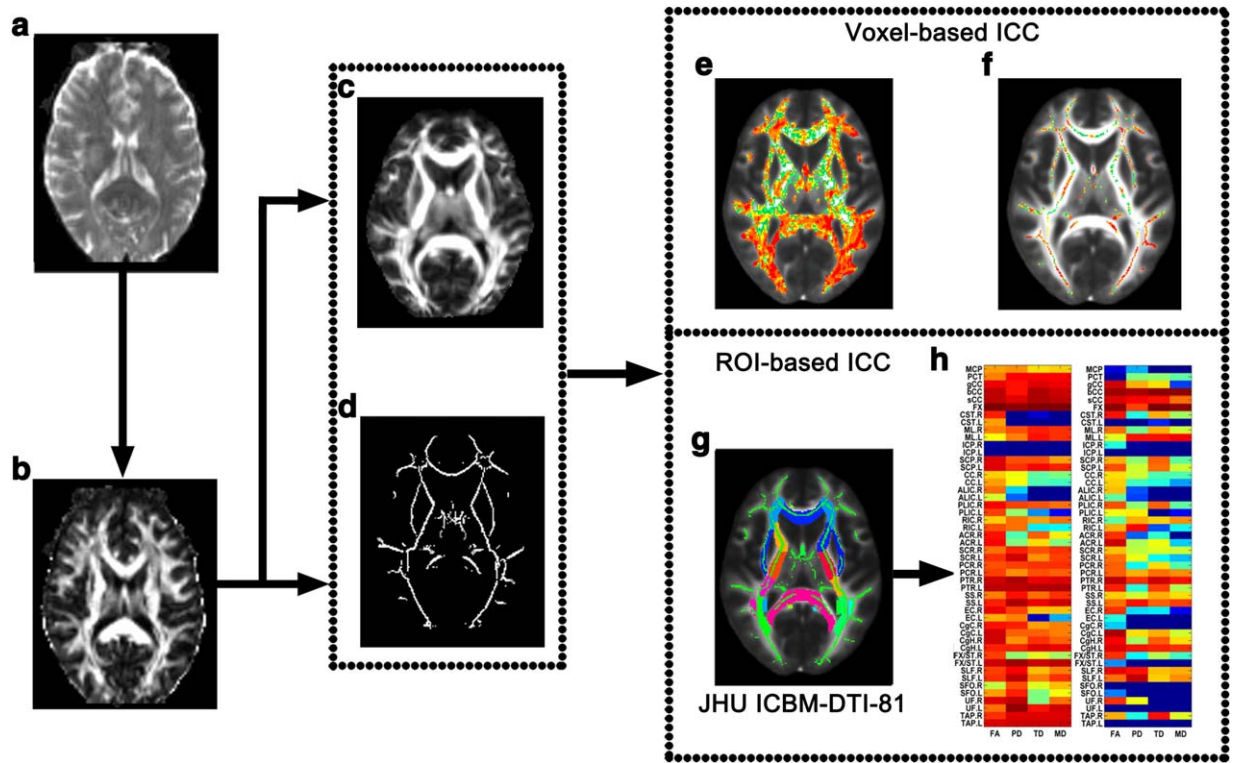


FIGURE 1: The flowchart of reliability analysis of diffusion measures in cerebral WM. First, the diffusion measures were calculated from the multiband diffusion MRI data (A). Then, the individual FA image (B) was normalized into the standard MNI space (C) for the VBA analysis and the WM skeleton (D) was extracted for the TBSS analysis. The same transformation used for the FA images was applied to the diffusivity images (PD, TD, and MD) to obtain the whole-brain and skeleton PD, TD, and MD images in the standard space. Then, voxel-wise ICC maps for whole-brain WM (E) and skeleton WM (F) were computed. For the ROI-based analysis, the JHU ICBM-DTI-81 WM atlas was applied to define different WM tracts (G). Then, the ICC values of the mean diffusion measures for each WM tract based on whole-brain and skeleton WM were calculated (H).

image. After this process, the diffusion tensor elements were estimated by solving the Stejskal and Tanner equation,¹ and then the reconstructed tensor matrix was diagonalized to obtain three eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) and three eigenvectors. Finally, the corresponding FA, MD, PD, and TD of each voxel were calculated according to the following formulas.³⁰ All preprocessing procedures of the DTI data were performed with the FDT toolbox in FSL (<http://www.fmrib.ox.ac.uk/fsl>).

$$FA = \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}}{\sqrt{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}, MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

$$PD = \lambda_1, TD = \frac{\lambda_2 + \lambda_3}{2}$$

VBA and TBSS

The VBA and TBSS analyses of FA, MD, PD, and TD images were performed using the FMRIB Software Library (FSL 4.1.4; www.fmrib.ox.ac.uk/fsl); for a detailed description of the methods, see Smith et al.³¹ Briefly, we performed the following steps on the FA images: (i) the FA image of each subject was aligned to a pre-identified target FA image (FMRIB58_FA) by nonlinear registration; (ii) all of the aligned FA images were transformed into the ICBM152 template by affine registration, thus the resultant FA

images in Montreal Neurological Institute (MNI) standard space were ready for VBA analysis. For the TBSS analysis, several following steps were further performed: (iii) the group-averaged normalized FA image in MNI space and its skeleton (mean FA skeleton) were created from the mean FA image of all subjects; and (iv) individual subjects' FA images were projected onto the skeleton, thus the individual FA skeleton maps were prepared for TBSS. Then, data for MD, PD, and TD were generated by applying the above FA transformations to additional diffusivity maps and projecting them onto the skeleton with projection vectors that were identical to the vectors inferred from the original FA data. Finally, for each participant, the voxel-wise FA, MD, PD, and TD maps for whole-brain and skeleton WM in standard space were calculated. The flowchart of the analysis procedure is presented in Figure 1.

