Magnetic Resonance in Medicine 53:1423–1431 (2005)

Cross-Subject Comparison of Principal Diffusion Direction Maps

Armin Schwartzman,[1]* Robert F. Dougherty,[2] and Jonathan E. Taylor[1]

Diffusion tensor imaging (DTI) data differ fundamentally from most brain imaging data in that values at each voxel are not scalars but 3 × 3 positive definite matrices also called diffusion tensors. Frequently, investigators simplify the data analysis by reducing the tensor to a scalar, such as fractional anisotropy (FA). New statistical methods are needed for analyzing vector and tensor valued imaging data. A statistical model is proposed for the principal eigenvector of the diffusion tensor based on the bipolar Watson distribution. Methods are presented for computing mean direction and dispersion of a sample of directions and for testing whether two samples of directions (e.g., same voxel across two groups of subjects) have the same mean. False discovery rate theory is used to identify voxels for which the two-sample test is significant. These methods are illustrated in a DTI data set collected to study reading ability. It is shown that comparison of directions reveals differences in gross anatomic structure that are invisible to FA. Magn Reson Med 53:1423–1431, 2005. © 2005 Wiley-Liss, Inc.

Key words: diffusion tensor imaging; principal diffusion direction; directional statistics; false discovery rate; multiple subjects

Diffusion tensor imaging (DTI) is a relatively new anatomic imaging technique that measures the diffusion of water molecules in tissue (1–4). The pattern of diffusion is an indicator of the microscopic properties of the measured tissue. In the brain, DTI reveals structure within the white matter that is unavailable in other imaging methods. For example, major white matter fiber tracts can be identified and measured in an individual subject (5). Further, certain aspects of the health of these white matter tracks can be measured, e.g., detecting demyelination in multiple sclerosis (6).

DTI data differ fundamentally from conventional imaging data in that values at each voxel are not scalars but 3 × 3 positive definite matrices, also called diffusion tensors. The diffusion tensor is in essence the covariance matrix of a 3D Gaussian distribution that models the Brownian motion of the water molecules within a voxel. Three important tissue properties can be derived from the tensors: isotropic diffusion coefficient (i.e., trace), fractional anisotropy (FA), and principal diffusion direction. The isotropic diffusion coefficient expresses an overall effect of diffusion magnitude and is useful for characterizing diseased tissue (6). FA is a normalized SD of the tensor eigenvalues and reflects the degree of anisotropy within a voxel; high FA values suggest the presence of highly directional diffusion such as that seen in white matter fiber tracts (2). The principal diffusion direction is the eigenvector corresponding to the largest eigenvalue of the tensor. It is generally assumed that diffusion is restricted in the direction perpendicular to the nerve fibers, and so the principal diffusion direction provides an estimation of the fiber direction within the voxel (for a review, see (3)).

Frequently, investigators restrict their analysis of DTI data to scalar quantities such as FA and trace (e.g., (6–8)). The main reason for proceeding in this manner is that statistical methods in the imaging field have been developed largely for scalar data. In DTI, such scalar summaries do not capture all the information available in the data. In particular, tensors might be oriented differently and yet have the same set of eigenvalues and thus the same trace and FA. Given the limitations of scalar summaries, new statistical methods are needed for analyzing vector and tensor valued imaging data.

In this article we introduce a method that is aimed at detecting statistical differences in fiber orientation between groups of subjects. In order to characterize the variability in fiber orientation across subjects, we propose a statistical model for the principal diffusion direction based on the bipolar Watson distribution (9). We chose this distribution because it is one of the simplest distributions on the unit sphere that possesses the property of being antipodally symmetric, giving to each direction and its negative the same probability. This is crucial because the diffusion tensor is invariant under sign changes of the principal eigenvector. We describe how to compute mean direction and dispersion of a sample of directions and present a test of whether two (or more) samples of directions have the same mean.

Given two groups of DTI images, the basic approach consists of two steps: (1) Consider each voxel to contain two groups of diffusion directions and apply the two-sample test to each voxel. (2) Identify voxels for which the test is significant. Here we use false discovery rate (FDR) theory to solve the detection task in step 2. FDR is sometimes used in brain imaging as an alternative to Gaussian random field theory for overcoming the multiple comparisons problem (10).

We illustrate these methods using the data set used in Ref. (8). In that study, certain FA differences were found between children with normal and poor reading abilities in a particular region of interest. We show that the proposed method reveals differences in gross anatomic structure that are invisible to statistical tests of FA.

THEORY AND METHODS

The Bipolar Watson Distribution

In DTI, the diffusion directions are not vectors but axes in the sense that the tensor eigenvectors $\mathbf{x}$ and $-\mathbf{x}$ are indis-

[1]Department of Statistics, Stanford University, Stanford, California, USA.
[2]Department of Psychology, Stanford University, Stanford, California, USA.

Grant sponsor: William R. and Sara Hart Kimball Stanford Graduate Fellowship; Grant sponsor: Schwab Foundation for Learning; Grant sponsor: NIH; Grant number: EY015000; Grant sponsor: NSF; Grant number: DMS-0405970.

*Correspondence to: Armin Schwartzman, 390 Serra Mall, Stanford, CA 94305, U.S.A. E-mail: armins@stanford.edu

Received 30 June 2004; revised 11 January 2005; accepted 11 January 2005.
DOI 10.1002/mrm.20503
Published online in Wiley InterScience (www.interscience.wiley.com).

