Clin Neuroradiol (2016) 26:199–208
DOI 10.1007/s00062-014-0349-8

ORIGINAL ARTICLE

Reproducibility of Tract-based and Region-of-Interest DTI Analysis of Long Association Tracts

N. Brandstack · T. Kurki · J. Laalo · T. Kauko · O. Tenovuo

Received: 29 May 2014 / Accepted: 14 September 2014 / Published online: 5 Oktober 2014

Abstract
Purpose Reproducibility of two different methods for quantifying fiber tracts by using a diffusion tensor imaging (DTI) sequence suitable for clinical magnetic resonance imaging (MRI) protocols was evaluated.
Methods DTI of 15 subjects was used to analyze intra-rater and inter-rater reproducibility. Another 10 subjects underwent MRI twice for assessment of between-scan reliability. Ten long association tracts were defined by fiber tracking using inclusion and exclusion regions of interest (ROIs). Whole-tract analysis and tractography-based core analysis were performed, and the effect of fractional anisotropy (FA 0.15/0.30) and turning angle threshold (27°/60°) on reproducibility was evaluated. Additionally, ROI measurements were performed in the core of the tracts.
Results For the tract-based methods, intra-rater and inter-rater reliabilities of FA and mean diffusivity (MD) measurements were excellent. Between-scan reproducibility was good or excellent in 127 of 130 of the measurements. There was no systematic difference in the reproducibility of the FA, MD, and volume measurements depending on the FA or turning angle threshold. For the cross-sectional ROI measurements, reliability showed large variation from poor to excellent depending on the tract.
Conclusions Compared with the commonly used cross-sectional core ROI method, the tract-based analyses seem to be a more robust way to identify and measure white matter tracts of interest, and provide a novel reproducible tool to perform core analysis.

Keywords Reproducibility · DTI tractography · Core ROI method · Tract-based analyses

N. Brandstack, MD, PhD (✉)
Department of Radiology, Helsinki University Hospital,
PL 340, 00029 HUS, Finland
e-mail: nina.brandstack@hus.fi

T. Kurki, MD, PhD
Department of Radiology, Terveystalo Turku Pulssi Medical Centre, Turku, Finland

J. Laalo, MD
Department of Radiology, Turku University Hospital,
PL 52, 20521 Turku, Finland

T. Kauko, MSc
Department of Biostatistics, University of Turku,
Turku, Finland

O. Tenovuo, MD, PhD
Department of Neurology, Turku University Hospital,
PL 52, 20521 Turku, Finland

Introduction

Magnetic resonance (MR) diffusion tensor imaging (DTI) is an imaging technique that can be used to measure and visualize the organization of white matter and to show abnormalities not visible on routine images [1]. Different tools for quantitative DTI analysis have been developed, but the reproducibility of the methods suitable for clinical protocols has been studied less.

In clinical DTI studies, tracts of interest are often identified on cross-sectional images, and the measurements are performed by drawing regions of interest (ROIs). Limitations related to the manual ROI definition include inconsistency in placing ROIs and contamination from the cerebrospinal fluid (CSF) or gray matter, as well as inclusion of white matter from different projections [2, 3]. Results on

reliability concerning the cross-sectional ROI measurements in a healthy population have been somewhat controversial [4–6], and the reproducibility of the ROI measurements has been shown to vary regionally [3, 5]. Another tool used to identify specific white matter tracts is quantitative DTI tractography that, based on an algorithm to group pixels according to their tensor properties, allows a given tract to be delineated [7], and provides tract-specific volume, mean fractional anisotropy (FA), and mean diffusivity (MD). The semi-automated tractography-based MD and FA measurements have been reported to provide good to excellent intra- and inter-rater reliability in several white matter (WM) tracts [8, 9], whereas the reproducibility of tract volume measurements is generally lower [9, 10]. The large inter-individual structural and volume variability of tractograms [1, 11] can decrease the utility of tractography in DTI analysis, and additional measurements from the core of the tract are commonly used to confirm the tract-based results [12, 13]. Tractography provides a novel alternative possibility to perform the central-part analysis in corresponding volumes representing the region with highest FA values in each subject [12], but the reproducibility of this approach is unknown.

Association tracts are fiber bundles that travel to other cortical areas in the same hemisphere and they are essential, for instance, for many cognitive functions of the brain [14]. Association pathways, especially frontal association trajectories, belong to the most commonly damaged tracts after traumatic brain injury (TBI) [1, 15–18]. The intra- and inter-rater reliability for the tractography-based FA and MD measurements of the association tracts has been shown to be mainly good or excellent [9, 10–19]. For example, for the uncinate fasciculus, one of the most studied association tracts, the inter-rater and intra-rater intra-class correlations for the tractography-based FA and MD measurements have varied between 0.81 and 1.00 [9, 10, 13, 19]. Some previous reliability studies on the cross-sectional ROI measurements of the association tracts have also shown rather good reliability [6, 20]. There are only few comparisons of the reproducibility of DTI analysis between the ROI method and tractography [2, 21–24], and the reliability of novel tractography-based core analysis is unknown. Our purpose was to evaluate the reproducibility of tractography-based measurements, including the tractography-based core analysis, compared with conventional core ROI measurements by using a DTI sequence easily achievable for clinical use.

Methods

Subjects

The subjects' informed consents were obtained. All human studies have been approved by the appropriate ethics committee and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Group A: DTI images of the 15 participants ($n=10$ patients with TBI sequels and 5 healthy subjects; mean age: 39.7 ± 12.2 years, range: 18–58 years) were randomly selected from a larger population that had been studied with DTI methodology.

Group B: Ten participants ($n=3$ patients with TBI sequels and 7 healthy subjects; mean age: 36.0 ± 11.4 years, range: 18–55 years) underwent DTI examination twice in the same scanner with the same imaging protocol.

