Published in final edited form as:

Psychiatry Res. 2011 September 30; 193(3): 191–198. doi:10.1016/j.psychres.2011.03.005.

A method to enhance the sensitivity of DTI analyses to group differences: a validation study with comparison to voxelwise analyses

Matthew D. Cykowski^{a,b,*}, Jack L. Lancaster^a, and Peter T. Fox^{a,c,d}

^aResearch Imaging Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX

^bDepartment of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK

^cVA Medical Center, San Antonio, Texas

^dDepartment of Radiology, University of Texas Health Science Center at San Antonio, San Antonio, TX

Abstract

Studies of white matter (WM) abnormalities in psychiatric and neurological disorders often use the analysis package Tract-Based Spatial Statistics (TBSS). However, with small samples and/or subtle effects, a study using the standard TBSS approach can be underpowered. For such cases, a new method is presented that summarizes *global* differences between TBSS-derived fractional anisotropy (FA) images with a single paired t-statistic, estimating the degrees of freedom using spatial autocorrelation. The sensitivity of the method is demonstrated by using well-known aging effects on FA as a proxy for disease effects. Sixty healthy subjects were divided equally into younger- (YA), middle- (MA), and older-aged (OA) groups and significant global differences were demonstrated in the YA vs. OA (all $N \geq 4$, FA difference ≈ 0.023), MA vs. OA (all $N \geq 4$, FA difference ≈ 0.017), and YA vs. MA (FA difference ≈ 0.005 at $N = 20$) comparisons. In contrast, no significant difference could be detected in the YA vs. MA comparison using voxelwise TBSS analysis with the full sample ($N = 20$ per group). This method should facilitate localizing analyses in the direction of a proven group difference while providing clinically relevant information about pathophysiologic processes globally affecting WM.

Keywords

spatial autocorrelation; fractional anisotropy; Tract-Based Spatial Statistics; diffusion tensor imaging

© 2011 Elsevier Ireland Ltd. All rights reserved.

Address for the corresponding author: Matthew D. Cykowski, M.D., Department of Pathology, The University of Oklahoma Health Sciences Center, 940 Stanton L. Young Blvd, Oklahoma City, OK 73104, matthew-cykowski@ouhsc.edu; Phone: 405-271-2422.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. INTRODUCTION

The *in vivo* detection of white matter (WM) abnormalities in psychiatric and neurological diseases has become an important aim of imaging neuroscience. The purpose of the present study is to demonstrate a new method that adds to the growing armamentarium of tools for detecting such abnormalities in diffusion tensor imaging (DTI) studies (Buffon et al., 2005; Smith et al., 2006; Pfefferbaum et al., 2009; Cykowski et al., 2010). The major goal of the method presented here is to enhance the sensitivity of case-control studies to disease effects on WM, particularly when such studies have limited power. The method uses a paired *t*-statistic to summarize the global differences between Tract-Based Spatial Statistics (TBSS) skeleton images (Smith et al., 2006) with the degrees of freedom (DOF) of the statistic estimated through spatial autocorrelation. The method is therefore referred to as the “Paired Evaluation of Skeleton Changes using Autocorrelation”, or “PESCA”.

The hypothesis underlying the development of PESCA is that in certain scenarios real WM changes can go undetected with corrected, whole brain voxelwise analyses. This includes scenarios wherein the sample size and/ or effect sizes are small (e.g., disorders with a low prevalence; pilot studies with limited funding). This also includes scenarios where significant effects at the global level are due to subtle and diffuse alterations that cannot be localized with conventional voxelwise analyses (Buffon et al., 2005). The current work presupposes that through the detection of *global* tissue differences between two groups (patient/ control) using PESCA, there is sufficient *a priori* justification to localize effects using a variety of alternative analysis strategies. For instance, a significant global difference detected by PESCA would facilitate more powerful, one-tailed voxelwise analyses (with correction) in the direction of the global difference, or allow a more limited tract-wise comparison using informed regions-of-interest (ROIs) (Mori et al., 2008). Additionally, the assessment of global changes and their impact on cognitive or functional impairment, independent of voxelwise findings, has been an important component of several DTI studies (Bednarczyk et al., 1998; Molko et al., 2002; Ragin et al., 2004). PESCA provides a method to assess such global differences in a robust and replicable way across studies.

The PESCA approach draws on two important concepts to enhance the sensitivity of DTI studies to global WM differences. The *first concept* is that of spatial (*not* within-subject) pairing as the basis for calculating a paired *t*-statistic on the mean difference between TBSS-derived images (e.g., FA skeletons). This is conceptually related to the pairing of data by features of environmental variation (Snedecor and Cochran, 1967) as used in a paired-site sampling design (Young et al., 2002). Specifically, this method is motivated by the hypothesis that these images, once aligned spatially to an invariant TBSS skeleton (Smith et al., 2006), exhibit spatial pairing at homologous sites. The non-independence of FA values at spatially aligned points along the TBSS skeleton (to be demonstrated below) suggests that a paired approach is therefore more appropriate and powerful than an unpaired approach (Zar, 1996) (Figure 1A). The *second concept* is that the number of *uncorrelated components* within an FA image should function as the degrees of freedom (DOF) for the paired *t*-test statistic. This requires accounting for spatial autocorrelation within the data used in significance testing (Legendre, 1993), an approach which results in a corrected (reduced) estimate of the degrees of freedom (Clifford et al., 1989). Therefore, an important feature of PESCA is the use of three-dimensional (3D) spatial autocorrelation to determine the number of uncorrelated components (i.e., the DOF) in TBSS FA skeleton images.

