Variability in diffusion kurtosis imaging: Impact on study design, statistical power and interpretation.

Szczepankiewicz, Filip; Lätt, Jimmy; Wirestam, Ronnie; Leemans, Alexander; Sundgren, Pia; van Westen, Danielle; Ståhlberg, Freddy; Nilsson, Markus

Published in:
NeuroImage

DOI:
10.1016/j.neuroimage.2013.02.078

2013

Title:

Variability in diffusion kurtosis imaging: Impact on study design, statistical power and interpretation

Author names and affiliations:

Filip Szczepankiewicz[1]
Jimmy Lätt[2]
Ronnie Wirestam[1]
Alexander Leemans[3]
Pia Sundgren[2,4]
Danielle van Westen[4]
Freddy Ståhlberg[1,4,5]
Markus Nilsson[5]

[1] Department of Medical Radiation Physics, Lund University, Lund, Sweden
[2] MR Department, Centre for Medical Imaging and Physiology, Lund University Hospital, Lund, Sweden
[3] Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands
[4] Department of Diagnostic Radiology, Lund University, Lund, Sweden
[5] Lund University Bioimaging Center, Lund University, Lund, Sweden

Corresponding author:

Filip Szczepankiewicz
Lund University, Department of Medical Radiation Physics
Barngatan 2, 22185, Lund, Sweden
E-mail: filip.szczepankiewicz@med.lu.se
Telephone: +46–46–178543

Manuscript info:

Word count – 6200 + figure/table captions
Figure count – 4
Table count – 4

Abstract:

Diffusion kurtosis imaging (DKI) is an emerging technique with the potential to quantify properties of tissue microstructure that may not be observable using diffusion tensor imaging (DTI). In order to help design DKI studies and improve interpretation of DKI results, we employed statistical power analysis to characterize three aspects of variability in four DKI parameters; the mean diffusivity, fractional anisotropy, mean kurtosis, and radial kurtosis. First, we quantified the variability in terms of the group size required to obtain a statistical power of 0.9. Second, we investigated the relative contribution of imaging and post-processing noise to the total variance, in order to estimate the benefits of longer scan times versus the inclusion of more subjects. Third, we evaluated the potential benefit of including additional covariates such as the size of the structure when testing for differences in group means. The analysis was performed in three major white matter structures of the brain: the superior cingulum, the corticospinal tract, and the mid-sagittal corpus callosum, extracted using diffusion tensor tractography and DKI data acquired in a healthy cohort. The results showed heterogeneous variability across and within the white matter structures. Thus, the statistical power varies depending on parameter and location, which is important to consider if a pathogenesis pattern is inferred from DKI data. In the data presented, inter-subject differences contributed more than imaging noise to the total variability, making it more efficient to include more subjects rather than extending the scan-time per subject. Finally, strong correlations between DKI parameters and the structure size were found for the cingulum and corpus callosum. Structure size should thus be considered when quantifying DKI parameters, either to control for its potentially confounding effect, or as a means of reducing unexplained variance.

Keywords:

Diffusion kurtosis imaging, Statistical power, Study design, Group size, Tractography

1. Introduction

Diffusion kurtosis imaging (DKI) is a technique that has been suggested to show higher sensitivity and specificity than diffusion tensor imaging (DTI) in detecting and differentiating alterations of tissue microstructure (Cauter et al., 2012, Cheung et al., 2009, Grossman et al., 2012, Wang et al., 2011, Wu and Cheung, 2010). Being an extension of DTI, DKI provides conventional DTI-based parameters, such as the mean diffusivity (MD) and the fractional anisotropy (FA), and unique parameters that describe the degree to which the water diffusion is non-Gaussian. This information is most commonly represented by the mean diffusional kurtosis (MK) and radial diffusional kurtosis (RK) (Jensen et al., 2005, Jensen et al., 2010), that can be related to properties of the tissue microstructure, for example, the axonal water fraction and the tortuosity of the extracellular space in white matter (WM) (Fieremans et al., 2011). In its application to clinical research, DKI has rendered promising results in studies of, for example, reactive astrogliosis (Zhuo et al., 2012), age-related diffusional changes (Falangola et al., 2008), and has been reported to outperform conventional DTI in the detection of Parkinson's disease (Wang et al., 2011) and in the grading of gliomas (Cauter et al., 2012). DKI has also been performed outside of the brain, for example, in the spinal cord (Hori et al., 2012, Szczepankiewicz et al., 2011).

In light of the emerging popularity of DKI, it is interesting to elucidate the statistical characteristics of the extracted parameters. Using a statistical power analysis, the variability of any parameter can be evaluated in terms of, for example, the minimal group size required to detect a true difference in means (effect size) at a predefined probability

(statistical power) (Cohen, 1976, Lenth, 2001, Maxwell et al., 2008). It may also inform better interpretation of experimental results by complementing statistical significance tests with information about the probability at which the test successfully rejects a false null hypothesis (Cohen, 1976).

A prerequisite to perform a power analysis is knowledge of the parameter variance and relevant effect size. Several studies have been dedicated to analyzing variability in DTI parameters. Heiervang et al. (2006) performed a statistical power analysis for several WM structures and various tracking methods, showing that inter-subject coefficients of variation (CV) for MD and FA were below 8% and 10%, respectively. Variations in the mean and standard deviation of DTI parameters have also been demonstrated *within* WM structures (Colby et al., 2012, Corouge et al., 2006, Wakana et al., 2007). Wakana et al. (2007) investigated the reproducibility in FA and structure size in several WM structures, and found that a 10% difference in fiber-bundle volume required a group size 10 times larger than that required to detect a 10% difference in FA, indicating a higher variance in the size parameter compared to FA. Variability is also introduced by the hardware and the post-processing of data. Pfefferbaum et al. (2003) compared within- and between-scanner reliability on two similar but not identical scanners, and reported a systematic mean bias across scanners with CVs of 7.5% and 4.5% for MD and FA, respectively. Few studies have analyzed the variability of DKI-specific parameters, however, data reported by Lätt et al. (2012), on the mean and standard deviations in 21 manually segmented structures, can be used to calculate CVs for the most frequently used DKI parameters. The CV, averaged across all structures, was lowest for MD and MK, with values of 5% and 8%, respectively, and highest in FA and RK with values of 10% and 14%, respectively. These

values indicate that the variability in MK and RK is larger but comparable to that found for MD and FA. However, more detailed information could improve study design and aid the interpretation of experimental results.

The aim of this study was, therefore, to evaluate three aspects of DKI parameter variability: the global and along-tract variability, the inter- and intra-subject variability, and the amount of variability explained by the WM structure size. The results were used to estimate the minimal group sizes required to find a physiologically relevant effect size, to quantify the advantage of increasing group size versus extending scan time per subject, and to estimate whether the introduction of additional covariates, such as the structure size, may lower demands on group size. The study was based on three major WM structures in the brain, defined using tractography-based segmentation.

2. Theory

2.1. Statistical power and group size

The power of a statistical test (π) represents its probability to correctly reject the null-hypothesis, i.e., "there is no significant difference in means between two groups". For a *t*-test, π can be estimated from the *t* statistic and the number of samples in each group, here referred to as the group size (n), given a predefined significance level (α) and an effect size defined as the absolute ($\Delta\mu$) or relative ($\Delta\mu/\mu$) difference in group means, respectively. The *t* statistic used for testing whether the means of two groups are significantly different is given by

$$t = \frac{\Delta\mu}{\mathrm{SE}(\Delta\mu)} = \frac{\Delta\mu}{\sqrt{2V/n}}, \qquad \text{Eq. 1}$$

where $\mathrm{SE}(\Delta\mu)$ is the standard error of the difference in group mean values, given by $\mathrm{SE}(\Delta\mu) = (2V/n)^{1/2}$ if the two groups are equal in size and have equal variance (V) (Vittinghoff et al., 2005).

Statistical power analysis may also be used to predict how a modification to an experimental protocol will influence the minimal group size. Below, we analyzed the influence on group size requirements from study-design alterations such as extending the acquisition time or correcting for hidden covariates.