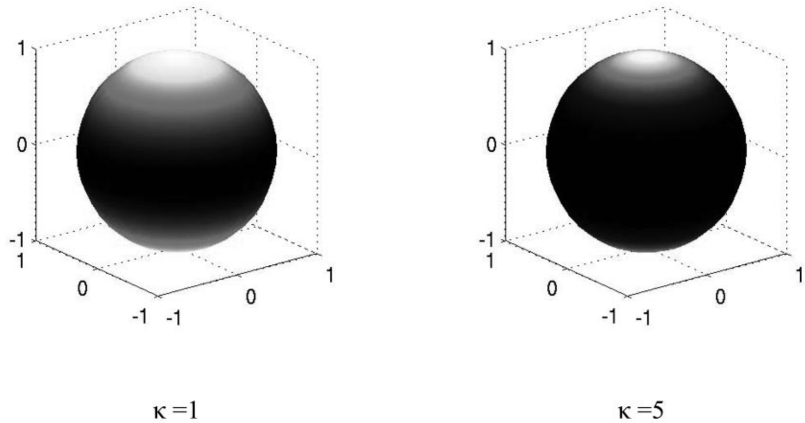

FIG. 1. The bipolar Watson density on the sphere for $\kappa = 1$ (**a**) and $\kappa = 5$ (**b**).

tinguishable. It is thus appropriate to consider probability density functions that are antipodally symmetric. That is, if $\mathbf{x}$ is a random unit vector in $\mathbf{R}^3$, we require that $f(\mathbf{x}) = f(-\mathbf{x})$. We use the notation $\pm\mathbf{x}$ to represent an observed random vector with unit length but undefined sign. One of the simplest models with this property is the bipolar Watson distribution, originally developed by G. S. Watson from Johns Hopkins University in 1960 (and independently by Dimroth and Sheidegger at about the same time) to solve a problem concerning the folding of a layer of rock. The bipolar Watson density is given by Ref. (9) as

$$f(\pm\,\mathbf{x};\mu,\kappa) = C(\kappa)\exp(\kappa(\mu^{\mathrm{T}}\mathbf{x})^2) \qquad [1]$$

$$= C(\kappa)\exp(\kappa\cos^2\theta), \qquad [2]$$

where μ is a unit vector called the mean direction and κ is a positive constant called the concentration parameter. The alternative expression in Eq. [2] uses $\cos\theta = \mu^{\mathrm{T}}\mathbf{x}$, the inner product of μ and $\mathbf{x}$, with θ being the angle between them. The density has maxima at $\pm\mu$ and becomes more concentrated around $\pm\mu$ as κ increases (Fig. 1). Notice that the density is rotationally invariant around $\pm\mu$. The constant $C(\kappa)$ ensures that the density integrates to 1 over the sphere. Computation of the normalizing constant is not necessary for the practical implementation of the comparison methods described here. It is included in the Appendix for completeness.

Mean Direction and Dispersion

Let $\pm\mathbf{x}_1,\ldots,\pm\mathbf{x}_N$ be a collection of observed principal eigenvectors from a single voxel for each of N subjects, or more generally, a collection of N unsigned random unit vectors. Because of the antipodal symmetry, these coordinates determine $2N$ points on the sphere, where each of the N pairs contains diametrically opposite coordinates. As a result, direct average of all $2N$ points is zero. On the other hand, if a sign is arbitrarily chosen for each direction, the direct average of the N signed directions depends on the signs chosen and thus is not unique.

An alternative way to define a measure of location for the sample together with a measure of dispersion is through the maximum likelihood estimators of the parameters μ and κ from the Watson distribution, as follows. Consider the data matrix $\mathbf{X}$ obtained by arranging the vectors as columns, i.e., $\mathbf{X} = (\mathbf{x}_1 \ldots \mathbf{x}_N)$. Define the scatter matrix $\mathbf{S}$ as

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{T}}, \qquad [3]$$

which may be interpreted as the empiric covariance of the points determined by $\mathbf{x}_1,\ldots,\mathbf{x}_N$ on the sphere. Notice that this definition is sign invariant. The sample mean direction $\bar{\mathbf{x}}$ is defined as the principal eigenvector of the matrix $\mathbf{S}$, or the eigenvector that corresponds to the largest eigenvalue. The intuitive rationale behind this definition is that if the points on the sphere have a preferential direction then, as a group, they are also further apart in space from their antipodes. The principal eigenvector of the scatter matrix points in the direction of maximal variance in space, which is the preferential direction for the points on the sphere. We show in the Appendix that $\bar{\mathbf{x}}$ is formally the maximum likelihood estimator of the location parameter μ in the Watson model.

If we denote the greatest eigenvalue of $\mathbf{S}$ by γ, then the quantity $s = 1 - \gamma$, which we call the sample dispersion, gives a measure of how disperse the sample is about the mean. Intuitively, when the sample is concentrated around the mean, the antipodes are far apart as a group and so the spatial variance γ is close to 1, giving a dispersion s that is close to 0. On the other extreme, when the sample is scattered uniformly on the sphere, all three eigenvalues are the same and equal to 1/3 (since their sum is equal to trace($\mathbf{S}$) = 1). The dispersion s in that case takes its maximum value of 2/3.

We show in the Appendix that s is formally the maximum likelihood estimate of $1/\kappa$ in the Watson model, asymptotically when $\kappa \to \infty$. Since κ controls concentration, this connection justifies the interpretation of s as a measure of dispersion. It is also shown in the Appendix that s is the average sine-squared of the angles the samples make with the mean direction, and therefore an interpretation of s in units of angle is obtained by computing the quantity $\sin^{-1}(\sqrt{s})$, which we call the angle dispersion of the sample. Notice that this definition results in a maximal angle dispersion of $\sin^{-1}(\sqrt{2/3}) = 54.74°$ in the case of uniformity.

It is worth mentioning that the quantities $\bar{\mathbf{x}}$ and s, despite having been derived from the Watson model, are descriptive statistics that may be computed for any collection of directions regardless of the underlying distribution. This is analogous to the regular sample mean and variance being generic measures of location and dispersion in scalar data regardless of whether the underlying distribution is Gaussian.

