Statistical Group Comparison of Diffusion Tensors via Multivariate Hypothesis Testing

Brandon Whitcher,[1]* Jonathan J. Wisco,[2] Nouchine Hadjikhani,[2] and David S. Tuch[2]†

Diffusion tensor imaging (DTI) provides a powerful tool for identifying white matter (WM) alterations in clinical populations. The prevalent method for group-level analysis of DTI is statistical comparison of the diffusion tensor fractional anisotropy (FA) metric. The FA metric, however, does not capture the full orientational information contained in the diffusion tensor. For example, the FA test is incapable of detecting group-level differences in diffusion orientation when the level of anisotropy is unaffected. Here, we apply multivariate hypothesis testing procedures to the elements of the diffusion tensor as an alternative to univariate testing using FA. Both parametric and nonparametric tests are proposed with each choice carrying specific assumptions about the diffusion tensor model. Of particular interest is the Cramér test, which works on Euclidean interpoint distances and can be readily adapted to a specific non-Euclidean framework by applying matrix logarithms to the diffusion tensors. Using Monte Carlo simulations, we show that multivariate tests can detect diffusion tensor principal eigenvector differences of 15 degrees with up to 80–90% power under typical design conditions. We also show that some multivariate tests are more sensitive to FA differences, when compared to a univariate test on FA, even if there is no principal eigenvector difference. The Cramér test, using the Euclidean interpoint distances, performed best under both simulation scenarios. When applying the Cramér test of the diffusion tensor in a clinical population with a history of migraine, a 169% increase was observed in the volume of a significant cluster compared to the univariate FA test. Magn Reson Med 57:1065–1074, 2007. © 2007 Wiley-Liss, Inc.

Key words: Cramér test; migraine; multivariate; permutation

INTRODUCTION

Diffusion tensor imaging (DTI) has identified white matter alterations in a wide number of neurological and psychiatric conditions including Alzheimer's disease, Parkinson's disease, schizophrenia, neurological complications of HIV infection, autism, multiple sclerosis, and numerous other conditions (1,2). DTI studies in clinical populations have thus far focused predominately on diffusion anisotropy; yet DTI provides information on both the degree and direction of the diffusion anisotropy in white matter (WM). Statistical hypothesis testing on the diffusion tensor, instead of scalar summaries derived from it, could provide greater statistical power, since there are more measurements available in the full tensor. Further, hypothesis testing on the tensor could detect pathologies that alter diffusion orientation but leave diffusion anisotropy preserved. This capability may be particularly effective in neurodevelopmental disorders characterized by aberrant WM patterning. In areas of fiber crossing, FA may be reduced if one of the tracts is undergoing demyelination. However, a decreased FA could also be observed if one of the tracts is more myelinated in the studied population than in controls. For example, intensive training of a motor skill such as piano playing has been reported to decrease FA values in the motor tracts (3).

The standard statistical tests for tensor-valued measurements impose distributional assumptions on the tensor, such as the Wishart distribution. Schwartzman et al. (4) described a parametric test based on the bipolar Watson distribution (5), yet this test is parametric and only operates on the mean diffusion direction and not the complete tensor. Multivariate hypothesis testing has been applied to a combination of scalar summaries derived from the diffusion tensor (6), where both the apparent diffusion coefficient (ADC) and fractional anisotropy (FA) were tested as a bivariate vector using the Hotelling T^2 test.

This article describes several tests, both parametric and nonparametric, for statistical group comparison of the estimated diffusion tensors. The selection of specific hypothesis testing procedures is not exhaustive and represents a selection of tests that are both statistically appropriate and relatively straightforward to implement. If one considers only the unique elements of the diffusion tensor as a six-dimensional vector, then the Hotelling T^2 test is available for making group comparisons. The Hotelling T^2 test assumes multivariate normality of the six-dimensional vector and identical group covariance matrices. This is not unreasonable given previous results based on Monte Carlo simulations (7) and statistical theory (8). One nonparametric alternative to the Hotelling T^2 test is the multivariate permutation test (9), where the only assumption is exchangeability between the multivariate vectors. A second nonparametric alternative is the Cramér test based on Euclidean interpoint distances between the vectors (10). No distributional assumptions are required for the Cramér test and it has been shown to perform well against both changes in scale and location between groups for a wide variety of probability distribution functions.

Applying statistical methodology to the multivariate vector derived from the diffusion tensor ignores key attributes of the tensor, such as positive definiteness. Recently, non-Euclidean metrics have been suggested and applied in

[1]Clinical Imaging Centre, GlaxoSmithKline, Hammersmith Hospital, London W12 0NN, United Kingdom

[2]A. A. Martinos Center for Biomedical Imaging, MGH-NMR Center, 149 13th Street, Room 2301, Charlestown, Massachusetts

†Present address: Novartis Pharma AG, Basel, Switzerland

Grant sponsor: National Institutes of Neurological Disorders and Stroke; Grant number: NS035611; Grant sponsor: National Center for Research Resources; Grant number: RR14075; Grant sponsor: National Alliance for Medical Image Computing, National Institute for Biomedical Imaging and Bioengineering (funded through National Institutes of Health Roadmap for Medical Research); Grant number: EB005149; Grant sponsors: GlaxoSmith-Kline, Swiss Heart Foundation, Athinoula A. Martinos Foundation, Mental Illness and Neuroscience Discovery Institute.

*Correspondence to: Brandon Whitcher, Clinical Imaging Centre, GlaxoSmith-Kline, Hammersmith Hospital, Du Cane Road, London W12 0NN, United Kingdom. E-mail: brandon.j.whitcher@gsk.com

Received 25 April 2006; revised 15 January 2007; accepted 16 February 2007.

DOI 10.1002/mrm.21229

Published online in Wiley InterScience (www.interscience.wiley.com).

registration and filtering of DTI data (11,12). In our case of multivariate hypothesis testing, the Cramér test works on Euclidean interpoint distances between the multivariate vectors and therefore it is well suited to work with the so-called Log-Euclidean metrics (13). Log-Euclidean metrics allow standard Euclidean computations to take place in the domain of matrix logarithms while retaining key properties of affine-invariant Riemannian metrics.

Monte Carlo simulations show that multivariate testing procedures can detect (at roughly 80% power) principal eigenvector differences of 15 degrees and FA differences of 0.10–0.15 under experimentally relevant SNR conditions. In DTI data from a migraine study, the Cramér test using a Euclidean metric provided substantially greater statistical power than the univariate FA test. Additionally, the Cramér test detected WM alterations that were not detected by the univariate test on FA. Statistical group comparison of tensors holds substantial promise to improve the statistical and anatomical sensitivity of DTI studies in clinical populations.