2.2. Parameter variance

Since the statistical power is related to the variance of the parameter under investigation, reducing the variance will reduce the required group size. The measured parameters can be modeled by a stochastic variable *Y,* described by the population mean (μ), the group-dependent deviation from the mean, that is the effect size ($\Delta\mu$), and a stochastic error term (E_{total}), according to

$$Y = \mu + \Delta\mu \cdot G + E_{\mathrm{total}}, \qquad \text{Eq. 2}$$

where G = [0,1] is a discrete index of group affiliation (G = 0 for controls and G = 1 for the experimental or patient group) (Vittinghoff et al., 2005). The error term can be described by a two-level random-effects model, where E_{total} is the sum of two independent error terms $E_{\mathrm{total}} = E_{\mathrm{inter}} + E_{\mathrm{noise}}$ (Clayden et al., 2006, Laird and Ware, 1982). Here, E_{inter} and E_{noise} represent the inter-subject variability and the variability introduced by imaging

and post-processing noise, with variances V_{inter} and V_{noise}, respectively. The total variance is thus the sum of the inter-subject and noise variances, according to

$$V_{\text{total}} = V_{\text{inter}} + V_{\text{noise}}. \qquad \text{Eq. 3}$$

Estimating the total variance in a new acquisition protocol (V'_{total}) is possible by studying how the noise component is modified, according to

$$V'_{\text{total}}(g) = V_{\text{inter}} + \frac{V_{\text{noise}}}{g^2}. \qquad \text{Eq. 4}$$

Two important factors affecting g are the signal-to-noise ratio per signal acquisition (SNR), and the acquisition time (T) of the new and the old protocol: $g \propto (T'/T)^{1/2} \cdot (\text{SNR}'/\text{SNR})$, assuming that T is proportional to the total number of acquired images. The factor g, and the new group size (n') both have an effect on the denominator in Eq. 1, according to

$$\text{SE}(\Delta\mu') = \text{SE}(\Delta\mu) \cdot \sqrt{\left(1 - RV_{\text{noise}} \cdot \left(1 - \frac{1}{g^2}\right)\right) \cdot \frac{n}{n'}}, \qquad \text{Eq. 5}$$

where $RV_{\text{noise}} = V_{\text{noise}}/V_{\text{total}}$ is the relative variance contribution from noise in the old protocol. Assuming large groups, the new and old protocol will have equal power if $\text{SE}(\Delta\mu') = \text{SE}(\Delta\mu)$, and the new group size will be given by

$$n' \approx n \cdot \left(1 - RV_{\text{noise}} \cdot \left(1 - \frac{1}{g^2}\right)\right). \qquad \text{Eq. 6}$$

Eq. 6 shows that an increase in g has the strongest effect on n' when RV_{noise} is relatively large, that is when most of the total variance is due to noise. In other words, for a fixed

statistical power, an increase in SNR or T can reduce the demand on group size n'. Likewise, a reduction in total scan time would increase the demand on the group size.

2.3. Parameter covariance

DKI parameters are influenced by properties of the tissue microstructure (Fieremans et al., 2011), but may also be affected by other factors, such as the partial volume effect (PVE) (Cao and Gold, 2008, Vos et al., 2011), image distortions, subject motion and post-processing, among many others (Jones and Cercignani, 2010). Some of these effects may be corrected for by expanding the model in Eq. 2 to include additional predictors. The addition of one predictor (x) to Eq. 2 results in

$$Y = \mu' + \Delta\mu' \cdot G + k \cdot x + E'_{\text{total}}, \qquad \text{Eq. 7}$$

where k denotes the regression coefficient of the predictor, and E'_{total} is the new error term (Vittinghoff et al., 2005). Identifying significant predictors means that their contribution to the variance of the error factor can be removed, resulting in a modified residual variance, according to

$$V'_{\text{total}} = V_{\text{total}} \cdot \frac{1 - R^2_{Y,[G,x]}}{1 - R^2_{Y,G}} \cdot \frac{2n - 2}{2n - 3}, \qquad \text{Eq. 8}$$

where $R^2_{Y,G}$ is the coefficient of determination for regression of Y on the group term G, and $R^2_{Y,[G,x]}$ is the coefficient of determination for regression of Y on G and the predictor x. The effect on the standard error of the estimated effect size is

$$\mathrm{SE}(\Delta\mu') = \mathrm{SE}(\Delta\mu) \cdot \sqrt{\frac{V'_{\mathrm{total}}}{V_{\mathrm{total}}} \cdot \frac{n}{n'} \cdot \frac{1}{1 - R^2_{G,x}}}, \qquad \text{Eq. 9}$$

where the term $(1 - R^2_{G,x})^{-1}$ is commonly referred to as the variance inflation factor, since it inflates the standard error of $\Delta\mu$ in cases where correlation between G and x exists, and may even outweigh the benefits of an additional predictor (Vittinghoff et al., 2005). However, if the groups are matched with respect to x, i.e., the two groups have equal mean values of x, the value of $R^2_{G,x}$ is zero, resulting in no inflation and a guaranteed reduction in the standard error of the estimated effect size. Assuming that the compared groups are large ($2n - 2 \approx 2n - 3$, in Eq. 8) and matched with respect to x ($R_{G,x} = 0$, in Eq. 9), the minimal group size after accounting for the additional covariate is given by

$$n' \approx n \cdot \frac{1 - R^2_{Y,[G,x]}}{1 - R^2_{Y,G}}. \qquad \text{Eq. 10}$$

In analogy with the improvements arising from increased SNR or extended acquisition times, Eq. 9 and Eq. 10 show that reducing the standard error of $\Delta\mu$, by accounting for covariates, can be translated into increased statistical power or reduced demands on group size.

3. Methods

3.1. Data acquisition and post-processing

In order to assess the variability characteristics of DKI parameters, DKI was performed on 31 healthy volunteers (12 male, 19 female, age 36 ± 13 years). The study was approved by the local ethics committee and informed consent was obtained from all volunteers. Imaging was performed on a Philips Achieva 3 T MRI scanner, with a maximum gradient

amplitude of 80 mT/m, using an 8-channel head coil. The DKI protocol consisted of one volume acquired with $b = 0$ s/mm^2, followed by 60 diffusion-weighted volumes in which the diffusion encoding was applied in 15 non-collinear encoding directions with *b*-values of 500, 1000, 2500 and 2750 s/mm^2. The selection of *b*-values was based on the protocol optimized by Poot et al. (2010). The image volume consisted of 35 contiguous axial slices at a spatial resolution of 2×2×2 mm^3, covering the CG, CC and CST (from the cerebral peduncle to the centrum semiovale). The echo time (TE) was 76 ms, repetition time (TR) was 7855 ms, half-scan factor was 0.78, SENSE factor was 2, and bandwidth was 2970 Hz, resulting in a scan time of 8:15 min. Motion and eddy current distortions were corrected in *ExploreDTI* (Leemans et al., 2009) where *ElastiX* (Klein et al., 2010) was used to register the images. The images were inspected for motion, ensuring that no image volume was rotated more than 2.5° during the acquisition. Parameter maps, including MD, FA, MK and RK, were calculated using in-house developed software, implemented in Matlab (The Mathworks, Natick, MA, USA). In this procedure, the diffusion-weighted images were modulated with the Jacobian determinant (Jones and Cercignani, 2010). In order to mitigate the potential effects of Gibbs ringing artifacts, all image volumes were smoothed using an isotropic 3D Gaussian kernel with a full width at half maximum of 2 mm (Veraart et al., 2012). This kernel size has little effect on sensitivity and specificity (Van Hecke et al., 2009), thus, it is not expected to significantly influence the parameter precision.

3.2. Bootstrapping

To estimate the variance component caused by noise, one oversampled set of data was acquired to facilitate a bootstrap analysis (Jones and Pierpaoli, 2005). This data was acquired in an extended imaging session for one of the volunteers, in which the DKI protocol was repeated in seven subsequent acquisitions with a total scan time of approximately 65 min. The subject was not repositioned between acquisitions. By randomly selecting one out of the seven image volumes for every combination of encoding strength and direction, 200 bootstrapped data sets were created, each with a composition corresponding to those acquired in the control group. This number of bootstraps, given the seven original data sets with 60 direction and *b*-value combinations in each, is expected to generate a reliable distribution of parameters (O'Goreman and Jones, 2006), where the CV of the relative noise contribution is given by $\mathrm{CV}(RV_{\mathrm{noise}}) = (2/N)^{1/2}$, i.e., 10% for $N = 200$. Individual post-processing and parameter calculation was performed on all of the simulated sets of data in a way identical to that performed in the control group. The bootstrapping generated unique noise realizations, allowing the resulting parameter variance to be attributed to imaging and post-processing noise only and thereby provide an estimate of V_{noise} in Eq. 3.

3.3. Structure definition

Three major WM structures were investigated: the superior cingulum bundle (CG), the medial motor corticospinal tract (CST) and the mid-sagittal corpus callosum (CC). These structures were selected to represent some of the structures most commonly investigated with diffusion tensor tractography, which also offer a variety of features, such as

proximity to CSF and gray matter (GM), and varying geometrical configurations. The structures were defined in native space using manually defined geometrical inclusion criteria (AND-gates, commonly referred to as ROIs) as shown in Figure 1. The structures were segmented from a whole-brain tractography (diffusion tensor was fit to b = 0, 500 and 1000 s/mm^2), generated in *TrackVis* (Wang et al., 2007), using a deterministic interpolated streamline algorithm. Track termination was based on a FA threshold of 0.2 and an angle threshold of 30°.