A Two-Sample Test for Directions

Here we address the problem of testing whether two collections of directions have the same mean direction. Consider two collections of directions of sizes N_1 and N_2, with respective sample dispersions s_1 and s_2. Under the null hypothesis H_0: $\mu_1 = \mu_2$, they can be viewed as one single collection of size $N = N_1 + N_2$ and corresponding sample dispersion s. Similar to an analysis of variance, the total dispersion Ns is decomposed as

$$Ns = (N_1 s_1 + N_2 s_2) + (Ns - N_1 s_1 - N_2 s_2),$$

where the first term in parentheses is the total intragroup dispersion and the second term corresponds to the intergroup dispersion.

The test statistic F is then defined as the ratio of the intergroup to the intragroup dispersion divided by the appropriate number of degrees of freedom, 2 for the intergroup term and $2(N - 2)$ for the intragroup term:

$$F = \frac{(N-2)\,(Ns - N_1 s_1 - N_2 s_2)}{(N_1 s_1 + N_2 s_2)}. \qquad [4]$$

The null hypothesis for this test is H_0: $\mu_1 = \mu_2$. If the underlying parameter κ is the same in both samples, then the test statistic F in Eq. [4] is F-distributed with 2 and $2(N - 2)$ degrees of freedom, asymptotically as $\kappa \rightarrow \infty$. Because of the asymptotic assumptions, this is called a high concentration test rather than a large sample test. This means that the test is valid for small sample sizes as long as the group dispersions are low. The Appendix gives a proof of the null distribution of the test statistic and treats the more general case of testing equality of means in a number of samples possibly greater than 2.

False Discovery Rates

The previous two-sample test applies to a single voxel. We overcome the multiple comparisons problem in a region of interest by controlling the FDR, or the proportion of false positives among the rejected voxels. The FDR concept was first introduced by Benjamini and Hochberg from Tel Aviv University in 1995 (11) and has quickly made its way into brain imaging (10). As an alternative to the FDR-controlling procedure described in those sources, we use an equivalent interpretation of the procedure taken from Ref. (12), as follows.

Let T be a test statistic connected to a one-sided test that rejects the null hypothesis at voxel v if the value of the test statistic T_v at voxel v is large. In our case, T is the statistic F from Eq. [4], but the following description applies more generally. For a given region of interest M and a threshold t, consider the actual number of voxels that have values of the test statistic above t and, on the other hand, the number of voxels that would have had values of the test statistic above t if the null hypothesis were true in the entire region. Then the ratio of these two quantities is an estimate of the FDR. Dividing both numerator and denominator by the total number of voxels in M, the FDR at threshold t is expressed in terms of probabilities as

$$\mathrm{FDR}(t) = \frac{P_{H_0}(T_\nu \geq t \,|\, \nu \in M)}{\hat{P}(T_\nu \geq t \,|\, \nu \in M)}, \qquad [5]$$

where $\hat{P}$ denotes a probability computed from the empiric distribution of the test statistic and P_{H_0} a probability computed from the exact distribution of the test statistic according to the null hypothesis H_0. In other words, the FDR is the ratio of the tail areas under the null density and the empiric density, respectively. Notice that this formula assumes that the null distribution of the test statistic is the same in all voxels.

Voxels for which true differences exist between the groups tend to have higher values of the test statistic than expected according to the null hypothesis. As a result, FDR(t) tends to decrease as t increases. In order to identify significant voxels, the FDR is computed for a range of values of t and the required threshold is selected as the lowest t for which the FDR is smaller or equal to the desired control level.

In comparison with the Benjamini and Hochberg procedure, the approach described above allows interpretation of the threshold in terms of the distribution of the data, a useful guide for the analyst. In addition, it avoids the task of sorting the P values, which is an advantage when the number of voxels in the region of interest is large.

Comparison of Direction Maps

The tools described above provide a solution to the problem of finding localized differences in fiber orientation between two groups of subjects. For a given set of DTI images, spatially normalized to a common template, the statistical methodology may be summarized as follows:

1. Select a region of interest.
2. At each voxel, replace the diffusion tensor by the quantity to compare, e.g., FA or principal direction.
3. Compute the value of the test statistic at each voxel across subjects. For FA, use the traditional t-statistic; for principal direction, use the statistic given by Eq. [4].
4. Estimate the FDR curve FDR(t) in the region of interest for a range of values of t.
5. Choose a threshold t that results in an FDR smaller or equal to the desired level.
6. Select voxels whose value of the test statistic is above the threshold. These are the significant voxels.

DATA EXAMPLE

In order to assess the practicality and usefulness of this approach, we applied this methodology to the data set used in Ref. (8). This study investigated anatomic differences related to reading development. The data set consisted of 14 DTI maps from children, of whom 7 were poor readers and 7 and were normal readers. Two subjects were removed from the analysis because the spatial registration to the template brain failed, leaving $N = 12$ DTI maps, 6 in each group.

The imaging methods involving human data were approved by the Stanford Institutional Review Board and conformed to institutional policies and US Federal law. Images were acquired on GE 1.5-T Signa LX. The DTI protocol involved four 3-min whole-brain scans that were averaged. The pulse sequence was a diffusion-weighted single-shot spin-echo, echo planar imaging sequence (TE = 63 ms; TR = 12 s; FOV = 260 mm with 128 × 128 matrix, voxel size = 2 × 2 mm, bandwidth = ±110 kHz, partial k-space acquisition). Thirty-eight axial, 3-mm-thick slices (no skip) were measured for two b values, $b = 0$ and $b = 800$ s/mm^2 (4).

Diffusion tensor maps were spatially normalized to the MNI EPI template as follows. First, linear and nonlinear transformation parameters were computed using the scalar T_2-weighted ($b = 0$) images using SPM99 (Wellcome Department of Cognitive Neurology, London, UK) (13,14). Both the linear and the nonlinear transformations were applied to the original sampling grid and the tensors were resampled in the new grid using a spline-based tensor interpolation algorithm (15). The tensors were then reoriented by means of the preservation of principal direction (PPD) algorithm described in Ref. (16), using the linear (affine) transformation.

