Power and variability analysis in diffusion kurtosis imaging: Sample size estimation in three white matter structures

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

Introduction: Diffusion kurtosis imaging (DKI) allows for more complete characterization of water diffusion in the brain than diffusion tensor imaging [1]. Prior to its application as a biomarker for disease, the variability of DKI estimates need to be assessed in order to calculate the number of subjects needed to find a significant difference in DKI metrics between two groups. In this study, we determine the group size required to detect an effect size of 10 % at a statistical power of 0.9 and a significance level of 0.05 in the mean value of four DKI metrics, mean diffusivity (*MD*), fractional anisotropy (*FA*), mean kurtosis (*MK*) and radial kurtosis (*RK*). Further, bootstrapping was used to identify the primary source of DKI variability by separating the measurement noise component from the total variability.

Materials and Methods: DKI was performed in 20 healthy volunteers using a Philips Achieva 3T employing 15 diffusion encoding directions with *b*-values of 0, 500, 1000, 2500 and 2750 s/mm^2; 35 contiguous axial slices; *TE*/*TR* of 76 and 7855 ms respectively; and a spatial resolution of 2×2×2 mm^3.

Three white matter (WM) tracts, i.e. cingulum (CG), corticospinal tract (CST) and corpus callosum (CC), were extracted using tractography-based segmentation as implemented in *ExploreDTI* [2]. The CC was represented by its body, the CST by its course from the cerebral peduncle to the motor hand area and the CG by its course along the body of the CC. Each tract was subdivided into three segments (Fig 1). Spatial evaluation along the tracts was performed using the in-house developed *QuTE* framework [3].

A statistical power analysis was based on a two-tailed t-test assuming equal sample size and equal variance in the healthy group and the hypothetical patient group, where the effect size was set to 10 %. Under these assumptions we calculated the sample size (n) required to obtain a statistical power of 0.9 and a significance of $p < 0.05$ for each WM tract and subsegment.

The total group variance (V_{tot}) was divided into biological (V_B) and measurement variance (V_M) where $V_{tot} = V_B + V_M$. Measurement variability is introduced by image noise, post-processing and data analysis. The value of V_{tot} was estimated from measured volunteer data, whereas the value of V_M was estimated by parametric bootstrapping that assumed normally distributed residuals with a spatially varying variance. The bootstrapping produced complete datasets, which were evaluated using the same method as was employed for volunteer data.

Results: DKI parameters were evaluated along the CG, the CST (Figure 1), and the CC in three subsegments of each structure (Table 1). The statistical power analysis showed that *MD* required the smallest group size, whereas *RK* requires the largest; detecting a 10 % change in *RK* requires 3 to 4 times as many subjects than for a similar change in *MD*. When the whole tract instead of the subsegments is analyzed, the required group size is reduced by approximately 30% (data not shown).

The variance analysis showed that the most prominent contributor to the total variance is the inter-subject biological variation (Table 2).

Discussion: The statistical power analysis indicated that DKI demands sample sizes of approximately 10 to 30 subjects, depending on which metric and which structure that is studied (Table 2). Evaluating complete tracts, instead of subsegments, requires smaller groups; however, local changes such as those provoked by focal diseases such as tumors or stroke, may well affect limited parts of tracts and may therefore have a low impact on the tract-averaged value of the metric of interest.

Our estimated values of *MD* in the CC were relatively high, possibly due to partial volume effects at the ventricle interface. This is also reflected in the high variability of the DKI metrics (Table 1) and in the group size calculations (Table 2).

Variance analysis showed that inter-subject variability contributed the most to the total group variability. This implies that the sensitivity of DKI metrics would benefit more from increasing the sample size than from increasing the SNR by, for example, increasing the scan time. Further, the analysis suggests that sub-segmentation of WM tracts does not significantly impair the sensitivity but rather increases it in cases of focal damage.

A possible shortcoming of this study was that intra-subject variability was addressed by bootstrap analysis and should be validated by performing measurements of inter-scan variability.

In conclusion, this study indicated that group sizes of approximately 15 subjects are required to reliably detect differences of 10 % in the group average of *MD*, *FA* as well as *MK*, and that subsegmentation the WM tracts is a viable way to increase local sensitivity.

References:
[1] Jensen, J. & Helpern, J. NMR Biomed. 2010, 23:698-710
[2] Leemans, A. *et al.* Proc. Intl. Soc. Mag. Reson. Med. 2009, 17:3537
[3] Mårtensson, J. *et al.* Proc. Intl. Soc. Mag. Reson. Med. 2011, 19:2335

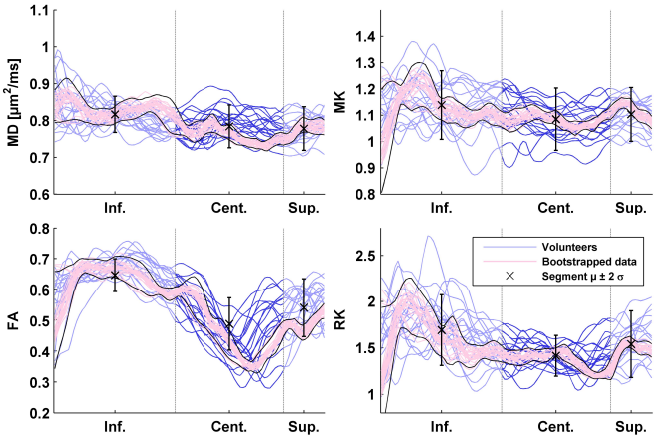

Figure 1 – DKI parameters as a function of position along the left hand side CST. Blue lines show results from the 20 volunteers and pink lines, framed by black lines, show results from 20 bootstrapped data sets based on the DKI from one of the volunteers. The simulated measurement variance, V_N, is seen as the variability of the pink lines and is much smaller than the total group variance V_{tot}.

Table 1 – Mean values of (μ) and standard deviations (σ) of four DKI metrics in three subsegments (anterior A, central C, posterior P, inferior I and superior S) of the CC, the left CG and the left CST, across 20 healthy volunteers.

		MD [μm²/ms]		*FA*		*MK*		*RK*	
		μ	σ	μ	σ	μ	σ	μ	σ
CC	A	1.06	0.06	0.63	0.03	0.91	0.04	1.48	0.15
	C	1.13	0.08	0.59	0.03	0.93	0.06	1.52	0.17
	P	1.05	0.08	0.68	0.03	1.10	0.06	2.09	0.27
CG	A	0.87	0.04	0.53	0.06	0.92	0.06	1.41	0.17
	C	0.84	0.04	0.61	0.04	0.98	0.05	1.62	0.17
	P	0.83	0.04	0.52	0.04	0.99	0.05	1.49	0.17
CST	I	0.82	0.02	0.65	0.03	1.14	0.07	1.69	0.19
	C	0.78	0.03	0.49	0.04	1.08	0.06	1.41	0.11
	S	0.78	0.03	0.54	0.05	1.10	0.05	1.54	0.18

Table 2 – Resulting values of the group size (n) and relative variance contribution. The value of n shows the group sizes required in order to obtain a statistical power of 0.9, assuming a 10% change in the average parameter value, with a significance threshold of $p < 0.05$. The relative variances are shown as the mean variance of all three segments in the structure.

	n			Relative Variance	
Metric	CC	CG	CST	V_B/V_{tot}[%]	V_M/V_{tot} [%]
MD	12	7	5	95	5
FA	7	18	14	90	10
MK	9	9	8	95	5
RK	30	30	25	96	4

Acknowledgements:
This research was supported by the Swedish Research Council, grant no.: 2010-36861-78981-35.