Statistical Assessment of the Effects of Physiological Noise and Artifacts in a Population Analysis of Diffusion Tensor MRI Data

L. Walker[1], L-C. Chang[1,2], E. Kanterakis[3], L. Bloy[3], K. Simonyan[4], R. Verma[3], and C. Pierpaoli[1]

[1]NICHD, NIH, Bethesda, MD, United States, [2]Dept of EE & CS, Catholic University of America, Washington, DC, United States, [3]Dept of Radiology, University of Pennsylvania, Philadelphia, PA, United States, [4]NINDS, NIH, Bethesda, MD, United States

Introduction: Corrupted DWI data is known to have a profound effect on diffusion tensor (DTI) derived quantities such as anisotropy (FA), and mean diffusivity (Tr(D)). Corruption can occur for many reasons including bulk subject motion and pulsatile motion from the cardiac cycle[1]. The use of a robust tensor fitting algorithm is an effective way of removing corrupted data on a voxel-by-voxel basis. The RESTORE[2] robust tensor fitting algorithm identifies corrupt data points and rejects them as outliers. These outlier data points are then removed from the tensor fitting calculations. It has been previously shown that outlier data points are regionally consistent across a population, and are heterogeneously distributed throughout the brain[3] (**Fig. 1**), and that this can have an effect on the statistical analysis of a patient population[4]. However, the exact nature of the effect is unknown. The presence of outliers could bias the value of your metric, causing a systematic increase or decrease. It could also increase the variance, modulating the statistical power in a heterogeneous way across the brain. There is also no guarantee that these effects are the same for different tensor derived quantities. We present an analysis of the effect of outlier rejection on the statistical analysis of a population.

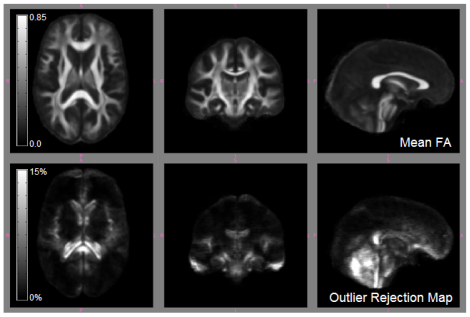

Fig 1. Top Row: Mean FA map (non-linear tensor data). Note the crisp structural detail indicating good spatial normalization. Bottom Row: Mean outlier map. Note the regional consistency of rejected data points on the average population.

Methods: ***Subjects*** – 20 healthy volunteers (age 52.7 ± 9.49 years, 8 male) were scanned on a 3.0T GE Excite scanner using an eight-channel coil. Whole brain single-shot echo-planar (EPI) DWI datasets were acquired: TE/TR = 73.4/13000ms, FOV=24cm, matrix=96 x 96 zero filled to 256 x 256, with 54 slices at 2.4mm thickness, b=1000s/mm^2 in 33 non-collinear directions, plus 3 images at a b=0s/mm^2, two replicates, SENSE factor=2. No cardiac gating was performed.

Post-processing – Images were corrected for motion, eddy current distortion[5] and EPI distortion[6]. Tensor fitting was performed twice on each subject, once using a non-linear least squares algorithm, and once using the RESTORE robust method to identify and reject outliers. An outlier map was calculated for each subject. One subject was selected as a template and all remaining subjects were spatially normalized to that template using a method that performs a fully deformable registration of the tensor data to the target data including reorientation of the tensors post normalization[7]. The deformation of each subject was applied to their respective FA, Tr(D), and outlier maps. Spatially normalized mean and standard deviation maps were calculated from the subject's outlier, FA and Tr(D) maps for both tensor fitting methods.

Analysis – Differences in the mean and variance of FA and Tr(D) between robust and non-robust are compared by a simple subtraction of the mean and standard deviation maps. Additionally, paired statistical analysis was performed on the FA and Tr(D) maps on a voxel-wise basis, using the threshold free cluster enhancement[8] option in Randomise[9].