Group differences were restricted to the intersection of white matter regions common to all the spatially normalized brains. These regions were determined by applying a Bayesian segmentation algorithm based on each subject's T_2-weighted ($b = 0$) image (17) and then thresholding the probability of a voxel belonging to the white matter at 0.8.

RESULTS

The intersection of white matter regions yielded a region of interest with a total of 9203 voxels. A histogram of the F statistics computed in these voxels is shown in Fig. 2a. Superimposed is the F density with 2 and $2(N - 2) = 20$ degrees of freedom. The apparent agreement between the curve and the histogram is first evidence that the Watson model is a good approximation to the distribution of principal directions across subjects within each voxel. Because of spatial normalization, most voxels correspond to the same anatomic structure across subjects and therefore the null hypothesis that there are no differences in direction holds in those voxels. Many other voxels, however, do not conform to the null hypothesis. The departure from the null density is more clearly seen in the quantile–quantile (Q–Q) plot of Fig. 2b. The agreement between the quantiles is good up to about $F = 3$, which accounts for about 93% of the voxels. The vertical dashed line corresponds to the threshold found in the FDR analysis described below, and so the selected voxels are the cases that lie to the right of this line.

The FDR analysis is summarized in Fig. 2c, which shows a plot of the estimated FDR as a function of threshold, computed from the histogram in Fig. 2a according to Eq. [5]. Notice that the FDR estimate has a general tendency to decrease for larger values of the test statistic. The lack of monotonicity is due to the discontinuous nature of the histogram. By selecting an FDR of 0.2 we obtain a threshold of 11.25, marked in the figure as a vertical dashed segment. As a reference, this threshold corresponds to an uncorrected P value of 5.3×10^{-4}.

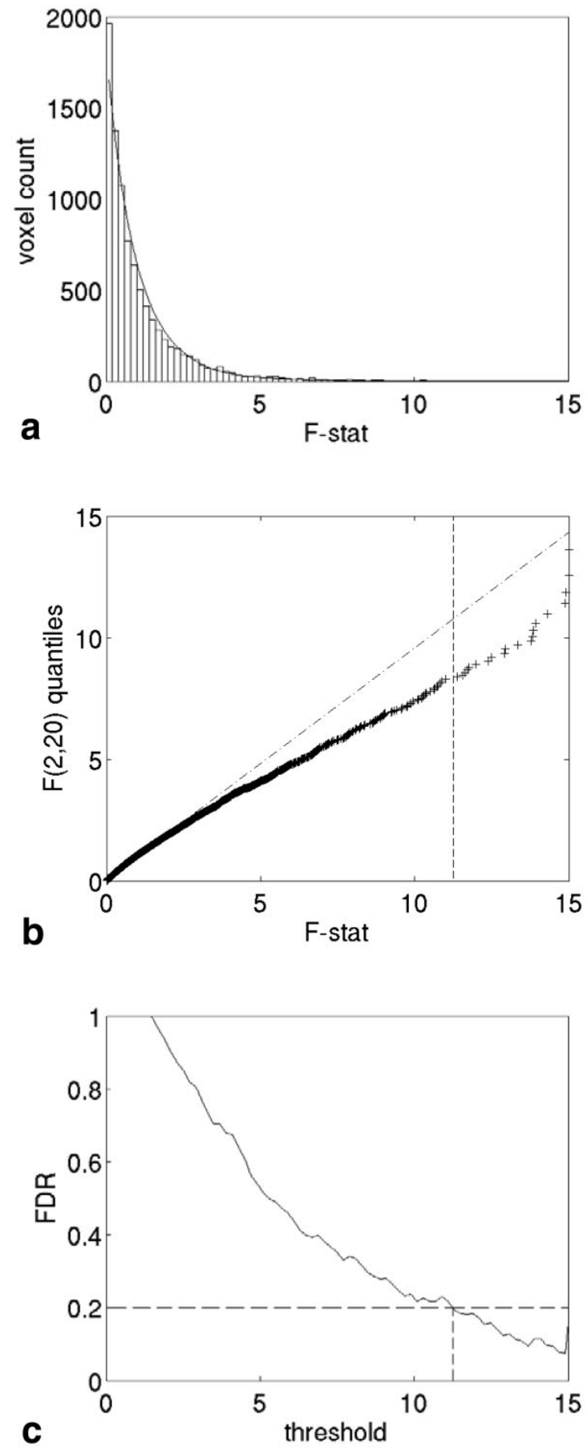

FIG. 2. (**a**) Histogram of the F statistic against the null density. (**b**) Q–Q plot of the F statistic against null quantiles. (**c**) Estimates of FDR as a function of threshold.

Of the total number of voxels in the region of interest, 21 voxels were found to have values of the F statistic larger

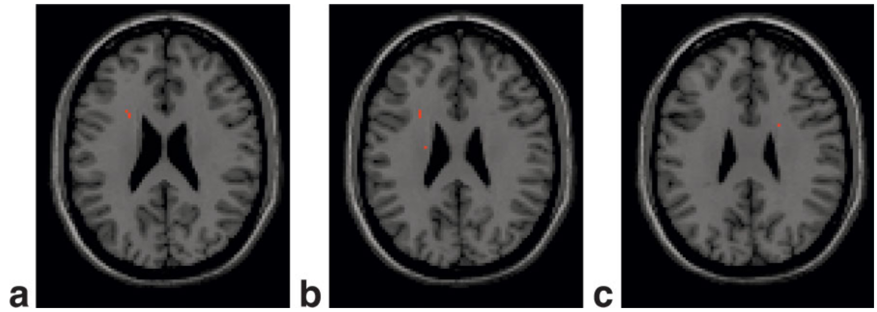

FIG. 3. Significant voxels at slices z = 23 mm (**a**), 25 mm (**b**), and 27 mm (**c**).

than the threshold. Although these selected voxels are located in several areas of the white matter, it is in slices $z = 23$ mm to $z = 25$ mm above the anterior commisure (MNI coordinates) that they are closer together and have the highest values of the F statistic. These three slices are shown in Fig. 3, with the corresponding subset of 8 voxels marked in red. The group difference in this region can also be seen as a local maximum in the F statistic map of Fig. 4a. As a reference, Fig. 4b shows a map of the pooled angle dispersions for all 12 subjects. In this slice, the largest F statistic is 14.90, which corresponds to an uncorrected P value of 1.1×10^{-4}.