The present study utilized PESCA to detect group differences in the FA data of younger- (YA), middle- (MA), and older-aged (OA) subjects. The TBSS-derived fractional anisotropy (FA) maps used in the study were acquired in 60 healthy adult subjects between 27 and 76 years old (divided into three equal YA, MA, and OA groups). In this way, the subtle

reduction of FA with normal aging (Pfefferbaum and Sullivan, 2003) was utilized as a proxy for unknown disease effects on FA in hypothetical case-control studies of psychiatric or neurological disorders. We gauged the sensitivity of PESCA for detecting age-related FA differences between all three groups and at a variety of sample sizes ($N = 4-20$). In all cases, the PESCA results were compared to those that would be obtained through unpaired testing on the same data. Further, standard voxelwise TBSS analyses were performed on the age groups using a highly sensitive multiple comparison correction (Smith and Nichols, 2009) to determine whether PESCA was even necessary. The present study also included several validation measures for critical components of PESCA (e.g., the use of spatial autocorrelation to estimate the DOF). Finally, we note that although the analysis and validation steps are described in detail below, PESCA has been fully implemented in Matlab (MathWorks, 2007) so that an investigator would only need to supply preprocessed TBSS skeleton data (for both groups) in order to generate the statistic summarizing global group differences.

2. METHODS

2.1. Participants

Sixty healthy individuals (39 females, 21 males) were participants in the study (mean age = 49.1, $sd = 14.4$). These consistently right-handed (Oldfield, 1971) participants were divided into 3 subgroups of 20 participants each (13 F, 7 M per group). These include YA (mean age = 31.9, $sd = 3.3$, range = 27–38), MA (mean age = 49.9, $sd = 2.7$, range = 44–55), and OA (mean age = 65.6, $sd = 4.6$, range = 57–76) groups. Imaging was performed at the Research Imaging Institute with the informed written consent of the participants and the approval of the Institutional Review Board at the University of Texas Health Science Center at San Antonio.

2.2. Image acquisition and DTI preprocessing

All imaging was done using a 3-Tesla Siemens scanner with an eight-channel head coil. A single-shot, spin echo EPI sequence ($TR = 7800$ ms, $TE = 88$ ms) was used to obtain diffusion-weighted images with the following parameters: in-plane resolution of 1.7×1.7 mm; slice thickness of 3.0 mm; diffusion directions = 55 with b -value = 700 sec/mm^2 ; three reference images with b -value = 0. The image matrix was 128×128 with 50 slices.

The FMRIB Diffusion Toolbox (Smith et al., 2004) was used to align diffusion-weighted and reference images, correct for eddy current distortion, perform the tensor fitting at each voxel, and produce FA images. The TBSS analysis package was used to perform nonlinear and affine alignment of FA images to a high-resolution, MNI-152 standard space image at 1 mm isotropic resolution (Smith et al., 2006). A study-specific mean FA skeleton image was created by skeletonizing the average of all 60 aligned FA images and thresholding the output at $FA = 0.3$. Individual FA skeleton images (FA_{indiv}) were generated by projecting nearby maximum FA values onto the mean FA skeleton (Figure 2 presents the mean values of FA_{indiv} by age). The remaining processing steps were implemented in Matlab® (MathWorks, 2007).

2.3. PESCA: a method for detecting global differences in FA skeleton images

Summarizing the PESCA method briefly (detail is provided in sections 2.3.1 – 2.3.5), spatially paired FA skeletons are first averaged to form groups of interest and an FA skeleton difference image (FA_{diff}) is generated as the difference between the group average images (e.g., $FA_{\text{group1}} - FA_{\text{group2}}$) (see Figure 1B). The histogram of the FA_{diff} image is converted to a probability density function from which the mean and standard deviation of FA_{diff} are estimated. To estimate the degrees of freedom (DOF) in the test, three-

dimensional (3D) spatial autocorrelation is utilized to determine the voxel size of *correlated* components in the input data (e.g., FA_{group}). The actual number of voxels in the TBSS skeleton ($> 1 \times 10^5$) is then divided by this correlated component size to return the number of *uncorrelated* components (i.e., the DOF). A paired *t*-statistic is then calculated as the mean of FA_{diff} divided by the standard error of FA_{diff} .

2.3.1. Generating FA Skeleton difference images (FA_{diff})—The exact alignment of TBSS-derived FA skeletons allowed group-specific average FA skeleton images (FA_{group} for $N = 20$ or FA_{subgroup} for $N = 1 - 10$) to be formed. From these, pairwise FA skeleton difference images (FA_{diff}) were generated by subtractions in these group or subgroup images (e.g., $FA_{\text{group1}} - FA_{\text{group2}}$) (Figure 1B).

As a validation step, *within-age group analyses* were performed to observe whether the standard deviation of FA_{diff} by input sample size (N) matched that expected from theory ($s_N = s_1 / \sqrt{N}$) (Zar, 1996). The FA_{diff} were generated by paired subtractions of within-group subgroup average images (FA_{subgroup}) generated by random permutation for $N = 1$ to $N = 10$. For example, the FA_{diff} at $N = 10$ in the YA group were formed as the paired subtraction of the averages (FA_{subgroup}) of two sets of 10 randomly drawn, YA FA_{indiv} (Figure 1B, *middle panel*, ‘Young – young’).

