Between session and between subject reproducibility of diffusion MR and tractography measures

E. Heiervang[1], T. E. Behrens[2], C. M. Mackay[3], M. Robson[3], H. Johansen-Berg[2]

[1]Department of Child and Adolescent Mental Health, Haukeland University Hospital, Bergen, Norway, [2]Oxford Centre for Functional MRI of the Brain, University of Oxford, Oxford, United Kingdom, [3]University of Oxford Centre for Clinical Magnetic Resonance Research, University of Oxford, Oxford, United Kingdom

Introduction

DTI can generate quantitative measures, such as fractional anisotropy (FA), that are thought to provide markers of tissue microstructure. As these measures are increasingly used to address clinical questions, it is important to fully characterise their reproducibility. Normal variability and test-retest stability are key issues for group and longitudinal study designs yet evidence so far is incomplete. Pfefferbaum et al (2003) reported a test-retest coefficient of variation (CV) of 1.9% for FA within the corpus callosum (*1*). It is clear from other work that variability measures can differ between tracts; Ciccarelli et al (2003) scanned four subjects on two occasions and found CVs of FA from within tractography-defined tracts of 6.2%, 7.1%, and 5% for callosum, optic radiation, and pyramidal tract respectively. Inter-subject CV from the same study was in the range of 6-9% for these tracts (*2*). Here, we assess the reproducibility of DTI measures from histogram, ROI and tractography with 8 subjects scanned three times with the same sequence and scanner.

Methods

We acquired MR data in eight healthy right handed adult subjects (4 men, 4 women, age range 21-34 years). Scans were obtained on three separate days within a three month period on a 1.5 T Siemens Sonata MR scanner with maximum gradient strength of 40 $mT \cdot m^{-1}$. Diffusion-weighted data were acquired using echo planar imaging (72x2mm thick axial slices, matrix 128x104, field of view 256x208, giving a voxel size of 2x2x2 mm). The diffusion weighting was isotropically distributed along 60 directions using a *b* value of 1,000 $s \cdot mm^{-2}$. Image analysis was carried out using tools from the FMRIB Software Library (FSL, www.fmrib.ox.ac.uk/fsl; (*3*)). FDT (FMRIB's Diffusion Toolbox) was used to fit a diffusion tensor at each brain voxel in the diffusion data and to calculate voxel-wise values for FA. Probabilistic tractography (*4*) was used to generate connectivity distributions from seed voxels/masks defined on FA images, or using standard space masks, to define the pyramidal tract, cingulum bundle, and optic raditions. The following quantitative measures were then dervied for each subject and each session: 1, Histogram measures for whole brain white matter FA. 2, Mean FA within regions of interest (3x3 voxel mask defined manually on FA images). 3, Mean FA along tracts of interest (with seed masks defined as: A) Single voxels in FA images; B) 3x3 voxel ROIs on FA images; C) Using a 2 ROI approach with standard space masks). 4, Volumes of tracts of interest (with seed masks defined as in 3). Co-efficients of Variation (CV) (100*mean/SD) for these measures were calclated between subjects and between sessions.

Results

Histogram analysis: We derived histogram measures (mean, mode, peak height, standard deviation) for FA within the whole brain white matter. Inter-session coefficients of variation (CVs) were less that 2% for all measures. Inter-subject CVs ranged from 3-11%.

Voxel-wise FA analyses: Inter-session CVs for mean FA within a 3x3 voxel ROI defined manually on each FA image were 4-6% and inter-subject CVs were 3-11%.

Tractography analysis: Voxels of interest were used as seed points for probabilistic tractography. Resulting tracts were thresholded at 10 (/5000 generated samples) and the mean FA and volume of the thresholded tract were found. The overlap of tracts across sessions and subjects for the pyramidal tract can be seen in Figure 1.