The CG was delineated using three AND-gates, combined in gate pairs, and positioned to include the superior CG bundle (Fig. 1A). Gates were defined in coronal projections and the mid-sagittal CC was employed as an anatomical reference. The gates were aligned with the anterior (Ant), central (Cent) and posterior (Post) part of the mid-sagittal CC body, and landmarks were placed at the center of each gate. The CST was delineated using two AND-gates (Fig. 1D), defined in axial projections and placed around the peduncle (Inf) and the medial motor area of the cortex (Sup). Landmarks were defined at each gate and at the level of the ventricles (Cent). The CC was extracted using two AND-gates, separated by 12 mm and centered on the mid-sagittal plane, that excluded the tracts outside of the intersections so that a truncated mid-sagittal segment was selected (Fig. 1G). The landmarks were placed at the inferior edges of the genu (Ant) and splenium (Post), as well as at the boundary between the body and the genu (PreA) and the splenium (PreP), respectively.

The sub-segments of each structure were defined by the intervals between landmarks, creating two sub-segments in the CG and CST, and three sub-segments in the CC.

Tractography and parameter extraction were performed independently on all of the bootstrapped data sets.

3.4. Parameter evaluation

Diffusion parameters were calculated as a function of position to retain spatial information along the tract, employing an evaluation method resembling that presented by Colby et al. (2012). The evaluation was performed in three steps. First, a single *mean track* was created to represent the geometrical features of the track bundle. Second, diffusion parameters were projected onto the mean track to create parameter vectors. In the final step, the parameter vectors were normalized across subjects using anatomical landmarks as points of reference. Figure 1 shows representative tractographies of the CG, CST and CC (Fig. 1A, D, G), along with the *point cloud* that makes up the tracts and constituted the cross-sections selected along the mean track (Fig. 1B, E, H). All calculations were performed using in-house developed software, implemented in Matlab, and details on the three steps are given below.

The first step was to calculate the mean track, which was represented by a number of consecutive points in 3D-space ($\mathbf{m}_i$), with each point placed at the center of mass of the cross section of the track bundle. Note that the mean track in the CG and CST is directed *along* the WM fibers, while in the CC it runs *perpendicular* to the WM fibers (Fig. 1).

In the second step, projection of the diffusion parameters to the mean track was performed by averaging the parameter values from all points in the cross section associated with $\mathbf{m}_i$. The cross section included at most one point per track, with the point selected being the one closest to a plane with normal $\mathbf{n} = \mathbf{m}_{i+1} - \mathbf{m}_i$, with its origin in $\mathbf{m}_i$. Only points within

1 mm distance from each plane were included in the cross section, resulting in a cross section thickness of 2 mm. The calculation of the apparent structure size (AS) was performed by determining the apparent *radius* (in the case of the CG and CST) or *thickness* (in the case of the CC) of the tract bundle mask at each cross section. The area of the mask was calculated by representing each point in the cross-section by a circle with radius 0.5 mm (Fig 1C, F, I). Only non-overlapping parts of the circles contributed to the AS.

In the third step, the individual parameter vectors were normalized in order to align them with respect to the anatomical landmarks. Each landmark was first associated with the point on the mean track closest to the landmark, which allowed the calculation of average interval lengths, i.e., the mean path track lengths between two landmarks. Next, the mean tracks and their associated parameter vectors were linearly interpolated so that the interval lengths of the individual mean tracks conformed to the average interval lengths. Further, the mean tracks were resampled to contain 100 equidistant elements per DKI parameter and WM structure, on which the final analysis was performed. To simplify the presentation of results for bilateral structures, the CG and CST estimates were evaluated as the average of both sides for each individual subject.

3.5. Statistical analysis

The statistical analysis comprised three aspects, all performed to improve the design of future DKI studies: first, calculating the group size required to find a subtle difference in group means, second, answering the question of whether to scan longer per subject or more subjects by analyzing the relative contribution of noise to the total variance, and

third, analyzing the potential reduction in group size requirement resulting from the addition of relevant covariates.

The group sizes required to obtain a statistical power of $\pi = 0.9$ at a relative effect size of 5% (i.e., absolute effect size was $\Delta\mu = 0.05 \cdot \mu$) were calculated for whole structures and sub-segments. We assumed that the difference in group mean values was tested using a two-tailed Student's *t*-test at a significance level of $\alpha = 0.05$, assuming that the control and experimental groups were of equal sizes. Furthermore, the analysis assumed equal variance in both groups, with a value given by that observed in the group of healthy volunteers. Even at a moderate departure from the assumption of equal group size and variance, the analysis is expected to produce robust estimates of the *t*-statistic and the statistical power of the study (Cohen, 1976). The effect size was chosen to represent a subtle but physiologically relevant change in DKI parameters, according to a survey of relevant DTI and DKI studies of the brain (Table 1). In this compilation, the approximate span of relative effect sizes is between 1 and 30%. However, it should be noted that the relative effect size can be much higher for more severe tissue alterations such as tumors and edema (Cauter et al., 2012, Harris et al., 2008, Jensen et al., 2011). Required group sizes were calculated by iteratively adjusting *n* until the desired statistical power was reached.

The total variance, measured in the control group, was separated into inter-subject variance and imaging noise variance in order to determine the effect of increasing scan time or group size (Eq. 6). The noise component (V_{noise}) was estimated from the bootstrapped data, by assuming that the variance in the simulated data was due to noise.

To obtain V_{inter}, the noise component was subtracted from the total variance according to Eq. 3.

DKI parameter correlation with the apparent structure size was assessed using Pearson's correlation coefficient (r). The effects of correlation on the statistical power were calculated according to Eq. 9, assuming that the two groups were matched with respect to AS, i.e., that there was no inflation due to predictor covariance ($R^2_{G,AS} = 0$).

4. Results

Figure 2 shows axial projections of the DKI parameter maps in one representative subject. Visually, the MK and RK maps are similar to the FA maps, with the highest values found in the WM. MK and RK maps are similar, since MK is partly determined by RK just as RD is partly determined by MD. The numerical values of the MK, RK, and FA maps are the lowest in the ventricles, as expected, due to the nearly unrestricted water diffusion in the ventricular cerebrospinal fluid (CSF).

Figure 3 shows the DKI parameters and AS, and their variability, as a function of anatomical position along each WM structure. The variability is represented by two components: the blue area shows two standard deviations from the mean of the inter-subject variability and the blue and red areas together show the total variability. The evaluation of parameters along structures allowed within-structure details to be resolved. For example, FA was reduced in the superior parts of the CST where the tract intersects with the CC. In the CC, MD was elevated and FA was reduced at the thinnest part (isthmus), probably due to stronger PVE with CSF at this location. This mode of visualization also supplies insight into the parameter covariance; MK generally showed

inverse correlation with MD, whereas the variation of RK exhibited similar patterns to FA and MK. The influence of noise and inter-subject variability was also dependent on position. For example, DKI parameters were more affected by noise and inter-subject variability in the inferior parts of the CST than in its superior parts (Fig. 3, center column). Table 2 presents these results in a condensed format, showing average parameter values with coefficients of variation in the sub-segments, compared with values from whole tract averages. Table 2 also shows the relative variability induced by imaging and post-processing noise, as calculated from the bootstrapped noise simulations. In most of the structures and parameters, less than 30% of the total variance was attributed to the influence of noise. The magnitude of the noise component was heterogeneous along the structures, indicated by a varying thickness of the red area in Figure 3. The value of RV_{noise} was found to be at its highest in the inferior segment of the CST, where it contributed with as much as 54% of the total variance in MD and approximately 35% of the variance in other DKI parameters (Fig. 3 and Table 2). The lowest relative noise contribution was found in the CC.

Table 3 shows the group size requirements in whole structures and in structure sub-segments, as calculated from the parameter variance. The most precise parameters, requiring the smallest group sizes, were MD ($n \approx 10$–40), followed by FA ($n \approx 10$–50). The kurtosis and structure size parameters generally demanded larger group sizes, where MK was the most precise ($n \approx 10$–70). The parameters RK and AS tended to require more than twice the number of subjects compared to any of the other parameters ($n \approx 30$–200, and $n \approx 80$–180, respectively). The worst case was found in the anterior CC where 200 subjects were required for detecting subtle group-wise differences in RK. Note that RK, in

this case, correlated strongly with MK (r = 0.93), suggesting that RK may not add substantially to the information already provided by the more precise MK. Evaluating whole structures, without dividing them into sub-segments, generally resulted in a lower group size requirement, although some combinations of structure segments and parameters exhibited behavior contrary to this generalization, for example, the MK in the posterior sub-segment of the CC. This indicates that sub-structures may exhibit smaller inter-subject variability compared to whole structures, despite having a smaller volume, thus increasing the statistical power when evaluated as a sub-structure.