MR Data Acquisition

MR imaging (MRI) was performed at 3 T (Achieva, Philips Medical Systems, Best, The Netherlands) using a sensitivity encoding (SENSE) eight-channel transmit–receive head coil. The 3DT1 and DTI imaging protocol consisted of the following:

1) sagittal 3DT1 turbo field echo images (repetition time (TR)/echo time (TE)/TI: 8.3/3.8/1032 ms, 165 slices with 1.0-mm thickness, 0.0-mm gap; 242 × 288r matrix, turbo factor: 240, flip angle: 8°, field of view (FOV): 244 mm, reduced field of view (RFOV): 105 %, number of excitations: 1, imaging time: 5 min 29 s);
2) transverse DTI images; diffusion-weighted turbo spin echo echo planar imaging (EPI) images (TR/TE: 5877/62, 60 slices with 2.0-mm thickness, 0.0-mm gap, 112 × 128r matrix, turbo factor: 59, EPI factor: 59, FOV: 224 mm, RFOV: 100 %, number of excitations: 2, imaging time: 3 min 52 s); b values of 0 and 800 s/mm^2 and 15 different gradient encoding directions were used, and isotropic images with 2.0 × 2.0 × 2.0 mm voxel size reconstructed into 2.0 × 1.75 × 1.75 mm voxel size were obtained. The images were postprocessed with the Philips Diffusion Registration Tool (Philips, Medical Systems, Best, The Netherlands) to remove distortions and misalignments due to shear and eddy current as well as head motion [25, 26]. All transverse images were obtained according to the line between the lower border of the genu and splenium of the corpus callosum. DTI images were fused with the 3DT1 images to obtain more exact anatomical landmarks before the subsequent DTI-based analysis.

MR Data Analysis

Group A: The tract-based and cross-sectional DTI analyses were assessed by three raters (TK, NB, and JL) to evaluate the inter-rater reliability. One rater (TK) was selected

to repeat the measures after a 2-week period to evaluate the intra-rater reliability.

Group B: To study the between-scan reproducibility, two repeated tract-based DTI analysis approaches using data from the DTI examinations at two different time point (8–210 days apart) were assessed for each of the 10 participants by one rater (TK).

One of the raters (TK with 14 years of experience in neuroradiology) had 4 years' experience in DTI data acquisition and analysis, and provided DTI-specific training for the other two raters (JL, neuroradiologist, and NB, radiology resident). Tract integrity was quantified by FA and MD values. The principal fiber orientation within each pixel was visualized by color-coded orientation maps (red: right–left, blue: superior–inferior, green: anterior–posterior). Regardless of the analysis approach used, the same fused DTI and 3DT1 images with a FA color map for each subject was used for each subsequent analysis step.

Semi-Automated Tract-Based Quantification

Deterministic DTI tractography (FiberTrak package, Philips) was performed, and the tracts were defined by means of two or three free-hand inclusion ROIs, placed in standard positions according to anatomical landmarks.

ROIs were drawn on the baseline image of the DTI scans. All ROIs were placed bilaterally. For uncinate fasciculus (UF), two ROIs were drawn on each side of the same coronal slice, one just anterior to the temporal stem where the frontal cortex begins and the other in the temporal pole. For the superior cingulum (SC) bundle, two ROIs were drawn in the antero-posterior direction on coronal slices around the cingulum on each side, posterior ROIs at the level of the upper back part of the aqueduct and anterior ROIs to between the anterior commissure and back part of the rostrum of the corpus callosum. For the superior longitudinal fasciculus (SLF), two ROIs were drawn on coronal slices. The anterior ROI was drawn in just behind the mamillary corpuses, and the posterior ROI was drawn in at the level of the upper rear part of the aqueduct. Connections to the temporal lobe were manually removed with exclusion ROIs at the level of the middle part of the splenium, and this removed part of the SLF formed the arcuate fasciculus (AF). For the inferior occipito-frontal fasciculus (IFOF), two ROIs were drawn on coronal slices, one just anterior to the temporal stem where the frontal cortex begins and one in the occipital cortex in the most anterior slice where the occipital cortex was visualized. After the ROIs were delineated and the fiber tracts were reconstructed, aberrant fibers were manually removed with exclusion ROIs to include only fibers within the desired tract. Fibers crossing the midline and fibers to thalami or brainstem were removed from all trajectories. Volume and FA and MD values from the remaining fibers were used for subsequent statistical analysis.

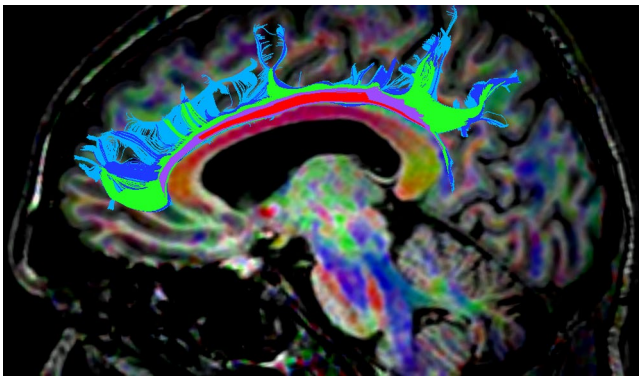

Fig. 1 An example of the tractography results in the superior cingulum of a research participant. A fusion of the DTI and 3D T1 images, where the fiber tracts were reconstructed with the following fractional anisotropy and turning angle thresholds: 15/60° (*light blue*), 15/27° (*blue*), and 30/27° (*green*), and the volume-based analysis in the central part of the tract (tractography-based core analysis) was performed with two predetermined sizes: 3 cm^3 (*purple*) and 1 cm^3 (*red*)

Group A: DTI tractography was performed using the following FA and turning angle threshold to terminate the tracking process: FA: 0.15/angle: 27° (UF, SC, SLF, and IFOF) or FA: 0.15/angle: 60° (AF). Additional volume-based analysis in the central part of the tract was performed, by repeating fiber tracking at increased FA thresholds with the same ROIs until the desired volume was achieved. Tract volumes closest to the predetermined size were used in the study: 3 cm^3 for UFs, SCs, and AFs; 6 cm^3 for SLFs and IFOFs.

Group B: The tractography was performed by using two different FAs and turning angle thresholds for each tract: FA: 0.15/angle: 27°; FA: 0.15/angle: 60°; FA: 0.30/angle: 27° (Fig. 1). The additional volume-based analysis in the central part of the tract was performed with two predetermined sizes for each tract: (3 cm^3 and 1 cm^3 for the both UFs, SCs, and AFs; 6 cm^3 and 2 cm^3 for the both SLFs and IFOFs; Fig. 1).

