Intersite Reliability of Diffusion Tensor Imaging on Two 3T Scanners

Koji Kamagata1*, Keigo Shimoji2, Masaaki Hori1, Akira Nishikori3, Kohei Tsuruta3, Mariko Yoshida1, Kouhei Kamiya4, Ryusuke Irie1, Michimasa Suzuki1, Shinsuke Kyogoku1, Yuriko Suzuki5, Noriko Sato2, and Shigeki Aoki1

1Department of Radiology, Juntendo University School of Medicine
2–1–1 Hongo, Bunkyo-ku, Tokyo 113–8421, Japan
2Department of Radiology, National Center of Neurology and Psychiatry Hospital
3Department of Radiological Sciences, Graduate School of Human Health Sciences, Tokyo Metropolitan University
4Department of Radiology, Graduate School of Medicine, The University of Tokyo
5Philips Electronics Japan, Ltd.

(Received July 24, 2014; Accepted September 19, 2014; published online February 12, 2015)

We report the intersite scan reliability of diffusion tensor imaging (DTI) parameters using identical 3T scanners and acquisition protocols at 2 sites. Voxel-based analysis revealed several regions with significant intersite differences. The intersite reliability of DTI measures showed coefficients of variation below 4% in tract-specific analysis (TSA) and below 6% in atlas-based analysis. Given the excellent reliability of TSA, our results suggest it as a promising and useful tool for multicenter DTI studies.

Keywords: diffusion tensor imaging, intersite, reliability, reproducibility, tract-based spatial statistics

Introduction

Diffusion tensor imaging (DTI) allows researchers to measure the characteristics of local microstructural water diffusion in the brain.1 DTI takes advantage of the macroscopic geometrical arrangement of white matter (WM) bundles that becomes clear through diffusion magnetic resonance (MR) imaging, which measures the translational displacement of water molecules. In addition, DTI enables evaluation of the microstructure of WM, so researchers can quantify the integrity of WM to investigate changes in fiber organization, density, diameter, or myelination. DTI has therefore been widely used to study the integrity of WM tracts in healthy brains as well as in a variety of neurological and psychiatric disorders.

Clinical imaging research often targets specific patient populations that could not be easily recruited independently by a single imaging center. By pooling patients, large-scale multicenter neuroimaging studies increase statistical power. However, differences in such factors as MR imaging scanner vendor and model (even for the same type of scanner) restrict comparison of imaging parameters across sites. Therefore, acquisition of intersite test-retest data in controls is necessary to assess reliability. Intersite test-retest studies enable estimation of the reproducibility of within-subject differences.

Recently, several researchers have evaluated the intra- and/or interscanner test-retest reliability of DTI measurements (Table),2–6 but they have mainly performed region of interest (ROI) analysis and in many cases only analyzed fractional anisotropy (FA). Therefore, to assess inter- and intrasite test-retest reliability in the same 3-tesla MR scanner (Philips, Achieva, Best, the Netherlands), we performed whole-brain voxel-based analysis (VBA) using tract-based spatial statistics (TBSS) and atlas-based analysis (ABA) as well as tract-specific analysis (TSA) for major WM tracts. Our study was unique in that we compared 2 identical 3T MR scanners across 2 centers and evaluated commonly used diffusion tensor parameters (FA and mean...
[MD], axial [AD], and radial [RD] diffusivity) using TBSS as well as ABA and TSA.

**Materials and Methods**

**Subjects**

Two sites, the National Center Hospital at the National Center of Neurology and Psychiatry and Juntendo University Hospital, each equipped with a Philips 3-tesla Achieva scanner (Philips Healthcare, Best, the Netherlands); gradient strength, 80 mT/m; slew rate, 200 T/m/s) were involved in this inter- and intrasite imaging study to evaluate DTI across centers. Seven healthy control subjects aged 28 to 52 years (one woman) were enrolled in the study after informed consent had been obtained in accordance with the requirements of the institutional review boards of each of the imaging sites.

**MR imaging**

For intersite analysis, all 7 subjects underwent imaging at each of the sites the same day. For intrasite analysis, each subject was scanned twice on the same day at the Juntendo University Hospital. The brains of all subjects were examined with a Philips 3T Achieva scanner and an 8-channel-array receiving head coil for sensitivity encoding parallel imaging. DTI was performed using the spin echo echo-planar technique (repetition time [TR]/echo time [TE], 8800/78 ms; matrix size, 112 × 112; field of

