Evaluation of within-site and cross-site accuracy and precision of DTI measurements through a multi-center human volunteer study

T. Zhu[1], M. Gaugh[2], X. Liu[3], M. Taylor[4], Y. Tso[5], G. Schifitto[2], C. Yiannoutsos[6], B. Navia[7], and J. Zhong[8]

[1]Biomedical Engineering, University of Rochester, Rochester, NY, United States, [2]Neurology, University of Rochester, Rochester, NY, United States, [3]Electrical Engineeering, University of Rochester, Rochester, NY, United States, [4]4University of California San Diego, San Diego, CA, United States, [5]Stanford University, Stanford, CA, United States, [6]Indiana University, Indianapolis, IN, United States, [7]Tufts University, Medford, MA, United States, [8]Imaging Sciences, University of Rochester, Rochester, NY, United States

Introduction: Diffusion tensor imaging (DTI) and tensor-derived parameters such as fractional anisotropy (FA) and mean diffusivity (MD), have been used to noninvasively depict pathological processes of diseases in brains. In a typical neuroimaging multicenter study, biases and variations in data due to differences in scanner hardware and software among sites need to be minimized for pooling data from multiple sites for conventional statistical inferences. This is one of unsolved critical issues we are facing for a multi-center DTI study. Only the precision of the DTI measurement has been previously studied using a within-subject design [1] and with cross-scanner variance [2] of DTI-derived parameters, while bias of DTI measurement in a multi-center study has not been reported. In this study, multiple DTI data of a healthy volunteer were acquired at three imaging centers. Precision of DTI measurement of each center was quantified by the bootstrap analysis of measurement uncertainty while the accuracy (bias) of DTI measurement was evaluated by comparing DTI parameters from each site to those from the "super" data set combining all DTI data together.

Methods: A healthy volunteer was scanned twice within 24hrs at each of the three centers and a total of six complete DTI datasets were acquired. Each imaging center was equipped with a GE 1.5T EXCITE scanner. DTI sequence parameters were: TR/TE=7s/74.7ms, 2x2x2.5 mm voxel (zero filling to 1x1x2.5 mm in the final image), ASSET acceleration factor=2, 30 diffusion gradient directions with b=1000 s/mm^2 and 3 averages, b=0 images with 15 averages. The total scan time was around 12 mins and the SNR levels in the resultant b0 image were approximately 20 for grey matter region and 15 for white matter region. Home-built software was used for post-processing with eddy-current corrections and tensor calculations. *Super dataset:* In order to evaluate the accuracy (bias) of DTI measurement at each center, a super dataset was created by the following steps. First, one DTI dataset was selected randomly as the reference image and a 12 degree-of-freedom affine registration was performed for each of the other five DTI datasets. Diffusion gradient vectors were adjusted according to the transformation matrix. Second, DTI datasets were pooled together to generate the super dataset. FA and MD maps of this super set were calculated as the gold standard. *Precision:* Precision of the DTI measurement with each DTI data was evaluated by the wild bootstrap analysis [3, 4]. 250 wild bootstrap samples were generated to calculate **σ(FA)** and **σ(MD)** as the measures for the measurement precision of each scan session. Coefficient of variation (CV) values, **CV(FA)** and **CV(MD)**, were further calculated. *Accuracy (Bias):* For FA and MD at each pixel location, differences between each session data and the corresponding values from the super set were calculated and further normalized by those values from the super set. These normalized difference values (**norm_Bias_FA** and **norm_Bias_MD**) were adapted as the measures for the accuracy. *Evaluation of multi-center DTI data:* Both within-site and cross-site comparisons of DTI precision (as CV values) and DTI accuracy (as normal_Bias values) were performed within four ROIs placed on the genu corpus callosum (GCC), the splenium corpus callosum (SCC), the putamen (PUT) and the centrum semiovale (CS).