Atlas-based Quantification of WM Tracts

To investigate the reproducibility of specific WM tracts, we used the digital WM atlas JHU ICBM-DTI-81 (<http://cmrm.med.jhmi.edu/>) to define the ROIs of major WM tracts. In the ICBM-DTI-81 WM labels atlas, 48 WM tract labels were created by hand segmentation of a standard-space average of diffusion MRI tensor maps from 81 subjects³² (Table 2). The tracts in the JHU ICBM-DTI-81 atlas can be classified into four types: projection tracts, association tracts, commissural tracts and tracts in the brainstem,³²

TABLE 2. Regions of Interest Defined in the JHU ICBM-DTI-81 Atlas

Index	Region	Abbreviation	Category
1	Middle cerebellar peduncle	MCP	Brainstem
2	Pontine crossing tract	PCT	Brainstem
3	Genu of corpus callosum	gCC	Commissural
4	Body of corpus callosum	bCC	Commissural
5	Splenium of corpus callosum	sCC	Commissural
6	Fornix	FX	Association
7	Corticospinal tract R	CST.R	Brainstem
8	Corticospinal tract L	CST.L	Brainstem
9	Medial lemniscus R	ML.R	Brainstem
10	Medial lemniscus L	ML.L	Brainstem
11	Inferior cerebellar peduncle R	ICP.R	Brainstem
12	Inferior cerebellar peduncle L	ICP.L	Brainstem
13	Superior cerebellar peduncle R	SCP.R	Brainstem
14	Superior cerebellar peduncle L	SCP.L	Brainstem
15	Cerebral peduncle R	CPR	Projection
16	Cerebral peduncle L	CPL	Projection
17	Anterior limb of internal capsule R	ALIC.R	Projection
18	Anterior limb of internal capsule L	ALIC.L	Projection
19	Posterior limb of internal capsule R	PLIC.R	Projection
20	Posterior limb of internal capsule L	PLIC.L	Projection
21	Retrolenticular part of internal capsule R	RIC.R	Projection
22	Retrolenticular part of internal capsule L	RIC.L	Projection
23	Anterior corona radiata R	ACR.R	Projection
24	Anterior corona radiata L	ACR.L	Projection
25	Superior corona radiata R	SCR.R	Projection
26	Superior corona radiata L	SCR.L	Projection
27	Posterior corona radiata R	PCR.R	Projection
28	Posterior corona radiata L	PCR.L	Projection
29	Posterior thalamic radiation R	PTR.R	Projection
30	Posterior thalamic radiation L	PTR.L	Projection
31	Sagittal stratum R	SS.R	Association
32	Sagittal stratum L	SS.L	Association
33	External capsule R	EC.R	Association
34	External capsule L	EC.L	Association
35	Cingulum (cingulate gyrus) R	CgC.R	Association
36	Cingulum (cingulate gyrus) L	CgC.L	Association
37	Cingulum (hippocampus) R	CgH.R	Association
38	Cingulum (hippocampus) L	CgH.L	Association
39	Fornix (cres)/Stria terminalis R	FX/ST.R	Association

TABLE 2: Continued

Index	Region	Abbreviation	Category
40	Fornix (cres)/Stria terminalis L	FX/ST.L	Association
41	Superior longitudinal fasciculus R	SLF.R	Association
42	Superior longitudinal fasciculus L	SLF.L	Association
43	Superior fronto-occipital fasciculus R	SFOF.R	Association
44	Superior fronto-occipital fasciculus L	SFOF.L	Association
45	Uncinate fasciculus R	UFR	Association
46	Uncinate fasciculus L	UFL	Association
47	Tapetum R	TAP.R	Commissural
48	Tapetum L	TAP.L	Commissural

therefore, the reliability of each category of tract can be quantified and compared.

Intraclass Correlation Coefficient

To evaluate the TRT reliability of diffusion measures between two sessions, a measurement of the intraclass correlation coefficient (ICC) was used. The ICC value was calculated as ³³:

$$ICC = \frac{\sigma_{bs}^2 - \sigma_{ws}^2}{\sigma_{bs}^2 + (m-1)\sigma_{ws}^2}$$

where σ_{bs} is the between-subject variance and σ_{ws} is the within subject variance, m represents the number of repeated measurements (here, $m = 2$). The ICC is a normalized measure which has a maximum of 1. The ICC values were categorized into five common intervals³⁴: $0 < ICC \leq 0.2$ (slight), $0.2 < ICC \leq 0.4$ (fair), $0.4 < ICC \leq 0.6$ (moderate), $0.6 < ICC \leq 0.8$ (substantial), and $0.8 < ICC \leq 1.0$ (almost perfect). Negative ICCs, suggesting negative reliability (i.e., completely nonreliable), are theoretically difficult to interpret.³⁵ Therefore, we set negative ICCs equal to zero, as suggested in other test–retest studies using the ICC.^{36,37}

For the VBA analysis, we calculated the voxel-wise ICC maps for whole-brain WM with an FA > 0.2. For the TBSS analysis, we calculated the ICC for each voxel within the WM skeleton. For the atlas-based ROI analysis, we calculated the mean ICC for each ROI from both whole-brain and skeleton levels by averaging the values across all WM voxels within each ROI.