Inspection of the original data helps explain the results in terms of anatomy. Figure 5 shows the mean principal directions of both groups at slice $z = 23$ mm. Color is used here to indicate coordinate directions, with blue corresponding to up–down, red to right–left, and green to anterior–posterior. The figure was constructed by taking the absolute value of the coordinates of each mean direction and mapping each coordinate to a scale in the corresponding color. Mixed colors represent directions that are oblique to the coordinate axes. Notice that the anterior region of the left hemisphere shows indeed different fiber orientations between the two groups. In particular, the largest angle difference among the selected voxels is 46.1°. Closer inspection reveals that angle differences as large as 88.9° exist in neighboring voxels. These voxels were not selected because the dispersion in these voxels was too high, resulting in a lower value of the test statistic.

For comparison, a similar FDR analysis was performed using FA as the criterion for comparison rather than principal direction. The methodology used followed the description under Comparison of Direction Maps, with the only exception that the standard t statistic was used as a

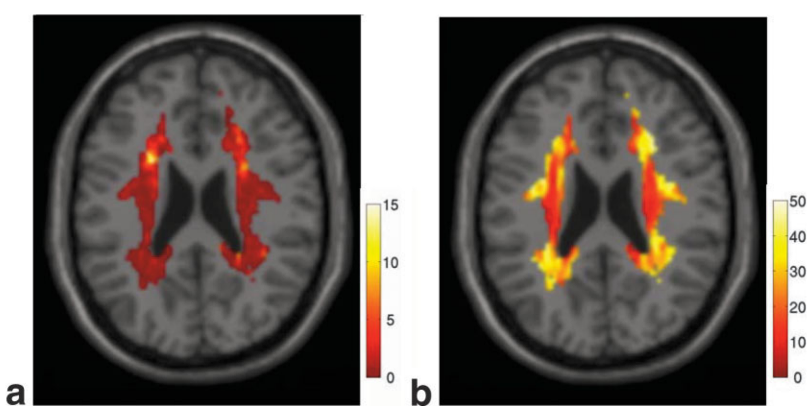

FIG. 4. The F statistic (**a**) and angle dispersion (degrees) of the principal direction (**b**) at slice z = 23 mm.

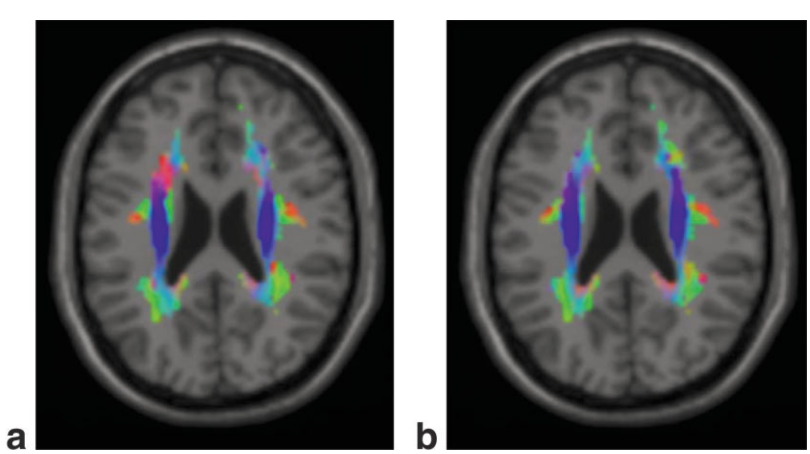

FIG. 5. Mean direction for control group (**a**) and dyslexic group (**b**) at slice z = 23 mm.

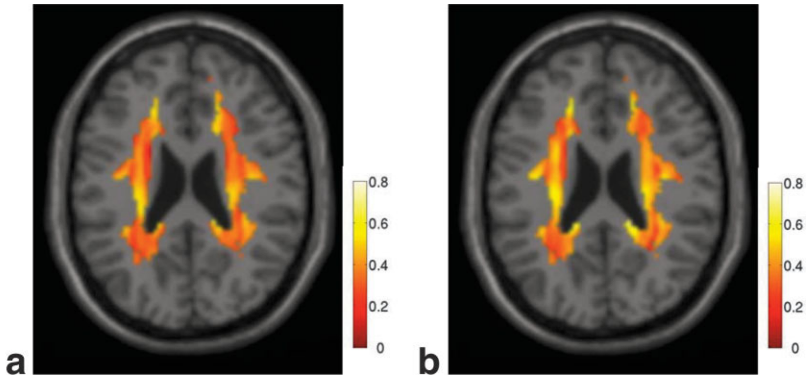

FIG. 6. Mean *FA* for control group (**a**) and dyslexic group (**b**) at slice $z = 23$ mm.

test statistic instead, given the scalar nature of the FA maps. Figures 6 and 7 show the FA maps and t statistic maps at the same slice as Figs. 4 and 5. Here the difference has been taken as control minus dyslexic. Not only are there no striking differences between the two groups, but also the region that was identified as being different using principal directions has very similar values of FA. An FDR analysis at the 0.2 level yielded no significant voxels.