THEORY

DT-MRI data is inherently multivariate; i.e., there are several measurements at each physical location (voxel) in the scan. Depending on the scanning protocol, at least six and up to several hundred gradient directions may be acquired in a single session. Physical models have been used to reduce the complexity of the observed data. For example, the diffusion tensor is one method of mapping $d \geq 6$ gradient directions to a second-order tensor with six unique elements related to the Gaussian model of diffusion in three dimensions (14). Even after this initial mapping, several rotationally invariant parameters have been derived from the diffusion tensor D (15,16). It has been argued that these parameters do not depend on the applied magnetic field gradient coordinate system and can be reproduced in successive examinations (given the same acquisition parameters). Fractional anisotropy or FA (15) is the most popular scalar invariant measure of the diffusion tensor. We introduce several hypothesis testing techniques that may be applied to a scalar summary of the diffusion tensor or to the diffusion tensor in its entirety. Both parametric and nonparametric testing schemes are provided, and are appropriate depending on what is assumed about the statistic of interest.

Univariate Hypothesis Testing

Group comparisons using FA may be performed using well-known univariate two-sample hypothesis tests. Although widely used and familiar to most researchers, we will introduce this parametric test to establish necessary notation that will be recycled in subsequent sections. Let $\mathbf{D}_1 = D_{1,1}, \ldots, D_{1,n_1}$ and $\mathbf{D}_2 = D_{2,1}, \ldots, D_{2,n_2}$ be the estimated diffusion tensors from the two groups of subjects. Then $\mathbf{FA}_1 = \mathrm{FA}_{1,1}, \ldots, \mathrm{FA}_{1,n_1}$ and $\mathbf{FA}_2 = \mathrm{FA}_{2,1}, \ldots, \mathrm{FA}_{2,n_2}$ denote the estimated values of fractional anisotropy from the corresponding diffusion tensors. If one assumes that FA from both groups of subjects follows a Normal probability density function (PDF), then the two-sample t-test may be used to detect differences in FA between the two groups at a voxel-by-voxel level. The formal statement of the hypothesis test is $H_0 : \mathrm{FA}_1 - \mathrm{FA}_2 = 0$ versus $H_1 : \mathrm{FA}_1 - \mathrm{FA}_2 \neq 0$. The test statistic depends on the average difference in FA between the two groups normalized by the pooled standard deviation of the Normal PDF; i.e.,

$$U_{n_1,n_2} = \frac{(n_1 + n_2 - 2)^{1/2}(\overline{\mathrm{FA}}_1 - \overline{\mathrm{FA}}_2)}{\left(n_1^{-1} + n_2^{-1}\right)^{1/2}\left(S_1^2 + S_2^2\right)^{1/2}}, \qquad [1]$$

where $\overline{\mathrm{FA}}_1, \overline{\mathrm{FA}}_2$ are the average within-group FA at a particular voxel and S_1^2, S_2^2 are the sums of squares of FA. The test statistic U_{n_1,n_2} will have a t distribution with $n_1 + n_2 - 2$ degrees of freedom. Two remarks about this testing scheme come to mind. First, the assumption of Gaussianity may seem odd given that there is Rician noise on the diffusion-weighted images, the log transform is used in the model fitting procedure and FA is a nonlinear combination of the diffusion tensor elements. By assuming the signal intensities from the data acquisition are multivariate Normal, it has been shown that FA is asymptotically Normal as the number of diffusion gradients goes to infinity (17). If one does not want to make such an assumption, it may be beneficial to investigate alternative hypothesis tests with weaker assumptions on the diffusion-weighted images; e.g., permutation, bootstrap or rank-based nonparametric tests. Second, the two-sample t-test will only detect group differences that manifest themselves in the scalar quantity FA and ignores potential information in the diffusion tensor.

If we make the assumption that the underlying distribution of errors about FA is symmetric, then a semiparametric alternative to the t-test is available by rearranging the labels of the observed FA values in a permutation framework (9,18). The null and alternative hypotheses may now be expressed via $H_0 : \mathrm{FA}_1 \stackrel{d}{=} \mathrm{FA}_2$ and $H_1 : \mathrm{FA}_1 \stackrel{d}{\neq} \mathrm{FA}_2$, respectively, where $\stackrel{d}{=}$ denotes "equal in distribution". Let $\pi(\mathrm{FA})$ be a permutation of the concatenated vector $\mathbf{FA} = [\mathrm{FA}_1, \mathrm{FA}_2]$, where $\pi_1(\mathrm{FA})$ is the vector of FA values that are associated with the first n_1 labels (group one) and $\pi_2(\mathrm{FA})$ is associated with the subsequent n_2 labels (group two). A suitable test statistic is given by the difference between within-permutation-group averages $U^*_{1,2} = \overline{\pi_1(\mathrm{FA})} - \overline{\pi_2(\mathrm{FA})}$. Thus, the observed value is $U_{1,2} = \overline{\mathrm{FA}}_1 - \overline{\mathrm{FA}}_2$. Standardization is not required because the test statistic $U^*_{1,2}$ is permutationally equivalent to its standardized form. The permutation sample space is the set of all permutations of the observed FA, with a cardinality of $n! = (n_1 + n_2)!$. For reasonably large n, the permutation sample space cannot be examined at all possible points and a random sample from the permutation space must be obtained. This is achieved by selecting B permutation sets $\pi(\mathrm{FA})$ and calculating the estimated p-value via $\hat{\xi} = (1 + \#\{|U^*_{1,2}| \geq |U_{1,2}|\})/(B+1)$ for the two-sided alternative hypothesis, where $\#\{A\}$ denotes the cardinality of the set A.

Multivariate Hypothesis Testing

One may argue that FA is not the most effective scalar summary of diffusion anisotropy in the human brain. In fact, no univariate measure may provide the best level of sensitivity to detecting subtle differences in WM microstructure.

Hence, we have decided to test all the unique elements that make up the diffusion tensor as a multivariate vector. Let

$$\mathbf{d} = [D_{x,x} D_{y,y} D_{z,z} D_{x,y} D_{x,z} D_{y,z}]^{\mathrm{T}} \quad [2]$$

denote the six-dimensional vector of elements from the diffusion tensor D, and let $\mathbf{d}_1 = \mathbf{d}_{1,1}, \ldots, \mathbf{d}_{1,n_1}$ and $\mathbf{d}_2 = \mathbf{d}_{2,1}, \ldots, \mathbf{d}_{2,n_2}$ be the random six-dimensional vectors from the two groups, respectively. The hypothesis test may now be stated as $H_0 : \mathbf{d}_1 - \mathbf{d}_2 = \mathbf{0}$ versus $H_1 : \mathbf{d}_1 - \mathbf{d}_2 \neq \mathbf{0}$. In addition, if $\mathbf{d}_1 - \mathbf{d}_2 \neq \mathbf{0}$, then which component means are different? Assuming the observations come from two multivariate Normal distributions with a difference in means of $\mathbf{0}$ and the same covariance matrix, the two-sample T^2-statistic is given by

$$T^2 = [\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2]^{\mathrm{T}} \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{\mathrm{pooled}} \right]^{-1} [\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2], \quad [3]$$

where $\bar{\mathbf{d}}_1$ and $\bar{\mathbf{d}}_2$ are the average within-group diffusion tensor elements at a particular voxel and $\mathbf{S}_{\mathrm{pooled}}$ is the pooled variance. The T^2-statistic will have a normalized F distribution given by