Statistical Analysis

To assess the reliability differences across diffusion measures and categories of WM tracts, a two-way mixed repeated measures analysis of variance (ANOVA) was performed with the category of tract treated as a categorical factor and the category of measure treated as a repeated measure. To further explore the difference in reproducibility among different types of tracts for each measure, one-way ANOVA was performed. Due to the limited sample size, the effects of possible variables, such as age and gender, on the diffusion metrics were negligible and not considered in the present study. All ANOVA analyses were performed with SPSS software

(version 13.0; SPSS, Chicago, IL). Moreover, the correlation of the average diffusion values between two repeated sessions across ROIs was calculated by Pearson's correlation using an in house Matlab program (The MathWorks, Inc.).

Results

Normal Values of Diffusion Measures in Cerebral WM

The group-averaged FA, PD, TD, and MD maps and the mean histogram distributions of four diffusion measures across WM voxels are shown in Figure 2. The normal values of FA of WM range from 0.2 to 0.94, PD ranges from 7.4×10^{-5} to 2.2×10^{-3} mm²/s, TD ranges from 7.0×10^{-6} to 1.6×10^{-3} mm²/s, and MD ranges from 1.4×10^{-5} to 1.8×10^{-3} mm²/s across whole-brain WM voxels.

Voxel-Based ICC

For each diffusion measure, the voxel-wise ICC maps for whole-brain and skeleton WM are shown in Figure 2. We found that most WM voxels (whole-brain > 75%; skeleton > 63%) exhibited moderate to high reliability (ICC > 0.4) (Table 3). Averaged across the voxels of the whole-brain WM, the ICC values ranged from 0.55 to 0.70. Averaged across the voxels of the skeleton WM, the ICC values ranged from 0.47 to 0.67 (Table 3), depending on the specific diffusion measure. High consistency of the ICC distributions across voxels between whole-brain and skeleton WM were observed (Fig. 2). Among different diffusion measures, FA exhibited higher reliability than other measures. The histogram distributions of the ICC values of each diffusion measure across WM voxels are also displayed in Figure 2.

ROI-Based ICC

For the atlas-based ROIs, high correlations of mean diffusion measures across ROIs were observed between the two sessions in both whole-brain and skeleton WM (all