DISCUSSION

The results described above can be interpreted in terms of brain anatomy. Referring to Fig. 5, the bright blue region in both hemispheres indicates the strong vertical directionality of the corona radiata, while the red anterior areas in the control group indicate the frontal projections of the corpus callosum, composed of laterally oriented fibers. In the dislexic group, that same area contains vertical rather than lateral fibers. This implies that the corona radiata extends more anteriorly in the dyslexic group than in the control group. The effect is more pronounced in the left hemisphere but it is also present in the right hemisphere, as indicated in Fig. 3c. This finding may relate to gross white matter differences between good and poor readers found in previous anatomic studies (e.g., (18)). Both the corona radiata and the corpus callosum are structures that possess medium to high FA values, which is why no differences in FA were observed. This is a perfect example of the importance of using directional information for analyzing DTI data.

The use of the F (2,20) density as null stems from the assumption that the principal directions are highly concentrated within each voxel for both populations. In order to determine the dependence of the results on the high-concentration assumption, we computed the null distribution for the test statistic F using Monte Carlo simulation for various values of the concentration parameter κ. The simulations showed that the distributions approached the F (2,20) distributions quickly as κ increased. For instance, for a marginal P value of 0.001, the corresponding thresholds were 8.5 for $\kappa = 5$ and 9.4 for $\kappa = 10$, compared to 9.9 for the F (2,20) distribution. The values $\kappa = 5$ and $\kappa = 9.84$ are the 25th and 50th percentiles of the distribution of κ among all 9203 voxels, so we may conclude the high concentration assumption indeed holds for most voxels. Figure 4b shows that high concentration regions tend to coincide with highly coherent anatomic regions, such as the corona radiata. For reference, $\kappa = 5$ and $\kappa = 10$ correspond to angle deviations of 29° and 19°, respectively. Notice that the threshold obtained from the F (2,20) distribution is higher than it would be if the true concentration were used instead, thus biasing the test toward selecting fewer voxels.

There is no universally agreed notion of power in the multiple comparisons literature. As a guideline, we consider the power at a single voxel, using the observed peak separation of 46.1° as the effect size. We used Monte Carlo simulation to compute the density of the test statistic under this alternative hypothesis and found that the power of a single test of the F (2,20) null at level $\alpha = 0.001$ was 0.180 for $\kappa = 5$ and 0.804 for $\kappa = 10$. The lack of power at low concentrations may explain why so few voxels were rejected in the FDR analysis, in addition to the threshold bias mentioned previously. The procedure can be adapted

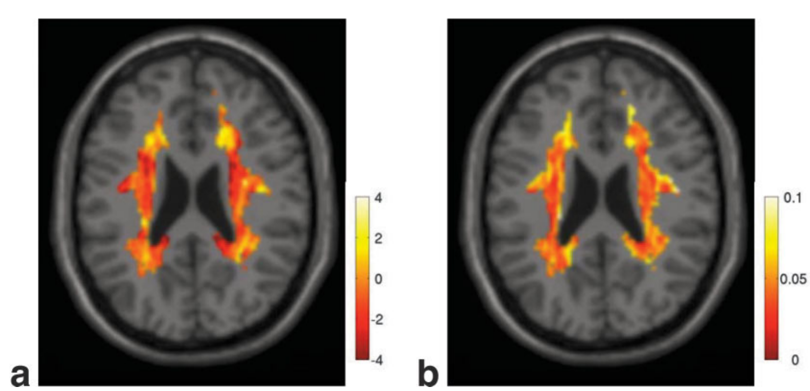

FIG. 7. The t test statistic (**a**) and pooled SD of FA (**b**) at slice $z = 23$ mm.

to low κ by computing the appropriate null density as described above.

For the selected threshold providing an FDR of 0.2, we expect 1 of every 5 significant voxels to be false positives. This is perhaps higher than desired as a compelling level of significance. Setting the FDR level at 0.1 would result in a threshold of 13.62. In this case 10 voxels would be rejected, but 5 of them would still be in the anterior region discussed above. It is this consistency and the proximity of the selected voxels to each other that suggest that the effect is real. Other significant voxels exist (see Fig. 3), but their being isolated makes it more likely that they are due to random fluctuations. Their proximity to the borders of the mask suggests that they might also be the result of small misalignments in the normalization process.

An inherent limitation of the procedure is that it does not take into account the relationship between neighboring voxels. Important differences in anatomic structure between groups would most likely involve many contiguous voxels simultaneously. Because of its simplicity, it is a great advantage of the FDR procedure that the dependency between voxels can be ignored, but at the same time this is valuable information that is not being used.

Another important issue to consider is that the method described in this paper depends heavily on having reliable normalization. A voxel by voxel comparison is only possible if voxels can be assumed to represent the same anatomic region in every subject. In addition, an inference about fiber directions is only valid if the normalization process does not introduce or remove the effects we are after. In this paper we have used current state-of-the-art normalization methods (SPM parameters in conjunction with the PPD algorithm). However, additional research is necessary to determine the effect of normalization on the inference process.

Having methods for comparing FA and principal direction across populations, the question remains of whether important information in the tensor data exists that is not captured by these two quantities. The diffusion tensor has 6 degrees of freedom. Geometrically, these can be expressed as 3 eigenvalues, corresponding to axis lengths of the associated ellipsoid, and 3 angles, corresponding to the orientation of the orthogonal frame of the ellipsoid. Fixing FA and trace establish one constraint each on the eigenvalues, while fixing the principal direction determines two of the angles. Together, these quantities account for 4 degrees of freedom. The remaining 2 degrees of freedom correspond to the angle that determines the direction of the second eigenvector and an additional constraint on the eigenvalues related to their relative ratios. These last 2 degrees of freedom would then relate to the oblateness of the ellipsoid and the orientation of the principal plane of the ellipsoid. Considering that mean diffusivity is generally uniform over healthy brains, it is possible that FA and principal direction are enough to describe most interesting aspects of DTI data as far as gross anatomy is concerned. Girdle distributions might be necessary, however, to describe variability among subjects in locations in which oblate tensors tend to be found, such as fiber crossings. These are questions for future research.