$$\frac{6(n_1 + n_2 - 2)}{n_1 + n_2 - 7} F_{6, n_1 + n_2 - 7}. \quad [4]$$

When the assumption of multivariate normality is not reasonable, hypothesis testing based on permutations offers one possible solution. Building on the notation already introduced for testing via rearrangements in the previous section, the null hypothesis of equality in multivariate distributions of the vector representation of the diffusion tensor is given by

$$H_0 : \{\mathbf{d}_1 \overset{d}{=} \mathbf{d}_2\} = \left\{ \bigcap_{i=1}^{6} H_{0,i} \right\}, \quad [5]$$

where each univariate null hypothesis $H_{0,i}$ tests of equality between the univariate distributions for the ith diffusion tensor element. Thus, H_0 is the global null hypothesis test and the global alternative may be represented by the union across all sub-alternatives

$$H_1 : \left\{ \bigcup_{i=1}^{6} H_{1,i} \right\}. \quad [6]$$

Basically, multivariate permutation testing performs all partial tests (i.e., tests on each element of the observed vectors) and combines P-values from the partial tests using a combining function. Several functions are available from the literature (9), and we use the Fisher omnibus combining function

$$\mathcal{C}(\xi_1, \ldots, \xi_6) = -2 \sum_{i=1}^{6} \log(\xi_i), \quad [7]$$

where ξ_i are the marginal P-values. Implementation of the multivariate permutation test involves two stages. The first stage involves applying B rearrangements to the multivariate observations and computing the marginal P-values $\hat{\xi}_i = (1 + \#\{|U_i^*| \geq |U_i|\})/(B+1)$, for $i = 1, \ldots, 6$. The $\hat{\xi}_i$ may be calculated using the procedure outlined for the univariate permutation test by replacing FA with the diffusion tensor elements $D_1, \ldots, D_6$ and utilizing the same permutation of labels across the elements in the vector; i.e., $\pi(D_1) = \cdots = \pi(D_6)$ for every rearrangement r. The combined observed value of the second-order test $\widehat{U} = \mathcal{C}(\hat{\xi}_1, \ldots, \hat{\xi}_6)$ is a function of the marginal P-values via the combining function. Every test statistic derived from the rth rearrangement of the labels ($r = 1, \ldots, B$) in the first stage is then compared with the permutation sample to provide the combined value of vector statistics given by $\widehat{U}_r^* = \mathcal{C}(\hat{\xi}_{1,r}^*, \ldots, \hat{\xi}_{6,r}^*)$, where $\hat{\xi}_{i,r}^* = (1 + \#\{|U_{i,r}^*| \geq |U_i|\})/(B+1)$, $i = 1, \ldots, 6$, $r = 1, \ldots, B$. The combined P-value for the test $\widehat{U}$ is then estimated via $\bar{\xi} = (1 + \#\{|\widehat{U}_r^*| \geq |\widehat{U}|\})/(B+1)$.

Testing Interpoint Distances

Recently, a new multivariate two-sample test was proposed (10). The test statistic is the difference of the sum of all Euclidean interpoint distances between the random variables from the two different samples and one-half of the two corresponding sums of distances of the variables within the same sample. The random vectors $\mathbf{d}_1$ are assumed to be identically distributed with distribution function G_1 and the random vectors $\mathbf{d}_2$ are assumed to be identically distributed with distribution function G_2 (univariate testing is also valid). The hypothesis to be tested is now $H_0 : \mathbf{d}_1 \overset{d}{=} \mathbf{d}_{2_d} \Leftrightarrow H_0 : G_1 = G_2$ versus the general alternative $H_1 : \mathbf{d}_1 \neq \mathbf{d}_2 \Leftrightarrow H_1 : G_1 \neq G_2$ by using the test statistic

$$T_{n_1,n_2} = \frac{n_1 n_2}{n_1 + n_2} \left[\frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|\mathbf{d}_{1,i} - \mathbf{d}_{2,j}\| - \frac{1}{2n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \|\mathbf{d}_{1,i} - \mathbf{d}_{1,j}\| - \frac{1}{2n_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} \|\mathbf{d}_{2,i} - \mathbf{d}_{2,j}\| \right], \quad [8]$$

where $\|\mathbf{x}\|$ denotes the Euclidean norm of a p-dimensional vector $\mathbf{x}$. The null hypothesis is rejected for large values of T_{n_1,n_2}. To obtain critical values for T_{n_1,n_2}, Baringhaus and Franz provide a nonparametric bootstrap procedure (18) or, alternatively, the critical values may be computed by approximating the asymptotic distribution of T_{n_1,n_2} based on the quadratic form of Normal random variables (19,20). For the latter case, the problem of obtaining a critical value for T_{n_1,n_2} is approximated via

$$\Pr\{T_{n_1,n_2} < x\} \approx \Pr\{\nu^{\mathrm{T}} Q \nu < x\}, \quad [9]$$

where Q is a symmetric matrix and ν is a length $N = n_1 + n_2$ vector of Normal random variables. In the case of Eq. [8], the matrix Q may be interpreted as the specific contribution of the average distance, between a single multivariate observation and all others, to the test statistic. Let H be an orthonormal matrix that converts Q to diagonal form $\Lambda = H^T Q H$, then Eq. [9] may be expressed as

$$\Pr\{Z^{\mathrm{T}} \Lambda Z < x\} = \Pr\left\{ \sum_{j=1}^{N} \lambda_j \xi_j^2 < x \right\}, \quad [10]$$

where $\lambda_j = \Lambda_{jj}$ and ξ is a Normal random variable. Applying the inverse Fourier transform to the characteristic function for the linear combination of χ^2 variables in Eq. [10] produces the critical value (20). We prefer the asymptotic approximation because it requires less computational effort and is reproducible, whereas the bootstrap procedure is iterative and induces an additional layer of stochastic variability in the resulting P-values. The algorithms of Baringhaus and Franz are available for download (21) and Fortran code to implement the Imhof procedure may be found in (20).

When pointwise distances between the two groups is large the first term in Eq. [8] will dominate the subsequent terms and produce a large test statistic. When the pointwise distances are similar, the second and third terms will suppress the first to produce a small test statistic. This is an example of a location alternative, where the difference between the two distributions is restricted to the location parameter. It has been shown that the Cramér test has power similar to that of the parametric t-test and Hotelling's T^2-test for location alternatives (10). The Cramér statistic is also sensitive to so-called dispersion alternatives where the variance-covariance matrix differs between groups and the location parameters are identical, but does not perform as well as parametric tests. An example of a difference in dispersions would be where the spread between points in the two groups, in six-dimensional space, differs even though a difference between the centers of the two groups is not apparent.

Non-Euclidean Metrics

As mentioned in the Introduction section, the use of non-Euclidean distance metrics have been developed to manipulate diffusion tensors. A particularly convenient Riemannian metric is the log-Euclidean metric which applies Euclidean calculations to the transformed diffusion tensors via matrix logarithms (22). Thus, the Cramér test may be applied to these transformed diffusion tensors. When considering the multivariate vector form of the diffusion tensor, we apply weights to the off-diagonal terms to produce a rotationally-invariant Euclidean metric using

$$\mathbf{d} = [D_{x,x} D_{y,y} D_{z,z} \sqrt{2} D_{x,y} \sqrt{2} D_{x,z} \sqrt{2} D_{y,z}]^{\mathrm{T}}, \qquad [11]$$

and a similarity-invariant log-Euclidean metric using

$$\log(\mathbf{d}) = [\log(D)_{x,x} \log(D)_{y,y} \log(D)_{z,z} \sqrt{2} \log(D)_{x,y} \times \sqrt{2} \log(D)_{x,z} \sqrt{2} \log(D)_{y,z}]^{\mathrm{T}}, \qquad [12]$$

as recommended in (22). Note, the term $\log(D)_{i,j}$ is the (i,j)th coefficient of the matrix logarithm of the diffusion tensor denoted by $\log(\mathbf{D})$.