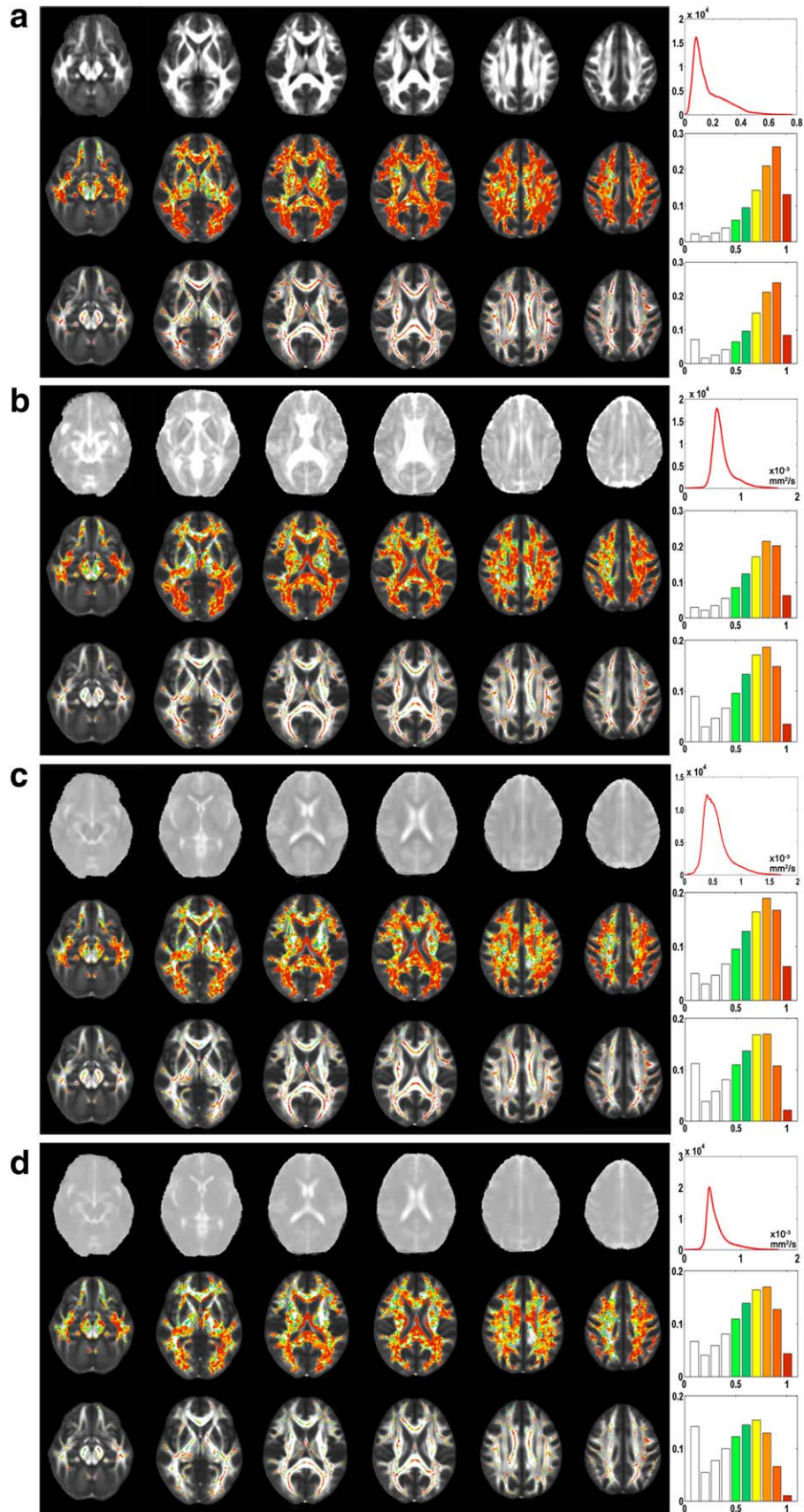


FIGURE 2: Voxel-wise reliability of diffusion measures for whole-brain and skeleton WM. For each diffusion measure, the first row represents the voxel-wise FA (A), PD (B), TD (C), and MD (D) maps across the whole brain. The second and third rows represent the voxel-wise ICC values for whole-brain and skeleton WM (FA > 0.2), respectively. Only WM voxels with an ICC > 0.4 overlaid on the mean diffusion measure maps are displayed. The histogram distributions of the values of diffusion measures and the distributions of ICC values across the whole-brain and skeleton WM are shown in the right lateral panel.

TABLE 3. Mean ICC Values of Diffusion Measures Across Whole-Brain and Skeleton WM and the Proportion of WM Voxels With Moderate to High ICCs

		FA	PD	TD	MD
Whole-brain	mean (std)	0.70 (0.21)	0.64 (0.23)	0.60 (0.25)	0.55 (0.26)
	proportion	90.0%	86%	81%	75%
Skeleton	mean (std)	0.67 (0.22)	0.59 (0.23)	0.54 (0.26)	0.47 (0.27)
	proportion	85%	77%	71%	63%