Results: Subtraction of the mean maps (non-linear minus RESTORE) shows regions of bias in both FA and Tr(D) which are consistent for the population of subjects. These areas are generally co-localized with areas of higher percentage outliers (**Fig. 1**). If we assume that the RESTORE fitting produces more accurate results, bright regions on these maps indicate areas where FA and Tr(D) are overestimated by the non-linear fit, while dark regions indicate areas where FA and Tr(D) are underestimated. In the cerebellum (**Fig. 2**) we see that both FA and Tr(D) are bright. This is consistent with previous reports of a pseudo-diffusion effect of increased intra-voxel incoherent motion caused by cardiac pulsation, leading to an over-estimation of both FA and Tr(D) in these regions[1]. In the genu of the corpus callosum we see underestimation of FA and overestimation of Tr(D) (not pictured) which is consistent with a shearing of the tissue, also likely caused by cardiac pulsation. All ventricular regions, including the lateral, 3rd and 4th ventricles (the latter is visible on **Fig. 2**), are dark on the subtraction maps, indicating higher FA and Tr(D) with the RESTORE algorithm. A subtraction of the standard deviation maps for both FA and Tr(D) shows increases in variance in these ventricular areas with the use of RESTORE; a result which was unexpected. This increased variance appears in all areas containing CSF. We attribute this to instability in the RESTORE algorithm with regards to rejection of the T2 weighted images. The above results are confirmed by the pairwise cluster-based statistics (**Fig. 3**). Orange clusters show areas of overestimation of FA and Tr(D) with the non-linear fit, while blue clusters show areas of underestimation. CSF containing regions have been masked out because of the above mentioned instabilities of the RESTORE algorithm in those regions.

Discussion and Conclusion: The presence of outliers in the DWIs originating from physiological noise and other artifacts has a statistical effect on the DTI derived metrics such as FA and Tr(D) in a population based analysis. The effect is regionally varying across the brain, and is not the same for different tensor derived metrics. When considering a statistical analysis of a population, one needs to consider that the effect of outliers may not be the same for both patient and control populations. If one population contains more physiological noise than the other, one could find statistically significant results that are attributable to the presence of outlier data points as opposed to the presence of the disease or pathology.

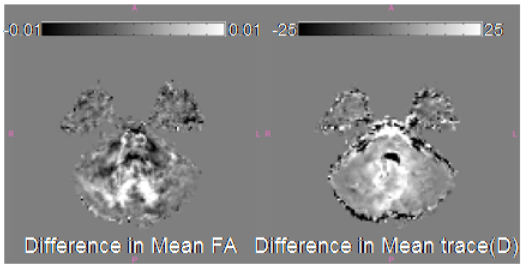

Fig 2. Difference in mean FA and Tr(D) at the level of the cerebellum

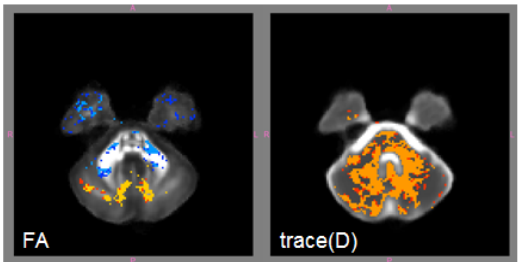

Fig 3. Orange: overestimation with non-linear fit, Blue: underestimation. p≤0.0005

References: 1. Pierpaoli C. et al. *ISMRM 11th Ann. Mtg*, Toronto, 2003, p70; **2**. Chang LC. et al. *MRM* 2005; 53:1088-1095; **3**. Walker L. et al. *ISMRM 16th Ann. Mtg*, Toronto, 2008, p139; **4**. Peterson D.J. et al. *ISMRM 16th Ann. Mtg*, Toronto, 2008, p1823; **5**. Rohde GK. et al. *Magn Reson Med* 2004; 51: 103-114; **6**. Wu M. et al. *ISMRM 15th Ann. Mtg*, Berlin, 2007, p1591; **7**. Yang J. et al. *MICCAI*,New York, 2008, p905-913; **8**. Smith S.M. and Nichols T.E. *NeuroImage*, 2009; 44(1): 83-98; **9**. Nichols T.E. and Holmes A.P. *Human Brain Mapp*, 2002; 15:1-25