In practice the Cramér test, based on interpoint distances, is applied to the two distinct collections of multivariate vectors (denoting the two groups for comparison) given by Eq. [11] for Euclidean distances and by Eq. [12] for log-Euclidean distances.

METHODS

Simulations

DTI data were simulated for two groups of participants ($n_1 = n_2 = 20$), where one of the groups represents a hypothetical control population, and the other group a test population. The sensitivity of the statistical tests was evaluated separately for two conditions between the groups: different PV (principal eigenvector) with identical FA, and identical PV with different FA. In both cases, the NMR signal was simulated by sampling the tensors from 60 directions of diffusion encoding (23) with $b = 700$ s/mm^2, 10 encodings with $b = 0$ s/mm^2, and a Rician noise distribution (SNR ≈ 20). The tensors were reconstructed using the standard least-squares method (24).

Intersubject variability was achieved by drawing the diffusion tensors from a Wishart distribution. The Wishart distribution is a multivariate analogue to the χ^2 distribution and is most commonly used to describe the covariance matrix from multivariate statistics, although it is more widely applicable to matrix-valued random variables (25). If $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are independent and identically distributed as three-dimensional Gaussian random variables with zero mean and covariance matrix V, then the maximum-likelihood estimator

$$S = \sum_{i=1}^{m} \mathbf{X}_i \mathbf{X}_i^T \qquad [13]$$

is a 3×3 random matrix and has a Wishart distribution with m degrees of freedom. The expectation of S is given by mV. Since we do not observe multivariate observations from the diffusion-weighted sequence it is difficult to match the degrees of freedom with any parameter from the acquisition. Values for the simulation studies were selected to provide different levels of randomness by imposing a distribution on the ideal diffusion tensor.

Participant Recruitment

Twenty-four patients with a history of migraine (age 35.0 $\pm$ 8.4 years, 16 F, 8 M) and 12 healthy matched controls (age 31.0 $\pm$ 8.0 years, 9 F, 3 M) were enrolled in the study. The migraine patients all met the ICHD-II criteria (26). Informed written consent was obtained for each participant before the scanning session, and the Massachusetts General Hospital Human Studies Committee approved all procedures under Protocol #2002P-000652.

Image Acquisition

DTI scans were acquired on a 3 T Siemens Allegra MRI scanner (Siemens, Erlangen, Germany). Head motion was minimized using tightly padded clamps attached to the head coil. The DTI scans used a single-shot, twice-refocused echo planar sequence (27). The protocol parameters were TR/TE = 9200/91 ms, slice thickness = 2 mm (0 mm gap), 64 axial slices, FOV 256 $\times$ 256 mm, matrix 128 $\times$ 128, 1 average, 60 directions of diffusion encoding (23) with $b = 700$ s/mm^2, and 10 encodings with $b = 0$ s/mm^2. High-resolution MPRAGE anatomical scans were also collected.

Table 1
Eigenvalues and Corresponding FA Values for the Simulation Results Using Different Angles

FA	$(\lambda_1, \lambda_2, \lambda_3)$
0.69	(1.5, 0.4, 0.4)
0.54	(1.3, 0.5, 0.5)
0.36	(1.1, 0.6, 0.6)

The units for the eigenvalues are $\mu m^2/ms$ with the same trace value of 2.3.

Preprocessing

Image preprocessing was performed using FreeSurfer (http://surfer.nmr.mgh.harvard.edu), FSL (http://www.fmrib.ox.ac.uk/fsl), and custom software developed at the A. A. Martinos Center. The diffusion-weighted images were corrected for head motion and residual eddy current distortions using the FLIRT program (12-dof, mutual information cost function, sinc resampling) (28,29). Diffusion tensor and FA images were reconstructed from the diffusion-weighted images (24,30) and spatially normalized to MNI (Montreal Neurological Institute) space. The spatial normalization was performed by registering the T2 image to a skull-stripped version of the MNI 152 T2 atlas. The skull-stripping was performed using the BET program (default values) (31). The registration was performed using FLIRT (12-dof, correlation ratio cost function). The atlas transformation was then applied to the diffusion tensor and FA images (nearest neighbor resampling). Nearest neighbor sampling was used to avoid averaging of the diffusion tensors. For the diffusion tensor images, the rotational portion of the affine atlas transform A was applied to the individual diffusion tensors: $D \leftarrow RDR^T$ where $R = A(A^TA)^{-1/2}$ (32).