Results: Comparing to FA maps calculated from one DTI dataset, the FA map from the super set (Fig.1A) created in this study shows cleaner anatomical details and a higher contrast-to-noise ratio. It thus provides a reasonable gold standard to evaluate the accuracy of individual DTI measurements. Precision of the DTI measurement with a typical clinical DTI sequence as used in this study can be depicted by CV maps from the wild bootstrap analysis (Fig.1B). The FA measurement is more reliable in more anisotropic tissues, as indicated in Fig 1B where CV(FA) decreases as the FA value increases, from over 20% when FA<0.2 to 6% when FA is between 0.8 and 0.9. On the other hand, similar levels of uncertainty in CV(MD) were observed for most white matter regions. The overall uncertainty associated with FA is worse than that associated with MD measurement. As for the accuracy of the DTI measurement, bias in FA measurement increases with the decrease of the anisotropy level, while overestimation were observed for low FA regions such as several GM structures and CSF-filled ventricles. In contrast, bias in MD measurement is more uniform in brain tissues and levels of bias are smaller than those of FA. The within-site and cross-site comparison results are listed in Table 1 for those from GCC and PUT as examples for typical WM and GM regions. There is no obvious difference in precision of FA and MD measurements, either within-site or cross-site. However, there are obvious differences in bias level of FA and MD measurements among datasets acquired among different centers, even though the scanners used in this study were from the same vendor with similar hardware and software. On the other hand, within-site differences in bias are smaller. These results imply that cross-site variation in hardware and software conditions, although small among same vendor's scanners, does affect accuracy of DTI measurement.

Discussions: Quantification of the accuracy and precision of DTI data acquired at different sites allows us to potentially increase the statistical power with a larger subject number for a multi-center DTI study. Our study suggests that, while precision level of DTI data from different sites is not significantly affected by differences in scanner's hardware and software conditions, the bias of DTI data from each site will vary and reduce the statistical power when combining data from multiple sites together. In order to facilitate the multiple-center DTI study, a routine calibration process to regularly measure the bias level is necessary. Further study using a DTI phantom to calibrate the bias will warrant the success of a multiple-center study. Precision estimated by the wild bootstrap analysis, such as CV(FA) and CV(MD) maps, can provide complementary information to evaluate the significance of DTI findings and can be integrated into a weighted t-test framework [5] to increase the statistical power.

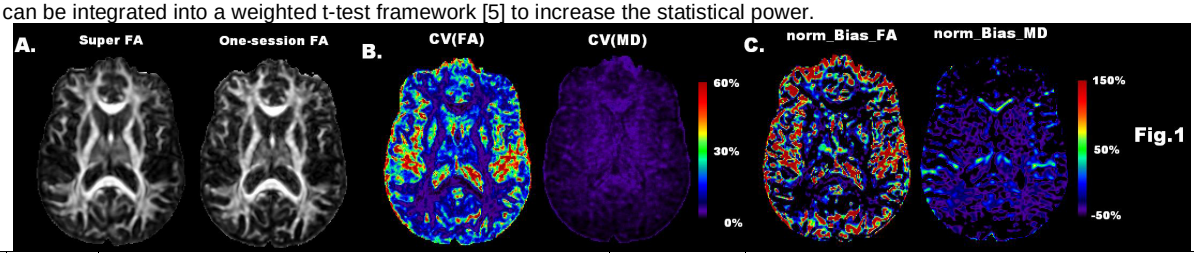

Table 1		Genu Corpus Callosum (GCC)							Putamen (PUT)					
		FA			MD				FA			MD		
		Site1	Site2	Site3	Site1	Site2	Site3		Site1	Site2	Site3	Site1	Site2	Site3
norm_Bias (%)	Scan1	3.32	-2.35	0.22	-2.83	7.52	2.89		51.35	34.09	46.42	-0.55	2.07	4.45
	Scan2	2.07	-3.15	0.93	-.364	7.60	3.19		43.71	26.71	35.58	-0.56	0.94	3.46
σ(FA)/ σ(MD)	Scan1	0.028	0.032	0.033	0.028	0.030	0.032		0.041	0.037	0.046	0.017	0.016	0.020
	Scan2	0.028	0.036	0.031	0.026	0.033	0.030		0.037	0.037	0.041	0.015	0.016	0.019

Reference: [1]. Marenceo et al. Psychiatry Res 2006; 147:69-79; [2].Pfefferbaum et al. J Magn Reson Imaging 2003; 18:427-433; [3]. Whitcher et al. Human Brain Mapping 2008; 29:346-362. [4]. Zhu et al. Neuroimaging 2008; 40:1144-1156. [5]. Bland et al. BMJ 1998; 316:129.