Reliability of FA and ADC measures in the healthy brain: implications for longitudinal DTI studies

R. Scheidegger[1,2], Y. Wu[1,2], R. Ochs[1], P. Storey[1,2], R. Edelman[1,2], A. Ragin[1,2]

[1]Radiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States, [2]Radiology, Evanston Northwestern Healthcare, Evanston, Illinois, United States

Introduction: Diffusion tensor imaging (DTI) [1] has been used to investigate central nervous system disorders and may have considerable potential in the diagnosis and monitoring of neurological diseases, as well as the evaluation of novel treatments. The validity and sensitivity of DTI are determined in part by the reliability of the measurements. While intra/inter-rater reliability [2] and inter-session reliability [3, 4, 5] of DTI measures have been estimated for select brain regions of interest (ROIs), limited information is available concerning the simultaneous influence of different raters, different post-processing tools and different imaging sessions on whole brain and localized DTI measurements. Inter-session reliability studies have examined the reproducibility of DTI measures in individual subjects across a three month period; however, neurological disorders evolve on much longer timescales. In this investigation, the inter-session reliability of DTI was analyzed across a one year period. In addition, intra-rater, inter-rater, and inter-tool reproducibility of DTI measurements for the whole brain, for white matter (genu, splenium, frontal white matter and centrum semiovale) and for gray matter (caudate and putamen) regions were systematically evaluated using three reliability coefficients.

Methods: Eight healthy volunteers (2 women and 6 men; mean age 40.4 years, range 22-50 years) were scanned on two occasions approximately one year apart. DTI was implemented on a 1.5 T GE MR scanner using an echo-planar spin-echo sequence with the following parameters: TE=82.3, TR=7000, b_{max}= 1000 s/mm^2, slice thickness=7 mm. One B=0 image and six diffusion-weighted images were obtained. Histograms of the apparent diffusion coefficient (ADC) and fractional anisotropy (FA) were derived from whole-brain maps calculated with custom MATLAB software. ROIs (43 mm^2) were acquired on an AWworkstation for genu, splenium, frontal white matter, centrum semiovale, caudate, and putamen. ROIs were first positioned on the B=0 reference image and then projected onto FA and ADC maps. For inter-tool reliability, ROIs were drawn with the AWworkstation and DPTools (http://fmritools.hd.free.fr). For the intra-rater and inter-rater reproducibility, AWworkstation ROIs and whole-brain MATLAB measurements were performed once by two separate raters and on two different occasions by an individual rater, respectively.

Results: The reliability of whole brain (Table 1) and ROI measures (Table 2) was assessed with intraclass correlation coefficients (ICC), coefficients of variation (CV=SD/mean, in %) and Bland and Altman repeatability coefficients (RC, in %) [6]. The reproducibility of ROI measures is presented as the median of each coefficient over all studied regions. ADC measures tended to be more reproducible than FA measures. Agreement for whole-brain measures was higher than for ROI measures. Although the second timepoint whole brain measures were lower than the first timepoint measures (Fig.1), the whole brain inter-session reliability was high (Table 1) and the systematic bias was not apparent in the ROI measures (Fig.2). For ROI measures, intra-rater reliability was high, while inter-rater, inter-tool and inter-session measures were more variable. No differences were observed between the reliabilities of white matter and gray matter regions (Fig.2).

Table 1: Whole Brain	ADC			FA		
	ICC	CV	RC	ICC	CV	RC
Intra Rater	> 0.99	<0.1	<0.1	> 0.99	<0.1	0.2
Inter Rater	> 0.99	0.4	0.6	0.71	2.1	8.6
Inter Session	0.70	9.1	3.7	0.84	6.7	14.4

Table 2: ROIs	ADC			FA		
	ICC	CV	RC	ICC	CV	RC
Intra Rater	0.90	0.6	2.9	0.87	2.6	10.8
Inter Rater	0.49	2.2	5.7	0.39	9.6	30.1
Inter Tool	0.16	6.9	16.8	0.27	21.5	35.6
Inter Session	0.12	9.0	11.7	0.37	14.3	36.6

Discussion: Findings of high reproducibility (ICC>0.7, CV < 10%) for whole-brain measures support the feasibility of longitudinal and multicenter DTI studies. Because they are less influenced by observer variability than ROI-measures, whole brain measures of ADC and FA would be more accurate measures of outcome. In order to minimize measurement error, ROI measures of ADC and FA should be obtained by the same rater with identical post-processing software and schemes. The inter-session reliability across a one-year period was not as high as estimates reported by short-term studies. Changes in the MR environment, including software and/or hardware updates, as well as physiological changes in individual subjects, may introduce variability in DTI measures that cannot be accounted for solely by inter-rater variability. These findings underscore the necessity of pre-study standardization of MR sequences and procedures, as well as inclusion of control subjects in longitudinal DTI studies of central nervous system disorders.

References:

1 Basser & Pierpaoli, *J Magn Reson B.* 1996; 111: 209-219
2 Bonekamp,et al.. *Proc. Intl. Soc. Mag. Reson. Med.* 2005; 13: 648
3 Lau & Goodyear. *Proc. Intl. Soc. Mag. Reson. Med.* 2005; 13: 1321
4 Cercignani, et al. *Am J Neuroradiol* 2003; 24: 638-643
5 Pfefferbaum, et al. *J Magn Reson Imaging* 2003; 18: 427-433
6 Bland & Altman. Lancet. 1986; 307-310

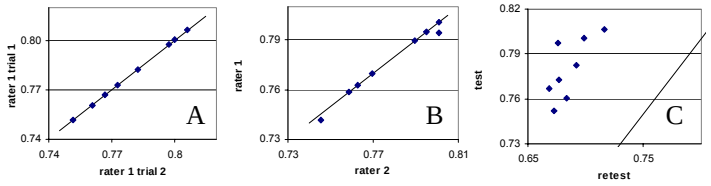

Figure 1: Scatter plots of reliability data for whole brain ADC measures: (A) intra-rater, (B) inter-rater, (C) inter-session. The line of equality is plotted for reference.

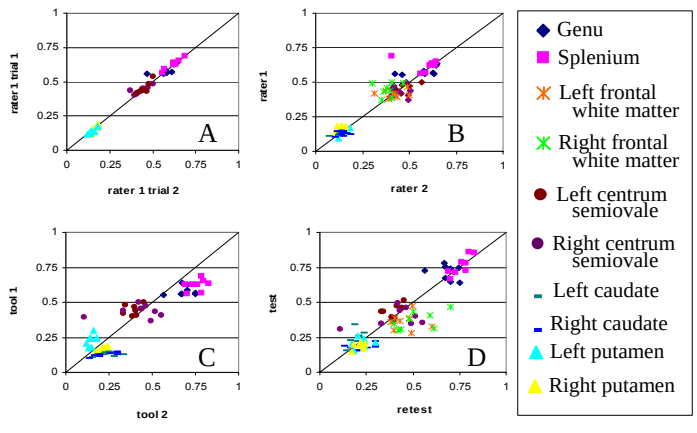

Figure 2: Scatter plots of reliability data for ROI FA measures: (A) intra-rater, (B) inter-rater, (C) inter-tool, (D) inter-session, shown with the line of equality.