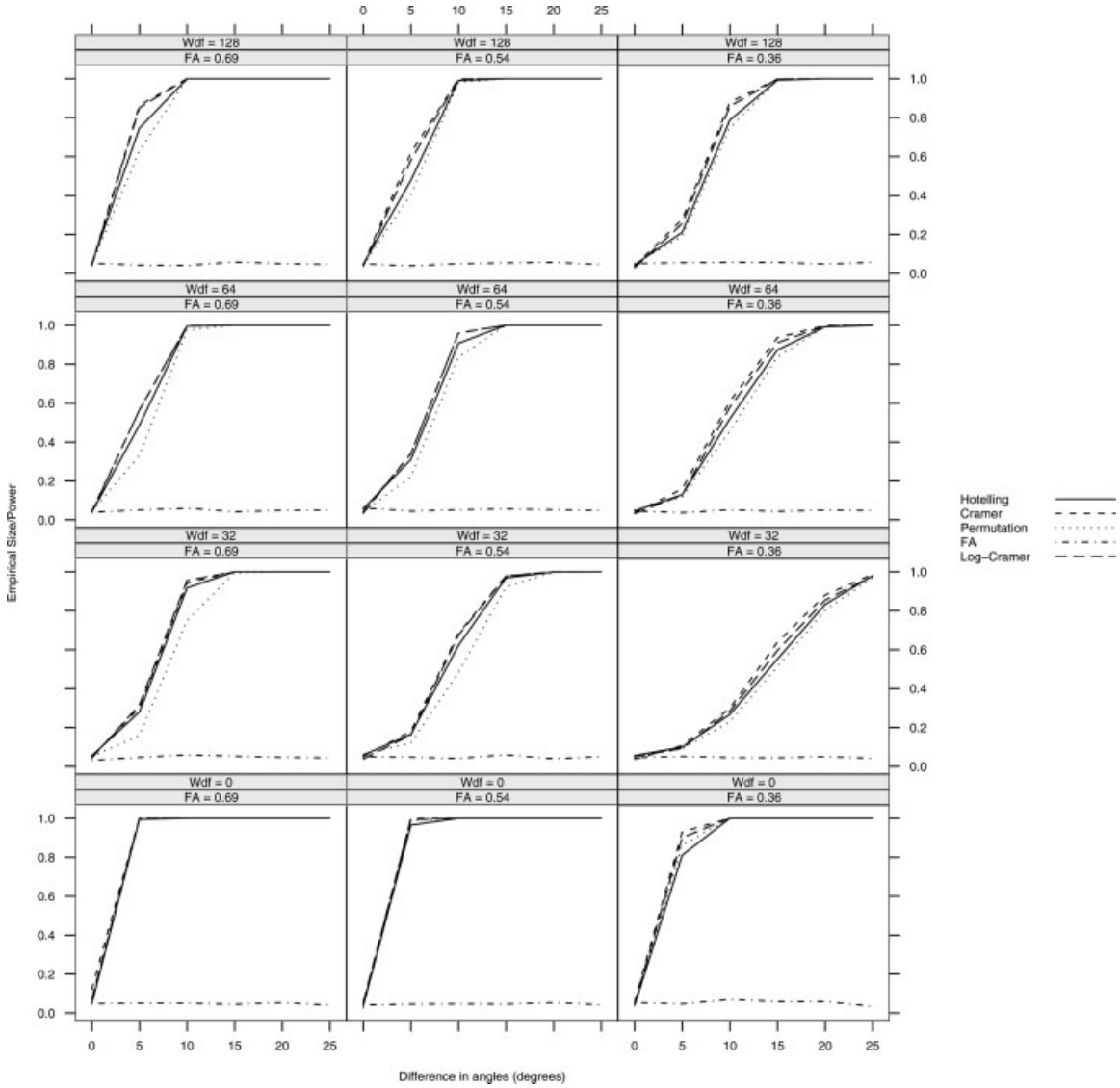

FIG. 1. Empirical size/power for simulated group comparisons ($n_1 = n_2 = 20$) of fiber orientation. FA is held constant (at 0.36, 0.54 and 0.69) while the fiber orientation is allowed to differ by 0°–25° between the two groups. The degrees of freedom for the Wishart distribution are 128, 64 and 32 and correspond to low, medium and high levels of subject-to-subject variability. The univariate test based solely on FA follows the empirical size of the test $\alpha = 0.05$.

Statistical Analysis

Comparison between the migraine and control groups was performed using a voxel-based analysis. The FA volumes were compared using a t-test and the diffusion tensor volumes were compared using the multivariate Cramér test (see Theory section). The diffusion tensors were normalized to unit trace; i.e., $D \leftarrow D/\mathrm{Tr}(D)$, to mitigate the effects of trace differences, for example, due to CSF partial volume and concentrate the comparison on differences in diffusion orientation and anisotropy. Without the trace normalization, the group statistical maps showed rim artifacts on the cortical surface and periventricular regions due to morphometric differences between subjects. Multiple comparisons correction was performed with the False Discovery Rate (FDR) method with $q = 0.05$ (33). Multiple comparisons control was performed at the cluster level using the Monte Carlo permutation method (34). The null distribution for the cluster size was computed at each voxel using a significance threshold of $P < 0.05$ and 18-neighbor voxel connectivity. The corrected significance of each cluster was then determined from the permutation distribution. The Monte Carlo calculation used 103 permutation trials.

Table 2
Eigenvalues and Corresponding FA Values for the Simulation Results Using Different FA Values

FA	$(\lambda_1, \lambda_2, \lambda_3)$
$\begin{pmatrix}0.52\\0.63\\0.69\\0.76\\0.80\\0.84\end{pmatrix}$	$\begin{pmatrix}1.00\\1.28\\1.50, & 0.4, & 0.4\\1.93\\2.31\\2.72\end{pmatrix}$
$\begin{pmatrix}0.41\\0.48\\0.54\\0.59\\0.64\\0.68\end{pmatrix}$	$\begin{pmatrix}1.00\\1.14\\1.30, & 0.5, 0.5\\1.46\\1.64\\1.82\end{pmatrix}$
$\begin{pmatrix}0.28\\0.32\\0.36\\0.41\\0.45\\0.48\end{pmatrix}$	$\begin{pmatrix}0.95\\1.03\\1.10, & 0.6, 0.6\\1.20\\1.29\\1.38\end{pmatrix}$

The units for the eigenvalues are $\mu m^2/ms$.

RESULTS

Simulations

Empirical size and power calculations were performed on the multivariate hypothesis testing procedures for group comparisons using simulated DTI data under two conditions. The first involved holding the FA constant between the two groups, using the FA values in Table 1, and varying the PV Euler angles. One thousand group comparisons ($n_1 = n_2 = 20$) were performed using each multivariate hypothesis test (Hotelling T^2, Cramér, log-Cramér, permutation) and the univariate test (Student's t-test) on FA. For the different PV conditions, the eigenvalues were fixed for all samples $(\lambda_1, \lambda_2, \lambda_3) = (1.5, 0.4, 0.4)$ $\mu m^2/ms$ and the PV Euler angles were fixed at $(0, 45^\circ, 0)$ for the first group of subjects. The PV Euler angles $(0, \alpha, 0)$ were allowed to vary where $\alpha \in \{45^\circ, 50^\circ, 55^\circ, 60^\circ, 65^\circ, 70^\circ\}$ for the second group of subjects, thus creating a difference in orientations between 0° and 25°. All hypothesis tests were performed with a signifiance level of 5%.

Figure 1 shows the empirical power of the four multivariate tests (Hotelling T^2, Cramér, log-Cramér, Permutation) and the univariate test on FA for the PV comparison where the two subject groups had identical FA but differing PV Euler angles. The four rows in Fig. 1 differ by the number of degrees of freedom used in the Wishart distribution when simulating the group data. From top to bottom, the degrees of freedom are 128, 64, 32, and in the last row no draw from the Wishart distribution was performed. These choices correspond to low, medium, high and no subject-to-subject variability. The performance of all tests decreases, although not uniformly, with a decrease in the number of degrees of freedom associated with the Wishart distribution. When the Wishart draw was omitted from the simulation procedure all variability is through the measurement of the MR signal only. The value of FA decreases from left to right in the columns.

The empirical power of the Cramér and log-Cramér tests, derived from simulations, were nearly identical for almost all angular differences under the three specific values of FA. Hotelling's T^2 test was slightly inferior, by 5–25% depending on the simulation conditions, when compared with the Cramér test regardless of the differing simulation conditions. The multivariate permutation test exhibited reduced power (by 5–50% depending on the simulation conditions) compared to with the Cramér test, most significantly for the smaller angular differences (5°–15°). The univariate two-sample t-test applied to FA was also calculated for all angular differences with an expected rejection rate of the null hypothesis equivalent to the significance level of the hypothesis test (i.e., 5%).