To gauge the sensitivity of PESCA to actual group differences, *between-age group analyses* were performed using the same approach at $N = 4, 6, 8, 10$, and 20. Specifically, FA_{diff} were generated as described above by subtracting randomly generated FA_{subgroup} data at $N = 4, 6, 8$, and 10, for YA versus MA, MA versus OA, and YA versus OA groups (Figure 1B, *right panel*, ‘Young – older subjects’). Three between-age group comparisons were also performed using the entire set of data per group (FA_{group} where $N = 20$).

2.3.2. Estimating the distribution parameters of FA_{diff} —An FA_{diff} histogram was calculated for each of the paired subtractions described in 2.3.1 (see also Figure 3A). Individual histograms were utilized to make a single composite average histogram for each analysis of interest (for example, histograms from randomly generated YA versus OA FA_{diff} at $N = 8$ were averaged). The composite average histograms were normalized to generate probability density functions from which distribution parameters were estimated. The same procedure to convert the distribution of FA_{diff} to a probability density function was repeated for group analyses at $N = 20$ (permutations and average histograms were not necessary with the full sample).

2.3.3. Estimating the degrees of freedom (DOF)—The DOF was estimated as the number of *uncorrelated components* in the entire skeleton as follows: $DOF = FA_{\text{indiv, subgroup, or group volume}} / \text{correlated } FA_{\text{indiv, subgroup, or group volume}}$. To obtain the correlated volume (e.g., for any $FA_{\text{indiv, subgroup, or group}}$), a fast Fourier transform of the image was taken and multiplied by its complex conjugate to generate a power spectrum (images were first forced to be zero-mean). The inverse fast Fourier transform of the power spectrum was then applied (Castleman, 1996). This resulted in a 3D spatial image of the autocorrelation function that peaked at zero displacement. Details of the characteristics of the autocorrelation functions observed are presented in Appendix Note 1. Details of the method used to estimate correlated volume of images from the autocorrelation function are presented in Appendix Note 2.

2.3.4. Validation of the PESCA autocorrelation method—Several validation measures were carried out to ensure that the 3D spatial autocorrelation approach was appropriate. First, a zero-mean input image (whether FA_{indiv} , FA_{subgroup} , or FA_{group}) was run through the 3D autocorrelation steps and it was verified that the peak of its

autocorrelation function at zero displacement matched the variance of the input image. Second, a “noise skeleton” was generated (TBSS skeleton image filled with random, uncorrelated noise) and it was verified that the 3D autocorrelation function resulted in a *correlated* volume of 1. In other words, in the absence of any correlation between voxels, the DOF (skeleton volume/ correlated volume) determined through 3D autocorrelation were equal to the number of voxels in the $1 \times 1 \times 1$ mm “noise skeleton”. Lastly, a “smoothed noise skeleton” was generated to artificially introduce correlation into the “noise skeleton” using a 5 mm FWHM Gaussian filter ($11 \times 11 \times 11$). Running the autocorrelation steps of PESCA on this artificially correlated input produced an autocorrelation function whose profile was the square of the point spread function in the “smoothed noise skeleton”. In other words, the PESCA autocorrelation step had recovered the exact size of the *correlated* volume (the size of the 5 mm FWHM Gaussian filter) in the “smoothed noise skeleton”. These results demonstrated the 3D spatial autocorrelation approach of PESCA was suitable for usage with TBSS skeleton data.

2.3.5. Paired t-statistic in PESCA: comparison with unpaired testing—The PESCA method was applied with group sizes of 4, 6, 8, 10, and 20 for YA versus MA, MA versus OA, and YA versus OA comparisons. Each of these PESCA comparisons was a simulation of a case-control scenario at a given N and certain effect size. In actual practice, only one such test would be performed with PESCA, taking advantage of all the subjects available. For each comparison, utilizing the relevant DOF, a paired *t*-statistic was calculated as the mean of FA_{diff} divided by the standard error of FA_{diff} .

These results were compared to the simpler approach of performing unpaired *t*-tests on mean FA values, which used the number of subjects to determine the DOF. The mean of FA_{indiv} were taken and the grand mean across these values was then calculated for N = 4, 6, 8, 10, and 20 per aging group. For samples < 20, permutation was used to randomly select N values from the 20 mean values available per group. The mean of standard deviations across resamples for a particular group and N was also calculated. The unpaired *t*-statistic was calculated for all age group comparisons at the relevant N assuming unequal variances (Zar, 1996).

2.4. Standard voxelwise TBSS analyses

To test whether a global assessment of FA differences extended conventional testing, whole brain voxelwise analyses were carried out using the general linear model implemented in FSL Randomise (www.fmrib.ox.ac.uk/fsl/randomise). This was done in three separate analyses in order to examine the sensitivity of this commonly used approach. In the first analysis, a simple voxelwise regression was run between age and FA values using the full complement of subjects available ($n = 60$). In the second analysis, the YA and OA groups were compared using a two-tailed, unpaired *t*-test at each voxel. Likewise, for the third analysis, the YA and MA groups were compared using a two-tailed unpaired *t*-test at each voxel. These results were corrected using the most sensitive procedure currently available in standard TBSS processing: the implementation of the threshold-free cluster enhancement algorithm (Smith and Nichols, 2009) with 2D optimization (www.fmrib.ox.ac.uk/fsl/tbss). Voxels where the two-tailed, TFCE-corrected $P \leq 0.05$ were considered significant.