---

### Table. Comparison of our results for DTI test–retest reliability with those of previous studies on 3.0T MRI

<table>
<thead>
<tr>
<th>Study</th>
<th>Field strength</th>
<th>No. of diffusion encoding directions</th>
<th>No. of subjects</th>
<th>Parameters used</th>
<th>Analytical method</th>
<th>Repeated scans/measures</th>
<th>Inter-site CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>3.0 T</td>
<td>32</td>
<td>7 (volunteers)</td>
<td>FA, MD, AD, RD</td>
<td>VBA (TBSS), ABA, TSA</td>
<td>Intra-site rescan ×2</td>
<td>0.4–3.2</td>
</tr>
<tr>
<td>Magnotta</td>
<td>3.0 T</td>
<td>30, 32, 71</td>
<td>5 (volunteers)</td>
<td>FA, MD, AD, RD</td>
<td>ROI</td>
<td>Intra-site rescan</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Takao (2012)</td>
<td>3.0 T</td>
<td>13</td>
<td>224</td>
<td>FA, MD, AD, RD</td>
<td>VBA</td>
<td>Longitudinal</td>
<td>Not reported</td>
</tr>
<tr>
<td>Vollmar (2010)</td>
<td>3.0 T</td>
<td>32</td>
<td>9 (volunteers)</td>
<td>FA</td>
<td>ROI</td>
<td>Intra-site rescan ×2</td>
<td>0.8–3.0</td>
</tr>
<tr>
<td>Bisdas (2008)</td>
<td>3.0 T</td>
<td>16</td>
<td>12 volunteers</td>
<td>FA</td>
<td>ROI</td>
<td>Intra-site rescan ×2</td>
<td>2%</td>
</tr>
<tr>
<td>Jansen (2007)</td>
<td>3.0 T</td>
<td>15</td>
<td>10 (volunteers)</td>
<td>FA</td>
<td>VBA</td>
<td>Intra-site rescan ×2</td>
<td>3–6.5%</td>
</tr>
</tbody>
</table>

ABA = atlas-based analysis, AD = axial diffusivity, CV = coefficient of variation (in %), FA = fractional anisotropy, MD = mean diffusivity, RD = radial diffusivity, ROI = region of interest, TSA = tract-specific analysis, VBA = voxel-based analysis

---

K. Kamagata et al.
vision [FOV], 224 × 224 mm²; section thickness, 2 mm with no gap). Images were obtained with both 30-direction diffusion encoding (b = 1000 s/mm² for each direction) and no diffusion encoding (b = 0 s/mm²). A total of 75 axial section images were obtained, covering the entire cerebrum. The scanning time was 10 min 42 s.

**DTI image processing with tract-based spatial statistics**

We performed voxel-based analysis of the DTI data with TBSS implemented in the software library (FSL) of the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRI B) (version 4.1.5, UK; www.fmrib.ox.ac.uk/fsl). We employed the eddy current function implemented in FSL to correct for distortions due to eddy currents using affine intrasubject registration to the respective individual b0 images; calculated FA, MD, AD, and RD maps for all subjects using the DTIFIT tool to fit a tensor model to each voxel of the raw diffusion data; eliminated non-brain structures using the brain extraction tool; and aligned FA maps of all subjects to standard Montreal Neurological Institute (MNI152) space using the FNIRT non-linear registration tool. The mean FA image was then generated and thinned to create the mean FA skeleton, which represented the centers of all tracts common to the groups. The mean FA skeleton was thresholded to FA > 0.20 to include the major WM pathways and exclude peripheral tracts and gray matter. The aligned FA map of each participant was then projected onto this skeleton by assigning to each point on the skeleton the maximum FA in a plane perpendicular to the local skeleton structure. The resulting skeletons were fed into voxel-wise statistics. By applying the original nonlinear registration of the FA of each subject to the standard space, the MD, AD, and RD maps were also projected onto the mean FA skeleton. The MD, AD, and RD data were used to calculate voxel-wise statistics as well. We analyzed voxel-wise statistics of the skeletonized FA data using Randomise (part of FSL) to perform paired t-tests. This program performed permutation-based testing with 5000 permutations and statistical inference using threshold-free cluster enhancement with a significance criterion of P < 0.05, which was corrected for multiple comparisons (family-wise error). The anatomical locations of regions with significant group differences in FA, MD, AD, and RD on the white-matter skeleton were identified with a WM atlas.

**Atlas-based analysis**

FA maps of all subjects were realigned to the FA template of the Johns Hopkins University International Consortium for Brain Mapping (JHU ICBM FA, one mm) using the FNIRT nonlinear registration tool, and the corresponding MD, AD, and RD maps were subsequently realigned to the transformation parameter obtained from the FA maps. Furthermore, atlas-based ROIs of the WM for the anterior thalamic radiation (ATR), corticospinal tract (CST), cingulum (CG), forceps minor, forceps major, inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), and uncinated fasciculus (UF) were generated using the JHU White-Matter Tractography Atlas of the FSL atlas tool.