For the variable FA case, the eigenvalues of group 1 were taken from Table 1 and applied to three choices of intersubject variability parameterized by varying degrees of freedom from the Wishart distribution. The first eigenvalue λ_1 was allowed to vary for group 2 to produce a range of FA values roughly centered around the three rows from Table 1. The specific FA values, and their corresponding eigenvalues, are provided in Table 2. Figure 2 shows the empirical power of the hypothesis tests from 1,000 simulated group comparisons ($n_1 = n_2 = 20$). The structure is identical to Fig. 1 in terms of degrees of freedom from the Wishart distribution. The empirical power of the Cramér and multivariate permutation tests were almost identical across the simulation conditions used and were substantially better than the alternative univariate and multivariate tests. However, the multivariate permutation test was the most powerful test for all differences in FA between the two groups. The three remaining tests, decreasing in power, were the univariate test of FA based on the parametric t-test, Hotelling's T^2 test and the log-Cramér test. When the two groups both exhibited low FA (third column in Fig. 2) the t-test of FA was more powerful than the Cramér test when the differences in FA between the two groups was small (0.04–0.05), but then the Cramér test became more powerful when the

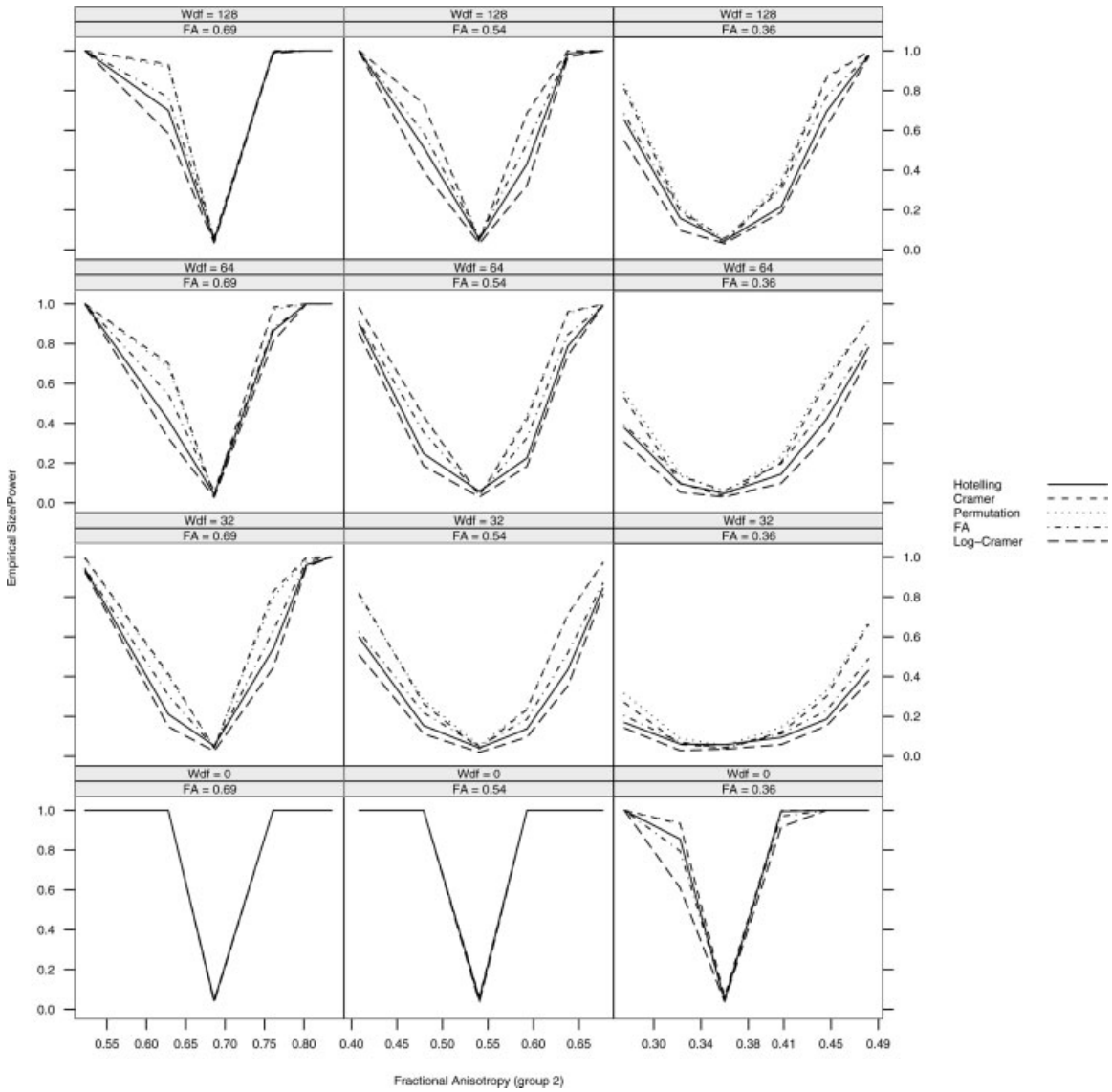

FIG. 2. Empirical size/power for simulated group comparisons ($n_1 = n_2 = 20$) of fractional anisotropy (FA). The Euler angles applied to the principal eigenvector are held constant while FA is allowed to differ by ± 0.15 between the two groups. When the FA associated with each group is the same, the hypothesis tests follow the empirical size of $\alpha = 0.05$. The degrees of freedom for the Wishart distribution are 128, 64, 32 and correspond to low, medium and high levels of subject-to-subject variability.

differences in FA became larger (0.08–0.15). The improvements in empirical power for the t-test of FA were modest at best (15–20%) and only occurred for a very small subset of simulated conditions. At larger differences in FA, the t-test based on FA resulted in up to a 30% decrease in power when compared with the multivariate permutation test. Hotelling's T^2 test performed slightly worse than the univariate test on FA with reductions in power ranging from 5 to 45% when compared with the multivariate permutation test. Finally, the log-Cramér test performed the worst in the ability to detect differences in FA between the two groups with reductions up to 70%, when compared with the multivariate permutation test, observed in simulations.

Additional simulations were run with duplicating the information provided so far, but with $(n_1, n_2) = (40, 20)$. No significant differences were observed when comparing the two groups under differences in FA or Euler angles.

Migraine Data

With respect to the migraine data, we focus on a significant cluster (cluster level $p < 0.05$) in the left postcentral gyrus (peak MNI $-28, -38, 52$) detected by both the univariate FA test and the multivariate tensor test (Fig. 3). The tensor-based significance cluster extended from the central sulcus to the crown of postcentral gyrus, whereas the cluster for the FA comparison was restricted to the body of the gyrus and had a smaller mediolateral extent. The volume of the FA-based cluster was 0.232 cc compared to 0.624 cc for the tensor-based cluster.