3. RESULTS

3.1. Degrees of freedom (DOF) in the paired test

3.1.1. Correlation of FA tract skeleton images—A large, positive distribution of correlation coefficients was seen between different FA_{indiv} 3D skeleton images (median value of $r = 0.68$; maximum $r = 0.77$, minimum $r = 0.49$) (Figure 4A). Likewise, a

significant linear trend was seen ($r = 0.83$) for the voxelwise FA values of a 2D slice between two randomly drawn FA_{indiv} (Figure 4B). This substantial spatial pairing across subjects supports the rationale for using paired and not unpaired testing (Snedecor and Cochran, 1967; Zar, 1996).

3.1.2. Variation of the DOF by sample size and across age groups—Figure 5A displays the DOF for all 60 FA_{indiv} using an autocorrelation function threshold method (see Appendix Note 3). To the far right of the plot, an asterisk marks the median value and a bar the range of values. The estimated DOF did not vary substantially across FA_{indiv} (Figure 5A) in spite of age differences. Neither was the estimated DOF affected by increased intersubject averaging beyond $N = 8$ (Figure 5B). This result is also presented in Table 1A, which lists the estimated DOF at a 10 % threshold of the autocorrelation function and at each N (the median of DOF values from permutation).

3.2. FA distribution parameters and t-test results

Table 1 summarizes the between-group distribution parameters and the performance of the paired and unpaired testing approaches. The standard deviation for the FA_{diff} decreased with increasing N , in a manner consistent with theory, where $s_N = s_1 / \sqrt{N}$ (Figure 3B). Table 1 also shows a drop in standard deviation with group size (3rd column) for paired (A) versus unpaired (B) methods.

Paired t -tests on the FA_{diff} were significant between all three age groups, though not at all N . The YA versus OA group comparison had the largest FA difference range of 0.021–0.024, which was significant for both one- and two-tailed testing at $N \geq 4$ (Table 1A). The MA versus OA group comparison had a slightly reduced FA difference range of 0.016–0.019, but this was also significant at $N \geq 4$ (both one and two-tailed testing). The YA versus MA group comparison had the smallest FA difference range of 0.005–0.006, which was significant in one-tailed testing at $N = 8$ and in one- and two-tailed testing at $N = 20$. The DOF in the paired approach (Table 1A) were substantially greater than those in the unpaired approach (Table 1B) at all N . Unpaired t -tests were considerably less sensitive to group differences in the FA_{diff} (see bold P -values in Table 1). The unpaired approach was not able to detect significant differences in the YA versus OA group comparison at $N = 4$ or in the MA versus OA group comparison at $N = 4$ and $N = 6$ (two-tailed testing). Nor could the unpaired approach detect significant differences in the YA versus MA group comparison at $N = 20$.

3.3. Whole brain, voxelwise analyses

The regression of age and FA ($n = 60$) revealed a number of WM tracts with significant voxels (Figure 6, *panel A*, $P \leq 0.05$ shown as a blue overlay, $P \leq 0.002$ shown as a green overlay). An unpaired t -test between the FA values of the YA ($n = 20$) and OA ($n = 20$) groups (Figure 6, *panel B*, $P \leq 0.05$, blue overlay) produced voxelwise findings that had a very similar spatial distribution to the association of age and FA across all 60 subjects (Figure 6, *panel A*).

In contrast, the standard TBSS analyses did not reveal any significant age effects in the comparison of YA and MA groups at $N = 40$. The age-associated FA changes demonstrated in the simple regression (Figure 6, *panel A*) were not demonstrable until the TFCE-corrected results of the YA and MA subject comparison were lowered to a threshold of $P \leq 0.4$ (Figure 6, *panel B*, red overlay onto the YA versus OA results).

4. DISCUSSION

The major result of the present study was that the PESCA method detected a global tissue difference in the YA versus MA comparison ($N = 20$, Table 1). A comparison using the same two groups and unpaired testing on mean FA values did not detect a significant global change. Likewise, a voxelwise TBSS analysis of the same two groups with two-tailed, TFCE correction found no significant effects even with a sample size of 40. This suggests that the PESCA method was the most sensitive to YA versus MA changes, even if not specific to the voxels contributing to that significant difference (Figure 7 demonstrates one approach to isolating such tracts/ voxels). This improved sensitivity is noteworthy because the YA versus MA comparison is likely the best proxy in this study for an actual DTI investigation of patients with a psychiatric or neurological condition matched with healthy controls (with obvious exceptions such as multiple sclerosis, dysmyelination syndromes, and stroke).

To provide spatial specificity, we examined the WM areas contributing to this PESCA-detected global difference by generating a voxelwise map of standardized effect sizes for the YA versus MA comparison. A threshold of Cohen's d of > 0.8 ("large" effects) was applied to the map and these suprathreshold voxels are shown in Figure 7. This revealed three important features that make the PESCA-detected difference between YA and MA groups highly plausible. The first feature is that these large effect size voxels are anatomically contiguous (i.e., they are not randomly distributed and thus suggest regional, tract-based FA differences). The second feature is that these suprathreshold voxels are in regions with a known age correlation in the entire sample (compare Figure 7, yellow overlay, with Figure 6A). These tracts include the cingulum, body and genu of the corpus callosum, forceps minor, and bilateral external capsules. The third feature that makes this YA versus MA difference plausible is that several of these large effect regions are consistent with the DTI literature on FA changes in young adulthood to middle age. For instance, using a limited set of ROI analyses, Yoon and colleagues reported that the cingulum, as seen in Figure 7, demonstrates the most rapid drop in FA between the third and fourth decades of life (Yoon et al., 2008). Comparing YA and MA subjects with similar group mean ages to the present study, Salat and colleagues found in uncorrected whole brain analyses that FA was significantly higher in the YA group in the genu and minor forceps (see Figure 5B of that study) (Salat et al., 2005). Again, in an investigation blinded to these effects, and proceeding directly to unconstrained, voxelwise analyses, it is likely that no significant effect would be reported if corrected comparisons were utilized.