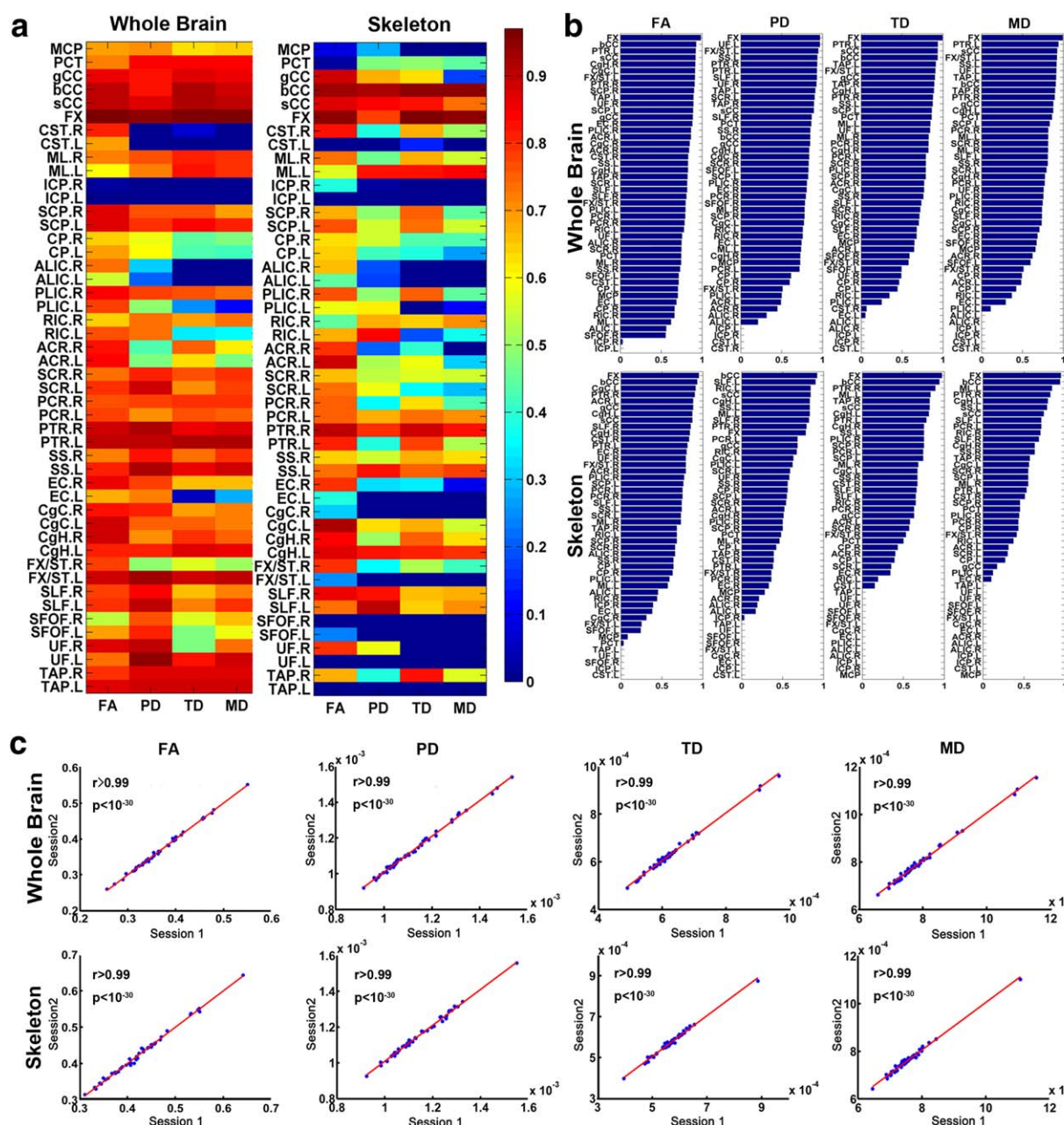


FIGURE 3: Reliability of diffusion measures across different WM tracts. For each diffusion measure, the ICC values of all 48 tracts in the JHU ICBM-DTI-81 WM atlas were quantified (A) and sorted in descending order (B). The upper panel shows whole-brain WM and the lower panel shows skeleton WM. High correlations of the mean diffusion measures across different tract ROIs were observed between two sessions for both whole-brain (top) and skeleton (bottom) WM (all $r > 0.99$) (C).

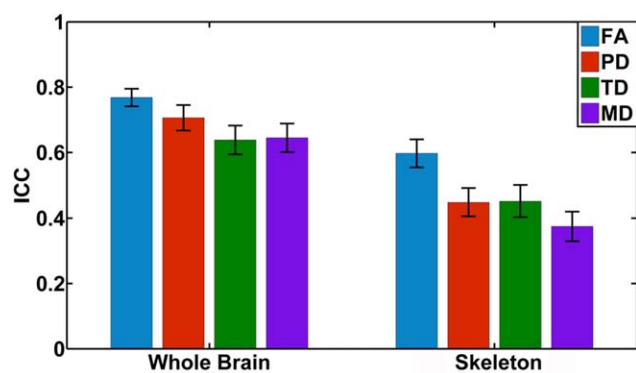


FIGURE 4: Reliability differences among different diffusion measures. The bars and error bars represent the mean values and standard deviations of the ICC values of all WM tracts, respectively.

$r > 0.99$) (Fig. 3C). For each diffusion measure, the ICC values of all 48 tracts in the WM atlas were quantified and shown in Figure 3A. Most of the tract ROIs exhibited moderate to high reliability, such as the fornix, corpus callosum, cingulum, superior longitudinal fasciculus, sagittal striatum, tapetum, posterior corona radiate, and posterior thalamic radiation, while some tract ROIs showed poor reliability, such as the inferior and superior cerebellar peduncle in the brainstem, corticospinal tract and anterior limb of the internal capsule (Fig. 3B). For the skeleton WM, more regions with poorer reliability were found compared with the whole-brain WM ($P < 0.001$) (Fig. 3A). The most reliable regions were located in the fornix, body and splenium of corpus callosum, cingulum and posterior thalamic radiation.

Comparisons of Reliability Across Measures and Tracts

In whole-brain WM, the ANOVA showed significant main effects of different diffusion measures ($P = 0.006$) (Fig. 4) and category of tracts ($P = 0.004$) (Fig. 5A). In skeleton WM, the ANOVA only showed a significant main effect of different measures ($P < 0.001$) (Fig. 4), but not for the category of tracts ($P > 0.1$) (Fig. 5B). No interactions of tract \times measure in either whole-brain or skeleton WM were found ($P > 0.1$). For diffusion measures, post hoc comparisons revealed that FA has a higher level of reliability than other measures both in whole-brain and skeleton WM (all $P < 0.03$) (Fig. 4). For different categories of tracts, the post hoc comparisons showed that commissural tracts have the highest reproducibility, while the tracts in the brainstem exhibit the lowest reliability as compared to other types of tracts (all $P < 0.05$) (Fig. 5).

Discussion

In the present study, we investigated the TRT reliability of diffusion measures (FA, MD, PD, and TD) from the multiband diffusion MRI dataset by several fully automated analysis approaches: two voxel-based analyses and an atlas

ROI-based analysis. Our results can be summarized as follows: (i) Moderate to high reproducibility of diffusion measures from the multiband EPI sequence were observed for both voxel-based and ROI-based analyses; (ii) FA exhibited higher reproducibility than other diffusion measures, while MD showed the lowest ICCs; (iii) ICC values varied significantly across different tract ROIs, especially for skeleton WM; and (iv) Commissural tracts showed higher reproducibility than other categories of tracts (projection, association and brainstem), while the tracts of the brainstem showed the poorest reliability.

