NeuroImage 76 (2013) 145–154

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

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

Filip Szczepankiewicz [a,*], Jimmy Lätt [b], Ronnie Wirestam [a], Alexander Leemans [c], Pia Sundgren [b,d], Danielle van Westen [d], Freddy Ståhlberg [a,d,e], Markus Nilsson [e]

[a] Department of Medical Radiation Physics, Lund University, Lund, Sweden
[b] MR Department, Centre for Medical Imaging and Physiology, Lund University Hospital, Lund, Sweden
[c] Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands
[d] Department of Diagnostic Radiology, Lund University, Lund, Sweden
[e] Lund University Bioimaging Center, Lund University, Lund, Sweden

ARTICLE INFO

Article history:
Accepted 25 February 2013
Available online 16 March 2013

Keywords:
Diffusion kurtosis imaging
Statistical power
Study design
Group size
Tractography

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.

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 and Helpern, 2010; Jensen et al., 2005), 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

* Corresponding author at: Lund University, Department of Medical Radiation Physics, Barngatan 2, 22185 Lund, Sweden.
E-mail address: filip.szczepankiewicz@med.lu.se (F. Szczepankiewicz).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.02.078

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 the lowest for MD and MK, with values of 5% and 8%, respectively, and the 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.

Theory

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

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

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.

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

$$Y = \mu + \Delta\mu \cdot G + E_{\mathrm{total}}, \tag{2}$$

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

$$V_{\mathrm{total}} = V_{\mathrm{inter}} + V_{\mathrm{noise}}. \tag{3}$$

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

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

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: $g \propto (T'/T)^{1/2} \cdot (\mathrm{SNR}'/\mathrm{SNR})$, 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

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

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

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

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.

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