Correlations between the investigated DKI parameters and the apparent structure size are shown in Table 4. Significant correlations between AS and several DKI parameters were found in the CG and CC. The most prominent correlation was found for FA in the CG ($r = 0.80$, $p < 10^{-7}$, for whole structure, Fig. 4) and for MD in the CC ($r = -0.53$, $p < 10^{-3}$, for posterior sub-segment). Adding the AS as a covariate could reduce the group size requirement by 30–60% in the CG and 20–30% in the CC (Eq. 10). No correlations between DKI parameters and AS were found in the CST.

5. Discussion

In this study, we investigated the group sizes required to find subtle differences in group means of DKI parameters in three WM structures with a statistical power of 0.9. The results, with respect to group sizes required, not only showed a large heterogeneity between the various DKI parameters and between the three WM structures investigated, but also heterogeneity between different sub-segments within the structures (Table 3). A similar heterogeneity in group size requirement has been found for DTI by Heiervang et

al. (2006) when comparing the CG, CST and CC. The heterogeneity in variability implies that, for a fixed relative effect size, the statistical power varies between structures and their sub-segments, as well as between parameters. For example, in the data presented, finding a difference in MK between two groups is more likely in the posterior CC than in its anterior part even if the relative effect size in these sub-structures is equal. Knowledge of this spatial and parameter-specific variation is expected to benefit studies aiming at early diagnosis, and it is critical when a pathogenesis pattern is inferred from the observation of significant alterations in one part of the brain before another. In other words, the conclusion that a disease did not have its origin in a given part of the brain must be accompanied by the knowledge that an effect was likely to have been discovered if, in fact, it was there. Thus, awareness of statistical power is crucial both for study design and for interpretation of results from DTI and DKI studies. Knowledge of these characteristics allows studies to be designed in a way that ensures sufficient power in all structures investigated, since the structure with the lowest statistical power defines the lower limit of the required group size. Such a procedure could result in some structures becoming overpowered, a potential downside for a study (Ferguson, 2009), hence all statistically significant group-wise differences should be scrutinized with respect to the effect size, considering its practical or physiological relevance using similar studies as a guideline (Table 1).

The analysis of the variations in variability along the structures could also be used to reduce group size demands, by sampling only those parts of a structure where the variability is expected to be low, assuming, of course, that homogeneous whole-structure alterations are expected. This conclusion is somewhat contra-intuitive, since inclusion of

larger volumes normally reduces the standard error of the mean. It should also be pointed out that the segment exhibiting minimal variability might vary depending on the evaluated parameter. An example of high group-wise variability can be seen in the superior part of the CST, where the group size for FA is three times larger than compared to the inferior part, which is likely due to the presence of crossing fibers in this region (Jeurissen et al., 2013, Vos et al., 2012). By contrast, group size demands for RK are a factor of two smaller in the superior part of CST. Thus, some WM structures could benefit from being subsampled, avoiding regions where variability is known to be high, resulting in favorable reductions in group size demands.

Two other strategies may also increase the power of a study or reduce the group size demands; first, to discern whether to prioritize longer scan times or to include more subjects when designing the study, and second, to incorporate hidden covariates in the data analysis (Vos et al., 2011). The first strategy was investigated by determining the portion of variability that could be attributed to effects other than the true differences between subjects, i.e., variability introduced by imaging and post-processing noise. This investigation was performed under the assumption that the variance of the noise component can be reduced by increasing the scan time dedicated to each subject (Eq. 5). In most structures, imaging and post-processing noise contributed with 5 to 25% of the total variance. In segments with $RV_{\text{noise}} \leq 25\%$, doubling the scan time for each subject would result in group size reductions of only 10%. In segments with higher values of RV_{noise}, such as the posterior CG, inferior CST and posterior CC, the corresponding reduction is 20%. The values of RV_{noise} reported herein are lower than those reported for a similar selection of WM structures by Clayden et al. (2009) for DTI performed at 1.5 T. In

that study, the noise component was generally dominant for both MD and FA, indicating that scan time extension could provide a viable power improvement at that field strength. By contrast, our study suggests that the gain in statistical power resulting from measuring twice as long per subject, for the DKI protocol employed here, is comparable with increasing the group size by no more than 5–20%. Therefore, it could be more profitable to invest resources in the inclusion of more patients rather than extending the individual scan time, provided that it is practically feasible.

The second strategy to increase the statistical power described in this report is to include hidden covariates in the analysis. The potential efficacy of this strategy was investigated by using the structure size as a covariate, which showed that correcting for correlations with AS could lower group size requirements by up to 60% for FA in the CG, and 30% for MD in the CC. We expected the correlation between AS and DKI parameters to be the highest for structures and parameters showing a high contrast to the surrounding tissue, as the probable mechanism responsible for the correlation is the variable amounts of partial volume effects induced by variations in structure size (Vos et al., 2011). Further, we expected this mechanism to be stronger for small structures, in which surrounding tissue comprises a larger partial volume fraction. In the data presented, the CG demonstrated these effects in accordance with our predictions in that FA, which exhibited the highest contrast between the WM of the CG and the surrounding GM (Fig. 1, coronal projections), had the strongest correlation to the size of the structure, followed by RK and MK. Further, MD did not correlate significantly with AS, again explained by the low contrast between the WM of the CG and the GM surrounding it. In the CC, MD was strongly correlated to AS, probably due to the large interface with the CSF-filled lateral ventricles. As expected,

correlations with size were absent in the CST, since its AS is highly dependent on the inclusion gate geometry rather than the structure size itself (Wakana et al., 2007). Although the strength and direction of correlation with volume may vary across the brain (Fjell et al., 2008), the presence of an association implies that any measured difference in diffusion parameters may be due to either alterations in tissue microstructure or in the amount of PVE. Disentangling these effects requires a correction for size, as described by Vos et al. (2011). For example, our results indicate that a 4% difference in FA may be induced by a radius difference of 10% in the CG, even if the microstructure is otherwise equal. Therefore, the search and correction for hidden covariates such as structure size, has the potential not only to increase the power of a given study, but also to allow for better interpretations of the results (Bendlin et al., 2010, Cao and Gold, 2008, Vos et al., 2011). Similarly the effects of age can be easily included by expanding the currently used methods. However, since the effects of aging are well documented elsewhere (Lebel et al., 2008, Löbel et al., 2009, Sullivan and Pfefferbaum, 2006), age was only considered as a possible confounder in the association between diffusion parameters and the structure size, and was found to have no significant correlation ($\alpha = 0.05$) with AS in any WM structure or sub-structure.

Finally, investigating group-wise AS differences would require much larger group sizes than for the DKI parameters, as it exhibits a large inter-subject variation ($CV \approx 10–15\%$ in all evaluated structures). This result is in agreement with multiple studies of the volumes of the healthy brain and individual structures, in which the CVs have been reported to be in the range of 10–20% (Choo et al., 2010, Flashman et al., 1997, Kristo et al., 2012, Pitel et al., 2010, Teipel et al., 2003). This indicates that a 5% effect in AS, as used in this

study, may be regarded to be small (Cohen, 1976) compared to the effect in diffusion parameters. The group size requirements in DKI as compared to DTI are expected to be higher, since diffusional kurtosis can only be probed at relatively high *b*-values with higher signal attenuation. Higher *b*-values also demands longer echo times. Taking this into account, DKI may still be preferable to DTI in tissue where the DTI model is invalid, for example, in regions with complex fiber organization. An example of this may be found in Alzheimer's disease, where the FA unexpectedly increases in areas of crossing fibers, probably due to the removal of one fiber population (Douaud et al., 2011). Notably, the MK maps are smooth in regions where the FA shows the characteristic reduction due to fiber crossings (Figure 2).

A limiting factor in the study is the bootstrapping procedure used to estimate the influence from noise since it is not exactly equivalent to repeated measurements. Although it is capable of assessing the contribution of specific sources of error (Jones and Pierpaoli, 2005), we believe that the reported magnitude of the noise component is slightly overestimated. This conclusion is supported by the observation that the variability between the seven repeated scans (data not shown), used as the base for bootstrapping, was generally lower than that found in the bootstrapped data and that it cannot be entirely explained by the expected precision in the estimation of the contribution from bootstrapping noise. For example, the seven repeated scans exhibited less of the elevated variance otherwise found in the inferior CST and posterior CG. The overestimation of variance in the bootstrapped parameter maps could be due to the large temporal spacing between images, resulting in exaggerated movement compared to a normal acquisition.

However, the conclusion derived from this evaluation, i.e., that increased group sizes improve the statistical power more than extended scan times, is still valid.