**DTI image processing using tract-specific analysis**

We created color-coded maps using 33 sets of images (32 sets with b = 1000 s/mm², one set with b = 0 s/mm²). On the color maps, red was assigned to the left-right direction, green to the anteroposterior direction, and blue to the craniocaudal direction. Fiber tracts were based on fiber assignments made using the continuous tracking approach⁷ to obtain 3-dimensional (3D) tract reconstruction. Identification of fiber tracts was initiated by placing a “seed” area and a “target” area in anatomic regions through which the particular fibers were expected to course.

Tractographies of the UF, CST, and posterior cingulum (PCG) bilaterally were performed using the 2-ROI method.⁸ The color-coded maps were used to place these ROIs into the WM tracts precisely and objectively. The UF, CST, and PCG were measured according to the report of Yasmin’s group.⁹ Two of the authors (K.S., K.K.), who were blinded to the subjects, measured tracts.

**Statistical analysis**

We analyzed all statistics using Statistical Package for the Social Sciences (SPSS) software for Windows (Release 20.0; SPSS, Chicago, IL, USA).

The coefficient of variation (CV) is defined as the ratio of the standard deviation of measurements divided by the mean measurement and multiplied by 100. It allows an intuitive estimate of the measurement variance expressed as a relative percentage regardless of the absolute measurement value. In previous studies of DTI test-retest reliability, the CV is the most commonly reported statistical measure. We calculated the CV for each tractogram and pair of scans (intra- and intersite) per subject and assessed inter-rater reliability using Pearson correlation coefficients.
Results

**Voxel-wise comparison with tract-based spatial statistics**

Figure 1 indicates areas of significant intersite change in FA (Fig. 1a), MD (1b), AD (1c), and RD (1d) obtained on the same scanner. There were no significant changes in FA (Fig. 1a). However, the voxel-wise analyses of skeletonized images revealed a number of regions with significant differences in MD, AD, and RD. For MD, the affected WM tracts included the left ILF, left IFOF, left CST, left ATR, left SLF, left thalamus, and corpus callosum (Fig. 1b). For AD and RD, they included the left ILF, left IFOF, left CST, left ATR, left SLF, and left thalamus (Fig. 1c, d).

Figure 2 indicates areas of significant intrasite change in FA (Fig. 2a), MD (2b), AD (2c), and RD (2d) in voxel-wise analyses of skeletonized images. For FA, the affected WM tracts included the forceps minor (Fig. 2a). For MD, AD, and RD, they included the right ILF, right IFOF, right CST, right ATR, right SLF, right thalamus, left ILF, left IFOF, and corpus callosum (Fig. 2b, c). Significant changes in FA, MD, AD, and RD were seen more extensively between sites than within sites.

**Coefficient of variation of atlas-based analysis**

Figure 3 summarizes the CVs of atlas-based analysis for intra- and intersite rescans. In atlas-based analyses, the intrasite CV ranged from 0.4 to 3.2% for FA, 0.4 to 1.8% for MD, 0.3 to 1.6% for AD, and 0.6 to 2.5% for RD (Fig. 3a), whereas the intersite CV ranged from 0.5 to 4.2% for FA, 1.0 to 3.6% for MD, 0.7 to 3.4% for AD, and 0.8 to 5.6% for RD (Fig. 3b). Comparing the examined regions, the highest intersite CV was found for the right CG in FA, MD, and RD as well as for the right SLF in AD.

**Coefficient of variation of tract-specific analysis**

Figure 4 summarizes the CVs of TSA for inter- (Fig. 4a) and intrasite (4b) rescans. In TSA, the intrasite CV ranged from 1.0 to 2.6% for FA, 1.0 to 1.9% for MD, 1.1 to 2.0% for AD, and 1.6 to 3.1% for RD (Fig. 4a), whereas the intersite CV ranged from 0.7 to 2.6% for FA, 0.6 to 2.2% for MD, 0.7 to 3.6% for AD, and 0.2 to 2.4% for RD (Fig. 4b).

**Inter-rater correlation coefficient of tract-specific analysis**

Reproducibility was expressed as an inter-rater correlation coefficient. For the FA analysis, the coefficient in the CST was 0.92, in the PCG, 0.98, and in the UF, 0.88. For the MD analysis, the coefficient in the CST was 0.80, in the PCG, 0.74, and in the UF 0.82. For the AD analysis, the coefficient in the CST was 0.86, in the PCG, 0.88, and in the UF 0.92. For the RD analysis, the coefficient in the CST was 0.88, in the PCG, 0.97, and in the UF 0.86.