PESCA also detected a global difference in all YA versus OA comparisons where $N \geq 4$. With such large effects, but small sample sizes in a hypothetical case-control pilot study (e.g., $N \leq 8-10$ per group), PESCA could therefore have an important role in a multi-step analysis. With the full sample, standard voxelwise analyses for the YA versus OA comparison (see Figure 6B) revealed WM changes consistent with the literature. This includes age-related reductions in the external capsule and gyral WM (Bhagat and Beaulieu, 2004), as well as the genu/ splenium of the corpus callosum and frontal/ parietal WM of the centrum semiovale (Pfefferbaum and Sullivan, 2003). Therefore, it could be argued that with a sufficient number of subjects ($N = 40$ here) and equally large effects (global FA difference of 0.024), PESCA would be less essential. Many DTI case-control studies, however, are more likely to deal with small-to-medium sized effects (YA versus MA) and large, homogenous groups with specific disease characteristics of interest are difficult to recruit. The detection of these smaller-to-medium WM differences more fully realizes the potential of diffusion imaging and PESCA is intended to enhance this sensitivity.

A second important result was that the PESCA method was more sensitive to detecting global FA changes than a simpler procedure of unpaired testing on mean FA values. This was true for all N and in all age comparisons. To detect significant differences with two-tailed *unpaired* tests required larger group sizes for all comparisons. Further, as judged by the probability values, a larger N was also required using unpaired testing to attain the same significance level found at a smaller N with paired tests. For instance, the level of significance found in the YA versus OA paired test at N = 4 was not seen in unpaired testing until N = 10 (Table 1). The paired testing of PESCA also uncovered significant YA versus MA global differences with the full sample that unpaired testing could not. These results demonstrate that the paired design is indeed effective for this data (Proschan, 1996) due to the substantial spatial correlation across TBSS-derived FA images (Figures 4A and 4B). This effective pairing along with the higher DOF it affords is why PESCA is more powerful than an unpaired approach (Zar, 1996).

One useful application of PESCA, to facilitate localizing steps at the discretion of the investigator, would be to generate an omnibus statistic as has been done in functional neuroimaging (Fox and Mintun, 1989). This should be helpful for small samples with larger effect sizes (here, the YA versus OA comparison), as well as for small effects even with relatively large samples in neuroimaging studies (here, the YA versus MA comparison). A significant PESCA omnibus statistic would justify an informed ROI-based comparison using well-defined tracts that reduces the number of comparison. For instance, the ICBM-81 atlas could be utilized to perform tract-wise (50 tracts are defined) (Mori et al., 2008) rather than voxelwise comparisons. A combination of the PESCA method with a ROI-based localizing approach, corrected for the number of ROI comparisons, should be preferable to uncorrected, voxelwise analyses. A review of the effect sizes detected here (see Table 1) might be helpful in designing such a test using PESCA.

A second application of PESCA would be to simply obtain information about the direction of changes in DTI parameters between two sets of data (e.g., globally reduced/elevated FA). A detection of significant global change would facilitate one-tailed voxelwise testing in the direction of the proven global difference. Alternately, confounds could be identified in cases where no global differences are expected (e.g., due to variations in data acquisition parameters). Pooling FA data from across scanning sites or even imaging protocols would be a logical example of this usage of PESCA.

Finally, PESCA might provide important information on global rather than regional tissue change. Global differences in DTI parameters could be evaluated in order to examine the relationship between global change and clinically relevant variables. This type of testing has proved useful in assessing DTI markers of HIV-related cognitive changes (Ragin et al., 2004) and in ischemic stroke (Molko et al., 2002).

In summary, PESCA was designed to provide a sensitive test for group-wise differences in DTI parameters. This study demonstrated that (a) PESCA is more sensitive than unpaired testing, (b) PESCA is capable of detecting plausible global differences where conventional TBSS analyses failed to identify voxelwise differences, and (c) the DOF in TBSS-derived FA image data varies little across “disease groups” (Table 1), across individuals (Figure 5A), or by the extent of inter-subject averaging (Figure 5B). The PESCA approach thus gains *specificity* from the alignment stages of standard TBSS processing and gains *sensitivity* from the higher DOF seen for TBSS skeletons. Though FA images were utilized in this study, other DTI parameter images (e.g., mean diffusivity images) could be used with sufficient inter-subject alignment.

Acknowledgments

Funding was provided by the National Institute on Deafness and other Communication Disorders/ NIH (F32-DC009116 and Loan Repayment Award to M.D.C.), the National Institute of Mental Health (R01-MH078143 to David Glahn, PI, who generously allowed access to subject DTI data), and the National Center for Research Resources grant for the Frederic C. Bartter General Clinical Research Center (M01 RR001346).

APPENDICES

Note 1: Features of the 3D autocorrelation function of FA tract skeletons

The peak location of the 3D FA tract skeleton autocorrelation functions represents the zero displacement point and the peak value is proportional to the variance in FA values. Autocorrelation functions for FA tract skeleton images drop off rapidly with distance, reaching a level of less than 1 % of the peak at a displacement of $\cong 25$ mm. These autocorrelation functions were radially symmetric about their peak, but extent varied by direction and did not conform in shape with common analytical functions (e.g. Gaussian). There were three major trends in shape beginning with a rapid drop off with small displacements (< 5 mm), followed by an intermediate drop (5–20 mm), and then leveling off with distances > 20 –25 mm.