Cross-sectional ROI Quantification

The UF, SC, SLF, AF, and IFOF were delineated bilaterally with defined standard ROI sizes (5–8 voxels for UF, AF, and IFOF; 10–15 voxels for SC and SLF) and locations (Fig. 2). The manually defined ROIs were drawn consulting FA maps, color maps, and a white matter atlas. The mean FA and MD values from the voxels within the manually defined ROIs were recorded for subsequent statistical analyses. Each ROI

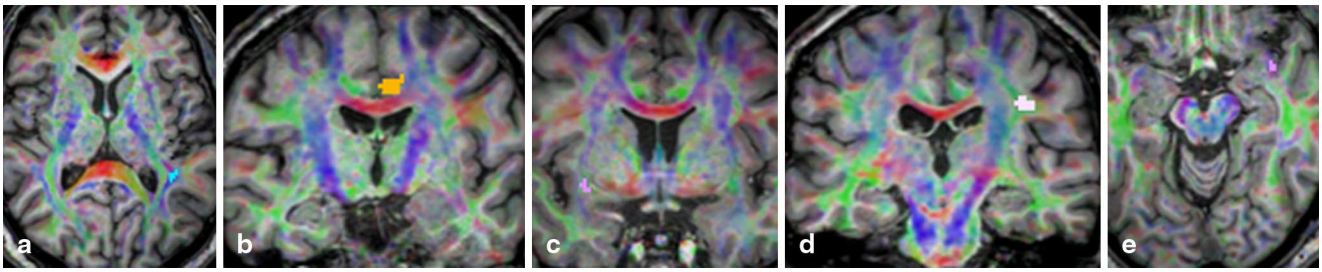

Fig. 2 Cross-sectional ROI quantification locations from *left* to *right*: **a** arcuate fasciculus (*light blue* ROI): *blue* in the axial slice, level of the middle splenium of the corpus callosum; **b** cingulum (*yellow* ROI): *green* in the coronal slice, level of the corpus mamillare; **c** inferior fronto-occipital fasciculus (*pink* ROI): *green* in the coronal slice, above the *blue* UF; **d** superior longitudinal fasciculus (*light pink* ROI): *green* in the axial slice, level of the upper back part of the aqueduct; **e** uncinate fasciculus (*pink* ROI): *blue* in the axial slice, level of the upper temporal lobe

measurement was repeated three times, and values from the ROI with the highest FA were used for further analysis.

Statistical Data Analysis

To assess the inter-rater and intra-rater reliability of each analysis method, intra-class correlation coefficients (ICCs) for mean tract volumes, for mean FA and for mean MD of each association tract were calculated. Separate two-way random effects models (i.e., ICC (2,1)), where raters are assumed to form a random subset of all possible raters, were used to calculate ICCs between raters (TK, NB, and JL) and within a rater (TK) for each approach [20]. This model was chosen because it is an appropriate model to assess inter-rater reliability with more than two raters, as well as intra-rater reliability with multiple values from one rater [27, 28]. ICC values greater than 0.75 were considered to demonstrate excellent reliability, while values between 0.40 and 0.75 indicated fair to good reliability, and below 0.40 poor reliability [21].

An additional Bland–Altman reliability analysis was done by calculating the coefficient of repeatability (CR), which represents the value below which the absolute difference between two repeated test results may be expected to lie with a probability of 95%. For comparison with literature data, the coefficient of variation (CV) was also computed according to the equation, where the CV is defined as the ratio of the standard deviation to the mean.

Results

Intra-rater Reliability (Group A)

Results from each ICC and Bland–Altman analysis are summarized in Table 1 and Fig. 3.

For both *tract-based methods*, intra-rater reliability of FA and MD measurements was found to be excellent (ICC range: 0.907–0.999) except for the whole-tract MD measurements in the right UF that showed good intra-rater reliability (ICC = 0.738). Bland–Altman CR values for all the tract-based FA and MD measurements were less than 5% (less than 3% in the central part of the tract) from the respective mean FA and MD values. Coefficient of variation estimates of the tract-based intra-rater variability in the MD were in the range of 0.0–0.8%, and 0.1–1.1% in the FA. Also, measurements for the whole-tract volume showed excellent intra-rater reliability (ICC range: 0.831–0.998).

Intra-rater reliability for the *cross-sectional method* was fair to excellent for FA measurements (ICC range: 0.418–0.813) in all tracts except for the left AF (ICC = 0.271, $p = 0.164$), and for MD measurements (ICC range: 0.515–0.810) in all tracts except for the left AF (ICC = 0.314) and the left UF (ICC = 0.228). All the CR values for the cross-sectional ROI-based FA measurements were more than 10% from the respective mean FA. The Bland–Altman CR values for the cross-sectional ROI-based MD measurements in SCs, SLFs, AFs, and in the right IFOF were less than 10% from the respective mean MD. Coefficient of variation estimates of variability in the MD were in the range of 1.2–3.9%, and 4.1–7.9% in the FA.

Inter-rater Reliability (Group A)

Results for each ICC and Bland–Altman analysis are summarized in Table 1 and Fig. 3.

For both *tract-based methods*, inter-rater reliability of the FA and MD measurements was found to be excellent (ICC range: 0.761–0.997). Bland–Altman CR values for the FA and MD were less than 5% from the respective mean tract FA and MD, except for the FA of the left AF (7.1%). All the CR values for the FA and MD measurements in the middle of the tract were less than 6% (FA) and 2% or less (MD) from the respective mean FA and MD. Coefficient of variation estimates of the tract-based inter-rater variability for MD were in the range of 0.1–1.0%, and 0.6–1.8% for FA. Also,

Table 1 Inter-rater and *intra-rater* reliability of each analysis method for uncinate fasciculus, superior cingulum, superior longitudinal fasciculus, arcuate fasciculus, and inferior fronto-occipital fasciculus