6. Conclusion

The variability in DKI parameters varies across the brain, and was seen to vary even within single WM structures. This implies that the statistical power is dependent on location, which could be a serious confound in studies aiming at early diagnosis of disease. Such studies typically focus on finding the region from which the alteration of cerebral microstructure originates. Lack of attention to the risk of being underpowered in some of the evaluated regions may lead to an incorrect interpretation of the results, i.e., the absence of significance may be interpreted as the absence of true effect. Although this study was based on the DKI model it should be noted that, since DKI includes the DTI model, these conclusions are also valid for conventional DTI.

An increase in statistical power can be achieved by extending the scan time per subject, although this was shown to be less potent than spending that time on scanning more subjects. Another strategy that may enhance the statistical power is to correct for hidden covariates, such as the size of the structure. In WM structures where the DKI parameters correlated significantly with the size of the structure, such a correction could reduce the group size requirements to approximately half of their initial size. In order to disentangle effects of variable PVE and alterations of underlying microstructure on group-wise differences in DTI and DKI parameters, correction for structure size should be performed in group comparisons, at least in the corpus callosum and cingulum.

Acknowledgements

This research project was supported by the Swedish Research Council, grants no. 2010-36861-78981-35 and 13514, and the Swedish Cancer Society grant no. CAN 2009/1076.

References

Bendlin, B. B., Fitzgerald, M. E., Ries, M. L., Xu, G., Kastman, E. K., Thiel, B. W., Rowley, H. A., Lazar, M., Alexander, A. L. & Johnson, S. C. 2010. White matter in aging and cognition: a cross-sectional study of microstructure in adults aged eighteen to eighty-three. *Dev. Neuropsychol.,* 35**,** 257-77.

Bozzali, M., Giulietti, G., Basile, B., Serra, L., Spanò, B., Perri, R., Giubilei, F., Marra, C., Caltagirone, C. & Cercignani, M. 2012. Damage to the cingulum contributes to alzheimer's disease pathophysiology by deafferentation mechanism. *Hum. Brain Mapp.,* 33**,** 1295-1308.

Cao, N. & Gold, B. 2008. Partial volume effect of cingulum tract in diffusion-tensor MRI. *Proc. SPIE,* 6916**,** 1U.

Cauter, S., Veraart, J., Sijbers, J., Peeters, R. R., Himmelreich, U., Keyzer, F., Gool, S. W., Calenbergh, F., Vleeschouwer, S., Hecke, W. & Sunaert, S. 2012. Gliomas: diffusion kurtosis MR imaging in grading. *Radiology,* 263**,** 492-501.

Cheung, M. M., Hui, E. S., Chan, K. C., Helpern, J. A., Qi, L. & Wu, E. X. 2009. Does diffusion kurtosis imaging lead to better neural tissue characterization? A rodent brain maturation study. *Neuroimage,* 45**,** 386-92.

Choo, I. H., Lee, D. Y., Oh, J. S., Lee, J. S., Lee, D. S., Song, I. C., Youn, J. C., Kim, S. G., Kim, K. W., Jhoo, J. H. & Woo, J. I. 2010. Posterior cingulate cortex atrophy and regional cingulum disruption in mild cognitive impairment and Alzheimer's disease. *Neurobiol. Aging,* 31**,** 772-79.

Clayden, J. D., Bastin, M. E. & Storkey, A. J. 2006. Improved segmentation reproducibility in group tractography using a quantitative tract similarity measure. *Neuroimage,* 33**,** 482-92.

Clayden, J. D., Storkey, A. J., Maniega, S. M. & Bastin, M. E. 2009. Reproducibility of tract segmentation between sessions using an unsupervised modelling-based approach. *Neuroimage,* 45**,** 377-85.

Cohen, J. 1976. *Statistical power analysis for the behavioral sciences, 2nd Edition*, Lawrence Erlbaum Associates, Publishers.

Colby, J. B., Soderberg, L., Lebel, C., Dinov, I. D., Thompson, P. M. & Sowell, E. R. 2012. Along-tract statistics allow for enhanced tractography analysis. *Neuroimage,* 59**,** 3227-42.

Corouge, I., Fletcher, P. T., Joshi, S., Gouttard, S. & Gerig, G. 2006. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Med. Image Anal.,* 10**,** 786-98.

Douaud, G., Jbabdi, S., Behrens, T. E. J., Menke, R. A., Gass, A., Monsch, A. U., Rao, A., Whitcher, B., Kindlmann, G., Matthews, P. M. & Smith, S. 2011. DTI measures in crossing-fibre areas: Increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. *Neuroimage*, 55**,** 880-90.

Falangola, M. F., Jensen, J. H., Babb, J. S., Hu, C., Castellanos, F. X., Martino, A., Ferris, S. H. & Helpern, J. A. 2008. Age-related non-Gaussian diffusion patterns in the prefrontal brain. *J. Magn. Reson. Imaging,* 28**,** 1345-50.

Ferguson, C. 2009. An effect size primer: A guide for clinicians and researchers. *Prof. Psychol.-Res. Pr.,* 40**,** 532-8.

Fieremans, E., Jensen, J. H. & Helpern, J. A. 2011. White matter characterization with diffusional kurtosis imaging. *Neuroimage,* 58**,** 177-88.

Fjell, A. M., Westlye, L. T., Greve, D. N., Fischl, B., Benner, T., Van Der Kouwe, A. J. W., Kouwe, A. J., Salat, D., Bjørnerud, A., Due-Tønnessen, P. & Walhovd, K. B. 2008. The relationship between diffusion tensor imaging and volumetry as measures of white matter properties. *Neuroimage,* 42**,** 1654-68.

Flashman, L., Andreasen, N., Flaum, M. & Swayze, V. 1997. Intelligence and regional brain volumes in normal controls. *Intelligence,* 25**,** 149-60.

Grossman, E. J., Ge, Y., Jensen, J. H., Babb, J. S., Miles, L., Reaume, J., Silver, J. M., Grossman, R. I., Inglese, M., Ge, Y., Jensen, J. H., Babb, J. S., Miles, L., Reaume, J., Silver, J. M. & Grossman, R. I. 2012. Thalamus and Cognitive Impairment in Mild Traumatic Brain Injury: A Diffusional Kurtosis Imaging Study. *J. Neurotrauma,* 29**,** 2318-27.

Harris, G. J., Jaffin, S. K., Hodge, S. M., Kennedy, D., Caviness, V. S., Marinkovic, K., Papadimitriou, G. M., Makris, N. & Oscar-Berman, M. 2008. Frontal White Matter and Cingulum Diffusion Tensor Imaging Deficits in Alcoholism. *Alcohol Clin. Exp. Res.,* 32**,** 1001-13.

Heiervang, E., Behrens, T. E., Mackay, C. E., Robson, M. D. & Johansen-Berg, H. 2006. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage,* 33**,** 867-77.

Hori, M., Fukunaga, I., Masutani, Y., Nakanishi, A., Shimoji, K., Kamagata, K., Asahi, K., Hamasaki, N., Suzuki, Y. & Aoki, S. 2012. New diffusion metrics for spondylotic myelopathy at an early clinical stage. *Eur. Radiol.,* 22**,** 1797-802.

Ito, S., Makino, T., Shirai, W. & Hattori, T. 2008. Diffusion tensor analysis of corpus callosum in progressive supranuclear palsy. *Neuroradiology,* 50**,** 981-5.

Jensen, J. H., Helpern, J. A., Ramani, A., Lu, H. & Kaczynski, K. 2005. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn. Reson. Med.,* 53**,** 1432-40.

Jensen, J. H. & Helpern, J. A. 2010. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed.,* 23**,** 698-710.

Jensen, J. H., Falangola, M. F., Hu, C., Tabesh, A., Rapalino, O., Lo, C. & Helpern, J. A. 2011. Preliminary observations of increased diffusional kurtosis in human brain following recent cerebral infarction. *NMR Biomed.,* 24**,** 452-7.

Jeurissen, B., Leemans, A., Tournier, J., Jones, D. K. & Sijbers, J. 2013. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum. Brain Mapp.***,** DOI: 10.1002/hbm.22099.

Jones, D. K. & Pierpaoli, C. 2005. Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magn. Reson. Med.,* 53**,** 1143-9.

Jones, D. K. & Cercignani, M. 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed.,* 23**,** 803-20.

Kim, S. J., Jeong, D., Sim, M. E., Bae, S. C., Chung, A., Kim, M. J., Chang, K. H., Ryu, J., Renshaw, P. F. & Lyoo, I. K. 2006. Asymmetrically Altered Integrity of Cingulum Bundle in Posttraumatic Stress Disorder. *Neuropsychobiology,* 54**,** 120-5.

Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. 2010. ElastiX: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging,* 29**,** 196-205.

Kristo, G., Leemans, A., Gelder, B., Raemaekers, M., Rutten, G. & Ramsey, N. 2012. Reliability of the corticospinal tract and arcuate fasciculus reconstructed with DTI-based tractography: implications for clinical practice. *Eur. Radiol.* DOI: 10.1007/s00330-012-2589-9.