Note 2: Estimating the DOF using an “equivalent correlated volume” method

The correlated volume was assumed to be proportional to the volume extent of the autocorrelation function. An “equivalent correlated volume” method was used to objectively estimate the DOF in FA tract skeletons. The first step was to determine a spatial range for the analysis of the FA skeleton autocorrelation function. This was done using a threshold of 1–1.5% of the peak. (Threshold values below these levels led to multiple, disconnected regions.) The purpose of the threshold was to form a single bounded ROI around the central portion of the autocorrelation function that best represented local positive spatial correlation. The volume of the autocorrelation function within this ROI was calculated. The size of a region, with correlation fixed at the peak value, but with the same total volume, was then determined. The volume of this *fixed correlation region* was used as an objective estimate for “correlated” size. This approach provides an equivalent volume at 100% correlation. This equivalent correlated volume, when divided into the volume of a FA skeleton, provides an objective estimate of the degrees of freedom for the skeleton.

Note 3: Estimating the DOF by thresholding the autocorrelation function

A simpler method than the “equivalent correlated volume” approach (see Note 2) was sought to estimate the DOF. Different fractions of the peak value of the autocorrelation function were used to create varying volumes of interest. Specifically, the correlated skeleton volume was initially taken at a specified percentage (5, 10, 15, and 20 %) of the maximum value in the autocorrelation image. These results were then compared to an estimate of the correlated volume in a subset of FA_{indiv} by the “equivalent correlated volume” method. The results from the thresholding method indicated that the 10 % threshold of the autocorrelation function provided the most robust estimate of the DOF. The results (at a 10 % threshold) were also highly correlated with the “equivalent correlated volume” method ($r = 0.99$, $P \leq 0.001$, $N = 9$). Therefore, we adopted this computationally simpler approach for determining DOF.

REFERENCES

- Bednarczyk EM, Remler B, Weikart C, Nelson AD, Reed RC. Global cerebral blood flow, blood volume, and oxygen metabolism in patients with migraine headache. *Neurology*. 1998; 50:1736–1740. [PubMed: 9633719]
- Bhagat YA, Beaulieu C. Diffusion anisotropy in subcortical white matter and cortical gray matter: changes with aging and the role of CSF-suppression. *Journal of Magnetic Resonance Imaging*. 2004; 20:216–227. [PubMed: 15269946]
- Buffon F, Molko N, Herve D, Porcher R, Denghien I, Pappata S, Le Bihan D, Bousser MG, Chabriat H. Longitudinal diffusion changes in cerebral hemispheres after MCA infarcts. *Journal of Cerebral Blood Flow and Metabolism*. 2005; 25:641–650. [PubMed: 15689956]
- Castleman, KR. Digital image processing. Englewood Cliffs, N.J.: Prentice Hall; 1996.
- Clifford P, Richardson S, Hemon D. Assessing the significance of the correlation between two spatial processes. *Biometrics*. 1989; 45:123–134. [PubMed: 2720048]
- Cykowski MD, Fox PT, Ingham RJ, Ingham JC, Robin DA. A study of the reproducibility and etiology of diffusion anisotropy differences in developmental stuttering: a potential role for impaired myelination. *Neuroimage*. 2010 (in press).
- Fox PT, Mintun MA. Noninvasive Functional Brain Mapping by Change-Distribution Analysis of Averaged Pet Images of (H₂O)-O-15 Tissue Activity. *Journal of Nuclear Medicine*. 1989; 30:141–149. [PubMed: 2786929]
- Legendre P. Spatial autocorrelation: Trouble or new paradigm? *Ecology*. 1993; 74:1659–1673.
- MathWorks. Matlab: version R2007b. Natick, Massachusetts: The MathWorks, Inc.; 2007.
- Molko N, Pappata S, Mangin JF, Poupon F, LeBihan D, Bousser MG, Chabriat H. Monitoring disease progression in CADASIL with diffusion magnetic resonance imaging: a study with whole brain histogram analysis. *Stroke*. 2002; 33:2902–2908. [PubMed: 12468789]
- Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, Hua K, Faria AV, Mahmood A, Woods R, Toga AW, Pike GB, Neto PR, Evans A, Zhang J, Huang H, Miller MI, van Zijl P, Mazziotta J. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage*. 2008; 40:570–582. [PubMed: 18255316]
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971; 9:97–113. [PubMed: 5146491]
- Pfefferbaum A, Rosenbloom M, Rohlfing T, Sullivan EV. Degradation of association and projection white matter systems in alcoholism detected with quantitative fiber tracking. *Biological Psychiatry*. 2009; 65:680–690. [PubMed: 19103436]
- Pfefferbaum A, Sullivan EV. Increased brain white matter diffusivity in normal adult aging: Relationship to anisotropy and partial voluming. *Magnetic Resonance in Medicine*. 2003; 49:953–961. [PubMed: 12704779]
- Proschan MA. On the distribution of the unpaired t-statistic with paired data. *Statistics in Medicine*. 1996; 15:1059–1063. [PubMed: 8783442]
- Ragin AB, Storey P, Cohen BA, Epstein LG, Edelman RR. Whole brain diffusion tensor imaging in HIV-associated cognitive impairment. *American Journal of Neuroradiology*. 2004; 25:195–200. [PubMed: 14970017]
- Salat DH, Tuch DS, Greve DN, van der Kouwe AJ, Hevelone ND, Zaleta AK, Rosen BR, Fischl B, Corkin S, Rosas HD, Dale AM. Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiology of Aging*. 2005; 26:1215–1227. [PubMed: 15917106]
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TE. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*. 2006; 31:1487–1505. [PubMed: 16624579]
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004; 23 Supplement 1:S208–S219. [PubMed: 15501092]

- Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*. 2009; 44:83–98. [PubMed: 18501637]
- Snedecor, GW.; Cochran, WG. *Statistical methods*. 6th ed.. Ames, Iowa: Iowa State University Press; 1967.
- Yoon B, Shim YS, Lee KS, Shon YM, Yang DW. Region-specific changes of cerebral white matter during normal aging: a diffusion-tensor analysis. *Archives of Gerontology and Geriatrics*. 2008; 47:129–138. [PubMed: 17764763]
- Young JA, Smith DR, Snyder CD, Lemarie DP. A terrain-based paired-site sampling design to assess biodiversity losses from eastern hemlock decline. *Environmental Monitoring and Assessment*. 2002; 76:167–183. [PubMed: 12108590]
- Zar, JH. *Biostatistical analysis*. 3rd ed.. Upper Saddle River, N.J.: Prentice Hall; 1996.

FIGURE 1.

(A) Paired versus unpaired testing. An FA skeleton is shown with 14 randomly selected voxels selected (red markers). The FA values for these 14 points are shown in the 14×3 matrix to the right with the white and green arrows indicating corresponding points for the skeleton and data matrix. The column labeled "s1" is the data from the FA_{indiv} of a 27 year-old subject. The column labeled "s2" is the data from the FA_{indiv} of a 53 year-old subject. The third column is the FA difference between these spatially aligned points (the mean of this FA difference column is 0.036). A paired t-statistic is substantially higher (paired $t = 1.41$) than an unpaired t-statistic calculated from the same data (unpaired $t = 0.62$).

(B) FA difference images. FA skeleton difference images (FA_{diff}) for intra-subject subtraction (*left panel*) following repeat DTI acquisitions, within-group subtractions (*middle panel*) for younger – younger age subjects, and between-group subtractions (*right panel*) for younger – older age groups. For the middle and right panel, each FA_{diff} image represents only one of an entire permutations set for that comparison at $N = 10$. The intra-subject subtraction (*left panel*) from the same subject over 2 scanning sessions reveals only random differences (e.g., due to head motion, pulsation artifact, and/or noise). In contrast, within- (*middle*) and between-group (*right*) subtractions exhibit tract-wise (spatially correlated) differences.

FIGURE 2. Mean FA of the individual FA skeletons (FA_{indiv})

Mean FA of the 60 individual FA tract skeletons (FA_{indiv}) for the younger- (circles), middle- (inverted triangles), and older-aged (squares) groups. The nonlinear correlation between age and mean FA was significant ($\eta^2 = 0.75$, $P < 0.002$). The fitted model is shown on the plot.

FIGURE 3.

(A) FA difference histogram. Distribution of FA_{diff} values for middle-aged subject, within-group subtractions. The plot of a single FA difference image is shown as individual points for $N = 4$. The solid line represents the composite histogram for $N = 100$ permutations at $N = 4$.

(B) Effect of sample size on standard deviation. Observed standard deviations of histograms calculated from difference images as a function of the number of images/subjects for middle-middle (squares), young-young (triangles), and older-older (circles)

subtractions. For all three groups the relationship shown follows sampling theory (i.e., $s_N = s_1 / \sqrt{N}$).

FIGURE 4.

(A) Whole brain correlation between individual FA skeletons. Distribution of Pearson linear correlation coefficients, generated from 1000 pairs drawn by a bootstrap procedure (permutation with replacement) from the 60 FA_{indiv} used in the study.

(B) Within-slice correlation for two subjects. Scatterplot of FA skeleton values in an axial section through the genu and splenium of the corpus callosum (MNI space $z = 21$) for two randomly drawn subjects (one younger- and one older-aged subject). The linear correlation coefficient for this data is $r = 0.83$. Note that spatial information in the axial plane is not conveyed by the adjacent points on the scatterplot, though the pairwise relationship between subject voxels is maintained.

FIGURE 5.

(A) Degrees of freedom in individual subjects. The estimated DOF in 60 FA_{indiv} (i.e., all $N = 1$), obtained by setting the threshold of the autocorrelation function at 10 %. Along the x-axis, the younger-aged subjects are positions 1–20, middle-aged subjects are positions 21–40, and older aged-subjects are positions 41–60. It is apparent there is no effect of age on the estimated DOF.

(B) Degrees of freedom for group average data. The effect on the estimated DOF of intersubject averaging (all $N \geq 2$) at a 10% threshold of the autocorrelation function. In this box and whisker plot, the central line is at the median value, and the bounding lines of the box are at the lower and upper quartiles. The whiskers of each box extend to the most

extreme point within 1.5 times the interquartile range (asterisks mark values beyond this range). This plot demonstrates there is little change in the estimated DOF beyond $N = 8$.

A) Negative association between age and FA (n = 60)

B) Standard voxelwise FA analyses: two age-group comparisons at n = 40

FIGURE 6.

