A critical consideration on the absence of significance and the impact of structure size when interpreting DTI and DKI results

Filip Szczepankiewicz[1], Jimmy Lätt[2], Ronnie Wirestam[1], Alexander Leemans[3], Pia Sundgren[2,4], Danielle van Westen[2,4], Freddy Ståhlberg[1,4], and Markus Nilsson[1]
[1]*Department of Medical Radiation Physics, Lund University, Lund, Sweden,* [2]*MR Department, Lund University Hospital, Lund, Sweden,* [3]*Image Sciences Institute, University Medical Center Utrecht, Utrecht, Netherlands,* [4]*Department of Diagnostic Radiology, Lund University, Lund, Sweden*

Introduction: Diffusion tensor imaging (DTI) and diffusion kurtosis imaging (DKI) parameters are commonly compared between groups to detect potential differences in the tissue of the brain. This analysis is frequently based on statistical tests where a significant difference between groups is interpreted as a difference in microstructure integrity. However, such a conclusion may be challenged for a number of reasons [1], among them the influence of the partial volume effect (PVE) on the diffusion parameters [2]. Furthermore, the interpretation of results may be biased if it rests solely on the level of significance, while the effect size (ES) and statistical power (π) of the test is disregarded, especially if pathogenesis patterns are inferred from the data. In this study, we show how the statistical power of mean diffusivity (MD), fractional anisotropy (FA), mean kurtosis (MK) and radial kurtosis (RK) varies as a function of the anatomical position along the cingulum (CG). We also evaluate the influence of PVE by correlating the parameters to the size of the CG, and demonstrate how a correction for covariance may improve statistical power.

Methods: The study was based on DKI data acquired in 31 healthy volunteers at 3T (Philips Achieva), employing 15 diffusion encoding directions with *b*-values of 0, 500, 1000, 2500 and 2750 s/mm^2, 35 contiguous axial slices, TE and TR of 76 and 7855 ms, respectively, and a spatial resolution of 2×2×2 mm^3. The cingulum bundle was defined by three ROIs placed on a color-FA-map (bilaterally), using the anterior (Ant), central (Cent), and posterior (Post) corpus callosum as anatomical reference (Fig. 1). Tractography-based segmentation was performed in *TrackVis* [3]. The CG size estimation was based on the tractography, and was quantified in terms of its apparent radius (AR), i.e., the radius of a circle with an area equal to that of the tract cross section [4]. The in-house developed *QuTE* framework was used for evaluation of diffusion, size as well as statistical parameters along the structure [4]. To simplify the presentation, parameters were averaged across the left and right CG bundle.

Analysis of how MD, FA, MK, and RK correlated with AR was accomplished by using Pearson's correlation coefficient (*r*). The variance of the diffusion parameters, induced by the AR, was removed by subtracting its regression component while retaining the group mean value, thus generating "corrected" data. The statistical power was calculated along the structure for both corrected and uncorrected data, based on a two-tailed *t*-test. Results were evaluated in terms of the statistical power that would be generated if the healthy group was compared to a hypothetical group with equal group size ($n = 31$) and variance, assuming a relative effect size of 5% (parameter mean value differs by 5% between groups), and a level of significance of $p < 0.05$.

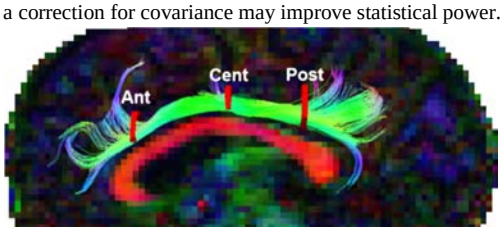

Figure 1 – Tractography of the left-hand side cingulum bundle in a healthy subject. The selected ROIs (Ant, Cent and Post) are shown as red vertical lines.

Results: Figure 1 shows a reconstructed CG, with the ROIs placed in relation to the corpus callosum. Figure 2 shows the average and standard deviation of the DKI parameters along the CG, as well as horizontal bars indicating the statistical power, in blue and red for uncorrected and corrected data, respectively. The positions of the bars indicate the regions in which the power is $\pi > 0.9$ (top row of bars) and $\pi > 0.6$ (bottom row of bars). Figure 3 shows the correlation between FA and AR.