Fasciculus	Analysis	ICC (±95% CI) Inter-rater	*Intra-rater*		ICC (±95% CI) Inter-rater	*Intra-rater*
UF dx	***Tract based***			**sin**		
	Mean V tot	0.971 (0.940–0.986)	*0.972 (0.920–0.990)*		0.681 (0.428–0.835)	*0.914 (0.769–0.970)*
	Mean FA tot	0.980 (0.959–0.991)	*0.968 (0.910–0.989)*		0.967 (0.932–0.984)	*0.989 (0.967–0.996)*
	Mean MD tot	0.943 (0.884–0.973)	*0.738 (0.391–0.902)*		0.882 (0.766–0.942)	*0.922 (0.790–0.973)*
	Mean FA std	0.987 (0.972–0.994)	*0.981 (0.947–0.994)*		0.948 (0.894–0.975)	*0.999 (0.996–1.000)*
	Mean MD std	0.988 (0.975–0.994)	*0.922 (0.789–0.972)*		0.973 (0.943–0.987)	*0.999 (0.996–1.000)*
	ROI based					
	Mean FA	**0.388** (0.035–0.654)	*0.813 (0.539–0.932)*		0.502 (0.176–0.728)	*0.612 (0.177–0.847)*
	Mean MD	**0.179** (−0.190–0.504)	*0.515 (0.036–0.802)*		**0.388** (0.035–0.654)	***0.228*** *(−0.293–0.645)*
SC dx	***Tract based***			**sin**		
	Mean V tot	0.892 (0.784–0.947)	*0.998 (0.995–0.999)*		0.893 (0.787–0.948)	*0.985 (0.958–0.995)*
	Mean FA tot	0.967 (0.931–0.984)	*0.986 (0.960–0.995)*		0.915 (0.829–0.959)	*0.969 (0.911–0.989)*
	Mean MD tot	0.989 (0.976–0.995)	*0.994 (0.982–0.998)*		0.971 (0.940–0.986)	*0.994 (0.984–0.998)*
	Mean FA std	0.984 (0.966–0.992)	*0.996 (0.989–0.999)*		0.911 (0.820–0.957)	*0.997 (0.992–0.999)*
	Mean MD std	0.988 (0.976–0.994)	*0.998 (0.994–0.999*		0.997 (0.993–0.998)	*0.991 (0.975–0.997)*
	ROI based					
	Mean FA	0.677 (0.422–0.833)	*0.765 (0.441–0.912)*		**0.361** (0.004–0.636)	*0.766 (0.444–0.912)*
	Mean MD	0.673 (0.415–0.830)	*0.810 (0.533–0.930)*		0.679 (0.425–0.834)	*0.621 (0.190–0.851)*
SLF dx	***Tract based***			**sin**		
	Mean V tot	0.911 (0.820–0.957)	*0.961 (0.891–0.986)*		0.639 (0.364–0.811)	*0.831 (0.578–0.939)*
	Mean FA tot	0.933 (0.863–0.968)	*0.982 (0.948–0.994)*		0.899 (0.797–0.951)	*0.969 (0.912–0.989)*
	Mean MD tot	0.974 (0.947–0.988)	*0.996 (0.988–0.999)*		0.980 (0.959–0.991)	*0.990 (0.971–0.997)*
	Mean FA std	0.969 (0.936–0.985)	*0.983 (0.952–0.994)*		0.864 (0.734–0.933)	*0.987 (0.962–0.995)*
	Mean MD std	0.995 (0.989–0.997)	*0.983 (0.951–0.994)*		0.941 (0.880–0.972)	*0.972 (0.920–0.990)*
	ROI based					
	Mean FA	0.690 (0.442–0.840)	*0.641 (0.222–0.860)*		**0.255** (−0.113–0.561)	*0.624 (0.196–0.853)*
	Mean MD	**0.188** (−0.182–0.511)	*0.742 (0.397–0.903)*		0.565 (0.261–0.768)	*0.762 (0.436–0.911)*
AF dx	***Tract based***			**sin**		
	Mean V tot	0.889 (0.750–0.953)	*0.929 (0.768–0.980)*		0.821 (0.656–0.911)	*0.931 (0.811–0.976)*
	Mean FA tot	0.886 (0.739–0.953)	*0.928 (0.765–0.979)*		0.761 (0.554–0.879)	*0.958 (0.883–0.985)*
	Mean MD tot	0.894 (0.755–0.956)	*0.964 (0.879–0.990)*		0.917 (0.833–0.960)	*0.957 (0.880–0.985)*
	Mean FA std	0.968 (0.923–0.987)	*0.982 (0.936–0.995)*		0.932 (0.861–0.967)	*0.997 (0.991–0.999)*
	Mean MD std	0.929 (0.832–0.971)	*0.937 (0.791–0.982)*		0.939 (0.875–0.971)	*0.988 (0.965–0.996)*
	ROI based					
	Mean FA	**0.182** (−0.265–0.565)	*0.418 (−0.189–0.794)*		**0.174** (−0.195–0.500)	***0.271*** *(−0.250–0.671)*
	Mean MD	0.447 (0.025–0.734)	*0.698 (0.223–0.905)*		**0.021** (−0.339–0.375)	***0.314*** *(−0.206–0.695)*
IFOF dx	***Tract based***			**sin**		
	Mean V tot	0.775 (0.555–0.887)	*0.978 (0.936–0.992)*		0.762 (0.555–0.880)	*0.977 (0.936–0.992)*
	Mean FA tot	0.939 (0.876–0.971)	*0.992 (0.976–0.997)*		0.933 (0.864–0.968)	*0.973 (0.923–0.991)*
	Mean MD tot	0.867 (0.738–0.935)	*0.987 (0.961–0.995)*		0.879 (0.761–0.941)	*0.907 (0.753–0.967)*
	Mean FA std	0.924 (0.847–0.963)	*0.988 (0.965–0.996)*		0.842 (0.693–0.922)	*0.974 (0.926–0.991)*
	Mean MD std	0.96 (0.924–0.983)	*0.996 (0.989–0.999)*		0.978 (0.954–0.990)	*0.993 (0.979–0.998)*
	ROI based					
	Mean FA	0.575 (0.274–0.773)	*0.735 (0.385–0.900)*		**0.262** (−0.105–0.566)	*0.473 (−0.020–0.781)*
	Mean MD	0.471 (0.136–0.709)	*0.721 (0.359–0.894)*		0.670 (0.412–0.829)	*0.578 (0.125–0.831)*

Results with poor reproducibility (ICC < 0.400) are given in bold

ICC intra-class correlation, *UF* uncinate fasciculus, *SC* superior cingulum, *SLF* superior longitudinal fasciculus, *AF* arcuate fasciculus, *IFOF* inferior fronto-occipital fasciculus, *dx* right, *sin* left, *std* standard volume in middle of tract, *tot* total

measurements for the whole-tract volume showed excellent or good inter-rater reliability (ICC range: 0.639–0.971).