To date, there are only two studies that examine the TRT reliability of multiband EPI data. Zuo et al²⁷ used the multiband resting-state fMRI dataset to examine the TRT reliability of functional homogeneity in the human brain, and their results showed that test-retest reliability was significantly improved by usage of a fast imaging sequence. Liao et al²⁸ used the same dataset to identify global hubs in the human voxel-wise functional brain networks and examine TRT reliability over scanning time. The present study acts as a necessary supplement focusing on diffusion MRI. The multiband EPI sequence with its relatively short scan time can reduce the effects of head motion,¹⁰⁻¹² and the large number of gradient directions can increase the reliability of tensor estimation. The high reproducibility of diffusion

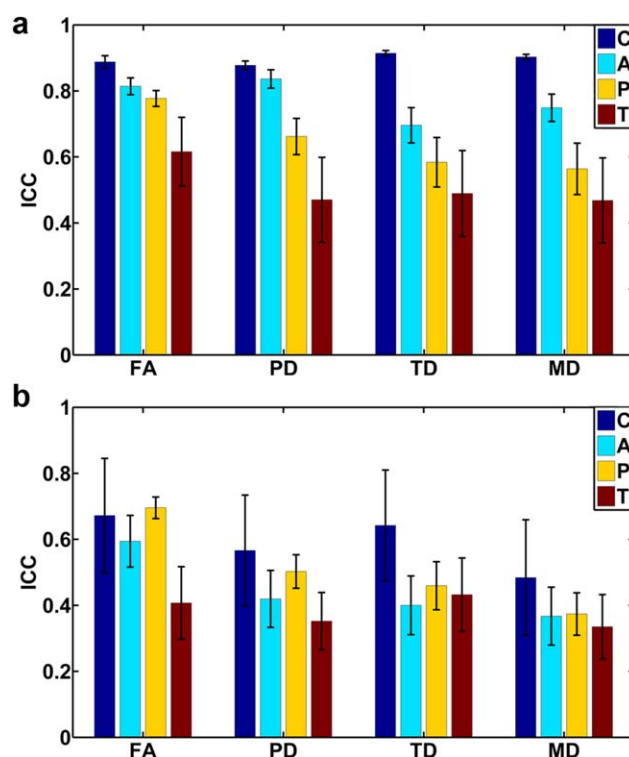


FIGURE 5: Reliability differences among different types of tracts in whole-brain (A) and skeleton WM (B). The bars and errorbars represent the mean values and standard deviations of the ICC values of all WM tracts, respectively. C, commissural tract; A, association tract; P, projection tract; T, tract in the brainstem.

metrics within WM suggests the applicability of the current sequence for clinical studies, where constraints on patient scanning time are an important issue.

First, the normal values of all diffusion measures across WM voxels from the multiband DTI dataset are comparable with the diffusion values from conventional DTI data.^{15,22,24,38,39} This confirms no much effect of the fast-collecting multiband EPI sequence on the normal values of diffusion measures in cerebral WM. From the reliability analysis of diffusion metrics at the voxel level, high consistency of the ICC distributions between whole-brain and skeleton WM were observed. The ICC results were also consistent with the findings of the previous reliability studies from conventional DTI sequences.^{40,41} Specifically, a recent TBSS study of WM reliability reported similar ICC values and distributions of the WM skeleton to our findings.⁴² Moreover, we found that the diffusion measures from TBSS are less reliable than those from VBA. Compared with the VBA analysis, the TBSS analysis focused on the main WM skeleton (major WM tracts) of each subject, with the advantage of minimizing the effects of tract misalignment encountered by conventional VBA methods.³¹ However, the step of skeleton extraction and projection may also bring in some biases for the estimation of diffusion measures, especially for the fiber-crossing areas. Madhyastha et al (2014) found that within the WM skeleton, there are a large percentage of voxels (> 16%) that do not correspond to each other, making it difficult to find reliable changes in these areas, which may account for the lower proportion of WM voxels with good reliability observed in TBSS analysis compared with VBA analysis.

In the present study, all of the analysis approaches were fully automated and normalized to the standard space. Transformation into template space may reduce the between-subjects variation, resulting in lower ICC values and indicating a possible reduced sensitivity.⁴¹ On the other hand, analysis in the standard space will avoid the bias from intra-rater reliability and effects of tractography protocol to some extent, but will include some errors from image registration accuracy. Different analysis methods have their own advantages and disadvantages and might provide complementary information.

For different diffusion measures, FA showed a higher reliability than the other diffusion measures, while MD showed the lowest reliability for both voxel-based and ROI-based analyses. These results were consistent with previous findings of the reliability differences in diffusion metrics from conventional EPI sequences.¹⁹ Given FA is a ratio measure and MD is the sum of primary and transverse diffusivity, one explanation for the lower reliability in MD is that because MD is the mean of the diffusion tensor eigenvalues, it can reflect variability due to changes in the overall

magnitude of diffusivity, anisotropy of diffusivity, or a combination of the two effects.