Laird, N. M. & Ware, J. H. 1982. Random-effects models for longitudinal data. *Biometrics,* 38**,** 963-74.

Lebel, C., Walker, L., Leemans, A., Phillips, L. & Beaulieu, C. 2008. Microstructural maturation of the human brain from childhood to adulthood. *Neuroimage*. 40**,** 1044-55

Leemans, A., Jeurissen, B., Sijbers, J. & Jones, D. K. 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. *Proc. Intl. Soc. Mag. Reson. Med. 17***,** 3536.

Lenth, R. 2001. Some practical guidelines for effective sample size determination. *Am. Stat.,* 55**,** 187-93.

Lätt, J., Nilsson, M., Wirestam, R., Ståhlberg, F., Karlsson, N., Johansson, M., Sundgren, P. C. & Van Westen, D. 2012. Regional values of diffusional kurtosis estimates in the healthy brain. *J. Magn. Reson. Imaging***,** DOI: 10.1002/jmri.23857.

Löbel, U., Sedlacik, J., Güllmar, D., Kaiser, W. A., Reichenbach, J. R. & Mentzel, H-J. 2009. Diffusion tensor imaging: The normal evolution of ADC, RA, FA and eigenvalues studied in multiple anatomical regions of the brain. *Neuroradiology*. 51**,** 253-63

Maxwell, S. E., Kelley, K. & Rausch, J. R. 2008. Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation. *Annu. Rev. Psychol.,* 59**,** 537-63.

O'Goreman, R. L. & Jones, D. K. 2006. Just how much data need to be collected for reliable bootstrap DT-MRI? *Magn. Reson. Med. 56***,** 884-90

Pfefferbaum, A., Adalsteinsson, E. & Sullivan, E. V. 2003. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J. Magn. Reson. Imaging,* 18**,** 427-33.

Pitel, A., S, Chanraud, R., Sullivan, E.V., Pfefferbaum, A. & Chanraud, S. 2010. Callosal microstructural abnormalities in Alzheimer's disease and alcoholism: same phenotype, different mechanisms. *Psychiat. Res.-Neuroim.,* 184**,** 49-56.

Poot, D. H., Dekker, A. J., Achten, E., Verhoye, M. & Sijbers, J. 2010. Optimal experimental design for diffusion kurtosis imaging. *IEEE Trans. Med. Imaging,* 29**,** 819-29.

Stenset, V., Bjørnerud, A., Fjell, A. M., Walhovd, K. B., Hofoss, D., Due-Tønnessen, P., Gjerstad, L. & Fladby, T. 2011. Cingulum fiber diffusivity and CSF T-tau in patients with subjective and mild cognitive impairment. *Neurobiol. Aging.,* 32**,** 581-9.

Sullivan, E.V. & Pfefferbaum, A. Diffusion tensor imaging and aging. 2006. *Neurosci. Biobehav. R.,* 30**,** 749-61

Szczepankiewicz, F., Nilsson, M., Mårtensson, J., Westen, D., Ståhlberg, F. & Lätt, J. 2011. Automated quantification of diffusion tensor imaging (DTI) and diffusion

kurtosis imaging (DKI) parameters along the cervical spine using tractography-based voxel selection. *Proc. Eur. Soc. Mag. Reson. Med. Bio. 27***,** 262-3.

Tang, J., Liao, Y., Zhou, B., Tan, C., Liu, T., Hao, W., Hu, D. & Chen, X. 2010. Abnormal anterior cingulum integrity in first episode, early-onset schizophrenia: A diffusion tensor imaging study. *Brain Res.,* 1343**,** 199-205.

Teipel, S. J., Schapiro, M. B., Alexander, G. E., Krasuski, J. S., Horwitz, B., Hoehne, C., Möller, H., Rapoport, S. I. & Hampel, H. 2003. Relation of corpus callosum and hippocampal size to age in nondemented adults with Down's syndrome. *Am. J. Psychiatry,* 160**,** 1870-8.

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., Zijl, P. & Mori, S. 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage,* 36**,** 630-44.

Van Hecke, W., Leemans, A., De Backer, S., Jeurissen, Parizel, P.M. & Sijbers, J. 2009. Comparing isotropic and anisotropic smoothing for voxel-based DTI analyses: A simulation study. *Hum. Brain Mapp.,* 31**,** 98-114.

Wang, J., Lin, W., Lu, C., Weng, Y., Ng, S., Wang, C., Liu, H., Hsieh, R., Wan, Y. & Wai, Y. 2011. Parkinson disease: diagnostic utility of diffusion kurtosis imaging. *Radiology,* 261**,** 210-7.

Wang, R., Benner, T. & Sorensen, A. 2007. Diffusion toolkit: a software package for diffusion imaging data processing and tractography. *Proc. Intl. Soc. Mag. Reson. Med. 15***,** 3720.

Veraart, J., Rajan, J., Peeters, R. R., Leemans, A., Sunaert, S. & Sijbers, J. 2012. Comprehensive framework for accurate diffusion MRI parameter estimation. *Magn. Reson. Med.***,** Nov 6. DOI: 10.1002/mrm.24529. [Epub ahead of print] PubMed PMID: 23132517

Vittinghoff, E., Glidden, D. V., Shiboski, S. C. & Mcculloch, C. E. 2005. *Regression Methods in Biostatistics*, Springer New York.

Vos, S. B., Jones, D. K., Viergever, M. A. & Leemans, A. 2011. Partial volume effect as a hidden covariate in DTI analyses. *Neuroimage,* 55**,** 1566-76.

Vos, S. B., Jones, D. K., Jeurissen, B., Viergever, M. A. & Leemans, A. 2012. The influence of complex white matter architecture on the mean diffusivity in diffusion tensor MRI of the human brain. *Neuroimage,* 59**,** 2208-16.

Wu, E. X. & Cheung, M. M. 2010. MR diffusion kurtosis imaging for neural tissue characterization. *NMR Biomed.,* 23**,** 836-48.

Zhang, A., Leow, A., Ajilore, O., Lamar, M., Yang, S., Joseph, J., Medina, J., Zhan, L., An, Kumar & Kumar, A. 2011. Quantitative Tract-Specific Measures of Uncinate and Cingulum in Major Depression Using Diffusion Tensor Imaging. *Neuropsychopharmacology,* 37**,** 959-67.

Zhuo, J., Xu, S., Proctor, J. L., Mullins, R. J., Simon, J. Z., Fiskum, G. & Gullapalli, R. P. 2012. Diffusion kurtosis as an in vivo imaging marker for reactive astrogliosis in traumatic brain injury. *Neuroimage,* 59**,** 467-77.