Inter-rater reliability for the *cross-sectional method* was fair to good for FA measurements (ICC range: 0.502–0.690) in the left UF, right SC, right SLF, and right IFOF, and for

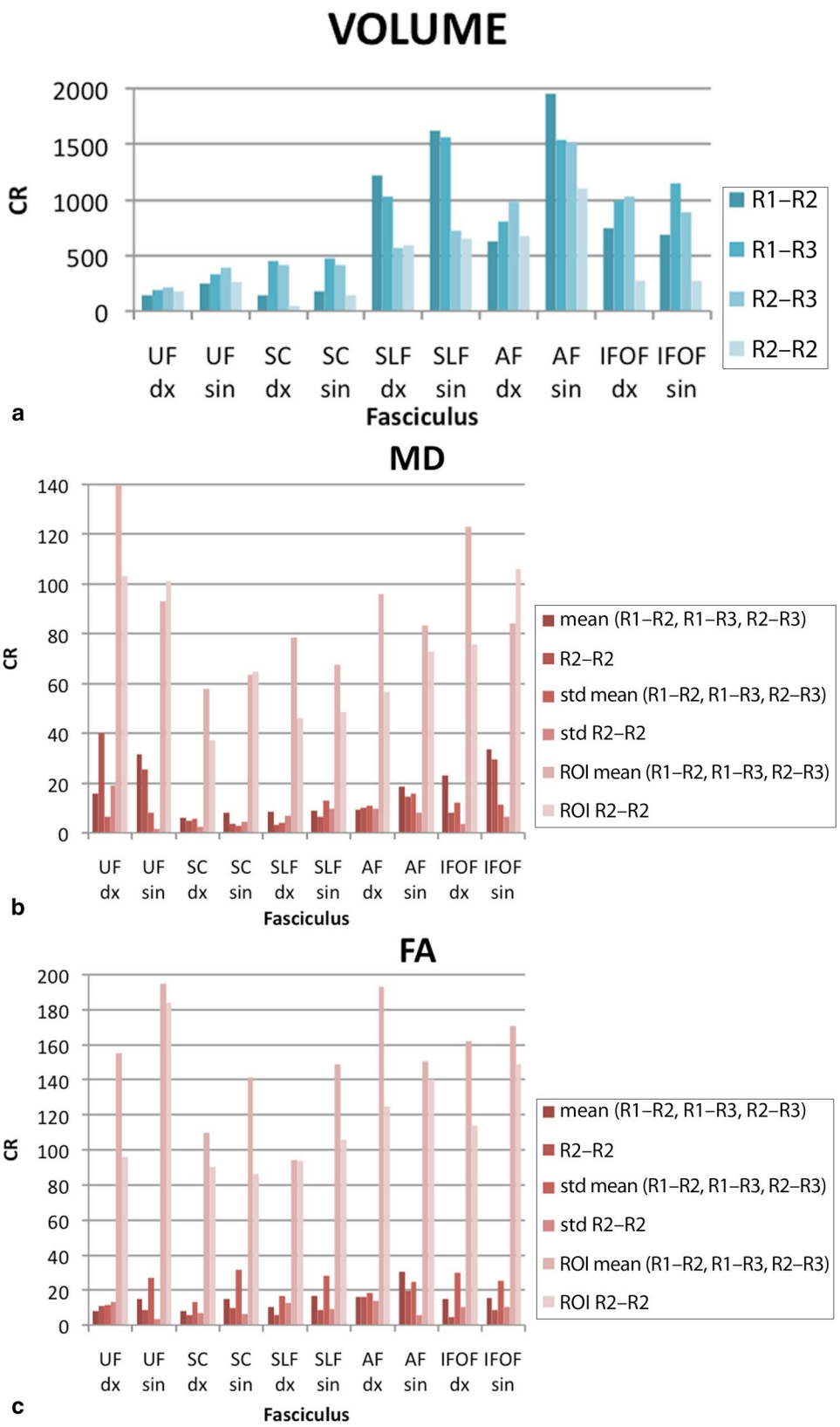

Fig. 3 Coefficient of repeatability (*CR*) values for the tract-based and cross-sectional region of interest (ROI)-based measurements for each tract (*UF* uncinate fasciculus, *SC* superior cingulum, *SLF* superior longitudinal fasciculus, *AF* arcuate fasciculus, *IFOF* inferior fronto-occipital fasciculus). **a** Inter-rater (*R1–R2*, *R1–R3*, *R2–R3*) and intra-rater (*R2–R2*) CR values for the tract-based volume measurements. **b** Inter-rater and intra-rater CR values for the tractography, tract-based central core, and ROI-based mean diffusivity (*MD*) measurements. **c** Inter-rater and intra-rater CR values for the tractography, tract-based central core, and ROI-based fractional anisotropy (*FA*) measurements. *R1* rater 1, *R2* rater 2, *R3* rater 3, *dx* right, *sin* left, *std* standard volume in middle of tract

MD measurements (ICC range: 0.447–0.679) in both SCs, left SLF, right AF, and both IFOFs. Inter-rater FA concordance was found to be poor (ICC≤0.388) in the right UF, left SC, left SLF, both AFs, and left IFOF. Inter-rater MD concordance was found to be poor (ICC≤0.388) in both UFs, right SLF, and left AF. All the Bland–Altman CR val-

ues for the cross-sectional ROI-based FA measurements were more than 10 % from the respective mean FA. Coefficient of variation estimates of variability for the MD were in the range of 2.2–5.5 %, and 5.1–11.0 % for the FA.

Between-Scan Reliability (Group B)

Results for each ICC analysis are summarized in Fig. 4.