DISCUSSION

We have investigated both parametric and nonparametric tests for comparing populations of tensors. Using Monte

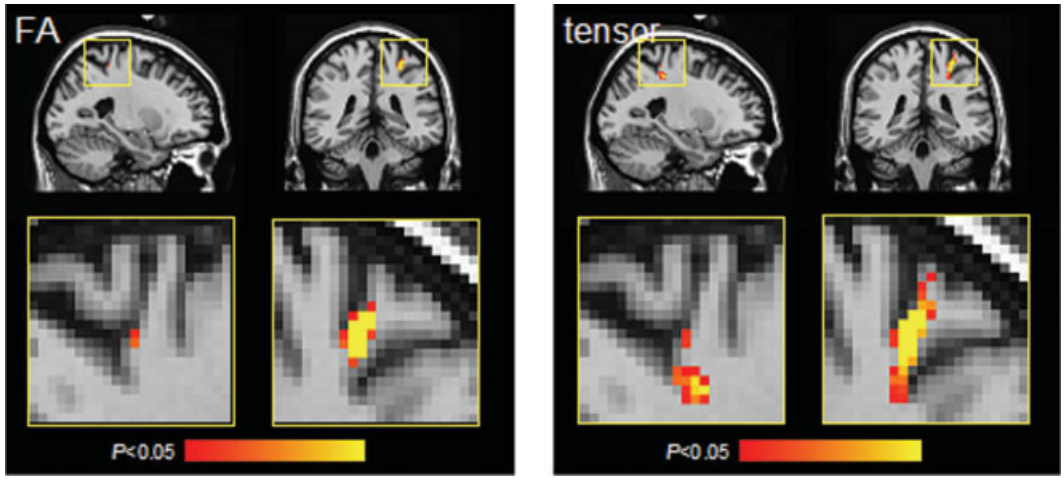

FIG. 3. Statistical parametric maps for (left) univariate FA comparison, and (right) multivariate tensor comparison. The displayed cluster was significant at the $p < 0.05$ cluster level (permutation method).

Carlo simulations, we have shown that the preferred method, the Cramér test on the unique diffusion tensor elements, can detect diffusion tensor principal eigenvector differences of 15 degrees with a power of 80% under typical design conditions. We have also shown that the test is more sensitive to FA differences even when there is no principal eigenvector difference. In the migraine data set, the multivariate tensor test provided a 169% increase in the volume of a signficant cluster compared to the univariate FA test. Statistical group comparison of diffusion tensors promises to boost the statistical sensitivity of DTI group comparisons. The tensor comparison approach will also extend the application of DTI to WM disorders characterized by a change in WM architecture without an associated change in FA.

In a simulation setting the multivariate permutation and Cramér tests were equally effective at detecting differences in FA, when the direction of anisotropy was fixed for the two subject groups, and superior to both all other multivariate hypothesis testing procedures considered in the experiments. There appears to be an asymmetry in the empirical power curves displayed in Fig. 2 implying that detecting differences between higher values of FA is easier than detecting differences between lower values. This is even more apparent when the FA for the reference group is smaller; e.g., 0.36 versus 0.69. A more exhaustive sampling of levels in FA for both groups in the simulation study is required to better characterize this apparent asymmetry.

As previously noted, the performance of all tests decreases with a decrease in the degrees of freedom associated with the Wishart distribution, which induces more variability between subjects. The adaptation of the Cramér test, which is based on Euclidean distances, to operate on a specific example of a Riemannian metric (i.e., log-Euclidean distances) did not improve the performance of the test. Whereas Riemannian metrics are advantageous in applications such as interpolation or regularization, from a hypothesis testing perspective there is no reason at this time to prefer the log-Euclidean distance for the specific scenarios simulated in this paper. Initial investigations into the consequences of applying matrix logarithms to the diffusion tensors indicate that the six-dimensional vectors between the two groups, when they are known to be substantially different, are brought closer together in six-dimensional space after the transformation. Thus, working in the log-Euclidean space reduces the group difference observed in Euclidean space. Given the relatively recent emergence of non-Euclidean metrics in DTI, elaborating on the differences in performance of the log-Euclidean distance metric for various applications, specifically statistical hypothesis testing, will provide a better understanding to their relative advantages and disadvantages.

In addition to the use of Euclidean or non-Euclidean metrics, the choice of parameterizations for the diffusion tensor could also be explored. The rotation-invariant parameterization in Eq. [11] or the multivariate vector form of the diffusion tensor in Eq. [2] have been used here, but others are equally plausible. For example, the eigenvalues and Euler angles after diagonalization were inputs to a Bayesian estimation scheme in (35). Such an alternative parameterization may or may not improve the performance of the hypothesis test, but may allow an easier interpretation of the results—especially in the case of the multivariate permutation test where the individual elements of the multariate vector are tested independently and then combined at the end to form a single test statistic.

It is important to note that the diffusion tensor is a model of the underlying diffusion properties of tissue, and thus, suffers from the assumptions imposed by the model. One such assumption is that a single tensor is sufficient to characterize the diffusion properties at every voxel. This is not valid for a variety of anatomical regions in the structure of WM and therefore all hypothesis testing procedures considered here may have difficulty detecting differences where the single tensor model breaks down. The application of multivariate hypothesis testing to more complicated models of white matter structure based on DTI acquisitions (36) and also alternative high-resolution angular diffusion imaging (HARDI) acquisition schemes

such as *q*-ball imaging (37) will be the subject of future study.

For the migraine patient data set, both the univariate FA test and the multivariate tensor test detected a significant cluster in the left postcentral gyrus. However, the cluster from the tensor test traversed the length of the gyrus, consistent with the expected geometry of the fiber pathway originating from the ventral posterolateral (VPL) nucleus of the thalamus and targeting S-I via the internal capsule and corona radiata (38–40). In contrast, the cluster from the FA test was restricted to the body of the gyrus.

Fibers entering the postcentral gyrus originate from the VPL and are destined for layer IV of Brodmann areas 3, 1 and 2, collectively known as S-I. Within S-I, fibers carrying information from muscle stretch receptors insert into area 3a, information from cutaneous innervation insert into area 3b, information from deep pressure receptors insert into area 2 and information from rapidly adapting receptors insert into area 1. Layer II/III of each area in turn projects to area 5 and S-II. It is known that areas 3 and 2 are located on opposite banks of postcentral gyrus, separated by area 1, which is located at the crown of the gyrus (38).

The apparent insertion of this fiber track delineated by both the FA test and the tensor test clusters into the bank of the postcentral gyrus is consistent with the known anatomical and functional organization of areas 3, 1, and 2. Since migraine is primarily referred cutaneous sensation, it is reasonable to conclude that the cluster identified by the multivariate tensor test appears to be inserting into area 3 of the postcentral gyrus. Furthermore, the finding of diffusion alterations in postcentral gyrus WM is consistent with the established involvement of somatosensory cortex in migraine pathophysiology (41,42).

The significant clusters from the FA and tensor-based comparisons both localized to left postcentral gyrus, but the tensor-based cluster traversed the length of the gyrus consistent with the projections of the corona radiata. The larger spatial extent of the tensor-based cluster is potentially attributable to two sources: differences in diffusion orientation which are not detectable on the FA test or the Cramér test's higher sensitivity to FA differences, as shown in the simulation studies (Figs. 1 and 2). The specific contributions of diffusion orientation and diffusion anisotropy to the tensor differences could be determined by testing for differences in the diffusion principal eigenvector without including anisotropy information. This test could also be performed using the Cramér method. The specific contributions of diffusion anisotropy and diffusion orientation to the tensor differences will be the subject of future study.