Figures

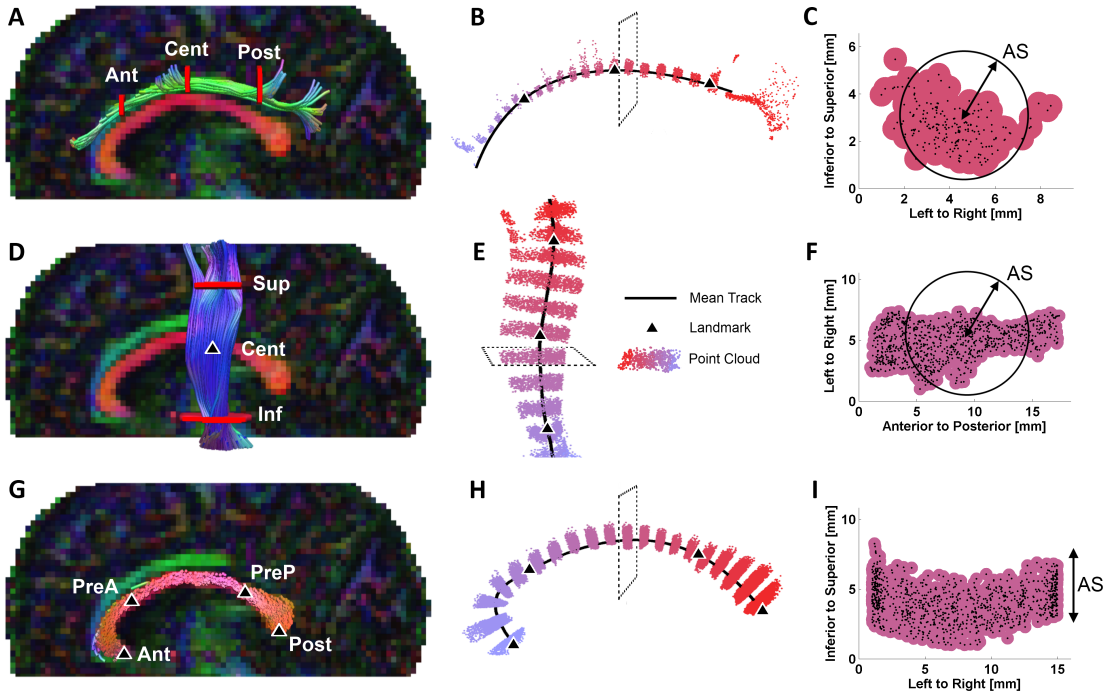

Figure 1 – The left column shows tractographies of the left hand side CG (A) and CST (D), as well as the mid-sagittal truncation of the CC (G), superimposed on a color FA-map. The AND-gates, used for structure delineation, are shown as red lines, and the anatomical landmarks are shown as black triangles (note that landmarks that coincide with AND-gates are not shown, and that the gates defining the CC are not displayed). The middle column (B, E and H) shows the mean track (black line), the point cloud that defines the tracts in 3D-space (red to blue dots), the landmarks (black triangles), and the selected cross section (dashed line) for display in the right column. Every other interval of the point cloud is omitted in order to visualize the path of the mean track (note that only points between the outermost landmarks were used in the evaluation and that the figures are not to scale). The parametric information contained within each sub-interval of the point cloud is projected onto the mean track, thus creating parameter vectors of MD, FA, MK and RK, that can be normalized across subjects with respect to the anatomical landmarks. The right column (C, F and I) shows cross sections of the point cloud, in a plane that is perpendicular to the mean track. Each point is the center of a circle with a radius of 0.5 mm. The area of the cross section, created in each interval, was used to quantify the apparent size (AS) of the structures. In the CG (C) and CST (F) the AS was defined as the radius of a circle with the same area as the structure. In the CC (I) the AS was defined as the thickness of the point cloud.

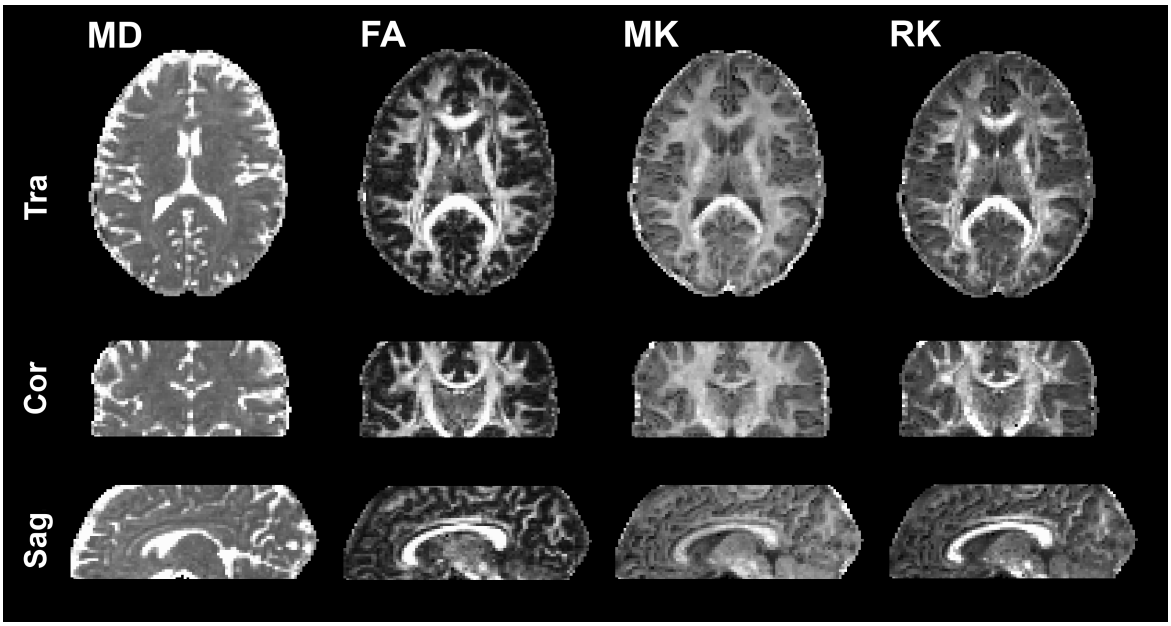

Figure 2 – The image depicts transversal (Tra), coronal (Cor), and sagittal (Sag) projections of the DKI parameter maps (MD, FA, MK and RK, respectively). The FA map displays the highest contrast between WM and GM, followed by RK and MK, in descending order. MD displays a high contrast when comparing CSF to WM and GM, but is low when comparing WM to GM.

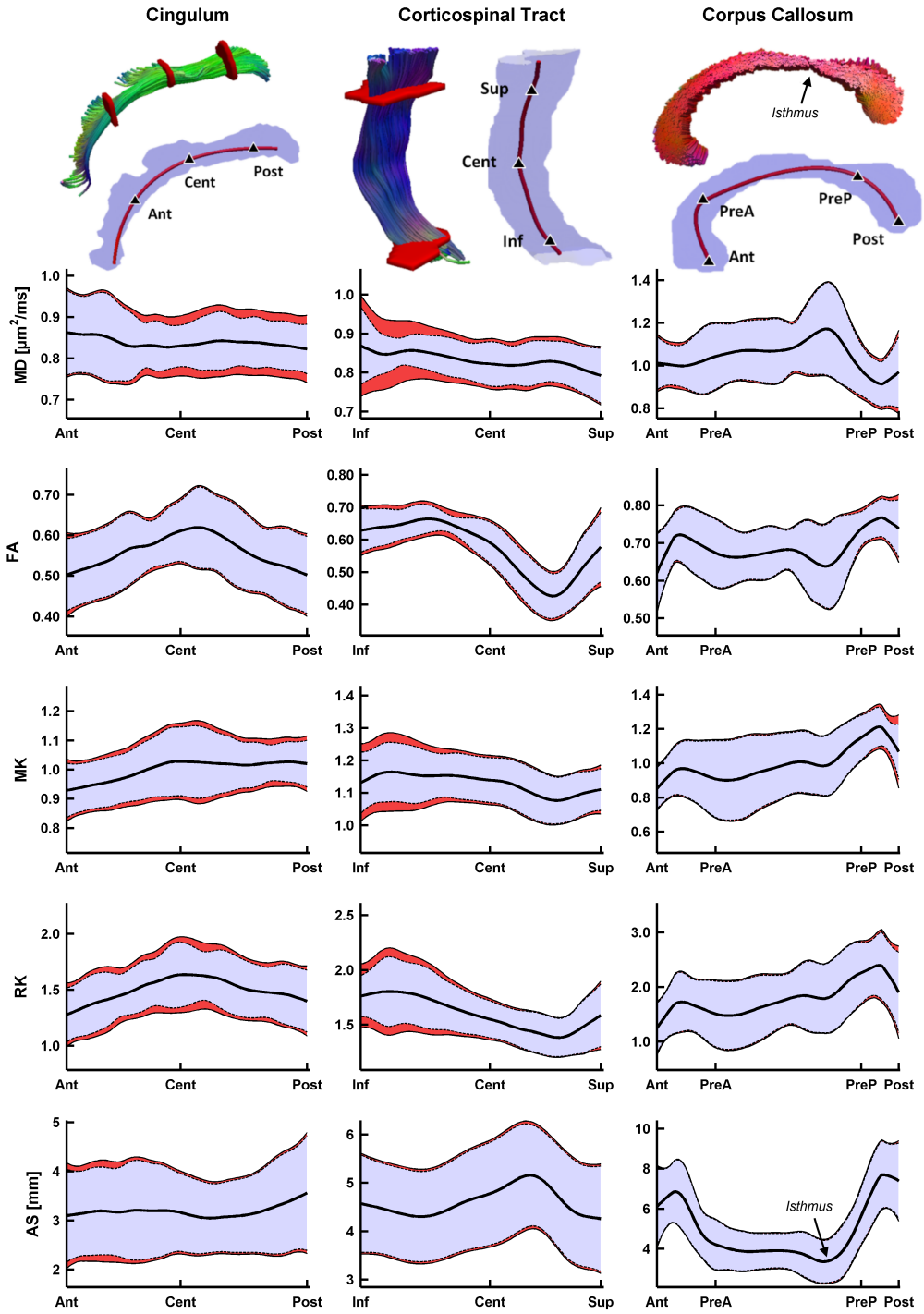

Figure 3 – The tractographies (top row) show a representative right-hand side CG (green tracts), CST (blue tracts), and a mid-sagittal truncation of the CC (red tracts) together with the AND-gates (red) used to segment the structures from the whole-brain tractography (not shown for the CC). The figure also shows a transparent representation of the same structures (blue) containing the mean track (red tract), and the landmarks (black triangles) used to normalize data. The plots show the group mean values (bold black line) of the apparent size (AS, bottom row) and the DKI parameters (MD, FA, MK and RK) as a function of anatomical position along the structures. The

parameter variability is visualized by thin black lines, where the solid lines show two standard deviations from the mean ($2V_{tot}^{1/2}$), and the dashed lines show two standard deviations from the mean after the contribution from noise has been removed ($2V_{inter}^{1/2}$, Eq. 3). The red field visualizes the variability contributed by noise. In the CG, MD displays a high inter-subject variability in the anterior regions, whereas MK has its highest variability in the central region. Both FA and RK peak at the center, tapering off towards the anterior and posterior endpoints. Parameter variations along the CST are most prominent for the FA, probably due to the crossing-fiber region in the superior segment. The variability of all parameters, except the FA and AS, is elevated in the inferior parts of the structure. Similarly to the CST, the CC displays significant parameter variation along the structure. In the thinnest region, the *isthmus* (black arrows), MD and FA are strongly elevated and reduced, respectively, probably due to the PVE at the WM/CSF interface. The CC also displays a much smaller relative dependence on noise (red area) compared with the CG and CST. It is also notable how the AS and the FA both follow the same trend, which showcases the modulating effect of PVE on diffusion parameters due to tract morphology.