Discussion

We report the intra- and intersite scan-rescan reproducibility of the DTI measures, FA, MD, AD, and RD using VBA, ABA, and TSA in 7 healthy volunteers tested with identical scanners and acquisition protocols at 2 different sites. The VBA of skeletonized images revealed a number of regions with significant inter-site differences in FA, MD, AD, and RD that were more extensive than the corresponding intrasite differences. Using the TSA approach, the intersite reproducibility of FA from a typical DTI protocol showed a CV below 4%. With the ABA approach, the corresponding CV was below 6%. These data on reliability are similar to those of previous studies of DTI reliability.2–6

The discrepancy between the CVs of ABA and TSA may be explained by image misregistration in the ABA. In VBA, the shapes of individual brains are normalized to a common coordinate space (e.g., ICBM-152 space), after which voxel-by-voxel analyses are performed. In the ABA approach, the shapes are normalized to a template space, after which the entire normalized brain is automatically segmented into many 3D structural components. In VBA and ABA, image registration is necessary for normalization to a common coordinate space. Therefore, misregistration is a major issue. Although we visually confirmed there was no gross misregistration, perfect image registration is extremely difficult because WM tracts usually have very sharp boundaries. Misregistration introduces partial volume effects in other WM tracts, cerebrospinal fluid, or cerebral cortices. TSA allows researchers to visualize the trajectories of specific WM fiber bundles and measure diffusion parameters more precisely than with manually drawn ROIs.10 In TSA (deterministic tractography), all analyses are performed in individual native spaces as opposed to a common coordinate space. Therefore, misregistration is nonexistent in TSA. Thus, the TSA approach may be a useful tool for intersite multicenter DTI trials. We performed all DTI scans with identical and isotropic voxels, which was important to measure reproducibility in TSA analysis, as pointed out in a previous study.11

Our study had a number of limitations. First, the small size of our samples may have limited the comparison of intersite scan reproducibility of DTI pa-

E-pub ahead of print  
Magnetic Resonance in Medical Sciences
Fig. 1. Intrasite differences based on tract-based spatial statistics (TBSS) analysis. Areas of significant intrasite differences in (a) fractional anisotropy (FA) and (b) mean (MD), (c) axial (AD), and (d) radial (RD) diffusivity. (a) There were no significant changes in FA. (b–d) Changes in white-matter tracts are reflected by MD in the left inferior longitudinal fasciculus (ILF), left inferior fronto-occipital fasciculus (IFOF), left corticospinal tract (CST), left anterior thalamic radiation (ATR), left superior longitudinal fasciculus (SLF), left thalamus, and corpus callosum as well as by AD and RD in the left ILF, left IFOF, left CST, left ATR, left SLF, and left thalamus. Red represents an increase and blue, a decrease.

Fig. 2. Intersite differences based on tract-based spatial statistics (TBSS) analysis. Areas of significant intersite differences in (a) fractional anisotropy (FA) and (b) mean (MD), (c) axial (AD), and (d) radial (RD) diffusivity. (a–d) Changes in white-matter (WM) tracts are reflected by FA in the forceps minor as well as MD, AD, and RD in the right inferior longitudinal fasciculus (ILF), right inferior fronto-occipital fasciculus (IFOF), right corticospinal tract (CST), right anterior thalamic radiation (ATR), right superior longitudinal fasciculus (SLF), right thalamus, left ILF, left IFOF, and corpus callosum. Red represents an increase and blue, a decrease.
rameters. Second, in the present study, we used FNIRT for the normalization algorithm, though the registration accuracy of FNIRT has been reported to be somewhat inferior to that of such other techniques as Automatic Registration Toolbox (ART), symmetric normalization (SyN), Image Registration Toolkit (IRTK), or statistical parametric mapping (SPM)-diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL). The application of other normalization algorithms to our study may afford lower CV values in atlas-based analysis. Third, in the TSA analysis, the seed and target ROIs were drawn manually, and the reproducibility of measurements was unclear. However, 2 of the authors drew all ROIs. Furthermore, blinding prevented rater bias, and the inter-rater correlation coefficients were 0.74 to 0.98.

Conclusion

Our results are promising for multicenter DTI studies in large populations. Given its excellent reproducibility, TSA is a useful tool for multicenter DTI studies.

Conflict of interest: We declare no conflict of interest.

Acknowledgments

We thank MR imaging technologists Nozomi Hamasaki and Syuji Sato for their skillful performance in acquiring data and Akira Nishikori, Tomo-mi Okamura, and Fumitaka Kumagai for their research assistance. We also thank Narisumi Cho for administrative assistance. This work was supported by Intramural Grant 24-S10 for Neurological and Psychiatric Disorders of the National Center of Neurology and Psychiatry (NCNP), a Grant-in-Aid for Scientific Research on Innovative Areas (Comprehensive Brain Science Network) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, and from the MEXT/Japan Society for the Promotion of Science KAKENHI (Grant Number 24591787).

References

1. Basser PJ, Mattiello J, LeBihan D. Estimation of the
Reliability of Diffusion Tensor Imaging