For both *tract-based methods*, between-scan reliability of the FA and MD measurements was found to vary between fair to good and excellent, except for poor reproducibility in MD measurements in the left IFOF with 0.15 FA and 27°

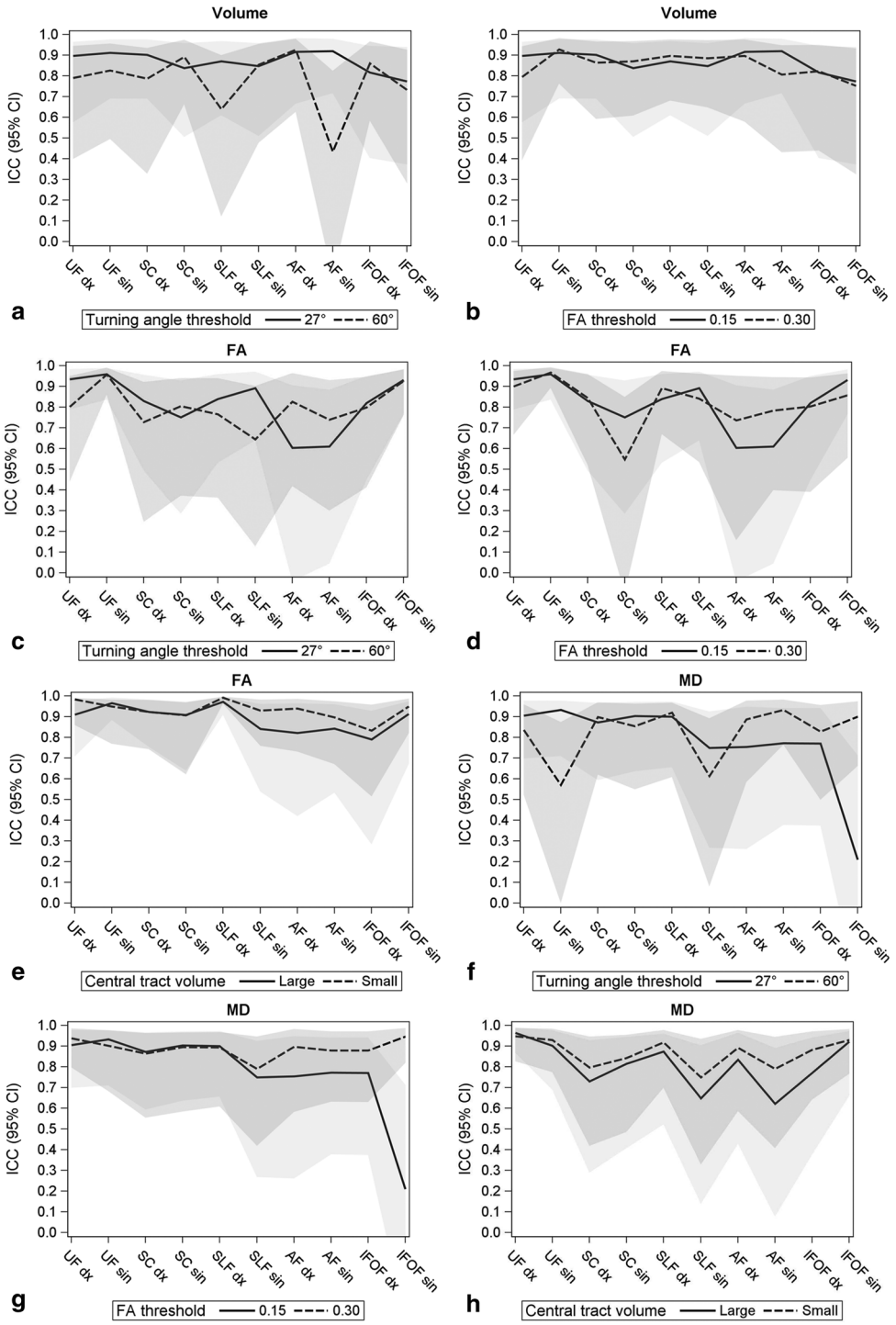

Fig. 4 Between-scan reproducibility of tract-based volume (**a**, **b**), fractional anisotropy (*FA*; **c**–**e**), and mean diffusivity (*MD*; **f**–**h**) measurements for right (*dx*) and left (*sin*) uncinate fasciculus (*UF*), superior cingulum (*SC*), superior longitudinal fasciculus (*SLF*), arcuate fasciculus (*AF*), and inferior fronto-occipital fasciculus (*IFOF*). Intra-class correlation (*ICC*) values with 95% confidence interval (*CI*) for each tract with different turning angle (27°/60°) and FA (0.15/0.30) threshold, as well as with two predetermined central-tract volumes, is presented

turning angle threshold. Measurements for the whole-tract volume with turning angle threshold 27° showed excellent reproducibility with both 0.15 and 0.30 FA threshold (ICC range: 0.787–0.918 and 0.766–0.932, respectively). There was no systematic difference in the reproducibility of the FA, MD, and whole-tract volume measurements depending on the FA or turning angle threshold. The reproducibility of the FA and MD measurements in the central part of the tract varied from good to excellent (ICC range: 0.746–0.989 and 0.634–0.964, respectively) in all the tracts with both the predetermined sizes (3 cm^3 and 1 cm^3 for both UFs, SCs, and AFs; 6 cm^3 and 2 cm^3 for both SLFs and IFOFs). However, there was a tendency toward better reproducibility of the central-tract measurements with the smaller predetermined sizes.

Coefficient of variation estimates of the tract-based between-scan variability in the MD were in the range of 0.9–2.5 % (FA/turning angle threshold: 0.15/27°)/0.6–2.1 % (FA/turning angle threshold: 0.15/60°)/0.7–1.3 % (FA/turning angle threshold: 0.30/27°), and 1.0–3.5 % (FA/turning angle threshold: 0.15/27°)/0.7–2.5 % (FA/turning angle threshold: 0.15/60°)/0.7–1.7 % (FA/turning angle threshold: 0.30/27°) for FA.

No systematic bias (i.e., one set of repeated measures is not consistently higher or lower than the other set) between the scans was found.

Discussion and Conclusions

The current study provided an explicit evaluation of the reproducibility of tractography measurements of ten association tracts using a clinical DTI protocol at 3.0 T. The examined fiber variables were the FA and MD together with the volumetric measurements including central-part analysis of corresponding volumes representing the region with the highest FA values in each subject. Different types of intra-rater, inter-rater, and between-scan reliabilities were compared to find important sources of variability and to evaluate whether increasing the FA threshold from 0.15 to 0.30 and decreasing the turning angle threshold from 60° to 27°, as well as decreasing predetermined central-part volume, improved reliability. Both high and low FA and turning angle threshold were chosen to test in a large scale the values that are commonly used in clinical DTI tractography analyses. In accordance with several previous studies [2, 21–24], the tract-based method showed better reproducibility of FA and MD measurements in all the studied tracts compared with the cross-sectional ROI method. We found no systematic difference in the reproducibility of the FA, MD, and volume measurements depending on the FA or turning angle threshold.