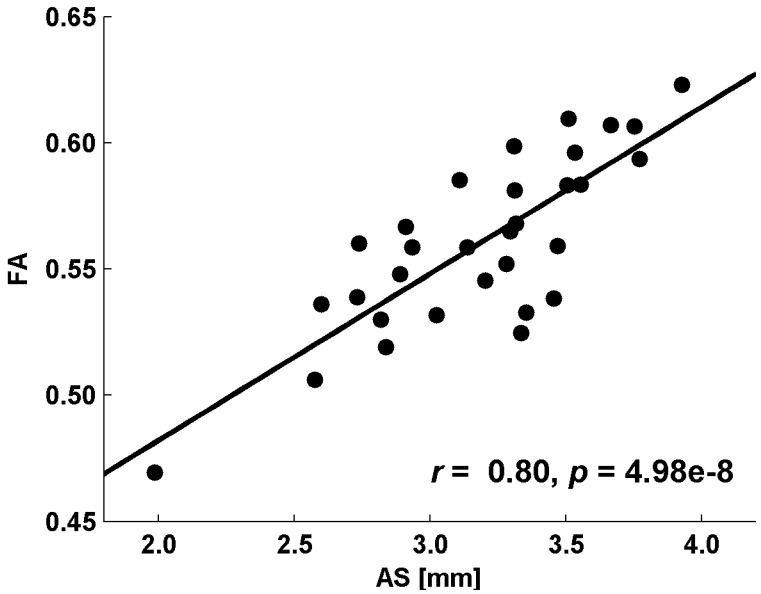

Figure 4 – Correlation between the mean FA and mean AS in the CG, for the 31 healthy subjects. The regression line (black line) shows that a CG bundle with a high AS is likely to exhibit a high FA. Note that the correlation coefficient value of $r = 0.8$ indicates that 64% of the variance in FA can be explained by its association to AS. If AS is known, this variance contribution can be removed (Eq. 10).

Tables

Table 1 – Relative effect sizes ($\Delta\mu/\mu$) of various conditions as observed in DTI and DKI parameters, and group sizes investigated (*n*, reported as size of control group + patient group). The values of $\Delta\mu/\mu$ are reported in regions where significant differences in group means were found. The coefficient of variation (CV) is the value reported for the control group specified for each parameter separately. In cases where the variability was not reported it is marked with a dash (-).

Source	Condition	Region	Parameter	CV [%]	$\Delta\mu/\mu$ [%]	*n*
Wang et al. (2011)	PD	Caudate, putamen, globus palidus, substantia nigra	MK	13	15 – 30	30 + 30
Grossman et al. (2012)	mTBI	Thalamus, internal capsule, splenium of the CC, centum semiovale	MK FA MD	1 – 2 2 1 – 4	2 – 3 3 1 – 2	14 + 22
Kim et al. (2006)	PTSD	CG bundle	FA	11 – 18	12 – 26	21 + 21
Zhang et al. (2011)	MDD	Right uncinate	FA RD	7 7	7 5	21 + 21
Ito et al. (2008)	PSP	Anterior CC	MD FA	17 8	15 – 34 12 – 17	19 + 7
Bozzali et al. (2012)	AD	Cingulum	MD FA	6 8	17 12	14 + 31
Stenset et al. (2011)	MCI	Cingulum, genu CC	FA RD	10 – 15 17 – 29	7 – 13 11 – 22	26 + 12
Tang et al. (2010)	EOS	Right anterior cingulum	FA	-	14	38 + 38

AD Alzheimer's disease, EOS early-onset schizophrenia, MDD major depressive disorder, mTBI mild traumatic brain injury, NAWM normal appearing white matter, PD Parkinson's disease, PSP progressive supranuclear palsy, PTSD posttraumatic stress disorder, RD radial diffusivity.

Table 2 – DKI parameter values in the group of healthy volunteers (n = 31), calculated in the cingulum (CG), corticospinal tract (CST) and corpus callosum (CC). The mean value (μ) is presented along with the coefficient of variation (CV in%) and the relative noise contribution to variance (RV_{noise} in%). Average whole-structure CVs were 4.2, 4.7, 4.9 and 8.8% for MD, FA, MK and RK, respectively. The most prominent contributor to variance was generally the inter-subject variability (reflected by a low RV_{noise}).

		MD [$\mu m^2/ms$]			FA			MK			RK		
		μ	CV	RV_{noise}	μ	CV	RV_{noise}	μ	CV	RV_{noise}	μ	CV	RV_{noise}
CG	Ant	0.84	4.6	20	0.56	6.6	8	0.98	4.7	20	1.47	8.4	25
	Post	0.84	4.2	47	0.57	7.6	5	1.02	4.2	28	1.53	8.2	25
	Whole	0.84	3.7	42	0.56	6.1	6	1.00	4.1	22	1.50	7.2	28
CST	Inf	0.85	4.2	54	0.64	4.1	35	1.15	3.8	36	1.72	7.9	35
	Sup	0.82	3.6	24	0.50	6.4	16	1.10	3.1	9	1.46	5.7	8
	Whole	0.83	3.6	42	0.57	4.1	33	1.13	3.2	25	1.60	6.4	25
CC	Ant	1.01	5.9	12	0.69	5.5	6	0.94	8.6	5	1.59	15.2	6
	Cent	1.09	6.3	5	0.67	4.2	5	0.98	8.8	5	1.75	14.5	4
	Post	0.93	6.5	17	0.76	3.6	19	1.17	4.8	35	2.27	13.0	14
	Whole	1.04	5.3	8	0.69	3.6	8	1.00	7.4	5	1.80	12.6	4

Table 3 – Calculated group sizes (n) for DKI parameters (MD, FA, MK and RK) and apparent structure size (AS), required in order to generate a statistical power of $\pi = 0.9$ at an effect size of 5% and a significance level of $\alpha = 0.05$. The group sizes shown the number of subjects needed in each group and were estimated for whole structures as well as sub-structures. The values of n mainly reflects the total parameter variability, meaning that a low V_{total} makes it easier to detect the proposed 5% change, making the required group size comparatively small.

		MD	FA	MK	RK	AS
CG	Ant	21	39	21	63	183
	Post	18	51	18	59	147
	Whole	14	34	17	47	148
CST	Inf	17	17	15	55	109
	Sup	13	38	11	30	108
	Whole	13	17	11	37	106
CC	Ant	32	28	65	199	100
	Cent	36	18	68	181	122
	Post	38	13	22	146	101
	Whole	26	14	48	137	85

Table 4 – Pearson's correlation coefficient (r) describing the association of DKI parameters (MD, FA, MK and RK) with the apparent structure size (AS). As expected, AS correlated with DKI parameters in the CG and CC which means that structure size may account for some of the measured variability. No significant correlation was found in the CST, as was expected due to the high AS dependence on AND-gate definition. No correction for multiple comparisons was done; however, no more than 5 significant correlations are expected on the 5% level for 40 independent comparisons.

		MD	FA	MK	RK
CG	Ant	–0.23	0.57‡	0.33	0.17
	Post	–0.31	0.69‡	0.37†	0.48‡
	Whole	–0.32	0.80‡	0.40†	0.45†
CST	Inf	–0.12	0.06	–0.10	0.01
	Sup	–0.11	0.17	0.02	0.29
	Whole	–0.12	0.17	–0.07	0.13
CC	Ant	–0.48‡	0.26	0.11	–0.14
	Cent	–0.44†	0.42†	0.18	0.15
	Post	–0.58‡	–0.12	–0.24	–0.44†
	Whole	–0.53‡	0.32	0.05	–0.09

(†) $p < 0.05$, and (‡) $p < 0.01$