CONCLUSIONS

We have shown that DTI principal directions provide insight into differences in anatomic structure between populations that may be invisible to statistical tests of FA. As presented in this paper, a formal methodology for comparison of principal directions exists based on the Watson model and false discovery rate theory.

APPENDIX

Computations for the Bipolar Watson Distribution

The following summary simplifies the language used in Ref. (9) (pp. 181, 202, 236–240), with understanding aided by the expositions in Refs. (19) and (20).

Integration Constant

Define a spherical coordinate system on the unit sphere so that the z axis coincides with the vector μ. Then Eq. [2] is the expression of the Watson density in this coordinate system, where θ is the co-latitude angle between x and the z axes. Denote the longitude angle by φ. Integrating the density over the sphere we get

$$1 = \int_0^{2\pi} d\varphi \int_0^{\pi/2} C(\kappa) e^{\kappa \cos^2\theta} \sin\theta = 2\pi C(\kappa) \int_0^1 e^{\kappa t^2}\, dt$$

and so

$$C(\kappa) = \left(2\pi \int_0^1 e^{\kappa t^2}\, dt\right)^{-1}.$$

We are particularly interested in the large concentration case. When κ is large, most of the probability density is concentrated around μ and $\mathbf{x}$ is close to μ with high probability. Using

$$\mathbf{x}^T\mu = 1 - \frac{1}{2}|\mathbf{x} - \mu|^2 \Rightarrow (\mathbf{x}^T\mu)^2 \approx 1 - |\mathbf{x} - \mu|^2, \quad [6]$$

the Watson density given by Eq. [2] may be approximated by

$$f(\pm \mathbf{x};\mu,\kappa) = C(\kappa)\exp(\kappa(\mu^T\mathbf{x})^2) \approx C(\kappa)e^{\kappa}\cdot\exp(-\kappa|\mathbf{x} - \mu|^2). \quad [7]$$

The region of the sphere close to μ looks locally like a 2D plane, and so Eq. [7] is a bivariate normal density with mean μ and covariance $1/2\kappa$ times the 2×2 identity matrix. By equating the multiplying constant with that of the corresponding normal density we obtain an approximation for $C(\kappa)$ given by

$$C(\kappa)e^{\kappa} \sim \frac{1}{2\pi(1/2\kappa)} \Rightarrow C(\kappa) \sim \frac{\kappa}{\pi e^{\kappa}}, \quad \kappa \to \infty. \quad [8]$$

Maximum Likelihood Estimates

Let $\pm\mathbf{x}_1, \ldots, \pm\mathbf{x}_N$ be an independent random sample from the Watson distribution. The log-likelihood function is

$$l(\mu,\kappa; \pm\mathbf{x}_1, \ldots, \pm\mathbf{x}_N) = \kappa\sum_{i=1}^{N}(\mu^{\mathrm{T}}\mathbf{x}_i)^2 + N\log C(\kappa)$$

$$= N[\kappa\mu^{\mathrm{T}}\mathbf{S}\mu + \log C(\kappa)], \quad [9]$$

where $\mathbf{S}$ is the scatter matrix given by Eq. [3]. For $\kappa > 0$, the maximizer $\hat{\mu}$ is the solution to

$$\max_{\mu}\{\mu^T\mathbf{S}\mu\} \quad \text{s.t.} \quad \mu^{\mathrm{T}}\mu = 1,$$

which is given by the eigenvector of S that corresponds to the maximum eigenvalue γ. At the maximum,

$$\hat{\mu}^T\mathbf{S}\hat{\mu} = \hat{\mu}^{\mathrm{T}}\gamma\hat{\mu} = \gamma. \quad [10]$$

Differentiation of Eq. [9] with respect to κ gives $\hat{\mu}^{\mathrm{T}}\mathbf{S}\hat{\mu} = A(\hat{\kappa})$, where

$$A(\kappa) = -\frac{C'(\kappa)}{C(\kappa)} = \frac{\int_0^1 t^2 e^{\kappa t^2}\,dt}{\int_0^1 e^{\kappa t^2}\,dt} \quad [11]$$

and $\hat{\kappa}$ is thus found by solving

$$A(\hat{\kappa}) = \gamma. \quad [12]$$

The function $A(\kappa)$ is monotonically increasing in the range [1/3,1) as κ increases from 0 to ∞. Replacing Eq. [8] in Eq. [11] we obtain the large concentration approximation

$$A(\kappa) \sim 1 - \frac{1}{\kappa}, \quad \kappa \to \infty.$$

By setting the dispersion $s = 1 - \gamma$ in Eq. [12] and using the previous approximation for $A(\kappa)$ we get that at the point of maximum likelihood $s \sim 1/\hat{\kappa}$, which justifies the interpretation of s as a measure of dispersion.

We now obtain an interpretation of s in terms of angle units. Replacing Eq. [3] in Eq. [10] we get

$$\gamma = \hat{\mu}^T\mathbf{S}\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}^{\mathrm{T}}\mathbf{x}_i)\,(\hat{\mu}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}} = \frac{1}{N}\sum_{i=1}^{N}\cos^2\hat{\theta}_i,$$

and so

$$s = 1 - \gamma = \frac{1}{N}\sum_{i=1}^{N}\sin^2\hat{\theta}_i.$$

In other words, s is the average sine-squared of the angles with respect to the mean direction. An interpretation of s in units of angle is obtained thus by computing the quantity $\sin^{-1}(\sqrt{s})$, which we call the angle dispersion.