As the large variability in WM tract volumes and structure determined by tractography can decrease the accuracy of FA and MD measurements, additional measurements from the core of the tract have been recommended to confirm the tract-based results [13]. The central-part volumes were chosen so that the bigger volume represents approximately one-third of the tract volume (defined with FA/angle: 0.15/27°) and the smaller approximately one-third of the bigger volume. According to our results, tractography provides a highly reproducible method to perform this central-part analysis, which promotes the use of tractography-based core analysis in future research and clinical applications.

Rater performance is a crucial source of variability in the reliability results. In our study, the only operator-dependent step in the analyses was the placement of the ROIs in the cross-sectional images for the tract delineation and for the ROI-based measurements. The degree of reproducibility in cross-sectional ROI approaches is related to the ROI shape and size; a small ROI is more prone to noise and partial volume effects, so increasing ROI size improves reproducibility, as long as contamination from surrounding structures with markedly different voxel values is avoided [6]. In our study, standard ROI sizes and positions were used. We used color-coded diffusion maps that show fiber orientation to improve the definition of the ROIs from the two-dimensional images [29], and placed the ROIs onto the shape of the structure to diminish the partial volume effects [3]. The most common methods for isolating fiber tracts, such as the method used in our study, include the use of multiple inclusion and exclusion ROIs. Tract-based analyses have several advantages compared with the cross-sectional ROI analyses, including the larger number of voxels that reduce the measurement variance [2], as well as the higher level of automaticity and lower level of subjectivity [21]. The need for many ROIs to isolate the specific tract can, however, diminish the reliability of the tract-based methods [9]. We tried to minimize the impact of the skills of the raters to follow the set rules for anatomic delineations by adequately training and supervising the raters; there were no systematic differences in our results depending on the experience of the raters.

In our study, the lowest agreement in the cross-sectional ROI measurements was found in the UF and AF, which may be related to the tract shape and relatively high amount of crossing fibers in these areas. Compared with our results, better reproducibility has been reported concerning MD and FA values of the cingulum [20] and FA values of the left uncinate fasciculus [6], which can be related to differences in ROI size, shape, and placement methods. In our study, both intra-rater and inter-rater comparisons revealed approximately twofold higher variability for FA values (mean CV: 5.5 and 8.2 %, respectively) compared with MD measurements (mean CV: 2.5, and 3.5 %, respectively),

which is in agreement with data reported by Bonekamp et al. [20].

For the tract-based method, the reproducibility of FA and MD values was comparable with earlier studies [9, 10, 19]. Measures of mean FA and MD along the tracts were more reproducible than measures of tract volume, which is in accordance with previous studies [9, 30].

Besides the rater performance, the reproducibility of DTI measures can be influenced by scanner parameters, including field strength and parallel imaging cf. non-parallel imaging [31], as well as by acquisition parameters such as voxel size [6], and use of motion correction [32], and cardiac gating [33], as well as the number of diffusion-weighted directions [19, 34]. Between-scan reproducibility may also be influenced by the changes in the brain water content and changes in the cohesiveness and compactness of the fiber tracts [20]. In accordance with previous studies [10, 19], intra- and inter-rater reproducibility was higher compared with the between-scan reproducibility in our study. For the present study, the intersession reproducibility for FA and MD measurements was, however, markedly higher compared with the previous results [19] when 15 gradient directions, such as in our study, were used. This may be related to the careful subject positioning and minor motion during the scan in our study.

Besides the excellent intra- and inter-rater reliability of the tract-based methods according to the ICC analyses, our study showed acceptably low CR values for the FA and MD measurements in all the studied association tracts, indicating reasonable reliability in clinical use. The large variation in volumes seems to partly explain the good ICCs for the whole-tract volumes in our study, as there was also relatively large variation in the measurements between raters, and the CRs of the volumes were large.

Even if the DTI tractography method with only 15 diffusion encoding directions is not optimal for searching small white matter pathways, it may be used for reliable FA and MD measurements of relatively large fiber tracts [19]. However, a study with different image acquisition parameters, different field strength, or non-parallel imaging could have yielded different results. Even if the relatively small study population can be considered as a limitation of our study, it is comparable to the previous reliability studies: $n=12$ [4], $n=9$ [6], $n=8$ [9], and $n=20$ [21]. Nevertheless, the CR values for the ROI measurements in our study could have been smaller, if our study population had been larger [35]. Evaluation of the difference between the diseased and healthy subjects is outside the scope of this study, which is why the variation in measurement of interest in a normal population was not separately calculated. As for accuracy of tracking results, we reconstructed tracts that are well documented in previous anatomical studies using ROIs based on a priori knowledge. Thus, the macroscopic configuration of these reconstructed tracts is likely to reflect true fiber bundles. However, it is possible that some parts of the trajectory may contain inaccuracies due to partial volume effects, noise, and crossing fibers.

According to our results, the tractography-based FA and MD measurements have both high reproducibility according to the ICC analyses and acceptably low CR values, indicating that these methods have a great potential in clinical use. In contrast, the cross-sectional ROI analyses in the study of the association tracts had significantly lower reproducibility and relatively high CR values. Our results show the primacy of tractography for DTI analysis and promote the use of tractography-based core analysis instead of the traditional ROI method, although the method is more time-consuming. The technique is recommended for the diagnostics of clinical patients with TBI and can be useful in other indications, e.g., various neurodegenerative diseases.

Conflict of Interest We declare that we have no conflict of interest.

References

1. Brandstack N, Kurki T, Tenovuo O. Quantitative diffusion tensor tractography of long association tracts in patients with traumatic brain injury without findings in routine MRI. Radiology. 2013;267:231–9.
2. Kanaan RA, Shergill SS, Barker GJ, Catani M, Ng VW, Howard R, McGuire PK, Jones DK. Tract-specific anisotropy measurements in diffusion tensor imaging. Psychiatry Res. 2006;146:73–82.
3. Snook L, Plewes C, Beaulieu C. Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment. Neuroimage. 2007;34(1):243–52.
4. Bisdas S, Bohning DE, Bešenski 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;29:1128–33.
5. Brander A, Kataja A, Saastamoinen A, Ryymin P, Huhtala H, Öhman J, Soimakallio S, Dastidar P. 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;7:800–7.
6. Vollmar C, O'Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, Duncan JS, Richardson MP, Koepp MJ. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 T scanners. Neuroimage. 2010;51:1384–94.
7. Mori S, Crain BJ, Chacko VP, van Zijl PCM. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann Neurol. 1999;45:265–9.
8. Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, Hua K, Zhang J, Jiang H, Dubey P, Blitz A, van Zijl P, Mori S. Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage. 2007;36:630–44.
9. Malykhin N, Concha L, Seres P, Beaulieu C, Coupland NJ. Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. Psychiatry Res. 2008;164:132–42.