The power is heterogeneous along the structure, as shown by the varying width of the standard deviation fields and the power indication bars. The diffusion parameters exhibit different whole- and within-structure power patterns. Power was highest for MD, followed, in descending order, by MK, FA and RK. For FA, the highest power was seen in central parts of the CG, while, in contrast, the highest power for MK was present at the anterior and posterior ends. The coefficient of correlation to AR was significant for all parameters except MD, with *r*-values of: –0.32 ($p = 0.08$), 0.80 ($p < 10^{-7}$), 0.40 ($p < 0.05$), and 0.45 ($p < 0.05$) for MD, FA, MK, and RK, respectively. For parameters that correlated significantly with AR, correction for the correlation reduced the parameter variance, which led to increased statistical power (seen as an extended coverage along the tract of the two red power bars in Fig. 2). This is most prominent for FA (Fig. 3) where the global coefficient of variation (ratio of standard deviation to mean value) was lowered from 6.1% to 3.6% after correction, lowering the effect size required to yield a statistical power of 0.8, from 4.5% to 2.7%, i.e., making it more sensitive.

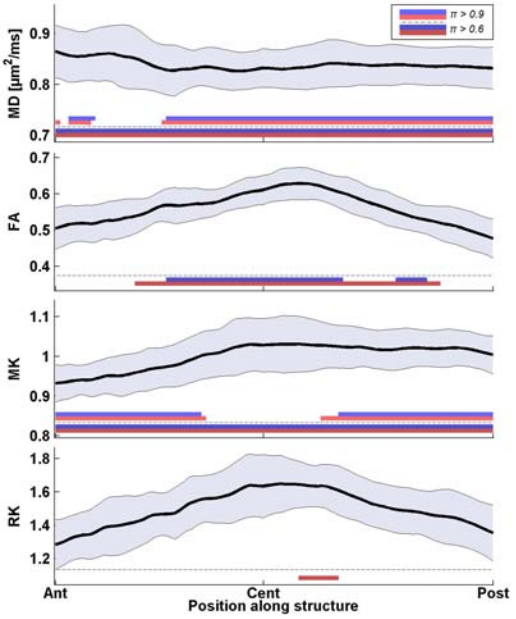

Figure 2 – Plots of group mean parameter value (black line) and one standard deviation from the mean (grey field) along the CG (anterior to posterior). Two power levels are depicted (top row of bars, $\pi > 0.9$, and bottom row of bars, $\pi > 0.6$) for data before correction (in blue) and after correction (in red). As expected, an increase in standard deviation can be seen to correlate with a decrease in power.

Discussion: Non-significant test results are often interpreted as evidence that no difference exists between groups, despite the fact that such a result can be due to either the absence of a relevant effect or to the lack of statistical power, i.e., the variance being too large to render statistically significant differences at a given effect size [5]. However, if the test is accompanied by a statistical power analysis, one can decide which of these situations is most probable, enabling a more correct interpretation.

Using the current premise, data showed that MD and MK would generate significant results more readily than FA and RK, even with a true 5% group difference in each parameter (Fig. 2). Hence, knowledge of the power allowed for the conclusion that this hypothetical study was underpowered, rather than the conclusion that no actual difference in FA and RK existed between the groups.

The correlation between FA and AR indicates a strong influence of PVE. This is to be expected in the CG, due to the high contrast in FA between white matter of the CG and the surrounding grey matter [2]. Correcting for correlation may reduce the unexplained variance, thereby reducing the group size required to generate a given statistical power. Correction may also be important when comparing groups that are not matched with respect to structure size, due to the relation between tract size and FA [2]. In such cases, group-wise differences in structure size may cause differences in observed diffusion parameters. In the data presented, the regression line in Fig. 3 indicates that a 10% difference in mean structure size could yield a difference in FA of approximately 4%. In conclusion, we demonstrated that awareness of statistical power is important when interpreting DTI and DKI results, and that the size of the evaluated structure is a potential confounder.

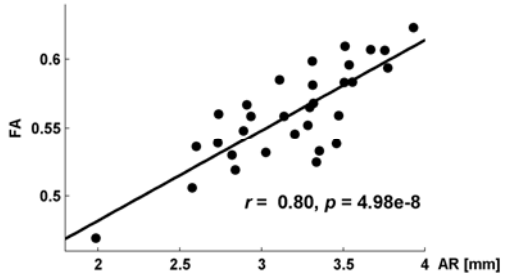

Figure 3 – Correlation plot between average FA and average AR for the 31 subjects. A strong correlation indicates that AR is a relevant predictor of the FA value. According to Vos *et al.* [2], the causal relationship between size and FA is mediated by a size-dependent partial volume effect.

References: [1] Jones, D. *et al.* Neuroimage. 2012, 10.1016/j.neuroimage.2012.06.081 – [2] Vos, S. *et al*. Neuroimage. 2011, 10.1016/j.neuroimage.2011.01.048 – [3] Wang, R. *et al.* Intl Soc Mag Reson Med. 2007, 15:3720 – [4] Mårtensson, J. *et al.* Intl Soc Mag Reson Med. 2011, 19:2335 – [5] Cohen, J. Psychol Bull, 1992, 112:155-159