Characterization of regional variation in measurement error for DTI is important for the interpretation of the results of group comparisons and longitudinal studies. Previous studies have explored the reproducibility of some major WM tracts, such as the corpus callosum, cingulum bundles, and uncinate fasciculus by fiber tractography or ROI-based approaches.^{41,43,44} In the present study, we investigated the reliability of all 48 tracts in the WM atlas of JHU ICBM-DTI-81. The most reliable measurements were for the fornix, corpus callosum, cingulum, superior longitudinal fasciculus, sagittal striatum, tapetum, posterior corona radiata, and posterior thalamic radiation. When the tracts were classified into different categories, we found that commissural tracts showed higher reliability than other categories of tracts. The commissural tracts are located in five ROIs: the genu, the body and the splenium of corpus callosum and the bilateral tapetum, which is the temporal component of the corpus callosum that is partitioned separately from the other corpus callosum regions. The high reliability of the corpus callosum is consistent with the previous findings from conventional DTI sequences.^{18,26,44} This may be because commissural tracts are composed of tightly packed fibers with highly consistent orientation and low anatomical variability,⁴⁵ are less prone to noise and partial volume effects and are less likely to be affected by “crossing” or “kissing” fibers. Additionally, the larger ROI sizes of commissural tracts result in a higher SNR and better reproducibility.

In contrast, some tiny WM tracts in brainstem showed poor reliability, such as the inferior and superior cerebellar peduncle. These tracts are relatively small with complex fiber architecture and have been seldom explored by diffusion MRI.⁴⁶ Previous studies have suggested that a small ROI is more prone to noise and partial volume effects from imperfect coregistration and interpolation and is more likely to show a greater variation.^{41,44} Additionally, the higher amount of variability encountered during the scanning process may occur because some tracts of the brainstem are located in deep gray matter regions with many different tracts passing through, and scans may be seriously affected by crossing fibers.⁴⁶ Increasing ROI size improves reproducibility, as long as contamination from surrounding structures with markedly different voxel values is avoided.⁴⁴ In future application studies, interpretation of the results from these WM regions or tracts should be undertaken cautiously.

There are some study limitations that need to be addressed. First, this study only used 11 subjects to calculate ICC. To obtain more accurate reliability results, future ICC reliability research should have a relatively large sample to obtain sufficient statistical power. Second, our findings suggest that whereas current techniques produce diffusion

measures capable of characterizing the genuine between-subject differences in WM connectivity, the moderate values of ICC indicate that future work must improve the acquisition protocols of multiband diffusion MRI to improve the data quality and the reliability. Third, investigation of the effects of different acquisition parameters, gradient sampling schemes and advanced diffusion modeling approaches, such as application of higher order models to disentangle crossing fiber structures,⁴⁷ on the reproducibility of diffusion metrics for this new sequence would be interesting, but was unfortunately outside the scope of this study. Finally, when monitoring longitudinal changes, it is important to consider the tradeoff between reliability and sensitivity of diffusion measures. In future studies, several measures (e.g., the coefficient of variation) can be further developed to comprehensively characterize the sensitivity of diffusion measures over scanning sessions.

In conclusion, our work demonstrated moderate to high reliability of diffusion measures in cerebral WM from the multiband diffusion MRI sequence. The findings suggest the potential utility of the multiband EPI sequence and may provide reference and guidance for future WM studies using the new sequence.

Acknowledgments

Contract grant sponsor: 973 program; Contract grant number: 2013CB837300; Contract grant sponsor: the Natural Science Foundation of China; Contract grant number: 81471732; Contract grant number: 31221003; Contract grant number: 81030028; Contract grant sponsor: the National Science Fund for Distinguished Young Scholars of China; Contract grant number: 81225012; Contract grant sponsor: the Beijing New Medical Discipline Based Group; Contract grant number: 100270569; Contract grant sponsor: the Fundamental Research Funds for the Central Universities; Contract grant number: 2013YB28.