10. Danielian LE, Iwata NK, Thomasson DM, Floeter MK. Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. Neuroimage. 2010;49:1572–80.
11. Hasan KM, Kamali A, Abid H, Kramer LA, Fletcher JM, Ewing-Cobbs L. Quantification of the spatiotemporal microstructural organization of the human brain association, projection and commissural pathways across the lifespan using diffusion tensor tractography. Brain Struct Funct. 2010;214:361–73.
12. Kurki T, Laalo J, Oksaranta O. Diffusion tensor tractography of the uncinate fasciculus: pitfalls in quantitative analysis due to traumatic volume changes. J Magn Reson Imaging. 2013;38(1):46–53.
13. Yasmin H, Nakata Y, Aoki S, Abe O, Sato N, Nemoto K, Arima K, Furuta N, Uno M, Hirai S, Masutani Y, Ohtomo K. Diffusion abnormalities of the uncinate fasciculus in Alzheimer's disease: diffusion tensor tract-specific analysis using a new method to measure the core of the tract. Neuroradiology. 2008;50:293–9.
14. Schmahmann JD, Smith EE, Eichler FS, Filley CM. Cerebral white matter neuroanatomy, clinical neurology, and neurobehavioral correlates. Ann N Y Acad Sci. 2008;1142:266–309.
15. Kraus MF, Susmaras T, Caughlin BP, Walker CJ, Sweeney JA, Little DM. White matter integrity and cognition in chronic traumatic brain injury: a diffusion tensor imaging study. Brain. 2007;130(10):2508–19.
16. Niogi SN, Mukherjee P, Ghadar J, Johnson C, Kolster RA, Sarkar R, Lee H, Meeker M, Zimmerman RD, Manley GT, McCandliss BD. Extent of microstructural white matter injury in postconcussive syndrome correlates with impaired cognitive reaction time: a 3T diffusion tensor imaging study of mild traumatic brain injury. AJNR Am J Neuroradiol. 2008;29(5):967–73.
17. Rutgers DR, Toulgoat F, Cazejust J, Fillard P, Lasjaunias P, Ducreux D. White matter abnormalities in mild traumatic brain injury: a diffusion tensor imaging study. AJNR Am J Neuroradiol. 2008;29:514–9.
18. Salmond CH, Menon DK, Chatfield DA, Williams GB, Pena A, Sahakian BJ, Pickard JD. Diffusion tensor imaging in chronic head injury survivors: correlations with learning and memory indices. Neuroimage. 2006;29:117–24.
19. Wang JY, Abdi H, Bakhadirov K, Diaz-Arrastia R, Devous MD. A comprehensive reliability assessment of quantitative diffusion tensor tractography. Neuroimage. 2012;60:1127–38.
20. Bonekamp D, Nagae LM, Degaonkar M, Matson M, Abdalla WMA, Barker PB, Mori S, Horskaỳa A. Diffusion tensor imaging in children and adolescents: reproducibility, hemispheric, and age-related differences. Neuroimage. 2007;34:733–42.
21. Borich MR, Wadden KP, Boyd LA. Establishing the reproducibility of two approaches to quantify white matter tract integrity in stroke. Neuroimage. 2012;59:2393–400.
22. Hong YH, Sung JJ, Kim SM, Park KS, Lee KW, Chang KH, Song IC. Diffusion tensor tractography-based analysis of the pyramidal tract in patients with amyotrophic lateral sclerosis. J Neuroimaging. 2008;18:282–7.
23. Partridge SC, Mukherjee P, Berman JI, Henry RG, Miller SP, Lu Y, Glenn OA, Ferriero DM, Barkovich AJ, Yigneron DB. Tractography-based quantitation of diffusion tensor imaging parameters in white matter tracts of preterm new-borns. J Magn Reson Imaging. 2005;22:467–74.
24. Wang S, Melhem ER. Amyotrophic lateral sclerosis and primary lateralsclerosis: the role of diffusion tensor imaging and other advanced MR-based techniques as objective upper motor neuron markers. Ann N Y Acad Sci. 2005;1064:61–77.
25. Netsch T, van Muiswinkel A. Quantitative evaluation of image-based distortion correction in diffusion tensor imaging. IEEE Trans Med Imaging. 2004;23(7):789–98.
26. Rohde GK, Barnett AS, Basser PJ, Marenco S, Pierpaoli C. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. Magn Reson Med. 2004;51:103–14.
27. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;2:420–8.
28. Portney L, Watkins M, editors. Foundations of clinical research: applications to practice. 2nd ed. Norwalk: Appleton & Lange; 2000.
29. Hermoye L, Saint-Martin C, Cosnard G, Lee S-K, Kim J, Nassogne M-C, Menten R, Clapuyt P, Donohue PK, Hua K, Wakana S, Jiang H, van Zijl PCM, Mori S. Pediatric diffusion tensor imaging: normal database and observation of the white matter maturation in early childhood. Neuroimage. 2006;29(2):493–504.
30. 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;33:867–77.
31. Alexander AL, Lee JE, Wu YC, Field AS. Comparison of diffusion tensor imaging measurements at 3.0 T versus 1.5 T with and without parallel imaging. Neuroimaging Clin N Am. 2006;16:299–309.
32. Ling J, Merideth F, Caprihan A, Pena A, Teshiba T, Mayer AR. Head Injury or head motion? Assessment and quantification of motion artifacts in diffusion tensor imaging studies. Hum Brain Mapp. 2011;33(1):50–62.
33. Chung S, Courcot B, Sdika M, Moffat K, Rae C, Henry RG. Bootstrap quantification of cardiac pulsation artifact in DTI. Neuroimage. 2010;49:631–40.
34. Ni H, Kavcic V, Zhu T, Ekholm S, Zhong J. Effects of number of diffusion gradient directions on derived diffusion tensor imaging indices in human brain. AJNR Am J Neuroradiol. 2006;27(8):1776–81.
35. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil. 1998;12(3):187–99.