ACKNOWLEDGMENTS

The authors thank Timothy Reese, Thomas Benner, Josh Snyder, Kevin Teich, and Alexandre DaSilva for assistance with this project. The authors would also like to thank the Associate Editor and two anonymous reviewers for constructive comments. This work was supported by National Institutes of Neurological Disorders and Stroke grant NS035611, National Center for Research Resources grant RR14075, GlaxoSmithKline, the Swiss Heart Foundation, the Athinoula A. Martinos Foundation, the Mental Illness and Neuroscience Discovery Institute, and the National Alliance for Medical Image Computing (National Institute for Biomedical Imaging and Bioengineering Grant U54 EB005149) which is funded through the National Institutes of Health Roadmap for Medical Research.

REFERENCES

1. Dong Q, Welsh RC, Chenevert TL, Carlos RC, Maly-Sundgren P, Gomez-Hassan DM, Mukherji SK. Clinical applications of diffusion tensor imaging. J Magn Reson Imaging 2004;19:6–18.
2. Taylor WD, Hsu E, Krishnan KRR, MacFall JR. Diffusion tensor imaging: Background, potential, and utility in psychiatric research. Bio Psychiatry 2004;55:201–207.
3. Bengtsson SL, Nagy Z, Skare S, Forsman L, Forssberg H, Ullén F. Extensive piano practicing has regionally specific effects on white matter development. Nat Neurosci 2005;8:1148–1150.
4. Schwartzman A, Dougherty RF, Taylor JE. Cross-subject comparison of principal diffusion direction maps. Magn Reson Med 2005;53:1423–1431.
5. Mardia KV, Jupp PE. Directional statistics. Chicester, UK: Wiley; 2000.
6. Ollinger JM, Kim J, Johnson SC, Alexander AL. Multivariate analysis of diffusion tensor data using the Hotelling T2 statistic. In Proceedings of the 13th Annual Meeting of ISMRM. Miami Beach, 2005. p. 1324.
7. Pajevic S, Basser PJ. Parametric and non-parametric statistical approaches in diffusion tensor magnetic resonance imaging. J Magn Reson 2003;161:1–14.
8. Basser PJ, Pajevic S. A Normal distribution for tensor-valued random variables: Applications to diffusion tensor MRI. IEEE Trans Med Imaging 2003;22:785–794.
9. Pesarin F. Multivariate permutation tests: With applications in biostatistics. Chichester, UK: Wiley; 2001.
10. Baringhaus L, Franz C. On a new multivariate two-sample test. J Multivariate Anal 2004;88:190–206.
11. Batchelor PG, Moakher M, Atkinson D, Calamante F, Connelly A. A rigorous framework for diffusion tensor calculus. Magn Reson Med 2005;53:221–225.
12. Pennec X, Fillard P, Ayache N. A Riemannian framework for tensor computing. Int J Comput Vis 2006;66:41–66.
13. Arsigny V, Fillard P, Pennec X, Ayache N. Fast and simple calculus on tensors in the log-Euclidean framework. Proc Med Image Comput Comput-Assist Intervent – MICCAI 2005; 3749:115–222.
14. Wheeler-Kingshott CAM, Barker GJ, Steens SCA, van Buchem MA. *D*: The diffusion of water. In: Tofts, P. editor, Quantitative MRI of the brain. Chichester, UK: Wiley; 2003. pp 203–256.
15. Pierpaoli C, Basser P. Toward a quantitative assessment of diffusion anisotropy. Magn Reson Med 1996;36:893–906.
16. Hasan KM, Basser PJ, Parker DL, Alexander AL. Analytical computation of the eigenvalues and eigenvectors in DT-MRI. J Magn Reson 2001;152:41–47.
17. Carew JD, Koay CG, Wahba G, Alexander AL, Basser PJ, Meyerand ME. The asymptotic distribution of diffusion tensor and fractional anisotropy estimates. In Proceedings of the 14th Annual Meeting of ISMRM, Seattle, 2006; P 1066.
18. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge, UK: Cambridge University Press; 1997.
19. Imhof JP. Computing the distribution of a quadratic form in normal variables. Biometrika 1961;48:419–426.
20. Farebrother RW. Algorithm AS 256: The distribution of a quadratic form in Normal variables. Appl Stat 1990;39:294–309.
21. Franz C. Cramer: Multivariate nonparametric Cramér-test for the two-sample-problem, 2004. R package version 0.7-1.
22. Arsigny V, Fillard P, Pennec X, Ayache N. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn Reson Med 2006;56:411–421.
23. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn Reson Med 1999;42:515–525.
24. Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. J Magn Reson 1994;103:247–254.
25. Arnold SF. Wishart distribution. In: Kotz S, Johnson NL, editors. Encyclopedia of statistical sciences, Volume 9. New York: Wiley; 1988; pp 641–645.

26. Silberstein SD, Olesen J, Bousser MG, Diener HC, Dodick D, First M, Goadsby PJ, Gobel H, Lainez MJ, Lance JW, Lipton RB, Nappi G, Sakai F, Schoenen J, Steiner TJ. The international classification of headache disorders, 2nd edition (ICHD-II)—Revision of criteria for 8.2 medication-overuse headache. Cephalalgia 2005;25:460–465.
27. Reese TG, Heid O, Weisskoff RM, Wedeen VJ. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. Magn Reson Med 2003;49:177–182.
28. Jenkinson M, Smith SM. A global optimisation method for robust affine registration of brain images. Med Image Anal 2001;5:143–156.
29. Jenkinson M, Bannister PR, Brady JM, Smith SM. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17:825–841.
30. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J Magn Reson B 1996;111:209–219.
31. Smith S. Fast robust automated brain extraction. Hum Brain Mapp 2002;17:143–155.
32. Alexander DC, Pierpaoli C, Basser PJ, Gee JC. Spatial transformations of diffusion tensor magnetic resonance images. IEEE Trans Med Imaging 2001;20:1131–1139.
33. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 2002;15:870–878.
34. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum Brain Mapp 2001;15:1–25.
35. Behrens TEJ, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM. Characterization and propogation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 2003;50:1077–1088.
36. Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn Reson Med 2002;48:577–582.
37. Tuch DS. Q-ball imaging. Magn Reson Med 2004;52:1358–1372.
38. Kaas JH, Nelson RJ, Sur M, Lin CS, Merzenich MM. Multiple representations of the body within the primary somatosensory cortex of primates. Science 1979;204:521–523.
39. Kaas JH, Nelson RJ, Sur M, Merzenich MM. Organization of somatosensory cortex in primates. In: Schmitt FO, Worden FG, Adelman G, Dennis SG, editors. The organization of the cerebral cortex, Proceedings of a Neurosciences Research Program Colloquium. Cambridge, MA: MIT Press; 1981. pp 237–261.
40. Kaas JH. How sensory cortex is subdivided in mammals: implications for studies of prefrontal cortex. Prog Brain Res 1990;85:3–10.
41. Rocca MA, Colombo B, Pagani E, Falini A, Codella M, Scotti G, Comi G, Filippi M. Evidence for cortical functional changes in patients with migraine and white matter abnormalities on conventional and diffusion tensor magnetic resonance imaging. Stroke 2003;34:665–670.
42. Lang E, Kaltenhauser M, Neundorfer B, Seidler S. Hyperexcitability of the primary somatosensory cortex in migraine–a magnetoencephalographic study. Brain 2004;127:2459–2469.