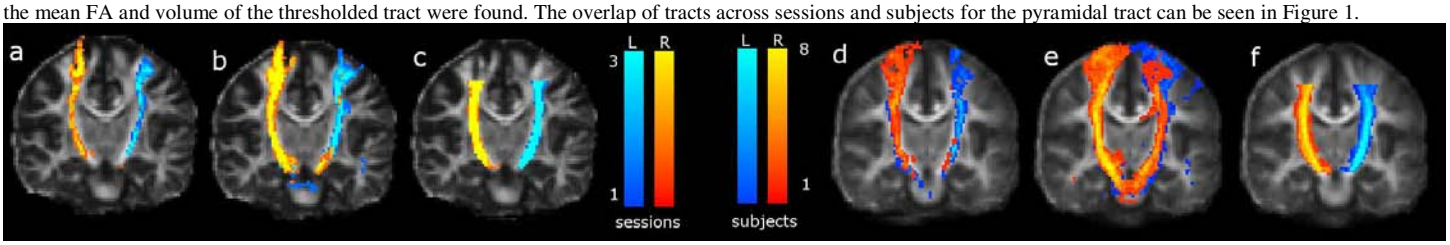

Figure 1: Overlap of pyramidal tract across sessions for a single subject (a-c) and across subjects (d-f). Tracts are generated using different methods of seed definition: a single voxel (a,d) or 3x3 voxel ROI (b,e) manually defined on FA maps or a variation on the two-ROI method, using standard space masks (c,f). For a-c colours represent number of sessions for which voxels were included in the tract generated; for d-f colours represent numbers of subjects.

The effects of different factors on inter-session CV were tested using repeated measures ANOVAs. CVs calculated from mean FA were lower (CV=4.9) than those based on tract volume (CV=14.3) (effect of measure $F=98.6$, $p<0.001$). The method of seed definition also influenced reproducibility with lowest CV for the two ROI standard space method (CV=5.9) followed by manually defined ROIs using a 3x3 ROI (CV=8.3) or a single voxel (CV=14.3) (effect of seed definition, $F=58.2$, $p<0.001$). An interaction between measure and method of ROI definition ($F=11.8$, $p<0.001$) reflected the finding that the effect of seed definition on reproducibility is more marked when measuring tract volumes than when measuring mean FA. An interaction between measure and tract ($F=6$, $p<0.03$) reflected the fact that the improved reproducibility of mean FA compared to tract volume varied between tracts, being most prominent in PT, followed by OR, then CB.

Effect of the number of diffusion directions: To test whether the number of diffusion directions influences inter-session CV we created 12 direction datasets from our 60 direction data and used the two ROI method in standard space to generate tracts. Although the mean CV was greater for 12 directions (CV=10.8, std error=4.0) than for 60 direction data (CV=6.0, std error=0.8), this effect was not significant ($F=1.3$, ns). It should be noted, however, that although the major fibre bundles studied here could be reproducibly generated in 12 direction data, the recovered volumes were not the same with 12 and 60 direction data, and the 60 direction data tended to generate larger pathways with greater sensitivity.

Implications for power calculations:

The values reported here should assist in estimation of required sample sizes to detect effects of a given size. Using the 60 direction data, and the two ROI method for defining tracts, we have calculated the required sample size to detect a change of 2-15% in mean FA for different tracts of interest with a significance level of 0.05 and power of 80% (Table).

Tract:	Cingulum Bundle		Optic Radiation		Pyramidal Tract	
Mean FA (SD):	Left 0.35 (0.026)	Right 0.32 (0.027)	Left 0.45 (0.024)	Right 0.45 (0.024)	Left 0.45 (0.014)	Right 0.45 (0.021)
Effect size	Required N	Required N	Required N	Required N	Required N	Required N
2%	171	221	89	89	31	68
5%	28	36	15	15	6	12
10%	8	10	4	4	2	4
15%	4	5	3	3	-	2

Discussion

By acquiring repeated DTI measurements in the same subjects we have been able to characterise within and between subject variability of quantitative measures such as FA and tract volume. Histogram and ROI based measures of mean FA were highly reproducible from session to session (under 6%). Probabilistic tractography was used to define tracts of interest from which quantitative measures were derived. With this approach we found that measures of mean FA from within tracts defined using a variation on the two ROI approach, were most reproducible, with an inter-session CV of under 5% for all tracts studies. The values reported here could be used for power calculations to determine the number of subjects required to detect an effect of a given size and should therefore assist in planning of clinical DTI studies.

References

1. A. Pfefferbaum, E. Adalsteinsson, E. V. Sullivan, *J Magn Reson Imaging* **18**, 427 (Oct, 2003); 2. O. Ciccarelli *et al.*, *Neuroimage.* **18**, 348 (2003).; 3.S. M. Smith *et al.*, *NeuroImage* **23 Suppl 1**, S208 (2004).; 4.T. E. J. Behrens *et al.*, *Magn Reson Med* **50**, 1077 (2003).