(A) Association between age and FA. Voxelwise regression in TBSS across all 60 subjects revealed significant, extensive, and symmetrical negative associations between age and FA. Tracts impacted at the significance level of *corrected*, two-tailed $P \leq 0.002$ are shown as a separate overlay (this green overlay appears aqua due to the voxels in common with the $P \leq 0.05$ threshold as represented in the blue overlay).

(B) Voxelwise, between-group FA analyses. A voxelwise, unpaired t -test in TBSS between younger and *older-aged* groups (total $n = 40$) demonstrated extensive group differences (blue overlay, *corrected* results). Expectedly, these findings recapitulate the age association across the entire cohort shown in panel A. However, also shown in panel B are the *uncorrected* results of a younger and *middle-aged* subjects comparison (red overlay). These findings are demonstrated at a threshold of $P \leq 0.4$, as no findings in this comparison reached the corrected significance threshold ($P \leq 0.05$). Importantly, the uncorrected threshold ($P \leq 0.4$, red overlay) reveals a plausible distribution of age effects that are largely in agreement with the known age-effects shown in the blue overlays of 6A and 6B (see Discussion). The PESCA method did detect a significant between-group difference in this same younger- versus middle-aged comparison at $N = 20$, facilitating alternative regionalizing approaches to explore these plausible age effects.

FIGURE 7. Large standardized effect sizes in the younger- versus middle-aged comparison

A voxelwise map of large effect sizes is shown for the comparison of mean FA values between younger and middle-aged subjects (placed in neurological orientation with left on the reader's left). The yellow regions represent areas where the mean FA values of younger subjects were greater than those of middle-aged subjects with a Cohen's $d \geq 0.8$. These regions are overlain on the TBSS analysis mask (blue) and the mean FA image from the study (grayscale). The regions meeting this criterion that are visually apparent include the right cingulum (top row, left panel), the body/ genu of the corpus callosum and forceps minor (top row, middle and right panels), and the bilateral external capsules (bottom row, middle and right panels). As discussed in the manuscript, these are plausible sites of FA decreases during aging.

Table 1

Results of Paired and Unpaired Tests Between Age Groups

(A) PAIRED t-TEST USING THE FA SKELETON DIFFERENCE IMAGE (FA _{DIFF})									
	N	FA _{DIFF}	FA _{DIFF}	DOF	t-stat	P-val ¹	P-val ²		
		MEAN	SD						
YOUNG-OLDER	4	0.021	0.060	134	4.0	< 0.01*	< 0.01*		
	6	0.024	0.051	128	5.3	< 0.01*	< 0.01*		
	8	0.024	0.047	125	5.8	< 0.01*	< 0.01*		
	10	0.023	0.043	124	6.0	< 0.01*	< 0.01*		
	20	0.023	0.035	121	7.2	< 0.01*	< 0.01*		
MIDDLE-OLDER	4	0.016	0.059	134	3.2	< 0.01*	< 0.01*		
	6	0.019	0.049	128	4.3	< 0.01*	< 0.01*		
	8	0.018	0.043	125	4.7	< 0.01*	< 0.01*		
	10	0.017	0.040	124	4.7	< 0.01*	< 0.01*		
	20	0.017	0.030	121	6.3	< 0.01*	< 0.01*		
YOUNG-MIDDLE	4	0.005	0.055	134	1.0	0.160	0.319		
	6	0.005	0.046	128	1.2	0.116	0.232		
	8	0.006	0.040	125	1.7	0.046*	0.092		
	10	0.005	0.036	124	1.5	0.068	0.136		
	20	0.005	0.027	121	2.2	0.015*	0.030*		
(B) UNPAIRED t-TESTS USING MEAN FA VALUES									
	N	Difference (in mean FA)	SD ³	DOF ⁴	t-stat ⁵	P-val ¹	P-val ²		
YOUNG-OLDER	4	0.023	0.017	6.0	1.9	0.06	0.11		
	6	0.023	0.018	10.0	2.3	0.02*	0.05*		
	8	0.023	0.018	14.0	2.5	0.01*	0.03*		
	10	0.023	0.018	18.0	2.8	0.01*	0.01*		
	20	0.023	0.018	38.0	3.9	< 0.01*	< 0.01*		

(B) UNPAIRED t-TESTS USING MEAN FA VALUES							
	N	Difference (in mean FA)	SD ³	DOF ⁴	t-stat ⁵	P-val ¹	P-val ²
MIDDLE-OLDER	4	0.017	0.013	5.6	1.6	0.08	0.16
	6	0.018	0.014	9.4	2.0	0.04*	0.08
	8	0.017	0.014	13.1	2.1	0.03*	0.05*
	10	0.018	0.014	16.9	2.4	0.01*	0.03*
	20	0.017	0.014	35.8	3.3	<0.01*	<0.01*
YOUNG-MIDDLE	4	0.006	0.017	5.5	0.5	0.31	0.62
	6	0.005	0.017	9.4	0.6	0.28	0.56
	8	0.005	0.018	13.1	0.7	0.26	0.52
	10	0.006	0.018	17.0	0.8	0.22	0.45
	20	0.005	0.019	35.9	1.0	0.15	0.31

¹ One-tailed *P*-value;

² Two-tailed *P*-value (significant comparisons are marked with an asterisk);

³ The 'sd' column represents the following: the mean standard deviations from 1000 resamples of 'young' FA values (first 5 rows), 'middle' FA values (middle 5 rows), and 'older' age group FA values (bottom 5 rows).

^{4, 5} DOF and *t*-statistic were calculated assuming unequal variances (Zar, 1996).