References

1. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J* 1994;66:259–267.
2. Le Bihan D, Mangin JF, Poupon C, et al. Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging* 2001;13:534–546.
3. Lebel C, Gee M, Camicioli R, Wieler M, Martin W, Beaulieu C. Diffusion tensor imaging of white matter tract evolution over the lifespan. *Neuroimage* 2012;60:340–352.
4. Mori S, van Zijl PC. Fiber tracking: principles and strategies - a technical review. *NMR Biomed* 2002;15:468–480.
5. Pierpaoli C, Basser PJ. Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med* 1996;36:893–906.
6. Pfefferbaum A, Sullivan EV, Hedehus M, Lim KO, Adalsteinsson E, Moseley M. Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. *Magn Reson Med* 2000;44:259–268.
7. McKinstry RC, Mathur A, Miller JH, et al. Radial organization of developing preterm human cerebral cortex revealed by non-invasive water diffusion anisotropy MRI. *Cereb Cortex* 2002;12:1237–1243.
8. Liu Y, Duan Y, He Y, et al. Whole brain white matter changes revealed by multiple diffusion metrics in multiple sclerosis: a TBSS study. *Eur J Radiol* 2012;81:2826–2832.
9. Li H, Liang Y, Chen K, et al. Different patterns of white matter disruption among amnesic mild cognitive impairment subtypes: relationship with neuropsychological performance. *J Alzheimers Dis* 2013;36:365–376.
10. Andersson JL, Skare S. A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. *Neuroimage* 2002;16:177–199.
11. Aksoy M, Liu C, Moseley ME, Bammer R. Single-step nonlinear diffusion tensor estimation in the presence of microscopic and macroscopic motion. *Magn Reson Med* 2008;59:1138–1150.
12. Rohde GK, Barnett AS, Basser PJ, Marengo S, Pierpaoli C. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn Reson Med* 2004;51:103–114.
13. Moeller S, Yacoub E, Olman CA, et al. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* 2010;63:1144–1153.
14. Van Essen DC, Ugurbil K, Auerbach E, et al. The Human Connectome Project: a data acquisition perspective. *Neuroimage* 2013;62:2222–2231.
15. Brander A, Kataja A, Saastamoinen A, et al. Diffusion tensor imaging of the brain in a healthy adult population: normative values and measurement reproducibility at 3 T and 1.5 T. *Acta Radiol* 2010;51:800–807.
16. Cercignani M, Bammer R, Sormani MP, Fazekas F, Filippi M. Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. *AJNR Am J Neuroradiol* 2003;24:638–643.
17. Ciccarelli O, Parker GJ, Toosy AT, et al. From diffusion tractography to quantitative white matter tract measures: a reproducibility study. *Neuroimage* 2003;18:348–359.
18. Pfefferbaum A, Adalsteinsson E, Sullivan EV. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J Magn Reson Imaging* 2003;18:427–433.
19. Heiervang E, Behrens TE, Mackay CE, Robson MD, Johansen-Berg H. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage* 2006;33:867–877.
20. Gao W, Zhu H, Lin W. A unified optimization approach for diffusion tensor imaging technique. *Neuroimage* 2009;44:729–741.
21. Jones D, Horsfield M, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn Reson Med* 1999;42:515–525.
22. Farrell JA, Landman BA, Jones CK, et al. 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* 2007;26:756–767.
23. Jones DK. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. *Magn Reson Med* 2004;51:807–815.
24. Landman BA, Farrell JA, Jones CK, Smith SA, Prince JL, Mori S. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Neuroimage* 2007;36:1123–1138.
25. Bisdas S, Bohning DE, Besenski N, Nicholas JS, Rumboldt Z. 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* 2008;29:1128–1133.
26. Marengo S, Rawlings R, Rohde GK, et al. Regional distribution of measurement error in diffusion tensor imaging. *Psychiatry Res* 2006;147:69–78.

27. Zuo XN, Xu T, Jiang L, et al. Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage* 2013;65:374–386.
28. Liao XH, Xia MR, Xu T, et al. Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. *Neuroimage* 2013;83:969–982.
29. Xu J, Moeller S, Auerbach EJ, et al. Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage* 2013;83:991–1001.
30. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson* 1996;111:209–219.
31. Smith SM, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 2006;31:1487–1505.
32. Mori S, Oishi K, Jiang H, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage* 2008;40:570–582.
33. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
35. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med* 2002;21:3431–3446.
36. Kong J, Gollub RL, Webb JM, Kong JT, Vangel MG, Kwong K. Test-retest study of fMRI signal change evoked by electroacupuncture stimulation. *Neuroimage* 2007;34:1171–1181.
37. Braun U, Plichta MM, Esslinger C, et al. Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 2012;59:1404–1412.
38. Pagani E, Hirsch JG, Pouwels PJ, et al. Intercenter differences in diffusion tensor MRI acquisition. *J Magn Reson Imaging* 2010;31:1458–1468.
39. Wakana S, Caprihan A, Panzenboeck MM, et al. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 2007;36:630–644.
40. Jansen JF, Kooi ME, Kessels AG, Nicolay K, Backes WH. Reproducibility of quantitative cerebral T2 relaxometry, diffusion tensor imaging, and 1H magnetic resonance spectroscopy at 3.0 Tesla. *Invest Radiol* 2007;42:327–337.
41. Vollmar C, O’Muircheartaigh J, Barker GJ, et al. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *Neuroimage* 2010;51:1384–1394.
42. Madhyastha T, Merillat S, Hirsiger S, et al. Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. *Hum Brain Mapp* 2014;35:4544–4555.
43. Danielian LE, Iwata NK, Thomasson DM, Floeter MK. Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. *Neuroimage* 2010;49:1572–1580.
44. Papinutto ND, Maule F, Jovicich J. Reproducibility and biases in high field brain diffusion MRI: an evaluation of acquisition and analysis variables. *Magn Reson Imaging* 2013;31:827–839.
45. Burgel U, Amunts K, Hoemke L, Mohlberg H, Gilsbach JM, Zilles K. White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *Neuroimage* 2006;29:1092–1105.
46. Stieltjes B, Kaufmann WE, van Zijl PC, et al. Diffusion tensor imaging and axonal tracking in the human brainstem. *Neuroimage* 2001;14:723–735.
47. Tournier JD, Yeh CH, Calamante F, Cho KH, Connelly A, Lin CP. Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* 2008;42:617–625.