A Multisample Large Concentration Test

Given q samples of sizes $N_1, \ldots, N_q$, we wish to test H_0: $\pm\mu_1 = \ldots = \pm\mu_q$ against the alternative that at least one of the means is different. For simplicity, we assume that all samples have the same unknown concentration κ.

Consider first the entire sample of size $N = N_1 + \ldots + N_q$, with common mean μ under the null hypothesis and pooled dispersion s. Putting $s = 1 - \gamma$, Eqs. [10] and [6] give

$$2\kappa Ns = 2\kappa N(1 - \hat{\mu}^{\mathrm{T}}\mathbf{S}\hat{\mu})$$

$$= 2\kappa[N - \sum_{i=1}^{N}(\mu^{\mathrm{T}}\mathbf{x}_i)^2] \approx \sum_{i=1}^{N}2\kappa|\mathbf{x}_i - \mu|^2.$$

When μ is known, $2\kappa Ns$ is the sum of N independent approximately χ^2_2 random variables. The estimation of μ reduces 2 degrees of freedom. Thus, under H_0, $2\kappa Ns$ is approximately distributed as χ^2 with $2(N-1)$ degrees of freedom, which we write as

$$2\kappa Ns \mathrel{\underset{H_0}{\dot{\sim}}} \chi^2_{2(N-1)}. \quad [13]$$

For q independent samples of sizes $N_1, \ldots, N_q$ and dispersions $s_1, \ldots, s_q$, $2q$ parameters are fitted and we have the intragroup sum of squares

$$2\kappa\sum_{j=0}^{q}N_j s_j \mathrel{\underset{H_0}{\dot{\sim}}} \chi^2_{2(N-q)}. \quad [14]$$

In the "analysis of variance" decomposition

$$2\kappa Ns = 2\kappa\sum_{j=0}^{q}N_j s_j + 2\kappa\left[Ns - \sum_{j=0}^{q}N_j s_j\right],$$

the asymptotics Eqs. [13] and [14] imply that the second term on the RHS is approximately χ^2 with $2(N-1) - 2(N-q) = 2(q-1)$ degrees of freedom and approximately independent of the first term. The second term represents the intergroup dispersion. Proceeding as in the analysis of variance for normal variables, we construct the test statistic F as the ratio between the intergroup and the intragroup terms divided by the appropriate number of degrees of freedom. Correspondingly, the test statistic is asymptotically F-distributed as

$$F = \frac{[Ns - \sum_{j=0}^{q}N_j s_j]/2(q-1)}{[\sum_{j=0}^{q}N_j s_j]/2(N-q)} \mathrel{\underset{H_0}{\dot{\sim}}} F_{2(q-1),2(N-q)}.$$

ACKNOWLEDGMENTS

The authors thank Brian Wandell for proposing the original research problem, Roland Bammer and Michael Moseley for developing the diffusion imaging techniques and their encouragement to pursue research in DTI, Gayle Deutsch for helping to collect and prepare the data for analysis, Debashis Paul for suggesting the appropriate statistical literature, and the reviewers for their valuable comments.

REFERENCES

1. Basser PJ. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. NMR Biomed 1995;8:333–344.
2. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J Magn Reson B 1996;111:209–219.
3. LeBihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H. Diffusion tensor imaging: concepts and applications. J Magn Reson Imaging 2001;13:534–546.
4. Bammer R, Auer M, Keeling SL, Agustin M, Stables LA, Prokesch RW, Stollberger R, Moseley ME, Fazekas F. Diffusion tensor imaging using single-shot SENSE-EPI. Magn Reson Med 2002;48:128–136.
5. Mori S, Kaufmann WE, Davatzikos C, Stieltjes B, Amodei L, Fredericksen K, Pearlson GD, Melhem ER, Solaiyappan M, Raymond GV, Moser HW, Van Zijl PC. Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. Magn Reson Med 2002; 47:215–223.
6. Bammer R, Augustin M, StrasserFuchs S, Seifert T, Kapeller P, Stollberger R, Ebner F, Hartung HP, Fazekas F. Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis. Magn Reson Med 2000;44:583–591.
7. Klingberg T, Hedehus M, Temple E, Salz T, Gabrieli JD, Moseley ME, Poldrack RA. Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. Neuron 2000;25:493–500.
8. Deutsch GK, Dougherty RF, Bammer R, Siok WT, Gabrieli JD, Wandell B. Correlations between white matter microstructure and reading performance in children. Cortex 2005;41:354–363.
9. Mardia KV, Jupp PE, Directional statistics. West Sussex, England: Wiley; 2000.
10. Genovese CR, Lazar NA, Nichols TE. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 2002;15:870–878.
11. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995; 57:289–300.
12. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J R Stat Soc B 2004;66:187–205.
13. Friston KJ, Ashburner J, Poline JB, Frith CD, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. Hum Brain Mapp 1995;2:165–189.
14. Friston KJ, Ashburner J. Nonlinear spatial normalization using basis functions. Hum Brain Mapp 1999;7:254–266.
15. Pajevic S, Aldroubi A, Basser PJ. A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue. J Magn Reson 2002;154:85–100.
16. Alexander DC, Pierpaoli C, Basser PJ, Gee JC. Spatial transformations of diffusion tensor magnetic resonance images. IEEE Trans Med Imaging 2001;20:1131–1139.
17. Friston KJ, Ashburner J. Multimodal image coregistration and partitioning: a unified framework. Neuroimage 1997;6:209–217.
18. Robichon, Habib. Abnormal callosal morphology in male adult dyslexics: relationships to handedness and phonological abilities. Brain Lang 1998;62:127–146.
19. Watson GS. Equatorial distributions on a sphere. Biometrika 1965;52: 193–201.
20. Best DJ, Fisher NI. Goodness-of-fit and discordancy tests for samples from the Watson distribution on the sphere. Austral J Stat 1986;28:13–31.